Rhythm Matters: The role of rhythmic attending for non-native learners of English

Ewa Wanat BA, MLitt

University of Glasgow

College of Arts

School of Critical Studies English Language & Linguistics

Submitted to the University of Glasgow in fulfilment of the requirements for the degree of Doctor of Philosophy

July 2019

Abstract

In English connected speech, reduction processes can dramatically affect the phonetic shape of words, especially function words, reducing their intelligibility for non-native listeners. There is a close connection between reduction and speech rhythm: metrically weak syllables reduce more, and may be cued only by subtle phonetic detail that non-native listeners struggle to detect. Despite growing evidence that attention to speech and music is rhythmically guided and that speech processing depends on language rhythm, little work has tested whether encouraging non-native learners to attend to rhythm might support their comprehension of casual speech. This thesis investigates whether comprehension of Glaswegian connected speech can be improved if English as a Foreign Language (EFL) learners receive training which relies on entraining the learners' attention to a rhythmic speech stimulus.

Three experiments were conducted. All three of them followed the same pre-test, training and post-test structure. The training differed across experiments as explained below, but in all cases, the pre-test and post-test assessed participants' comprehension of reduced unstressed function words and morphemes in casually-spoken sentences. All pre-tests and post-tests involved listening to fast casual sentences spoken by Glaswegian English native speakers while reading the sentences on a computer screen where they appeared with gaps, corresponding to reduced unstressed function words. Participants' task was to fill these gaps with the words they heard. The participants' score on these tests was the dependent variable.

Experiment 1 investigated whether listening to rhythmically organized speech would improve learners' comprehension as opposed to speech that was not rhythmically organized. In its training phase, EFL learners in the experimental group listened to speech of high rhythmicity, i.e. sentences of regular metrical structure which had been recorded by asking the speaker to align their speech to a regular metronome beat. Each sentence was presented four times, with its rate increasing from slow to fast, so that participants were exposed to a range of degrees of phonetic reduction, within a rhythmically predictable frame. A control group of learners listened to speech of low rhythmicity: metrically irregular sentences which had been recorded without a metronome beat, as part of a read

story; again they heard four presentations ranging from slow to fast rate. To maintain attention, participants' task during training was to count the number of times food was mentioned. Experiment 1 did not show a significant difference between the experimental and control group in terms of improvement in pre- and post-test comprehension scores, though the numerical result showed more improvement in the experimental than in the control group, i.e. in the expected direction.

Experiment 2 tested whether sensorimotor synchronisation with the beat in speech would improve the learners' connected speech comprehension, as opposed to control training. Participants in the experimental group received training in which they performed sensorimotor synchronisation (i.e. finger tapping) with the beat they perceived in the speech they heard, while the control group received training in which they listened to the same stimuli but did not tap to the beat. Instead their task was to listen and tap their finger when they heard a randomly placed click sound. Three listener groups took part in this experiment: Chinese EFL learners, native English speakers of a different variety than the speech presented (Canadian and US English), and Glaswegian English native speakers. The results showed that only the Chinese EFL learners improved in the post-test, compared to the control group. In the Canadian/US group, an interesting interaction was found suggesting that the listener's musical ability affected whether they could benefit from the training. No improvement was found in the Glaswegian native group, whose performance was highest overall.

In the third experiment, Chinese EFL learners, who were attending a preparatory course to study at the University of Glasgow, took part in a short 4-week course, with one 40-minute session per week. In this training phase they learned, as a group, to drum the rhythm of Glaswegian English utterances. The control group continued their curriculum as usual. Neither group showed a significant improvement in comprehension from pre- to post-test, which may relate to aspects of the group training setting, or to the level of English of the participants. The results allowed further exploration of the role of prior musical ability, but this did not appear to affect performance, unlike in Experiment 2.

Taken together, these results are interpreted in the light of Dynamic Attending Theory (Large & Jones 1999) as well as of previous research on perceptual learning and the role of

musical ability in learning a second language. Overall, the results of the three experiments offer only limited support for the idea that rhythmic training helps comprehension. The mix of negative and positive findings is interpreted in the light of Dynamic Attending Theory (L&J 1999), as well as previous research on perceptual learning and the role of musical ability in learning a second language. The conclusion argued for here is that benefits of rhythmic training are seen under specific circumstances, where training is set up in such a way as to optimize the possibility of entrainment to speech.

Contents

List of Figures

List of Tables

Acknowledgements

Firstly, words cannot describe how grateful I am to my supervisors Rachel Smith and Jane Stuart-Smith for guiding me through this PhD project, for discussing all the ideas, making suggestions and comments, answering all my (sometimes stupid) questions, and for always making me work harder.

Huge thanks also to Colin Hamilton for his enthusiasm and help with any technical issues as well as working together on making use of my research. For valuable advice and guidance during as well as after my time in Montreal and for letting me use their lab and teaching me how everything worked, I'm grateful to Caroline Palmer and Frances Vera Spindle.

For their practical advice on materials for my experiments 2 and 3 I am grateful to Clare Thompson, Gemma Archer and Jane Bentley.

At the Glasgow University Laboratory of Phonetics members, I am grateful to Vijay Solanki for all his help with R and any other programming-related issues, Clara Cohen, Rachel Macdonald, Duncan Robertson, Robert Lennon, Fabienne Westerberg and Julia Moreno for all their support and putting up with all the questions I had for all of them.

For their moral support, proofreading and always being there for me - enormous thanks to my wonderful friends Sophie Casimira Kromholz and Sreemoyee Roy Chowdhury. To my dance community, for being there when I most needed breaks from all the thinking, for all the dances that took my mind off working, or worrying - huge thanks!

Finally, thanks to my family. My mum, Wladyslawa Wanat, thank you for your support in all the various ways. And my husband, Michael Callan who was with me from the very start of this project, who spent hours letting me record his voice, saying all the strange sentences I asked him to say, in the way I needed him to say them (he's so good at rhythm!), who helped me figure out some difficult things at times when I panicked, and who was there, always.

Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature _____ Printed Name: Ewa Wanat

For Michael.

Chapter I

Rhythm and connected speech in second language speech comprehension learning

1.1. Introduction

In 2003 David Crystal had estimated the number of learners of English as a Foreign Language (EFL) in the world at 750 million. Fifteen years later, the number of learners is likely to have increased; however, difficulties with speech comprehension and, relatedly, with pronunciation are still common. The learners' struggle is especially clear when they are listening to an everyday, casual, conversational speech style, i.e. connected speech. This kind of speech is characterised by a variety of phenomena, e.g. elision, reduction and assimilation, rather than by carefully uttered words. The situation is not helped any further when the speech is fast, or when a less familiar regional accent is used. Despite this, both connected speech and regional accents (e.g. Levis 2005, Archer 2017) are typically covered only briefly, or not at all, in EFL classrooms.

Many connected speech phenomena relate closely to lexical and utterance-level prosody: i.e. to the stress, prominence, and rhythmic patterns of English speech. Rhythmic patterns vary widely across languages, and a growing literature on the rhythmically guided nature of attention to speech and music (e.g. Large 2008, Dilley & McAuley 2008) allows the inference that rhythmically organised contexts might support pronunciation learning in an L2 (second language). Supporting this idea, there is some evidence that entrainment with the rhythm of speech may enhance listeners' speech processing (e.g. Quené & Port 2005) and that rhythmically synchronized imitation of a target L1 (first language) speaker's speech may have positive effects on L2 (e.g. Wang et al 2016). As well as this, rhythmic training can be used as therapy for speech and language impairments (e.g. Thomson et al 2006, Corriveau & Goswami 2009), and dyslexia (e.g. Overy 2000, 2003).

The main goal of this project is to investigate ways in which rhythmic entrainment can be used as a tool to improve EFL learners' comprehension of native English connected speech. The approach to the question was experimental. A series of experiments with L2 learners was conducted to test their comprehension before and after interventions in which an experimental group was offered a type of rhythm-based training, and a control group was not. All three experiments in this thesis have the same structure, i.e. pre-test, training and post-test.

The thesis is organized as follows. Chapter I, the literature review, presents the key theoretical and methodological context for the particular focus on investigating the role of rhythm for language learners, covering four main areas. First, it reviews literature on rhythm in general, speech rhythm in particular, attempts to quantify speech rhythm, as well as what role it plays in our speech processing. Second, the chapter gives an overview of what connected speech is, and how it relates to speech rhythm. The third area covered is Second Language Acquisition (SLA), with a brief overview of theories, as well as factors affecting the success of SLA. One aspect that will be looked into a little more detail is the role of musical ability, or musical training in SLA. Then a more specifically relevant sub-area of SLA will be discussed, namely research into methods of training the learner's perception of their target language speech. As well as this, a brief overview of Computer Assisted Language Learning (CALL) will be offered. Finally, the fourth area of the chapter seeks to tie all these areas together, by reviewing research on acquisition of L2 rhythm and connected speech.

Chapter II will present Experiment 1 of this thesis. The experiment aims to investigate whether auditory entrainment with speech that is highly rhythmic can be useful for improving L2 learners' comprehension of connected speech. Highly rhythmic speech was achieved by asking the speaker to synchronise their speech with a metronome and by designing the materials to have regular metrical structure. The results in the group trained with these materials showed a numerical change in the expected direction compared to a control group. However this was not a significant result.

Therefore a decision was made to increase the opportunity for entrainment in the second experiment, by asking participants to listen and perform a simple motor synchronisation task, i.e. tap their finger to the beat in speech. Chapter III reviews additional literature on sensorimotor synchronisation, then presents Experiment 2. This experiment compares the comprehension of Glaswegian English by three listener groups: Chinese EFL learners, Canadian/US native English speakers and native Glaswegians. The Chinese EFL learner group who were trained using sensorimotor synchronisation (SMS) showed an improvement compared to the control group who was not trained using an SMS task. As well as this, a possible link was found between the Canadian group's musical training and improvement. This link was found via a self-reported variable, rather than a more objective measurement of the participants' musical training, or musical ability.

Chapter IV will review literature on measuring musical and rhythmic ability. This is because a decision was made for the following experiment to include a more objective test of rhythmic ability.

Chapter V will present the third and final experiment of this thesis. This experiment was a practical application of using SMS (sensorimotor synchronisation) in class, with a group of Chinese EFL learners. This experiment did not show significant improvement in the expected direction, and a number of possible explanations for this negative result are considered.

Chapter VI will consist of a general discussion of the results of the three experiments in the light of previous theories and research areas, such as Dynamic Attending Theory, perceptual learning, sensorimotor synchronisation, P-centres and language rhythm. As well as this, it briefly outlines new research directions geared towards a practical implementation of the findings contained in this thesis.

1.2. What is rhythm?

Rhythm is generally seen as a pattern of events that recur in a regular manner in time. We observe rhythm in nature every day, e.g. oscillations in our brains and beating of our hearts. We create rhythm for entertainment and aesthetic purposes through art, poetry, music or dance. These rhythms can have a common property which is called periodicity, i.e. "a pattern repeating *regularly* in time" (Patel 2008: 96). We could, however, also say that even walking or speaking have rhythm. For Patel (2008) periodicity may or may not occur in these. A widespread belief is that periodicity, or strict isochrony is important for musical rhythm; however, as Large & Jones (1999) demonstrate, the external rhythm of music itself is actually rarely periodic (see Fig. 1.1 below). In Figure 1.1 we are shown the musical notation in A), pitch sequence when this musical notation is actually played by a piano player in B), and in C, we can observe the onsets of the key presses that were performed for this musical notation by the player. Here, the time intervals between these onsets demonstrate very clearly that they vary, i.e. are not in fact regular, or periodic.

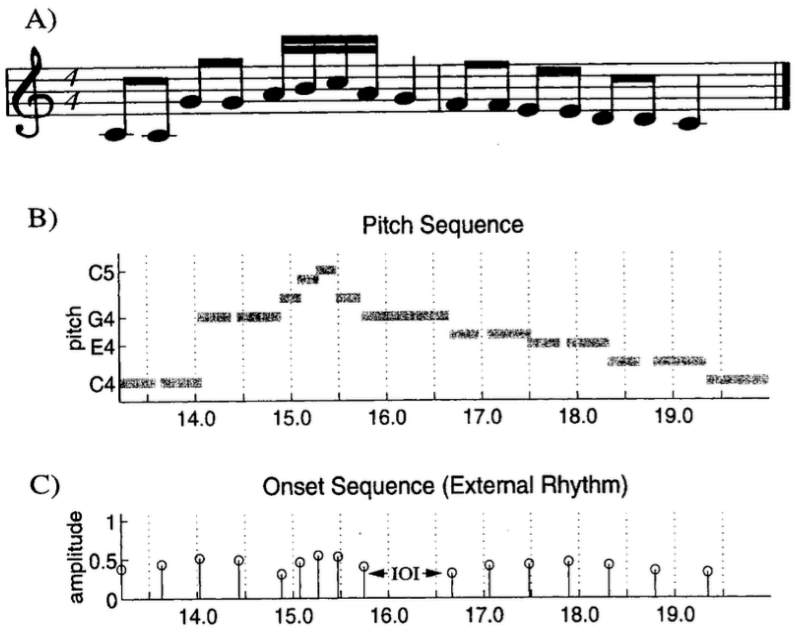


Fig. 1.1. A) represents musical notation, B) represents onset times on x axis and absolute pitches as the musical notation is actually performed by a musician on the y axis, and C) shows the inter-onset intervals, i.e. the onset timings of successive piano key presses. From: Large and Jones (1999: 120).

The dimension that is, however, always important for rhythm is time, as all rhythmic events unfold in time. Apart from time, or timing, other elements of rhythm are prominence and grouping (Patel 2008: 96). Prominence is the element which tells us which sound stands out from its surroundings. Grouping is the element which tells us which sounds belong together, in groups.

Whether we think of speech or of music, they will both have other elements that contribute to the patterning, such as pitch and loudness (Jones 1976, Kohler 2009, Barry et al. 2009). These interact to form a rhythmic pattern.

Two main concepts which relate to rhythm and which will frequently recur in this thesis need to be first defined. These are *pulse* or *beat* and *metre*. For the purposes of this thesis, these key terms, which have attracted substantial discussion in the literature, are defined and operationalised as follows.

Beat and *pulse* are often used interchangeably and they "refer to regularly recurring articulations in the flow of musical time" (London 2001). As London (2001) says, pulses do not necessarily need to be present physically, and the sense of pulse arises "through the listener's cognitive and kinaesthetic response to the rhythmic organization of the musical surface" (London 2001). Following Patel (2008: 99), *beat* is a stable, perceptually

isochronous *pulse* with which we can synchronise periodic movements of e.g. limbs, steps, finger taps etc. In Western music, beats can be strong (accented) or weak (unaccented), i.e. they are organised hierarchically according to their strength (Patel 2008: 103). Similarly in speech, there can be weak or strong syllables, e.g. the sentence *He is fat for a cat in a box*, when said out loud could have the *wwSwwSwwS* structure (where *w*-weak syllable, *S*-strong syllable).

Metre is the organisation of these beats into groups. This is the case for both speech, and music. In music we can have e.g. a Waltz meter, i.e. strong-weak-weak (which in poetry would be called dactylic), or a March meter, i.e. strong-weak alternating (which in poetry would be called trochaic), or various other kinds of metre. In musical notation, metre is typically indicated by the time signature which tells us about the grouping and accentuation patterns (Patel 2008: 107). To go back to the example we used for speech, in the sentence *He is fat for a cat in a box* there are three groups of *wwS,* i.e. two weak syllables and one strong syllable.

It is important to remember that both beat (or pulse) and metre are defined as perceptual elements of rhythm, i.e. they exist in listeners' perception. Even if not completely regular in the physical domain, beats are frequently perceived as falling in equal intervals of time; and metre is organisation of these beats into groups in listeners' perception, whether or not cues to the grouping exist in the physical domain.

1.3. Rhythm in speech

The idea that different languages may have different kinds of rhythm was first written about in 1940 by Lloyd James. He wrote that languages such as Spanish sound like a 'machine gun' and languages such as English sound like 'Morse code'. Since then, the traditional idea that languages of the world can be grouped into rhythm classes has been discussed over and over again in many studies by a great number of researchers, e.g. Lloyd James (1940), Roach (1982), Dauer (1983), Ramus et al (1999), Grabe and Low (2002), Dellwo (2003) & (2006), White and Mattys (2007), Arvaniti (2010), (2012), to name a few. These will be reviewed below. As well as this, some have asked the question whether it may be in our perception, rather than production, of speech that rhythm is found; a brief look at the literature on this point will also be taken. Next, an attempt will be made to answer the question of whether or not speech is rhythmic at all. A possible alternative to the traditional linear view of speech rhythm will then be discussed. Following on from that, Dynamic Attending Theory will be introduced. Finally, there will be a brief review of the

roles that speech rhythm plays in communication.

1.3.1. Discrete rhythm classes and isochrony

Two related ideas have been researched extensively. One is that there is isochrony between some basic elements of speech and that is why the speech signal is rhythmic. Isochrony is the state of being divided into intervals of equal durations. For language, this would mean that there is a unit, e.g. stresses or syllables, which occurs regularly in time. The other key idea is that there are discrete types of rhythm classes in languages of the world and each language can be assigned to one particular rhythm class.

1.3.2. Rhythm classes

First of all, according to Lloyd James' idea, languages can be put into two categories: syllable-timed and stress-timed. Syllable-timed languages are the ones with a "machine-gun-like" rhythm, and stress-timed languages are those with a "Morse-code-like" rhythm (Abercrombie (1967), Pike (1945)). Syllable-timed languages are those for which syllables, whether stressed or unstressed, "occur at regular time intervals and the time between stressed syllables will be shorter or longer in proportion to the number of unstressed syllables" (Roach 2010: 107-108). Examples of such languages would be Spanish, French and Telugu. Stress-timed languages, in the traditional view, would be the ones in which "stressed syllables occur at regular intervals, whether they are separated by unstressed syllables or not" (Roach 2010: 107). For example, English, German and Russian would be classed as stress-timed languages.

Later on, the need for another group was identified since languages such as Japanese would not fit into either of the two categories. In order to include those languages, the so-called mora-timed category was added to the classification, in which the basic rhythmic unit would be the mora (Fletcher 2010: 552). The mora can be defined as "a subconstituent of a syllable [which] dominates either a vowel segment or segments in the syllable rhyme" (2010: 529).

Such assigning of a specific rhythm class to all languages suggests that what occurs is simple isochrony, i.e. "regular recurrence in time of some given unit of speech rhythm [which] is a pervasive attribute of all spoken languages" (Laver 1994: 523). Researchers tested the hypothesis that isochrony exists in speech, e.g. Dauer (1983) and Roach (1982). However, no such evidence in the physical speech signal was found, as discussed below.

1.3.3. Isochrony in the speech signal

Roach (1982) tested the two claims that were the base of the rhythm class hypothesis, i.e. first that there is a considerable variability of syllable duration in a stress-timed language, but in a syllable-timed language the syllable duration tends to be equal, and second that "in syllable-timed languages stress pulses are unevenly spaced," whereas in a stress-timed language the intervals between successive stress pulses are regular (1982: 73). He measured the standard deviation of syllable durations, and these measurements did not support the first claim. In order to test the second claim Roach first calculated the hypothetical figure that would be expected in a perfectly stress-timed language and compared the "range of percentage deviations in inter-stress intervals" of syllable- and stress-timed languages. He hypothesised that the range would be higher for syllable-timed than for the stress-timed languages. However, the calculated percentage variance showed that the stress-timed languages showed much higher values than the syllable-timed, which was against the hypothesis. Roach then also compared the correlation coefficient for the association between percentage deviation and number of syllables in an inter-stress interval in each type of language. Here, the syllable-timed languages' intervals would be expected to be longer when measured in proportion to the number of syllables they contained, than the intervals in the stress-timed languages. This was also not the case, therefore as Roach (1982) concludes, it was "not possible to separate the two languages on this basis".

Dauer (1983) noted that the phenomenon of vowel reduction occurs widely in languages which were traditionally seen as stress-timed, whereas in languages labelled as syllable-timed vowel reduction does not occur. Another difference that Dauer noticed between stress- and syllable-timed languages is to do with language-specific syllable structure, i.e. the permissible types of syllable in e.g. English and Spanish differ, in that English can have a large variety of vowel and consonant combinations whereas Spanish is more limited (1983: 55).

1.3.4. Rhythm metrics

Both Dauer and Roach conclude that it is not possible to put languages into the two traditional categories based on physical measurements. They also suggest that it may well be that the difference is of a perceptual nature, i.e. we tend to hear isochrony in languages where there is no isochrony in the actual acoustic rhythm of the languages. We will return to this idea later (see section 2.5 below).

After a substantial gap from the early 1980s until the mid-1990s in the search for discrete rhythm classes, researchers became interested in this topic again, for various reasons. Some, e.g. Ramus et al (1999) were motivated by developments in psycholinguistics, e.g. Cutler's work showing that speech processing can depend on the language-specific timing of events (reviewed in section 2.9 below) as well as by the fact that infants seemed to be able to perceive the difference between stressed and syllable-timed languages. Others, e.g. Nolan, Grabe and Low (for details, see review in 1.3.4.2 below), were motivated by the desire to establish ways of describing the prosody of different languages and language varieties.

This research generated various kinds of rhythm metrics, e.g. Ramus et al.'s (1999) %V-ΔC, Grabe and Low's (2002) Pairwise Variability Index (PVI) and Dellwo's VarcoC (2003) and later on VarcoV. These metrics attempt to capture the rhythmic, or to be more precise, timing differences in various languages by using the acoustic signal directly rather than phonological units such as interstress intervals or syllables.

Ramus et al.'s (1999) %V-ΔC metric looks at the proportion of time in an utterance that is taken up by vowel intervals (%V), and the variability of consonantal intervals' duration (ΔC). In contrast, Grabe and Low's (2002) Pairwise Variability Index (PVI) looks at successive intervals rather than at all available intervals. In other words, PVI measures the average difference in duration between successive pairs of vocalic or consonantal intervals (excluding pauses). Grabe and Low (2002) say that this is important to do because, for example, three successive long vowels and three short vowels would give the same ΔV as three short and long vowels alternating; and it is important to capture this alternation between strong and weak elements because it is often what is thought to be the key dimension of rhythm.

Dellwo & Wagner (2003) and White & Mattys (2007)) looked into how speech rate influences the %V and ΔC measures. They found, for example, that variability of the duration of consonantal intervals is affected by a faster speech rate (or shorter mean consonant interval duration (meanC)), i.e. that speech rate can affect the ΔC. This caused them to devise VarcoC and VarcoV – new measures that normalise for speech rate.

1.3.5. Evaluation of rhythm metrics

When the different rhythm metrics were being investigated, a range of languages were used to test the metrics on. It turned out that the %V and ΔC measures were related to the

syllable structure a language permits, i.e. languages with low %V and high ΔC corresponded to the traditional stress-timed languages, and languages with lower ΔC and higher %V corresponded to the traditional syllable-timed languages. A division into stress- and syllable-timed languages was also achieved using the PVI measures (by plotting the normalised PVI (nPVI) and raw PVI (rPVI) on y and x axes respectively (Grabe & Low 2002:12). However, as the authors themselves conclude, that distinction between rhythm classes was weak, because degrees of syllable- and stress-timing were also found in the languages tested, and because not all languages could be categorised that way.

Generally, it seems that rhythm metrics are sometimes, though not always, able to find ways in which languages can be categorised into different rhythm-based groups. Thus, methodology and reliability have been criticised by some researchers while others have made more conceptual criticism of metrics.

To begin with the methodological critiques, Grabe and Low (2002) find that one language can be assigned to two different rhythm classes depending on which metric has been used. Furthermore, there were a substantial amount of languages that could not be assigned to any class with help of these metrics (Grabe and Low 2002). As well as this, as Arvaniti (2009) says, the results can vary according to the segmentation practices that are used, and it is not unreasonable to assume that the same can also happen depending on ways in which data were elicited. Barry et al (2009) showed that using the same measure (PVI), different results were obtained depending on speech style, e.g. when speakers produced poetic utterances vs when speakers modified the duration of a different element in an utterance for accentual purposes.

Moving beyond the issues with reliability, researchers have also critiqued the metrics' focus on duration alone. Barry et al (2009) showed that other parameters, such as f0, intensity and vowel quality, can also contribute to the perceived rhythm and that the same rhythm can be achieved if different parameters contribute to it to a different extent. To show this, Barry et al (2009) had their subjects listen to nonsense (*dada*) lines which differed in the values of those parameters, and asked them to rate their rhythmicity. They found that changes in not only duration but also f0 contribute to the perceived rhythm. Arvaniti (2009) agreed with this, saying that what these metrics provide are "crude measures of timing" without being based on all the other appropriate parameters as well. She then tested ΔC, %V, PVIs and Varco on a variety of languages (English, German,

Greek, Italian, Korean and Spanish) and using different methods of speech elicitation and found that these metrics did not show consistent results, i.e. the rhythm class a language could be assigned to could depend on e.g. the speaker, speech elicitation method and syllable complexity (Arvaniti 2012). Further she says that, due to these metrics being based on durations only, the relationship between these metrics and the definition of rhythm is circular (Arvaniti 2009).

Arvaniti (2009) therefore turns to the field of psychology for a definition of rhythm that relies on repetitive patterning of groups of series of stimuli. The grouping aspect is the key, because this relates to intensity, duration as well as temporal spacing of elements. Furthermore, as Sluijrer and van Heuven (1996) and Campbell and Beckman (1997) say, the acoustic cues on which native listeners of various languages rely for their perception of stress can vary. The perceived prominences can result from a combination of various parameters rather than just one, depending on the language. Therefore, according to Arvaniti (2009) classifying languages based on impressions or based on only a single chosen acoustic parameter cannot be successful.

Other researchers, e.g. Shattuck-Hufnagel and Turk (2013), and Kohler (2009) agree with Arvaniti's (2009) view that rhythm metrics are not always able to identify the appropriate rhythm class of a language (Shattuck-Hufnagel and Turk 2013: 108) and that "the global temporal bracketing of the speech signal into chunks that have recurring phonetic characteristics over and above the syntactic and semantic organisation, but interacting with it" (Kohler 2009: 34). Kohler also suggests that speech rhythm should possibly be looked for in patterns of f0 and energy that recur with regularity, e.g. how high-low pitch patterns, are perceived as rhythmic high-low progression in a word (or short phrase), just as high-low "marks conclusion in phrasal utterances" (2009: 34).

To sum up, while absolute physical isochrony cannot be found in naturally occurring speech, it needs to be remembered that even in the definition of rhythm in music, isochrony does not often occur physically (see section 1.2 above), but may still be perceived. The sense of regularity of the acoustic elements arises in our perception, whether in music or speech. Therefore, while the acoustic correlates of speech rhythm are still being investigated, we know that it is not simple isochrony that we are perceiving. As for rhythm metrics' success at classifying languages, no timing-only-based metrics have been successful at clearly classifying all languages into distinct rhythm classes. More research has been conducted since, which includes other parameters, (e.g. Tilsen 2013) and

it is possible that the different perceived rhythms can be found through such metrics. While this thesis itself does not directly utilise any of the metrics, it was important to discuss these to show the complexity of speech rhythm, as well as to show that while no absolutely conclusive evidence has been found for the different rhythm classes that languages have, languages are still widely thought of as belonging to different classes, and, that learning to understand a language from a different rhythm class could be different, than learning to understand a language from the same class.

1.3.6. The production vs perception dichotomy for speech rhythm

Since no clear acoustic evidence for isochrony or for discrete rhythm types among languages has been found, researchers have also started considering the fact that speech rhythm is perceptual phenomenon (Roach 1982, Dauer 1983, White and Mattys 2007, Arvaniti 2009, Shattuck-Hufnagel & Turk 2013), i.e. that there may be isochrony in our perception, rather than in production. Lehiste (1970) showed that isochrony seems to occur in our perception more than in the physical world, because listeners exhibited a so-called just noticeable difference for duration – i.e. they failed to notice differences in duration of stimuli under a certain threshold. As well as this Lehiste (1977) showed that listeners performed better at the task of judging durations of metric feet with non-speech stimuli than with speech stimuli, suggesting that when it comes to speech, listeners are less sensitive to purely durational differences and may therefore be more likely to perceive regularity in speech, even if physically it is not present. This is what Shattuck-Hufnagel and Turk call "periodicity-based perception" (2013: 96).

On the other hand, Arvaniti and Ross (2010) asked speakers of Korean, Greek and American English to rate English, Italian, Korean, Spanish, German and Greek utterances, which had been low-pass filtered in order to ensure that the language itself would not be recognisable. They were asked to rate those utterances on how similar they are to a series of non-speech trochees, i.e. one strong element followed by one weak element (as in e.g. the word *otter* in English). These listeners' impressions did not support the idea that there are impressionistic rhythm differences between languages either. Thus even if a tendency to perceive isochrony exists, it is still not clear that perceptual behaviour supports the idea of discrete rhythm classes.

1.3.7. Is speech rhythmic?

My own very short answer to this question would be: "It depends." As Shattuck-Hufnagel & Turk (2013) suggest, it depends on how we define speech rhythm. Giving a clear and definitive answer to this question is beyond the scope of this thesis; however, a few general points can be made.

First of all, if surface periodicity, i.e. isochrony in the acoustic signal, is a part of the definition, then there is no such thing as speech rhythm in normal conversational speech. This kind of rhythm can possibly be achieved but only in certain circumstances, e.g. repetition of a phrase (more detail on this in section 1.3.7 and in Chapter 2, section 2.7 below). Some, e.g. Nolan and Jeon (2014) suggest that the idea that there is rhythm in speech is a metaphor, just as calling a piece of land a chessboard from a window of an airplane is a metaphor. They do not mean to discourage further research on the relation of speech and rhythm but, rather, encourage researchers to re-think the questions they ask, and they point to those specific situations in which speech becomes rhythmic, such as e.g. when it is aligned with a metronome beat (2014: 8). This way of thinking about rhythm can be a useful construct for analysts to work with, and is used in this thesis.

However, periodicity may not always be a part of the definition of rhythm, even in music (see section 1.2 above). If that is the case, and all the other elements of the definition of rhythm, such as timing, grouping and prominence are all present in speech, then it can be said that speech does have rhythm. In other words, since speech unfolds in time and since some elements of speech are grouped together, and some are more prominent than others, then in this sense, speech can be seen as rhythmic.

The situation can, however, get more complex. There is the question of whether we perceive periodicity where there is none, as mentioned above, or whether there may be periodicity in human motor control which would be responsible for our tendencies for producing (near-) periodicity, evidence for which can be found in speech in e.g. polysyllabic shortening, i.e. the vowel portion in the word *hamster* being shorter than in *ham* (Shattuck-Hufnagel & Turk 2013).

Another possible way of trying to capture the differences in speech rhythms which languages, or even varieties of the same language, exhibit, is to do with the global differences between various timing and prominence patterns languages show (Shattuck-Hufnagel & Turk 2013). Languages exhibit these differences due to the fact that they

employ various timing and prominence-placement phenomena differently; these patterns, or aspects of segmental and prosodic phonology, can be seen as giving rise to distinct rhythms languages can have.

To sum up, there is no short answer to this very big question of whether or not speech is rhythmic. If surface periodicity is needed for speech to have rhythm, then it can only have rhythm in certain situations. It can also be that surface periodicity is not needed in rhythm and only perceiving periodicity is enough. Yet another option is that even without surface periodicity, we perceive some general, global pattern, or rules, of the interaction between timing, prominence and grouping that characterise different languages, and this is enough for us to say that speech has rhythm and that rhythm varies between languages. This thesis will proceed on the assumption that we are able to perceive different rhythms that speech can have, and that there is a specific way of producing speech in which rhythmicity emerges, such as e.g. speaking in time with a metronome and repetition.

1.3.8. An alternative view of speech rhythm: "circles in time"

From the information contained in the previous sections (1.3.1-1.3.6) as well as from the sections on connected speech (see section 1.4) below, it is clear that spoken languages differ in their temporal patterning. As Port (2007: 503) says, native listeners notice the foreign accent because, among other things, these temporal patterns of non-native speech may be different.

Port (2007) notes that traditional representations of speech are linear. For example, phonetic transcription is a serially ordered sequence of symbols. As he says, it is true that we know the difference between [tæn] vs. [ænt] vs. [næt] because of the order these symbols are arranged in. However, as he further points out, one of the main issues with this type of representation of speech is that it does not make it possible to transcribe segment durations, relative to each other and relative to the utterances they are in. Languages differ in their patterns of sounds in time, not just in ways these symbols are arranged sequentially, but also how their timing relates to each other. Therefore Port's (2007) main criticism of using segmental representation of speech is that a segmental representation does not capture these timing relationships, as it focuses only on linear serial order.

The speech rhythm metrics discussed above also attempt to measure time in linear ways. Port (2007), however, suggests a different way of representing speech, using a "circle in time", or a cycle. This, Port claims, allows us to observe events together with their timing

relations to each other. He compares this to how meter is represented in music, where a pattern of nested cycles are located in time somewhere relative to the basic pulse. Further, Port describes rhythm in terms of phase angles of repeating cycles. As he notes, there are various ways of representing this, such as e.g. degrees (360) to represent a complete cycle, or the interval {0,1}, division of which into fractions results in phase angles (for four beats, these angles would be 0, 0.25, 0.5, 0.75). A simple illustration of this can be seen in Figure 1.2 below. Change of rate need not affect these patterns, due to the fact that they are relative to this basic regular pulse.

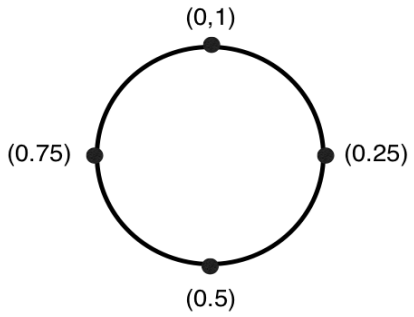
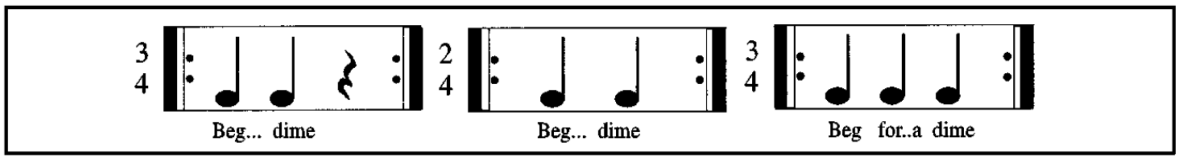


Fig. 1.2. A graphic representation of phase angles of 0 or 1 and its fractions of 0.25, 0.5 and 0.75. This is a simple case in which the events which are nested within the cycle are perfectly isochronous.

The question arises how far this idea can be applied to speech. Cummins and Port (1998) conducted experiments in which English native participants repeated short phrases (e.g. *beg for a dime*), and their task was to align the stressed syllables of the phrase with a regular metronome beat. The relative timing of the metronome beeps was varied (for a more detailed explanation of the paradigm see Chapter 2 below) in order to test whether there were any regularities in the way participants aligned elements of the phrase within the cycle which was defined by this metronome beat. Which elements might they align in a stable way, if any, and to which points in the cycle?

In short, Cummins and Port (1998) showed that subjects never placed the stress beats freely while repeating the phrase, but only at three specific points, which were at 0.33, 0.5 or 0.67 of the cycle. That is, the second stressed syllable, e.g. *dime*, even though participants were given the opportunity to place it anywhere within the cycle, occurred only around those three stable points. These three preferred points are easily illustrated with musical notation, as in Figure 1.3 below.

A



B

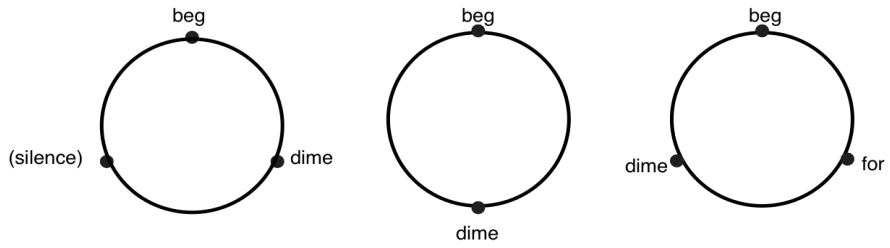


Figure 1.3. A. From Cummins and Port (1998): Musical notation illustrating the three stable patterns produced. B. The same three productions of the repeated phrase represented as fractions of a cycle, corresponding to the above musical notation.

In the dynamical systems approach, these points are called attractors – "privileged states of the system towards which the system will tend, and at which the system is most stable" (1998: 165). Cummins and Port (1998: 165) called this division of the cycle into simple integer ratios, the Harmonic Timing Effect. As Port (1999) suggests, this may arise from the subjects' inability to either perceive or produce the phase lags that are other than 0.33, 0.5 or 0.67.

Thus, as Cummins and Port (1998) state, the stressed syllables, i.e. the elements lower in the hierarchy were nested an integral number of times within the whole phrase, which was the cycle, or the element higher in the hierarchy. This concept therefore offers an alternative way of examining speech rhythm - as organised hierarchically rather than linearly. However, it needs to be remembered that Cummins and Port (1998), and Port (2003) use speech that is rhythmic because it is made to be so, by asking people to speak to a metronome. Understanding speech in these circular terms in which one element lower in the hierarchy is nested within another element higher in hierarchy is more complicated when considering naturally occurring spontaneous speech. As well as this, it may be worth noting that in the example from Cummins and Port (1998) above, the third option i.e. the one in the third notation, or circle the syllable *for* which is usually weak in conversational speech is made to be a strong syllable and thus became an element which is higher in the hierarchy (compared to the other weak syllable *a*).

The application of the concept to naturally occurring speech can be a little better understood using an example. For a certain syllable to be uttered, e.g. *to* ([tə]), the task of achieving a certain configurations of our vocal tract has to occur. This in itself can be thought of as involving the production of two target gestures for the two target configurations (i.e., a specific type of constriction for [t] plus a specific kind of opening for [ə]). The hierarchical nature of speech production can be noted from the fact that two gestures are nested within one syllable, then syllables form words or feet, words/feet form phrases, etc.

If we were going to simply repeat this syllable at a constant rate, this same pattern would keep occurring, i.e. these two events nested inside the bigger event that is the syllable. The case of repeating the same syllable over and over again is a simple one, in the way that it is easy to imagine how a repetition of it (at a constant rate) creates a cycle with a repeating pattern. In this dynamical systems approach (Cummins and Port 1998), on one level each element of the syllable would be seen as two oscillators coupled, or phased, with each other, where the first gesture appears at a certain point of the whole cycle, and the second gesture appears at a different point in the cycle, and these points are stable as the repetition goes on. On the next level, higher in the hierarchy, each element of the syllable is also coupled, or phased with the whole syllable.

In naturally-occurring conversational speech a simple CV syllable being repeated regularly is a rather rare occurrence. Not only do syllables vary in their complexity (especially in a language like English), but it is also uncommon to produce the same syllable, or even different syllables with similar structure, over and over again. However, what is important is that the gestures that contribute to a syllable are phased with each other, or coordinated with each other in stable ways (Byrd & Saltzman 2003), and these gestures overlap in time creating various coarticulatory effects. The various gestural timing properties that arise within and between syllables, words, phrases and sentences (Byrd & Saltzman 2003) are handled in complex ways by dynamical systems approach details of which are beyond the scope of this thesis. In essence, Byrd and Saltzman state that that when there is coupling of oscillators and when there are some slight perturbations of the phase which oscillators are in with respect to each other (for example, caused by the syllable, or word, being at the end of a phrase), this is dealt with by adjusting that phase locally and constantly at the point it is needed rather than changing the phase of the overall utterance. So as Byrd & Saltzman

put it, structural features such as phrase boundaries, which is what they specifically investigated, "warp[…] the temporal fabric of an utterance" (2003: 176).

To return to the speech cycling task itself, researchers turned to asking what would happen if the task was given to speakers of other languages than English: Would cross-linguistic differences in speech cycling behaviour be found, would any such differences confirm or disconfirm traditional assumptions about rhythm typology, and how might they be captured using the hierarchical approach? Chung and Arvaniti (2013) carried out two experiments, in which they tested whether Korean speakers can perform the speech cycling task at all and, if so, which prosodic element they would use in place of stressed syllables (which is used by English speakers). Korean would be classed as a syllable-timed language due to not having stress or foot structure. Chung and Arvaniti's (2013) results showed that Korean speakers are able to perform the speech cycling task, and that the role of a stressed syllable is taken on by the accentual phrase-initial syllable. The accentual phrase is described as a unit between prosodic word and phrase, which is typically marked by a Low-High intonation. Chung and Arvaniti therefore conclude that, since Korean speakers' privilege certain syllables based on the syllables' position, their behaviour is no different to that of English speakers because that may be interpreted as a more general strategy which speakers use and that it does not depend on the rhythmic classification of a language.

It could be the case that speech cycling can be carried out by speakers of any language, if they are given the opportunity to place the prominences in ways that would be convenient for their native language, and doing so could reveal some rhythmic differences which can be approached in the hierarchical, rather than linear way. Nolan and Jeon (2014) speculate that there could "be greater diversity of solution [to the way speakers align their speech to a rhythmic stimulus] for languages without a steep prominence gradient" (9). A steep prominence gradient is "a large disparity between stronger and weaker elements" (2014: 1), i.e. it captures the difference between successive syllables not only in terms of duration but also other qualities, e.g. amplitude or vowel spectral dispersion (2014). They also suggest that in order to gain a better understanding of speech rhythm, a wider variety of languages would need to be tested. Arguably, investigating the ways in which L2 speakers would perform this task might also reveal some areas in which language rhythms differ.

In sum, an alternative approach to speech rhythm has been offered. Rather than looking at linear serial relationships between segments, rhythm is viewed as hierarchical relationships

between various elements in speech. Here, a larger unit, e.g. phrase would contain a variety of smaller elements, i.e. syllables, and syllables would be made of even smaller units etc. Each of the lower-level units is nested within the higher-level unit. As well as this, speech is not necessarily inherently rhythmic, but rhythm in speech emerges in certain situations, e.g. speaking in time with a regular metronome beat. So far, research has shown that speakers of English and Korean are able to perform this task, and that there are certain stable ways in which speakers align their speech to such a metronome beat (e.g. Cummins & Port 1998, Chung and Arvaniti 2013). This method of eliciting speech will be utilised in the present work and the details of the method will be outlined in section 2.3 below. The following section will look into a related area, i.e. Dynamic Attending Theory and will draw connections between it and speech rhythm viewed in this way.

1.3.9. Dynamic Attending Theory

As Cummins and Port (1998: 167) concluded, "[i]n the presence of repetition the speech production system necessarily becomes coordinated, such that a higher level dynamic emerges within which the timing of subordinate processes are constrained" which is no different from other coordinative or rhythmic tasks, like walking. In walking we coordinate our steps so that one foot is always on the ground while we lift and start moving the other foot. That way, they are always coordinated so that there is a 0.5 phase relation between the two legs (Cummins 2009).

Dynamic Attending Theory was first proposed by Jones (1976), who noted that the outside world has a pattern structure with which humans interact. This external pattern structure has a temporal dimension. Jones assumes it is rhythmical, also having a nested, hierarchical structure. In her theory of perception and attention she emphasises the relational nature of time, pitch and loudness. She further says that such "structure of relations in time is important in determining people's expectancies about other serial relations" (1976: 327) such as pitch and loudness.

Human attention is also rhythmical. Our attention is described as "a behaviour of internal oscillations, called attending rhythms" (Large & Jones 1999: 119), which we then synchronise with the external world's rhythms. So, for example, if timing of a stimulus, e.g. a speech stimulus, is regular, the listener will entrain with that external rhythm.

A simple example of two rhythms whose phases and are perfectly synchronised, or entrained, can be seen in Fig 1.4. below.

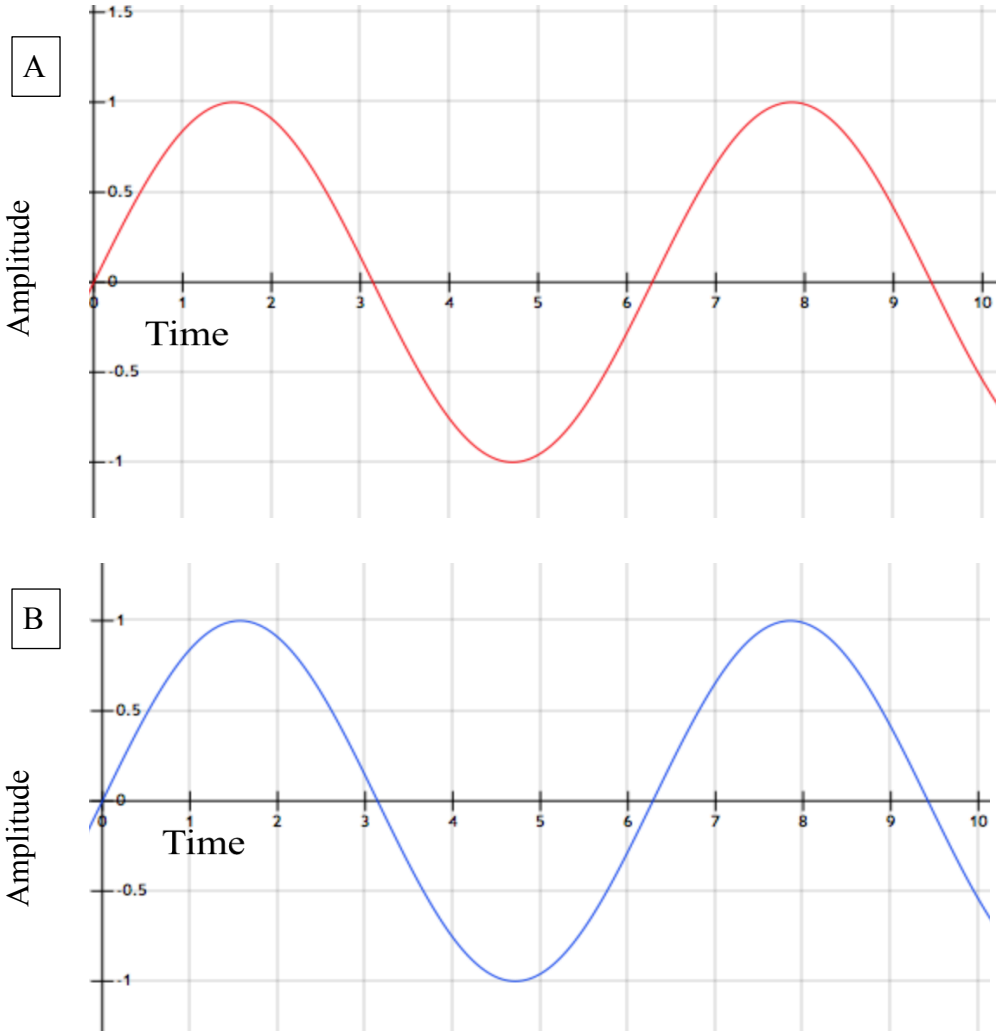


Fig. 1.4. Panels A and B represent two perfectly entrained rhythms.

Such entrainment, or coupling of the two rhythms, results in the listener being able to form expectations of what is going to happen next (Jones and Boltz 1989). Because of this expectation, the listener can focus on the words more quickly (Kohler 2009: 38).

In other words, when the attentional rhythm and external rhythm are entrained, they are said to create a system in which these coupled rhythms are able to return to synchrony after a perturbation. As Large and Jones (1999: 127) say, "coupling exerts a force which pulls the two rhythms towards a synchronous relationship". When a system is coupled, the coupling strength is a parameter by which the system is made resistant to perturbations; it is so because an attractor now exists in that system towards which the system is drawn. Coupling strength determines "the speed with which the coupled system relaxes to the attractor" (1999: 128). There can be situations due to which their synchrony can be

destroyed, i.e. when either the phase and the period, or both, change. For clarification, these are illustrated in Fig. 1.5 and Fig 1.6. below.

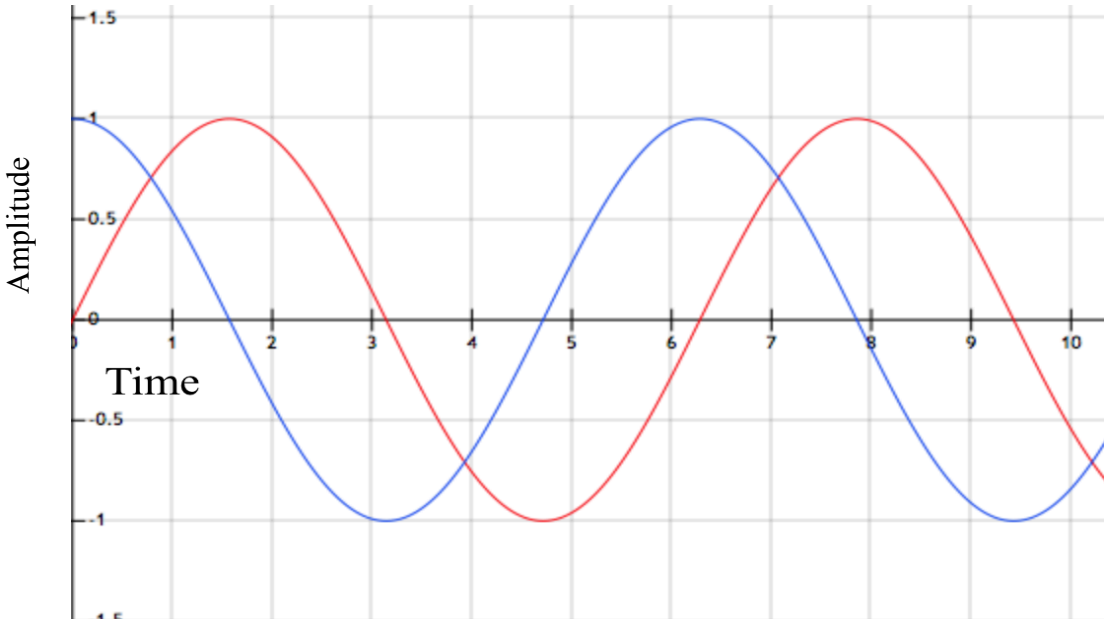


Fig. 1.5. Illustration of two rhythms with an out of phase relationship. They are both the same in terms of their amplitude and period, but differ in timing such that the red rhythm lags behind the blue rhythm. This illustrates phase shift.

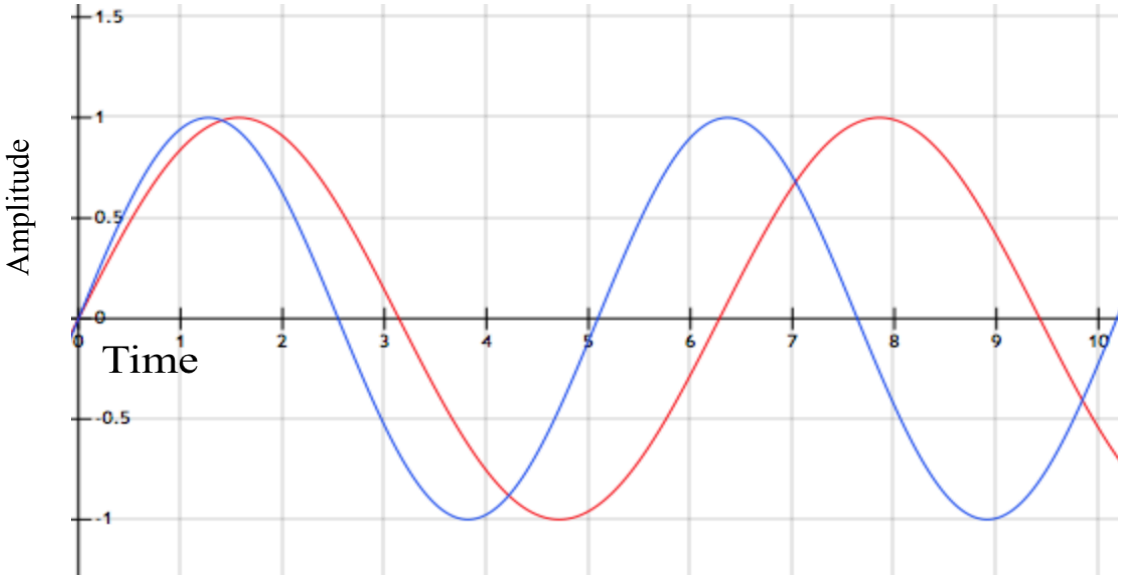


Fig. 1.6. Illustration of two rhythms with an out of period relationship, they differ in timing such that the red rhythm requires more time to complete its cycle than the blue rhythm, i.e. has a longer period. This illustrates period shift.

To deal with these changes in the external rhythm, the attentional rhythm must be able to

track such time-varying events (1999: 128). For that, both period and phase of the internal oscillation need to be able to adapt. Here the period adaptation rate is equivalent to the coupling strength which is used in dealing with perturbations in phase synchronisation. When the phases are entrained, they are synchronised with each other. This is when the two rhythms, which may start out out of phase with each other, as for example as in Figure 1.6 above, gradually change their phases to become synchronised. Period adaptation happens when the period of one rhythm gradually adapts towards the period of the other rhythm.

Together, the concepts of phase entrainment and period adaptation describe how an attending rhythm can generate periodic expectations that are well coordinated with an external rhythm. They also allow us to explain how this kind of system deals with temporal fluctuations of rhythms found in nature. That is, if an external rhythm has a regular pulse, the attentional rhythm will also have a regular pulse, and that pulse will be more regular, the more regular the external rhythm is that it is entrained to. The more regular such a pulse, the stronger the coupling, and the narrower the temporal focus of the pulse. This can be illustrated as a probability density function as in Figure 6.1 below, where $\kappa = 4$ shows a narrower, i.e. more concentrated focus, and $\kappa = 1$ illustrates a broader focus, which is less concentrated. The flat line, $\kappa = 0$, indicates a uniform dispersion of attentional energy over time.

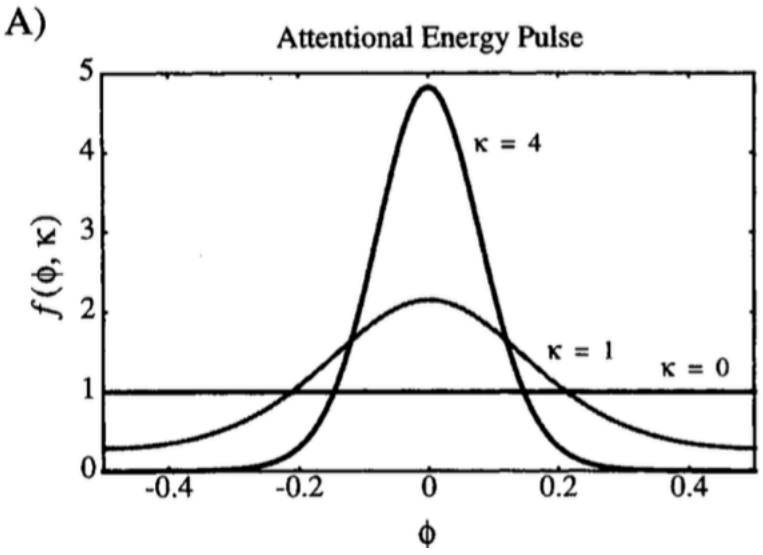


Fig. 1.7 The attentional pulse A: Distribution of energy […] as a function of ϕ for the three different focus (κ) values. Larger κ values correspond to a greater concentration of energy about the expected phase $\phi = 0$. Reproduced from Large & Jones (1990: 130).

The relevance of the Dynamic Attending Theory to speech is that if there are rhythms in speech, then listeners can be expected to entrain to those rhythms, just as they do to e.g. music. It would therefore not be unreasonable to think that the more stable or stronger the beats in speech, the more entrained the listeners can become. Or, that the more experience a listener has with a particular speech rhythm, the better able they will be to entrain to it. Therefore, it would also seem plausible that perception of speech in a particular language or variety can be facilitated by regularity in that speech, as well as by familiarity with its rhythm.

To apply the ideas of Dynamic Attending Theory to speech, Quené & Port (2005) investigated the effect of regular timing on speech perception using a phoneme monitoring task, and measured the participants' reaction times. They looked at whether the perception of target phonemes in English is made easier for the listener if words containing these phonemes are spoken with constant inter-stress intervals because a "good entrainment of the attentional oscillator to salient time points in the auditory stimulus" (2005: 4) would take place. They take into account that metrical expectancy might play a role in the listeners' reaction time to the stimulus as well. For this reason they included words with either trochee-like meter or iamb-like meter. They found that if the words in the list were in time with the regular inter-stress intervals, i.e. when timing regularity occurred (2005: 9), then the perception of those words was easier for the listeners, in comparison to irregular timing which slowed the listeners' reaction time. This supports the Dynamic Attending Theory. The metrical expectancy, however, had no effect on the listeners' reaction times, i.e. unexpected changes in the word metre (e.g. from iamb to trochee) did not make the listener's perception task more difficult.

Quené & Port suggest that this study could be repeated with connected speech stimuli in order to create a more natural experimental environment than word-lists. They conclude that "regular timing of stressed syllables, or speech rhythm, helps to get a spoken message across to the listener, because such regularity helps the listener to attend to the speech signal and to its linguistic content" (2005: 11). Quené & Port proposed that this would be the case for native speakers of a language, however, the present work is interested in whether such speech timing regularity would also aid the non-native listener's comprehension.

In summary, speech cycling (Cummins & Port 1998 and Chung & Arvaniti 2013) experiments show that speakers do a very good job at placing the prominent beats of a

repeated phrase regularly within the phrase, which as, Quené & Port (2005) say can be seen as coupled oscillators, one (slower) producing phrase repetition, and the other (faster) producing the time points that attract the onsets of stressed vowels (2005: 1). The listener who perceives such rhythmic speech entrains their attention with the stimulus, which in turn, is in line with the Dynamic Attending Theory and as Quené & Port (2005) suggest, "regular speech timing improves speech communication". This theoretical approach as well as the speech elicitation method will be used in the present work.

1.3.10. Functions of speech rhythm

We have looked at existing attempts to capture the nature of speech rhythm. However, the question of what role speech rhythm plays in everyday communication is yet to be answered. A review of the main functions of speech rhythm is offered in this section.
Speech rhythm plays important communicative roles, such as e.g. grouping and chunking of the speech signal (e.g. Cutler 2012), taking turns in conversation, guiding attention, i.e. predicting when an important event will happen, and in synchronising speech or repetition. Some of these will be looked at in a little more detail below.

Grouping and chunking means finding where word or phrase boundaries are. Listeners have a variety of tools on their disposal to group and chunk the speech signal, such as phonotactics of the language and their knowledge of it, stress and their knowledge of how it's applied in the language they're listening to, and the overall rhythm of the utterance and the knowledge of how that is applied in speech.

Phonotactic knowledge means the listener knows which sequences of sounds are permitted in the given language, e.g. in English the /lb/-cluster does not occur at the start of a word. This language-specific knowledge aids the listener be more efficient at recognising the sequence [ɪlbɛə] as *ill bear*, rather than the first word finishing with an [ɪ] and the next one starting with [lbɛə] (Cutler 2012: 118).

Stress also plays a role. In fact, Cutler and Norris (1988) used a word-spotting paradigm to investigate how listeners use stress in word segmentation. They asked their participants to listen to short strings of nonsense words, in which actual English words could be embedded, e.g. *mintayf* or *mintef*. In both words the word *mint* was embedded, but the participants' reaction times in spotting that word were much faster in *mintef* than in *mintayf*. The difference between these is that in *mintef* the second syllable has a reduced

vowel, a schwa, and *mintayf* has an unreduced vowel. Cutler and Norris (1988) interpret this difference in reaction times to be the due to the fact that listeners have the knowlegde that strong syllables are usually word-initial and therefore they segmented them as strong-weak for *mintef* and strong-strong for *mintayf*. This knowledge that native listeners have is also the reason why mishearing a phrase such as "She's a must to avoid" can be misheard as "She's a muscular boy" as "boy" starts with/is a strong syllable, and avoid starts with a weak one (Cutler & Butterfield 1992). The above-mentioned studies look at stress in English, however, it needs to be remembered that stress and the degree of reduction can differ in different languages. For example, in English there is a high degree of vowel reduction compared to e.g. Polish, and only slightly higher compared to e.g. Dutch (Cutler 2012). This is likely to have an effect on how native listeners of other languages will use the particular cues to segment the speech stream. This relates to how native speech works when produced in a natural conversational way, i.e. what we call connected speech.

To move on and look even a little further beyond stress and reduction itself, Dilley and McAuley proposed "a perceptual grouping hypothesis, namely that prosodic (i.e., fundamental frequency and duration) cues distal from the locus of segmentation or lexical access of a word affect the unfolding process of perceiving prosodic constituents, thereby influencing word segmentation and lexical recognition" (2008: 296). In three experiments, they showed that listeners pay attention to the overall rhythm, i.e. what they call "distal" prosodic characteristics of the utterances they are exposed to rather than to "proximal" prosodic characteristics of the utterances. They designed 8-syllable-long sequences ending in syllables which could be grouped either as disyllabic or monosyllabic in an ambiguous way, e.g. *footnote bookworm or foot notebook worm*) and manipulated the prosodic characteristics of the first 5 syllables of these utterances (the F0 and duration separately, as well as F0 and duration at the same time). The task was to report the last word heard in the sequence. The result showed an effect of both F0 and duration separately and an increased effect of these features joined together. In the second experiment they removed 4 syllables from the beginning of each utterance in order to test whether the effect in Experiment 1 was found due to the immediately adjacent word's characteristics. This truncation reduced the magnitude of the 'distal prosodic effect'. In their third experiment using the same stimuli, they also showed that subjects were better able to recognise previously heard lexical items if the distal prosodic context matched the item rather than when there was a mismatch. As the authors say, these results together provide evidence in support of the perceptual grouping hypothesis and they interpret this to be evidence that what listeners

pay attention to is the overall rhythm of an utterance.

As well as this, Dilley and Pitt (2010) showed that the overall speech rate has an effect on whether or not listeners perceive function words. They achieved this by testing whether changing the rate of the overall sentence rate in English sentences with respect to the target function word (i.e. the duration of that words was kept the same) would affect whether or not it is perceived. From this experiment it was clear that timing information is also crucial for speech perception.

Further work using related paradigms has revealed that the overall rhythm of sentences also contribute to listeners' spoken word recognition in English (Morrill et al 2014). This was concluded from an experiment in which participants were presented with sentences in which the pitch pattern at the end of a sentence either matched or did not match the rest of the sentence. Here the participants were more likely to perceive a function word contained in the target part of the sentence when the pattern matched the preceding pitch rather than when it did not match it (i.e. H-L-L-H-L-L-H-L-L-H vs H-L-H-L-H-L-H-L-L-H). Another study which supports the perceptual grouping hypothesis conducted by Morrill et al (2014b) showed that word learning in an unknown artificial language is facilitated when learners are "exposed to syllable sequences in which the embedded words were congruent" rather than when they were incongruent with the distal prosody of a carrier phrase. As well as this, if the preceding context was shortened, the effect was weaker, suggesting that listeners pay attention to the overall rhythm of utterances and that this aids their learning.

These studies together show that prosody plays a great role in the way listeners segment speech as well as in how they learn languages. This suggests that listeners can and do very quickly entrain to those patterns and are able to use them for more efficient language learning. Relatedly, languages use these prosodic patterns in their own language-specific ways, e.g. English uses reduction more than other languages, which relates to its overall rhythm. Because languages largely exist in form of connected speech, these patterns also emerge best in connected speech, rather than e.g. in production of words in isolation. Connected speech is discussed in more detail in sections 1.4 and 1.7 below.

Overall, there is evidence for a variety of communicative functions of speech rhythm. These studies indicate that listeners pay attention to detail of not only the order in which segments unfold in time, but also to the stress of syllables relative to their neighbours as well as the overall sentence rhythm. As well as this, there is evidence that learning a new language can be aided by a more rhythmic organisation of pitch patterns (Morrill et al

2014b). These studies are of significance for this thesis, because the perceptual grouping hypothesis points towards the importance of the overall rhythm of the utterances listeners are listening to. In the authors' words, the distal properties of the speech signal matter for how well, or how quickly, the utterance is processed, such that the more rhythmically regular, be it due to duration, pitch, or a combination of these, the better the processing. This means, it is not only important to show the listener the difference between two neighbouring elements in the speech signal, but to bring their attention to the overall prosody of the utterance as a whole. This directly motivates some decisions made in this thesis, and will be brought up again in the relevant subsections.

1.3.11. Summary

This section provided an overview of what rhythm is and definitions of some of the key concepts that go with the notion of rhythm, i.e. beat and metre. It moved on to ways in which rhythm in speech has been researched by looking at the attempts at quantifying and describing it, and, after presenting the key issues with establishing discrete rhythm classes the languages of the world could be assigned to, it briefly attempted giving an answer to the big question of whether rhythm is a property of speech. It then showed what alternative, non-linear, approach to representing speech rhythm, and reviewed some of the key work in that area. This area is linked to Dynamic Attending Theory, i.e. synchronising our attention with an external stimulus. This kind of entrainment is easily achieved especially if entraining to a regular stimulus. Research (Quené & Port 2005) showed that such entrainment to a regular stimulus made the listeners' perception of speech easier.

Finally, this section also showed the main functions rhythm takes on in speech and why it is an important aspect of speech. As mentioned above, connected speech is where these various rhythmic properties of languages emerge and become important for listeners. Languages vary in ways they employ the phenomena that give them their specific rhythmic properties, and that is one of the reasons why learning (to comprehend) second language speech is difficult. As well as this, as suggested by Quené & Port (2005) and demonstrated by Morrill et al (2014b) connected speech comprehension and learning may be made more efficient if the utterances are organised more rhythmically.

1.4. Connected Speech

The main idea that will be tested in this thesis is that attending to speech rhythm will help EFL listeners to understand connected speech. This section reviews what is meant by

connected speech, and the processes by which the pronunciation of words—especially function words—is altered in connected speech. Research on training and teaching of connected speech and of rhythm in a foreign language setting will be looked at towards the end of this chapter.

Connected speech is "spoken language, when analysed as a continuous sequence, as in normal utterances and conversations" (Crystal 1980: 81). The key aspect of connected speech is the difference between linguistic units spoken as words in isolation, and when produced as a part of an utterance (Cruttenden 2001:266).

The way that syllables in an utterance are organised in terms of prominence is crucial to understanding how connected speech works. As already outlined in the section on rhythm in this chapter (section 1.2), pitch, loudness, quality and quantity are the key factors which make a syllable more prominent than its neighbours (Cruttenden 2001: 242). *Word stress* is the organisation of syllables in a word in terms of prominence; e.g. in the word *otter* there are two syllables and the first is more prominent than the second. There is also *sentence stress*, organisation of prominences within a sentence, e.g. in the sentence *I love monkeys but I love otters more!,* the word *more* would normally be made more prominent than other words, though other words may still have their own word-internal stresses. Sentential prominence is achieved mainly through pitch, but also through duration and amplitude. The alternation of weaker (less prominent) and stronger (more prominent) elements gives speech its rhythm, as discussed above; crucially, it also affects many aspects of the segmental pronunciation of words in connected speech.

An understanding of the distinction between *content words* and *function words* is also helpful for understanding the phonetics of connected speech. Content words, which are also called lexical words, are verbs, adverbs, nouns, adjectives, and demonstrative pronouns. Function words are auxiliary verbs, conjunctions, prepositions, pronouns, relative pronouns, and articles. In speech, content words are more likely to be accented and function words are more likely to be unaccented (Cruttenden 2001: 249). Content words usually remain prominent to some extent, i.e. the vowel in their stressed syllable(s) is still a full vowel, even if there is no pitch prominence (2001: 252). Function words can be accented or unaccented. Accented versions of function words are called *strong* forms, in which the vowel stays unreduced. Unaccented forms of function words are known as *weak* forms, and may be quite different from strong forms in terms of segmental composition. For example, the word *and* in *otters and ducks*, would probably usually be pronounced as

[ənd], [ən], or even just [n̩] in normal conversational speech: these are the word's weak forms. However, if the speaker wanted to stress the fact that it is both otters and ducks that they were talking about, they might use the word's strong form, i.e. [ænd].

Connected speech processes are the types of segmental changes that the citation forms of words undergo when the words are produced in connected speech. Both prominent and non-prominent syllables, and both content and function words, can undergo various types of connected speech processes. That said, the most drastic changes are typically seen in syllables that lack prominence, and since function words are more likely than content words to be produced without prominence, unstressed (weak) forms of function words are often particularly affected. They may "show reduction of the length of sounds, obscuration of the vowels towards /ə, ɪ, ʊ/, and the elision of vowels and consonants" (Cruttenden 2001: 252). Connected speech processes are tightly linked to the rhythm of a language: it is rhythmically weaker syllables that are most likely to undergo connected speech processes, while at the same time, the application of these processes can itself give rise to differences in perceived rhythm, e.g. when weak syllables are dropped from an utterance altogether.

1.4.1. Types of connected speech process relevant for this thesis

For the present thesis, some connected speech processes are more relevant than others. The discussion below focuses on those that featured in the experiments. For a fuller treatment, see e.g. Cruttenden (2001).

Reduction is a process in which phonemes are changed, minimised, or eliminated. Both consonants and vowels may be reduced (Simpson 2013). Vowels may be reduced by taking on a shorter duration, or by taking on a less peripheral quality. Consonants may be reduced in terms of duration, i.e. made shorter, and/or by a reduction in the extent of movement of the articulators. This can result in a different manner of articulation, e.g. a bilabial fricative [β], rather than a bilabial plosive [b], if the closure is not complete (Simpson 2013: 274).

As Simpson (2013) says, the most extreme form of reduction is *elision*, i.e. when a phoneme is completely eliminated, e.g. the word which in its citation form would be /'laɪbrərɪ/, often turns into /'laɪbrɪ/ in connected speech in English. As Ernestus and Warner (2011: 255) note, connected speech phenomena occur with different phonetic gradience, i.e. they span a wide range. This means that a vowel, or even a full syllable, can be reduced completely, i.e. elided, or only subtly weakened by employing only a slight reduction of

duration, or e.g. centralisation in the vowel space for vowels, or absence of a burst for a plosive consonant (Ernestus & Warner 2011: 255).

Another broad class of connected speech process involves sounds changing to become more similar to an adjacent sound. Such processes include *assimilation* and – of most relevance for the present thesis, *coalescence,* where two phonemes *merge* into one phoneme, which is different from the two that it resulted from. E.g. *would you* in connected speech can be produced with [dʒ] at the boundary between the two words, rather than in the sequence [dj].

Other types of connected speech phenomena, such as *liaison* and *intrusion,* are not directly relevant for the present thesis and are therefore not discussed here, but see Cruttenden (2001: 294, 343) for an account of them.

1.4.2. Connected speech processes in unstressed function words and affixes

The experimental stimuli used in this thesis included common English function words and affixes that are prone to various types of connected speech processes when they are unstressed. In this section, the types of CSP that these function words and affixes undergo are briefly presented, along with a discussion of the ways these might give rise to the potential for ambiguity for listeners, especially for L2 listeners unfamiliar with the specific ways connected speech processes operate in English.

1) The definite article *the* can undergo lenition, such that its dental fricative can be produced without frication. For example, in a phrase like *in the,* it may be realised simply with dentality on the preceding nasal i.e. as [in̪:ə] (Manuel 1995). *The* therefore has the potential to be confusable with the indefinite article, *a*.
2) The pronouns *he* and *she* can undergo elision: in particular *he* can be pronounced simply as [i] (with no [h]), and *she* simply with postalveolar frication (with very little vowel; Baker, 2008).
3) The *is* and *was* forms of the verb *to be* can undergo elision and/or coalescence. In particular, a possible weak form of *is* is simply [z]; the weak forms of *was* are theoretically [wəz] or [wz], but in extreme cases, *was* can be realised simply as [yzʷ] or [zʷ]. This can create the potential for perceptual ambiguity between *he is* and *he was*, if *he is* is produced as [iz] and *he was* as [yzʷ].

4) Negative contracted forms of auxiliary verbs, such as *can't* and *shouldn't*, can undergo reduction. The [t] in *can't,* or the [nt]-portion of *shouldn't* can be reduced to a great degree, and this cause miscomprehension and confusion between *can* vs *can't* (in dialects such as Scottish English where both have the same vowel), and *should* vs *shouldn't*.
5) The personal pronoun *her* can undergo a variety of reduction processes, e.g. both the initial [h] and the final [ɹ] can be reduced to a smaller or greater extent, the greatest being complete elision of one or both of them. If both get elided, it would leave a schwa on its own, or a slightly rhotacised schwa. This could lead to confusion of the pronoun with an article, e.g. *a*.
6) In the participle present ending *–ing,* the velar nasal [ŋ] can get substituted by an alveolar nasal [n]. This results in some cases in the verb sounding almost identical in present and past participle, e.g. *taken* and *taking* (when pronounced as something similar to [tɛɪkən] vs [tɛɪkɪn], respectively).
7) For prepositions in their weak forms, e.g. *in* or *on*, the vocalic parts are often elided or nearly-elided, such that e.g. the preposition *in* contains a very weak [ɪ], or the nasal only.

1.4.3. Summary

To sum up, speech undergoes a variety of processes in natural everyday communication. These processes can change the phonetic shape of speech. These changes can vary from being only very slight to being major. However major they are, native speakers do not normally struggle with these, if given context (Ernestus & Warner 2011: 255). Non-natives however struggle to recognise the highly reduced forms without context, e.g. they struggled to identify *he already* pronounced as [iɹɾɪ] out of context (2011: 255). As Ernestus and Warner (2011) say, comprehension of speech when it has undergone such connected speech processes is even more difficult for non-native listeners, who have not been told in their EFL classroom, that [iɹɾɪ] is a possible pronunciation of *he already*.

1.5. Overview of the study of second language speech learning

Before looking into acquisition of connected speech specifically in an L2, a brief overview of the study of L2 speech learning will be given along with an overview of the most influential models of L2 speech acquisition.

Our ability to learn second language speech as adults is relatively poor (e.g. Strange & Shafer 2008: 153, Escudero 2007: 109, Mennen & de Leuw 2014), even if we consider individual differences, which can lead to either extremely high or extremely low success. Much research concentrated on the factors affecting our ability to achieve native-like speech. Piske, MacKay & Flege (2001) provide a review of the research into this area. As they say, the factors that have always been considered in previous studies have been age of L2 learning, length of residence in an L2-speaking country, gender, formal instruction, motivation, language learning aptitude and amount of native language (L1) use. Piske et al (2001) found that indeed age of L2 learning had an effect, however, the amount of continued use of L1 also had an effect on the degree of foreign accent. The sole amount of time spent in the L2-speaking country did not have an effect, and neither did gender or the self-estimated ability in the native language (Piske et al 2001). As Strange (2007: 7) says, some studies in this area (e.g. Flege, Munro & MacKay 1995 and Flege, Yeni-Komshian & Liu 1999) show that age of arrival "strongly predicts strength of perceived foreign accent". Apart from age of arrival, a few other factors are important here, such as the "quantity and quality of L2 input" (Strange 2007: 10), which have to do with language experience rather than purely the length of time spent in the country where the target L2 is spoken.

Other possibly important factors mentioned by Piske et al (2001) that could have an effect were an inherent language learning aptitude, i.e. a "special aptitude for producing unfamiliar speech sounds" (2001: 202), and musical ability (2001: 202). However, as they say, at the time there was no evidence from the very few studies to show that either of those variables had affected the degree of L2 foreign accent (2001: 202). The role of musical aptitude/training is an area with a considerable amount of research and is therefore discussed separately and in more detail in section 1.6 below.

The main concern of each of the three experiments in this thesis is, however, the role of training. Therefore after a brief introduction to the main theories on second language

speech learning, training will be discussed in some more detail in sections 1.5.1 and 1.5.2 below.

1.5.1. The Second Language Speech Learning Models

The most influential models in the field of foreign language speech sounds acquisition are the Speech Learning Model (SLM) and Perceptual Assimilation Model (PAM). There is a very large number of studies looking at acquisition of segments in a foreign language but the number of studies looking into suprasegmentals is much smaller. Therefore, only broad patterns in the SLM and PAM models' evidence base when it comes to segmental categories will be outlined, and more attention will be paid to studies focusing on prosody in later sections (1.5.3.3 and 1.7).

1.5.1.1. Perceptual Assimilation Model (PAM)

Best's (1994, 1995) PAM proposes three ways non-native segments can be assimilated into an L2 learner's phonemic inventory, by trying to use the similarities and dissimilarities the new segments would have with those in their native language. Best argues that these similarities and discrepancies are perceived in terms of articulatory gestures, i.e. that the listener will not perceive the differences between the native and non-native sounds if they perceive them to be articulatorily similar to their native phoneme (1994: 190). This would then lead to assimilation of the non-native sound to the phoneme that is closest to it. If however the listener does not perceive a correspondence between the articulatory-gestural properties of the native and non-native sounds, then they have perceived the difference and no assimilation takes place. However, Best also notes that not all assimilation is the same, i.e. even if a non-native sound is assimilated into the native category to which it is most similar, the differences can still be recognized by the listener and the non-native sound can still be perceived as "less than nativelike" (1994: 190). Thus, there are three ways in which non-native sounds can assimilate into native sound categories based on how they are perceived. As well as accounting for how individual non-native sounds are assimilated perceptually, Best also considers the consequences for non-native *contrasts.* There are also ways in which non-native contrasts could assimilate into native contrasts, based on how close or deviant these contrasts are from the existing native contrasts. In other words, depending on how well the non-native sounds can assimilate to native categories, non-native contrasts will then be discriminated differently, and this ranges from perfect to poor discrimination (1995: 195).

Best conducted a number of studies on non-native consonant and vowel contrasts, upon which she supports PAM (to name a few: Best 1995, Harnsberger 2001, Polka 1991, 1992, 1995 Polka et al 2001, Polka & Bohn 1996, Best & Hallé 2010, Hallé & Best 2007, Hallé et al 1998, Tyler et al 2014). However, not many studies have so far concentrated on the relevance of this model for L2 rhythm acquisition.

Two studies that have aimed to investigate aspects of L2 prosody acquisition in the light of PAM directly are Nagao et al (2003) and Alexander and Wang (2016). Nagao et al (2003) tested Japanese learners of English on the identification of non-native syllable structures and voicing in non-native syllables presented at a slow and fast rate. The study found that the Japanese listeners could correctly identify non-native syllable structures, concluding that the Japanese listeners' performance was accurately predicted by the PAM, and that PAM can be applied at a prosodic level as well as segmental a segmental one.

An unpublished study by Alexander & Wang (2016) has recently investigated ways in which Cantonese, Thai, Mandarin, and Yoruba listeners identify CV syllables bearing the six phonemic Cantonese tones. They found support for PAM for Suprasegmentals which they based on their experiment that showed that listeners were able to assimilate novel tones L1 tone categories by attending to those cues which were relevant to the phonetic and phonological properties of their L1s (2016: 2161). Therefore, there is research suggesting that the strong evidence supporting PAM for segments can also be applied to suprasegmentals, but this topic has been little developed to date. For more studies on prosody in a foreign language, see section 1.7 below.

1.5.1.2. Speech Learning Model (SLM)

Flege developed the Speech Learning Model (SLM) which aimed to account for how some individuals learn – or fail to learn – to accurately produce and perceive phonetic segments (i.e. vowels & consonants) in a second language (L2) (Flege 2005). Flege's SLM is based on five basic ideas, outlined in Flege (2005). The first is that phonetic properties of L2 speech sounds can be perceived accurately by L2 learners if given enough time. The second is that L2 learning takes time, and is affected by the nature of input received, and the third is that production depends on perceptual representations. The fourth key idea is that the ability to acquire languages which we have as children remains intact and accessible across the life span, and the fifth is that sounds which our L1 and L2 use exist in a common phonological space and therefore they mutually influence each other. Based on

these assumptions Flege formed hypotheses on how particular L2 sounds will be learned, depending on factors such as the perceptual closeness of the new L2 sounds to the learners' L1 sounds, and the learner's stage of linguistic development. Many studies have tested and found support for these hypotheses (e.g. Flege 1987, Flege 2005, Flege and Eefting (1988), Flege & Efting (1986), Flege & MacKay (2004), Mora et al (2011 & 2015), Flege & MacKay (2011). Some of them took into account such things as e.g. language input (Flege & MacKay 2004, Flege & MacKay 2011) and patterns of language use (Mora et al 2011 & 2015). However, the studies on SLM do not focus on acquisition of prosodic L2 features, and are therefore not discussed in detail in this thesis.

1.5.2. SLM vs PAM?

Both the SLM and PAM are models which aim "to account for the difficulties listeners have with sounds from a language other than L1" (Strange 2007: 6). However, there are some major differences between them. One is that PAM focuses on non-native speech perception (i.e. unfamiliar non-native distinctions) in naïve listeners, whereas SLM tries to account for phonetic L2 acquisition by more experienced learners over time.

As Strange points out (2007: 37) both SLM and PAM are looking at cross-language phonetic similarity as a concept that "plays a central role in determining how well L2 learners will perceptually differentiate non-native speech". However, SLM describes L2 speech production, based on the perceived auditory-acoustic differences and similarities between languages, while PAM argues that it is the articulatory gestures that the L2 learner perceives and does not directly address L2 speech production.

Generally, they both use the perceived relationship between sounds in L1 and a target L2 as an explanation for the processes which drive L2 production and perception. Best & Tyler (2007) revised PAM so that it would also take more experienced L2 learners into account, as does SLM. This model was called PAM-L2. To develop PAM-L2 Best & Tyler (2007) reviewed and evaluated SLM postulates from a PAM perspective. They agree that the ability to perceive L2 sounds remains in an adult learner, but they say that it is the gestures rather than the acoustic-phonetic categories that are perceived. In other words PAM does not assume mental representation of the phonetic categories, but rather direct perception of the articulatory gestures. Best & Tyler (2007) also state that PAM-L2 sees as phonetic categories as the sub-lexical information that does not distinguish between

different words (which PAM sees at the phonological category), but would let an experienced L2 listener know about the speakers' identity, region, etc.

According to PAM then, the L2 learner continues to refine their perception of speech gestures rather than categories throughout their lifespan. However, in PAM-L2 listeners may assimilate L1 and L2 sounds, i.e. treat them as phonologically equivalent, but still perceive a phonetic difference between them, e.g. between a French and an English /r/. As far as contrasts between the categories are concerned, the refined version of PAM uses both phonological and phonetic levels of L2 perceptual learning. It would assume that L2 and L1 categories exist in a common space, but that the listener is able to keep the phones separate in that space, as different phonetic realisations of the same category.

While the present work does not concern itself directly with testing any of the hypotheses of these models, and while L2 connected speech learning may be quite different to learning L2 categories, some of the key principles from these models may still apply. For example, the quality and quantity of input (i.e. how long have they been studying their L2 for, and at which point of their life they began learning their L2) could be a significant factor affecting how well a learner learns to comprehend L2 connected speech. As well as differences in segmental categories, it may matter what difference there is between their L1 and L2 in terms of rhythm, e.g. whether the languages are traditionally stressed-timed or syllable-timed could play a role. Some of these factors are taken into account when designing the experiments in this thesis.

1.5.3. L2 speech training

An aspect of previous research on SLA study that is directly relevant for this thesis is training. Training is a widely studied area of SLA. This study of training effects is concerned with the usefulness of various training methods, i.e. how well can L2 learners learn the various L2 aspects using different types of training. Training can concentrate either on the learner's perception or production of their target language speech. Another way in which different kinds of training can vary is whether the training targets the segmental or the suprasegmental aspects of the target language. As well as this, training can provide explicit instruction to the learner on what is being trained, or it can attempt to make changes in a more subtle way, without explicitly stating the purpose of the training. The elements relevant to the present work are perceptual training and suprasegmental training.

1.5.3.1. Production vs perception training for SLA

Being able to produce L2 sounds as they are produced by the native speakers of that language is a goal of many L2 teachers and students. Since the current work is not concerned primarily with training the leaners' L2 speech production abilities, but with their perception, only those studies which point to links between L2 perception and L2 production will be reviewed. As Munro and Bohn say "studies of L2 production are [also] indispensable in the development of a full understanding of speech learning phenomena" (2007: 9). A large amount of studies have been conducted in this area. To name a few, Adank et al (2010) and Bradlow et al (1997) showed that perceptual training improves production, but other work suggests that articulation training can also improve production, possibly before perception improvement occurs (Wilson & Gick 2006).

Bradlow et al (1997) showed that there was an "improvement in the Japanese trainees' /r/-/l/ spoken utterances as a consequence of perceptual training" (1997: 2299). The training their participants were given consisted of 45 sessions, over a period of 3–4 weeks, of perceptual identification with feedback. Their pre-test and post-test productions were compared by English native speakers using two methods: rating of these /r/ and /l/ productions and identification of these productions via a two-alternative minimal pair test (1997: 2299). The same type of training also showed the improvement in other studies in adults (Bradlow et al 1999, Iverson et al 2005), as well as in children (Shinohara & Iverson 2015).

In Adank et al's (2010) study, groups of native Dutch subjects were taught an unfamiliar made-up accent of Dutch using such methods as listen only, listen and repeat using their own variety of Dutch, listen and repeat imitating the target accent, listen and repeat without hearing their own productions and no training. The post-test measuring the subjects' comprehension showed improvement only in the group who imitated the unfamiliar accent.

On the other hand, production training using ultrasound to train the Japanese learners on the articulation of the English /r/ and /l/ sounds showed improvement of production without improvement in perception. Gick et al (2008) conducted a small pilot study in which they used three native speakers of Japanese learning English as a foreign language. They assessed their ability to produce the English /r/ and /l/ sounds in pre-test and gave them training using their own ultrasound images (in the best and worst version) and let

them compare these to the native speakers' productions in a 30 minute session). Here, the learners were able to see directly what the position of the tongue was and compared it to the position of the native speaker's tongue when producing target sounds. In the post-test, all three subjects were able to produce the sounds successfully (2008: 319). Wilson and Gick cite two studies (Goto 1971 and Sheldon and Strange 1982) which "found that at least some Japanese adults had more accurate production of the difference between English /r/ and /l/ than their perception of that difference in their own speech or that of others" (2006: 151). They conclude that using this method, it may also be possible to train speakers to produce a consistent difference between English /r/ and /l/ before they are able to perceive the difference (2006: 151).

As Escudero (2007) says, the majority of the studies point towards the priority of perception in L2 phonology over the production, i.e. that perception develops first and that the difficulties an L2 learner has with production have a perceptual basis (2007: 111). As well as this, there is some more recent evidence suggesting that production accuracy improvement lags behind perceptual accuracy in L2 learners (Nagle 2018).

1.5.3.2. Perceptual training for SLA

As well as the likely priority of perception over production in L2 learning, there are some more recent, general views that should be considered. While the ability to produce native-like sounds still is most certainly a goal for at least some L2 learners (and teachers), the more recent suggest that being comprehensible to listeners would be the desirable outcome for second language learners, rather than being able to produce sounds "perfectly", or "like a native speaker" (e.g. Jenkins 2000, Levis 2005, Walker 2010). The comprehension of L2 speech could possibly, therefore be seen as more important to the leaners' communicative success in everyday interactions with native speakers of the target language. Therefore, some of the main techniques that train L2 learners' perception along with studies that have used them are reviewed below. The two most researched types of training, though not mutually exclusive, are the High Variability Perceptual Training and Perceptual Fading.

1.5.3.2.1. High Variability Perceptual Training

High Variability Perceptual Training (HVPT) is a kind of perceptual training "which involves having subjects give identification judgments with feedback for natural recordings of words produced by multiple talkers, with target phonemes in multiple syllable positions" (Iverson et al 2005: 3268).

Logan et al (1991) investigated the role of a specific technique aiming at training Japanese learners of English to distinguish between the /r/ and /l/ phonemes. This specific training was minimal pair identification which utilised multiple natural exemplars, i.e. they used six different speakers to produce their stimuli and a variety of phonetic environments in which /r/ and /l/ occur. They found that their participants' overall identification accuracy showed an improvement after training. They concluded that these results demonstrate the importance of stimulus variability. However, this study did not directly compare the results of training using stimuli that came from various speakers with results that came from a training that used one speaker only.

Therefore, to investigate this further, Lively et al (1993) conducted a study in which they compared the effects of training leaners using either one or multiple speakers. They found that both groups improved in the post-test, but only the group who heard multiple speakers in the training phase were able to generalise to both new words produced by a familiar talker and to novel words produced by an unfamiliar talker (1993: 1242).

The studies mentioned at the start of this section (Bradlow et al 1997, Bradlow et al 1999, Iverson et al 2005, Shinohara & Iverson 2015), also investigated whether Japanese learners' production of the English /r/ and /l/ can also benefit from HVPT training. They concluded that the Japanese learners' ability to correctly identify and to produce the English /r/-/l/ contrast can be improved after receiving HVPT.

1.5.3.2.2. Perceptual Fading

Perceptual Fading is another training technique which was first used, though not called by this name by Terrace (1963), to teach parrots colour discrimination. It was then adapted by Jamieson and Morosan (1986) as a training method for the voiced and voiceless fricative sounds /ð/ and /θ/ to adult francophone learners of English in Canada. The training itself involves a continuum of voiced to voiceless sound recordings, manipulated so that the sounds at each end of the continuum were either extremely voiced or extremely voiceless, and there would be a range of voicedness/voicelessness in between these extremes. Participants would take part in a pre-test which involved identification as well as discrimination between these sounds, and after a training period, they would be asked to perform a post-test. The results from the experimental group were compared to a control group and showed that this training improved the Perceptual-Fading-trained groups'

identification and discrimination ability in both natural and synthetic speech task. Later on, Jamieson and Morosan (1989) also tested whether this method is as effective when participants are tested on novel words, i.e. whether generalization occurs. This proved successful, however, it was not as successful in teaching the participants the /ð/-/d/ distinction.

The above studies used stimuli which had a combination of one vowel and then the consonant (i.e. CV). Jamieson & Moore (1991) therefore, searching for increased generalizability for this type of training, used VCV clusters in order to provide the listeners with more phonetic environment for the occurrence of the target sounds, as they would in natural speech. Following this training, the results showed improvement and generalization on post-test also took place.

1.5.3.3. Training for segmental vs suprasegmental L2 features

Most of the studies reviewed above investigate the learning of difficult L2 segmental contrasts. However, other aspects of second language speech also cause difficulties for the L2 listener. These are be prosodic, or suprasegmental aspects, composed of such elements as e.g. intonation, durations, rhythm, stress and tempo and are important for the present work. As Mennen & de Leuw (2014) say, these are aspects that are often extremely difficult to acquire for L2 learners. Prosodic features, perceived as duration, pitch and loudness are used by native listeners for a variety of purposes, from extracting the information on e.g. turn ends, or distinguishing certain words from one another, e.g. *forebear* vs *forbear*, which would sound the same if there was no different in stress placement (Mennen & de Leuw 2014), to chunking information, e.g. in a sentence such as *When you learn gradually you worry more* – whether the learning or the worrying is happening gradually depends on when the speaker paused, which syllable they lengthened and when they changed pitch (Mennen & de Leuw 2014), as well as what type of syllable structures are permitted. Durational differences have been widely studied with a variety of different rhythm metrics and this aspect of speech rhythm has been discussed in section 1.3.4 above.

As well as this, as stated above, various languages can use these different features, in different ways for different purposes, causing further difficulties for second language learners. As Mennen and de Leuw (2014) say, studies such as Altmann (2006), Dupoux et al (1997), Dupoux et al (2008), and Tremblay (2008) revealed that these difficulties in the learners are often "perceptually motivated" (Mennen & de Leuw 2014: 188). However,

there are also studies that suggest that these difficulties are not exclusively due to perceptual struggles (Archibald 1997 & Altmann 2006).

Some studies looking into training of connected speech and speech rhythm in L2 are reviewed in section 1.7 below. Before that, though, another area of interest, Computer Assisted Language Learning, is briefly reviewed below.

1.5.3.4. Computer Assisted Language Learning

Computer Assisted Learning is a growing part of language learning. This area is very briefly looked into below because of the collaborative nature of the current PhD project, i.e. the collaboration with Micro-phonics (http://micro-phonics.com). A theoretical point that is relevant to this area is that this kind of self-study tool enables the learner to learn on their own in the absence of a more direct personal training.

Micro-phonics makes use of computer assisted learning which combines testing (such as frequent formative testing) with some features of one-to-one learning, immediate feedback, without being as costly as one-to-one tuition. This is achieved through computers allowing personalisation, and a personalised curriculum can be better tuned to the individual student's needs. Therefore, Micro-phonics' *Pronunciation Coach* involves self-paced quizzes, which have also been proven to be an important learning tool, not just a tool for assessment (cf Larsen et al 2008, Roediger et al 2011).

Computer Assisted Learning (CAL), can be very broadly defined as "the learning procedures and environments facilitated through computers" (Schittek et al 2001). As Schittek et al (2001) further state, there can be two types, or levels, of interaction involved here, one is where the learner interacts with the content of the material to be learned and another one is where the computer is used to interact with other learners or tutors (members of the learning community) (2001: 93). Here, we are considering Computer Assisted Language Learning (CALL) specifically.

Even though language is a specific area for a learner to study, the definition of the concept can still be quite broad, similarly to the definition of the more general CAL, as Beatty (2003: 7) puts it, it "is *any process in which a learner uses a computer and, as a result, improves his or her language*''. This definition is purposefully broad, as Beatty (2003) further states, because the spectrum of current practice of learning a language using a computer is also broad and it allows us look at the whole its variety.

Some mention that one of the critique points of CALL has been that it is technology-driven rather than focussed on research, pedagogy and theory (e.g. Egbert and Hanson-Smith 1999). However, CALL has also been defended as being useful from the "interactionist perspective of second language acquisition" (Hubbard 2009: 3) as well as becoming a usual and normal form of learning, rather than something special. As Clifford says, "computers will not replace teachers, teachers who use computers will eventually replace teachers who don't" (1987: 13). Some important features of CALL are, e.g. autonomy in learning, meaning that the student can use a computer to study on their own, at their own pace, and can to revisit the material as many times as required (Beatty 2003: 11-12). As well as this, it is now no longer linear, as it was at the very start, "requiring each learner to follow the same steps in the same fashion" (Beatty 2003: 21), but can be tailored to the individual student's needs or interests.

In 1984 Bloom published a paper reporting studies comparing conventional teaching (i.e. 30 students per class with periodical tests), with Mastery Learning (30 students per class with frequent formative tests and feedback) and Tutoring (one-to-one teaching). One to one tutoring was found to be most effective, i.e. "the average tutored student was above 98% of the students in the control class'' (1984: 4). Mastery Learning was also found to bring better results than the conventional teaching method; here the students "were above 84% of the students in the control class" (1984: 4). These results lead the researchers to conclude that all students have the potential to achieve excellence.

Larsen et al (2008) talk about two functions of testing, one being the traditional method for assessment and the other one being aiding retention of studied material. As they say, it has been shown that if students are being tested on material, they tend to remember the material better than when they are not tested (2008: 959). If students are being tested frequently, they are forced to study in more spaced out sessions rather than having periods of intense study but all at one time. They conclude that "[m]uch research shows that spaced out study sessions aid memory performance" (2008: 960). They call this test-enhanced learning.

Roediger et al (2011) conducted experiments to test whether quizzing is more effective than not quizzing and whether quizzing is more effective than being re-exposed to the material by reading it. These experiments showed that quizzing improved students' performance and that "the rereading condition did not differ from the nontested condition" (2011: 391). In their final experiment Roediger et al (2011) sought to answer the question

whether learners would benefit from self-testing rather than being tested in a classroom environment and whether quizzing practice would also have a positive influence on the students' short answers, which is a different kind of testing than a quiz. Students were encouraged to use an "Internet-based quizzing Website outside of class to practice retrieval on their own" (2011: 391). The results showed that students who self-tested improved and that this improvement also generalised to the other kind of testing, which was writing short answers to questions.

In sum, there is a large amount of work on the role of CALL and its strengths and weaknesses. A full review is not attempted here, as the main focus of this research is on the role of rhythmic attending, not on the mode of delivery. That said, there is at least some evidence to suggest that having test enhanced learning with feedback also enables learners to correct errors and gives them the opportunity to process and absorb information without embarrassment or judgment from a human teacher, which would be in line with the Classroom Assessment Environment theory (Brookhart 1997: 329). As well as this, practice is easier with a computer because a computer does not tire of presenting material multiple times or calculating a grade for the same quiz repeatedly.

1.5.4. Summary

In summary, various types of training have shown promising results for both perception and production of L2 sounds. For training perception, using High Variability Perceptual Training in order to increase generalisability, or Perceptual Fading technique, which uses extremes on a continuum and incrementally decreasing the difference, are some examples. An additional training method would be musical training, which as many studies showed, is also beneficial in many aspects of L1 as well as L2 learning. This aspect is discussed in more detail in the following section.

As mentioned above, there is a large amount of studies which investigate L2 segmental learning, however, the suprasegmental aspects of L2 speech learning has not yet been studied by many. This area is looked into more detail in section 1.7. An area that has been briefly looked into was CALL which is of interest because of the collaborative nature of this project, the collaboration being with Micro-phonics Ltd.

1.6. The speech-music link in L2 learning

The link between speech and music is also of interest to this thesis. As Patel (2008) says, there are differences and commonalities between speech and music. Perhaps the main

commonality is that both speech and music are sound systems that (healthy) humans are born with. As well as this, elements of both are learned to be perceived as the categories that our environment, or native culture, perceives them (Patel 2008). The common elements in both speech and music are the use of pitch and timbre, and both speech and music unfold in the temporal domain with a variety of durations according to a variety of rules, and therefore both (for some - arguably, in the case of speech) use rhythm. These are all perceived and learned in specific ways by the human brain. Therefore, as Patel (2008: 79) says "[i]f there is overlap in the mechanisms that the brain uses to convert sound waves into discrete sound categories in speech and music, then it is conceivable that exercising these mechanisms with sounds from one domain could enhance the ability of these mechanisms to acquire sound categories in the other domain." Patel (2008) reviews two studies which show such links in both children (Anvari et al 2002) and adults (Slevc and Mikaye 2006), some more detail of both of which will be given below. A related point, though from a slightly different angle, would also be that, if there are shared mechanisms in the brain for processing speech and music, then having a good ability in one, or – in fact – training one, could increase the ability in the other. Many researchers have tackled this idea from a variety of perspectives.

1.6.1. Role of musical ability in second language learning

Phonological awareness is linked to reading ability because reading requires mapping the written code onto its phonemic equivalent (Anvari et al 2002). Anvari et al 2002 found that music perception skills, and especially pitch processing rather than rhythm processing in 5 years olds were linked to the children's phonemic awareness skills (2002: 120). In the 4 year olds, rhythm production and rhythm discrimination tasks both correlated with the musical pitch tasks and with reading. Therefore, as far as rhythm skills are concerned, this study was not as clear as with the children's pitch skills. However, the main conclusion of the study was that music perception skill predicts reading.

Slevc & Miyake (2006) conducted a study aiming to investigate whether musical ability can be the reason why there are individual differences in adult L2 learners' phonological ability. They studied Japanese learners of English and included a wide variety of factors and found that perception and production of L2 sounds was better in those who performed better at analysing, discriminating, and remembering simple musical stimuli (2006: 679).

A review of studies which look at the potential role of music in second language learning

has been provided by Zeromskaite (2014). For the following section, some of the studies reviewed by Zeromskaite, and some additional studies, will be looked at in more detail. She divided her literature review into the four themes according to which areas of L2 music can enhance. These areas are tonal variations, processing of utterance duration, pronunciation and L2 comprehension. These themes are discussed here in turn. An additional area of musical training and language disorders is also discussed. Therefore, the sub-themes here are Tonal Variation (1.6.2), Perception of L2 speech timing cues (1.6.3), Pronunciation of L2 segments (1.6.4), L2 comprehension (1.6.5), and Musical training vs Language disorders (1.6.6).

1.6.2. Tonal Variations

As far as processing of tonal variations is considered, Gottfried (2007) investigated this aspect of the effect of musical training. He conducted two experiments with English learners of Mandarin. In the first experiment musicians' and non-musicians' ability to identify sine wave tones (rising, falling, or constant) was tested in one task and their ability to identify Mandarin tones was tested in another task. In both tasks musicians performed significantly better than the non-musicians.

In the second experiment musicians and non-musicians were asked to perform a discrimination task (i.e. are the stimuli same or different?), rather than an identification task. As well as this, they were asked to perform an imitation task, in which they listened to the stimuli and repeated what they heard. Again, in both of these tasks musicians performed significantly better than non-musicians (Gottfried 2007: 234).

To sum up, Gottfried (2007) showed that those English-native learners of Mandarin who were musically trained were able to identify and imitate Mandarin tones better than those who were not musically trained. This is in line with the studies reviewed by Zeromskaite (2014), e.g. Marques et al (2007), Delogu et al (2008) and Marie et al (2011).

Marques et al (2007) showed that musicians were better and faster than non-musicians at detecting small pitch deviations in an unfamiliar language (subjects were French native speakers and the language was Portuguese). Marie et al (2011) tested French musicians and non-musicians on how well they could discriminate between tones and segments (consonant or vowel) in an unfamiliar language, Mandarin. Event Related Potentials (ERPs) were measured in the subjects and the results showed that musicians were more

accurate at detecting both tonal and segmental variations in a foreign and unfamiliar language. This finding contradicted Delogu et al's (2010) results, which showed that musicians performed better than non-musicians at tasks involving detecting tonal variations but not better on phonological ones.

In summary, these studies among others have shown that lexical tone variation detection is easier for musicians than non-musicians (Zeromskaite 2014: 79).

1.6.3. Perception of L2 speech timing cues

As Zeromskaite (2014) notes, utterance duration processing can also be enhanced by musical experience. This was shown, e.g. by Marie et al (2010). In this study, they asked musicians and non-musicians to listen out for either metric or semantic incongruency in sentences. The metrically incongruent sentences had the last two syllables lengthened as opposed to the last one, which is the usual case in their native language (French), while in the semantically incongruent sentences the last word was replaced so that semantically the sentences made little sense. The results showed that musicians made fewer errors in the metrically incongruous task than in the semantically incongruous task and that the musicians' reaction times were faster in the metric than in the semantic task. Comparing the behavioural data between musicians and non-musicians, the results showed that musicians outperformed non-musicians on both metric and semantic task. Marie et al (2010) also collected ERP data, which showed that "musical expertise increased both the automatic and controlled aspects of auditory processing" (2010: 303). From this, the authors concluded that musicians have a greater sensitivity to the acoustic information in both music and speech (Marie et al 2010: 303).

Sadakata and Sekiyama (2011) compared Japanese and Dutch native musicians and non-musicians performance in two perceptual tasks: discrimination and identification of timing and quality of Japanese consonants (singleton and geminate), and quality of Dutch vowels. Stimuli were used for both groups to test their performance in L1 and L2. In the first experiment, the results showed that musicians responded more accurately and quickly in two-alternative forced choice test than non-musicians, in both L1 and L2. In their second experiment they used a discrimination task. Here, both Dutch and Japanese musicians also performed better. The authors therefore concluded that musical training also enhances processing of consonant timing information and spectral information, which in their case was the quality of vowels and nasals (Sadakata & Sekiyama 2011: 6), and that "there is a

strong link between musical training and accuracy of the identification of timing features of linguistic signals, for example, musicians may be able to represent the incoming acoustic timing features of speech signals in an automatic and stable manner" (2011: 9).

François et al (2012) conducted a longitudinal study over 2 years with children and tested in pre- and post-test whether musical vs painting training aided their development of speech segmentation in their L1, and more specifically their ability to extract meaningless words from a continuous flow of nonsense syllables (François et al 2012: 2038). They measured the responses using both behavioural (two-alternative forced choice test) and electrophysiological (EEG) methods. The results showed that speech segmentation skills improved in the music group only (2012: 2038).

Herrera et al (2011) conducted a similar study with children. These children were either native Spanish or native Tamazight speakers. The stimuli were all in Spanish. The training the children were given included either phonological plus musical training, or musical training only, or phonological training only. The pre- and post-test were designed to test the children's phonological awareness and naming speed. The phonological training taught the children to identify sounds (in all possible word positions), and segment words into syllables, etc (2011: 73). The results of the study showed that "phonological training with and without music [was] effective in improving two predictors of reading readiness in young children: phonological awareness and naming speed" (2011: 77). However, they add that children who received combined musical and phonological training were better at the post-test which looked at their ability to identify sounds at word endings.

Generally, results of these studies indicate that music training is beneficial for perception of speech sounds and speech timing in an L1 as well as L2.

1.6.4. Pronunciation of L2 segments and musical ability

Links between pronunciation of L2 segments and musical skills have also been investigated. Posedel et al (2011) looked at pronunciation of Spanish by English native speakers. Correlations were found between musical training and pitch perception, as well as musical training and working memory. Pitch perception in turn was found to be a significant predictor of how well the subjects would produce Spanish sounds.

Milovanov et al (2008) study looked at Finnish children learning English and compared the musical aptitudes of those who had superior English pronunciation with those with worse English pronunciation skills. They first measured the children's musical aptitude and

collected data on their listening discrimination ability for both chords and phonemes. All children were they given an 8-week pronunciation training "to ensure adequate amount of pre-training in pronunciation" (2008: 83). They then performed a post-test on the listening discrimination of chords and phonemes and also collected ERP data to measure the children's sound processing accuracy. The more advanced pronunciation group outperformed the less advanced group on both phoneme and chord discrimination tests. As well as this, the ERP data supported the hypothesis that children with good pronunciation skills showed "more pronounced sound-change evoked activation with the music stimuli" than the children with worse pronunciation skills. As the authors say these "results imply that musical and linguistic skills could partly be based on shared neural mechanisms" (2008: 83).

A more recent study by Milovanov et al (2010) compared Finnish adult learners' of English pronunciation of English, specifically phonemic discrimination of English phonemes and musical aptitude. The results of their previous study on children were consistent with the results of this study as well. The main finding of their study was that "participants with higher musical aptitude were able to pronounce English better than the participants with less musical aptitude" (2010: 56).

In sum, there are studies showing that L2 pronunciation is positively influenced by musical training/musical aptitude. The distinction, or connection, between musical training and musical aptitude, as it will later turn out may be of importance. This is discussed in more detail later, in Chapter 4. However, for the purpose of Experiment 1, this distinction will not be made.

1.6.5. L2 comprehension and musical ability/training

Finally, L2 comprehension was the last sub-theme of Zeromskaite's (2014) literature review. However, it should be made clear that here, the term 'comprehension' is different to purely semantic comprehension of speech. It means either being able to read (as in Herrera 2011), or being able to give coherent answers to questions that were asked (as in Swaminathan & Gopinath 2013).

Herrera et al (2011) showed that the two variables which were positively correlated with being musically trained were phonological awareness and naming speed regardless of whether Spanish was their L1 or L2. Since these are two predictors of reading readiness (2011:69), this result was taken to demonstrate that musical training improves reading

ability. Swaminathan and Gopinath (2013) researched the question whether musically trained children performed better than untrained children on various L2 English tasks. They found that being musically trained had a positive effect for these children's L2 comprehension and vocabulary skills.

In short, these studies therefore showed that being musically trained enhances L2 learning aspects relating to phonological and reading abilities.

1.6.6. Parallels with the role of musical training in other areas of language processing

Apart from literature on the correlations between musical aptitude and second language learning, there is also research on the usefulness of musical training and speech and language disorders and for dyslexia. This is important for the present work, because if musical training can be useful as therapy for speech and language disorders as well as dyslexia, then it may also be possible that musical training can be useful for training speech and language in healthy individuals who are learners of foreign languages.

To give some examples of such training for speech and language disorders, Fuji and Wan (2014) reviewed studies which suggest that rhythm-based therapy could be beneficial for treating a variety of disorders which affect speech and language, such as Parkinson's disease, stuttering, aphasia and autism. Fuji & Wan (2014) argue that synchronisation and entrainment to pulse would work as therapy by stimulating those brain networks that underlie human communication.

As well as the ones listed above, more direct research by Overy (2000: 218) showed that musical training focused on timing skills significantly improves dyslexic children's literacy and language skills (2000: 225). As Overy (2000) says, some theories concerning the causes of dyslexia proposed that these are problems with neurological timing, or temporal processing and that these problems can be overcome through training. Since being musically trained requires developing temporal processing skills, as Overy (2000: 220) says, music can provide the perfect medium for such training. Some previous studies also revealed that engaging in a motor activity, such as tapping along to syllables in a word (Thomson 1993) or using nursery rhymes (McLean 1987) are a helpful tool in improving dyslexic children's language skills. Overy (2000) conducted a study investigating rhythm, melody, tempo and timbre perception, metre and rhythm production, and note detection by children with a high risk of dyslexia, compared with children with no risk of dyslexia. The high-risk children were worse than the low-risk children at rhythm and at tempo perception

and at metre production as well as other temporal processing tasks (such as identifying the number of notes played within a short time), and at timbre discrimination (Overy 2000: 225). They were however, not worse than low-risk children at melody perception. Later, Overy (2003) investigated the effects of musical training for dyslexic children and compared the results of those who took part in music class for 15 weeks with those who did not. She compared their phonological, spelling and reading skills scores from a pre- and post-test. The result showed that dyslexic children's phonological and spelling (but not reading) skills can be improved using rhythmic training (Overy 2003: 497).

Relatedly, Przybylski et al (2013) tested whether children with developmental language disorders improve in their judgments of sentence grammaticality after listening to regular or irregular musical prime sequences. The experiment showed that not only the speech and language impaired or dyslexic children's judgments improved but also those of the unimpaired control children.

1.6.7 Summary of links between language and musical training/aptitude

One other area mentioned in Zeromskaite's (2014) literature review is second language speech comprehension. Musical aptitude generally seems to have a positive effect on a variety of speech and language related skills: recognising tone variation, L2 sound production, L2 reading abilities and speech sound quality and timing. The fact that being musically trained, or having a musical aptitude, is linked to being better able to recognize tonal variations is probably not surprising as the relationship of musical and speech pitch variations due to the melodic component of musical skill. Similarly, the rhythmic component would be linked to the processing of temporal information, whether in speech or music. In English connected speech, reduction processes affect the phonetic shape of words, especially function words, reducing their intelligibility for non-native listeners compared to words as they're heard by listeners in their citation form. There is a close connection between reduction and speech rhythm: metrically weak syllables reduce more, and may be cued only by phonetic detail that non-native listeners struggle to detect. Relatedly, as reviewed below (sections 1.6.1-1.6.5), musical training is beneficial for second language acquisition and rhythm-based training seems to be increasingly promising as a therapeutic means for a variety of speech and language-related, as well as other motor, disorders.

1.7. Connected Speech and L2

If connected speech can sometimes cause comprehension problems for native speakers (see review in 1.4 above), then both the comprehension and production of those forms by L2 learners are also likely to be affected. Research indicates that language learners have problems in producing as well as understanding connected speech (Brown & Kondo-Brown (2006: 5), and literature review in section 1.7.1 below). These comprehension problem are most likely due to the fact that connected speech processes (reviewed in sections 1.3.9 and 1.4 above), are language-specific. In other words, these comprehension issues would not occur if the same kind of connected speech processes were at work in the same ways in every single language of the world. As we know, however, that is not the case. For example, the intrusive /r/ is known to occur in some non-rhotic varieties of English, e.g. in RP (e.g. *law [r] and order*), but not in a number of other varieties, and not in other languages; vowel reduction is very common and happens to a great extent in English, syllable and vowel reduction are features of German, Dutch, but not of e.g. Polish.

1.7.1. Connected Speech Perception and Speech Rhythm in L2

In this section studies showing that L2 learners have difficulties with connected speech comprehension in a variety of L1-L2 combinations will be reviewed. Since connected speech processes are one of the factors that give languages their rhythms, studies on speech rhythm acquisition in L2 are also included in this section.

Henrichsen's 1984 study on reduced forms in listening comprehension, who used a slightly modified version of Bowen's (1976) Integrative Grammar Test (which will be discussed in more detail in Chapter 2 below), showed that there was a significant difference between native and non-native speakers' comprehension of reduced vs non-reduced forms. In a comprehension task, native speakers performed well in both the condition with and without reduced forms. The non-natives, however, performed significantly better if there were no reduced forms in their comprehension task (1984:117). Ito (2006) also showed that in absence of reduced forms the non-natives scored significantly higher. Nouveau (2012) showed that Dutch learners of French struggle with comprehension of the so-called 'unstable e' (which can exist as a full vowel, e.g. *Sax[o]foon* or be reduced to schwa), which improves when the word is known to the learner and which is affected by L1 transfer effects. Ten Bosch (2015) showed that L2 learners of Dutch from a variety of L1 backgrounds struggle with connected speech comprehension in Dutch.

Then the question arises of what the cues are that the learner uses (as opposed to the native speaker) that cause these issues. Several studies have looked at various cues that native vs non-native listeners use for connected speech comprehension. These are briefly reviewed below.

Tuinman, Mitterer & Cutler (2012) conducted a study on resolving ambiguity in familiar and unfamiliar speech, using the intrusive /r/ in RP (in such utterances as, e.g.: *Canada aided* vs *Canada raided)*. They found that in native listeners the phenomenon does not cause problems in the interpretation of the meaning. The difference between the intrusive /r/ and word initial /r/ in English is that intrusive /r/ is shorter than a word-initial /r/ (2012:532) and native speakers are able to use this cue to distinguish between them. They conducted experiments with Dutch listeners as the non-native L2 participants. In both experiments participants listened to ambiguous phrases and choices as to which words were presented, e.g. *raided* or *aided*. Here, the English-native listeners used the phoneme duration as a cue for word boundaries, but the non-native listeners did not (2012:539).

Following Tuinman et al (2012) who showed that the non-native listener has more difficulty with a process that does not occur in their L1 (in line with Best and Tyler's 2007 Perceptual Assimilation Model), Mitterer & Tuinman ask the reverse question, i.e. whether L2 connected speech processes are easier for the learner if they exist in the learner's L1 (2012:2). They looked at/t/-reduction in Dutch, and how it is perceived by German learners of Dutch.

They first established that /t/-reduction is more likely after /s/ than after /n/ in both languages. However, there is a morphological difference between the two languages, namely in Dutch this is true for both nouns and verbs, but in German verbs the /t/-reduction did not depend on the preceding phonological context, i.e. it was the same after /s/ as it was after /n/ (2012: 4). In subsequent experiments they tested German and Dutch listeners who they presented with Dutch words with five levels of the /t/-Ø continuum in the two contexts (after/s/ and after /n/) in adjectives, nouns and verbs. The results showed a tendency for German, L2, listeners to give more /t/-responses when there is little evidence for a /t/ (2012:8). This showed that German listeners to Dutch deal with the Dutch /t/-reduction well.

Next, in order to investigate whether this is because of the similarity of L2 and L1 or whether this is because German learners of Dutch acquired the Dutch /t/-reduction patterns, another experiment was conducted. Here Mitterer & Tuinman (2012) investigated

whether the German listeners learned the Dutch pattern in verbs, i.e. the phonological rule which is that more reduction occurred after /s/ than /n/. Here, clear between-group differences were shown. An L1 transfer effect from German was found, such that German learners of Dutch did not perform as well on the unfamiliar cue from L1, which was the phonological rule, i.e. L2 comprehension is easier if the connected speech process exists in the listeners' L1.

Ernestus et al (2017) found that the cues listeners use for connected speech comprehension depends on the listener's L1. This was demonstrated by Mandarin Chinese EFL learners' bias towards perceiving what was not phonotactically illegal in their L1 (i.e. *can* rather than *can't*, /nt/-cluster being illegal in Mandarin). They also found that perceiving the cues used by the target language can be learned, because their Spanish native participants who were at a higher level of English did not show such a bias (the /*nt*/-cluster being illegal in Spanish as well).

Wong et al (2017) tested whether perception of reduced forms in English predicts listening comprehension in Chinese learners of English. The results showed that general listening comprehension and reduced forms dictation were correlated, and that general listening comprehension was also correlated with larger receptive vocabulary. In the general listening comprehension test, the participants' comprehension of CNN/BBC news reports was measured, in form of a multiple-choice test, and the reduced forms dictation was a dictation task consisted of sentences with various reduced forms embedded in them (the reduced forms were contraction, juncture, elision, vowel weakening, assimilation, intrusion, flapping, glottalization, and palatalization). Since these tests showed a high correlation, the conclusion was that the leaners' ability to perceive phonologically reduced forms was a significant predictor of comprehension of native English connected speech.

To summarise, studies have shown that L2 learners do exhibit difficulties with comprehension of connected speech in their target language. These issues can be due to a variety of language-specific cues that native speakers of those languages rely on. These cues can be lexical, morphological, phonotactic, segmental and sub-segmental. As well as this, they can combine in complex ways. Native speakers rely on this knowledge, and therefore, to achieve target comprehension skills in a foreign language, the learner is faced with a challenge.

Phonotactic, segmental and sub-segmental cues that native listeners rely on when parsing

connected speech, also contribute to the overall rhythm of a language. A few studies that approach the issue from the perspective of acquisition of speech rhythm will be looked at below.

White and Mattys (2007) applied rhythm metrics to investigate how L1 rhythm would influence L2 rhythm, using Dutch and Spanish learners of English and vice versa, i.e. English learners of Dutch and Spanish. They expected that those learners who had to switch between rhythm classes would show rhythm scores different from both their native and target languages. No difference was found in the VarcoV measures between English and Dutch L1 and L2. However, those did differentiate between L2 learner groups. That is, English natives speaking Spanish grouped with English natives, and Spanish leaners of English generally grouped with Spanish L1 speakers. This study showed that L2 rhythm can be difficult for learners to produce if 'switching' rhythm classes.

Previous studies by Ordin & Polyanskaya (2014, 2015) suggested that stress-timed speech rhythm is gradually acquired by early learners of various first language backgrounds, whether it is in their L1 or L2. Polyanskaya & Ordin (2015) set out to test whether this means that stress-timing is a marked feature, as had been suggested by Allen & Hawkins (1978) before. They used strictly controlled materials, which were the same for all age groups with respect to the lexical, phonotactic and segmental composition. They used rate-normalized rhythm measures (nPVI and Varco) in order to make sure that between- and within-speaker consistency is as high as possible. Their results showed that child speech in the L1 does become more stress-timed with time. They interpret this as a confirmation of syllable-timing being a "default setting" (2015: 203), and stress timing being marked, i.e. it has to be developed in the course of L1 and L2 language acquisition.

Astruc et al (2013) studied the acquisition of prosody in children of English, Catalan and Spanish which are typologically different (English - stress-timed, Spanish - syllable-timed and Catalan in-between stress and syllable-timed), and found that rhythm develops gradually as the different properties of the L2 are acquired. A year later, Li & Post (2014) investigated how German and Mandarin learners of English at different proficiency levels acquire English speech rhythm. The results showed that the different L1 groups followed similar paths of acquiring of vocalic variability and accentual lengthening. However, they differ in the %V, showing transfer from their L1s. Li & Post (2014) conclude that L2 rhythm acquisition is multisystemic, i.e. different properties of an L2, such as e.g. syllable

structure, vowel reduction and accentual and final lengthening, are acquired gradually as by the learner. As well as this, this acquisition process is also influenced by the differences between L1 and L2, i.e. there are L1 transfer effects.

1.7.2. Connected Speech & Rhythm Training Research

L2 rhythm acquisition may be difficult, because the cues that native listeners of various languages use for connected speech comprehension differ (e.g. Ernestus et al 2017, Mitterer & Tuinman 2012, Tuinman, Mitterer & Cutler 2012). Training of L2 rhythm, or particular aspects of rhythm and connected speech, has not yet been investigated by many.

In such languages as, for example English or Dutch, one key aspect of their rhythm is that vowels can undergo reduction, and especially so in faster and more casual speech, rather than careful speech. One paper has investigated the learning of this feature in the context of SLA. Morano et al (2015) looked into the role of exposure to schwa-reduced words in L2 French learners (L1 Dutch). In French, words can be reduced so that the schwa is completely omitted and pronounced as e.g. /pluz/, whereas the non-schwa-reduced word would be /pəluz/ (for *pelouse*, 'lawn'). In their experiment they exposed the subjects to words with schwa and without (i.e. the reduced forms) in French equally often. The words were all new to the subjects, i.e. not previously learned. This was done to test whether learners would learn both forms if they heard these forms equally often. If they did Morano et al (2015) aimed to test whether the rule of reducing the schwa sound would be generalized to other words. For English, this would be e.g. the word *ambulance* /ˈæmbjələns/ as it is reduced to /ˈæmbjləns/. The training involved learning new French words by looking at pictures while listening to four repetitions of each word, twice with the schwa sounds and twice without it, i.e. schwa-reduced. They performed a lexical decision post-test in which they had to decide whether or not a word they were presented with was a French word or not.

Morano et al (2015) found a significant difference between the accuracy scores for full and reduced forms, i.e. the full forms were recognised better. However, the difference was only 6%, and the authors attribute it to the reduced forms being shorter, i.e. leaving less time for the subjects to process them and to the troublesome onsets of reduced French words (illegal consonant clusters).

Their second research question was whether this can be generalised, i.e. whether the subjects would be able to apply the reduction rule to words they had already learned in

their full form before. They found that this was not the case, i.e. that the fact that the subjects had learned the reduced form for some words does not aid the learners' processing of the reduced forms of other words, which they had previously learned in their full form. Therefore, generalisation did not occur. As the authors say, the possible reasons for this could be that learners need "more examples or also examples for words they already know in order to deduce a generalization" (Morano et al 2015), as well as being exposed to the same pattern of reduction on the words they had already known for a longer period of time in order to be able to process these patterns, in order to learn - or generalise.

As far as training of the overall speech rhythm in EFL learners is concerned, there is only one fairly recent study by Wang et al (2016), who developed an automatic tool to help Chinese learners of English learn to produce English rhythm. The software (not available online) converts written sentences into music by extracting which are the weak and which are the strong syllables, assigning primary and secondary stress durations, loudness and pitch and displaying the sentence on the screen. The learner can then choose which instrument they want to hear play out the automatically generated rhythm, and play that while simultaneously seeing which word is playing on the screen, as the words get highlighted while their rhythm plays. Learners can use this to speak in synchrony with the musical rhythm.

Wang et al (2016) took the syllabic and vocalic PVI measures of the learners' speech before the training and the speech the learners produced while using the tool, i.e. their simultaneous productions of sentences while they were listening to the rhythms produced by the tool and reading the sentences from the screen. They also took PVI measures of the same sentences produced by American English native speakers and they compared these to the learners' productions. The PVI measures showed that the speech produced by the learners while they were using the tool was closer to that of native speakers than the speech they produced before they used the tool. As well as this, they asked native English speakers and trained phoneticians to rate the speakers on the native-likeness of their speech rhythm before they used the tool and while they were using it. This showed that those learners whose rhythm was generally poor were rated better during using the tool than before. However, those whose rhythm was not originally rated as poor, were rated as worse during using the tool. The authors suggest that this is possibly "due to [the tool's] structural rigidity, following the generated musical rhythm can reduce the naturalness of the speakers' speech and make their [utterances in rhythmic style] sound worse" (2016: 78)

than the utterances they produced in their natural-style. There is also the possibility that some initial improvement is easy to achieve, but then a plateau may have been reached.

Wang et al.'s approach showed some evidence of improvement in early learners of English. However, the study seems to have compared the learners' speech production before they used the tool with their productions while they used the tool. They did not test whether the training had any lasting impact on the learners' speech rhythm when they were not speaking simultaneously with the musical rhythm. Since there was no post-test, we do not know what the effect of this training is on a more long-term basis, i.e. whether the participants were able to hold on to the speech rhythm production benefits of this training after the training. As well as this, the study looked at speech rhythm production only, and therefore, whether the participants' perception of English rhythm improved as well is unknown.

1.7.3. Connected Speech in L2 Teaching

We have established that comprehending native English speakers' connected speech, as well as producing it, is a known problem for EFL learners. As Ito says, connected speech is important because it is used every day by native speakers, regardless of the register and speech rate (Ito 2006b: 19), and because learners are tested on connected speech when taking examinations in schools, universities etc. Therefore, when considering connected speech in L2 teaching, the two questions that come to mind are 1) it is it taught in a foreign language class, and -2) what kind of teaching, or training, is effective?

Brown and Hilferty (1986) tested effects of instruction on reduced forms. The instruction involved presenting them with reduced forms so that they understand them (Brown and Hilferty 1986: 63). The exact procedure of how this was achieved is, however, not specified in the paper. They used Bowen's Integrative Grammar Test (reviewed in more detail in Chapter 2 below), a multiple choice and a dictation test. The aim of the dictation was also for the students to review and practice comprehending the reduced forms. The group who received 4 weeks of instruction had comprehension scores that were significantly higher than the control group who received "placebo" lessons instead (i.e. 10 minutes drilling on phoneme contrasts).

In another, later study, Saber Khaghaninezhad & Jafarzadeh (2013) showed positive effects of instruction on reduced forms in English for Iranian EFL students. The instruction

involved explaining the phonological rules of English reduction and focusing the learners' attention on such sounds and how they change. However, there is also evidence that such instruction does not aid the learners' comprehension of reduced forms. Carreira (2008) had Japanese EFL students attend a course where they listened to connected speech forms such as contraction, assimilation and reduction, and then performed dictation tasks. No significant improvement was found.

In sum, there is some evidence that instruction on aspects of connected speech may be effective in improving L2 learners' comprehension, but also some evidence that it may not be. This could be a result of various factors impacting the learners' comprehension. These factors can depend on the learner (such as e.g. their L1 background, motivation to learn, musical training, etc, as shown in section 1.5. above), or on the teacher/teaching.

From a teaching perspective, there are also some other aspects which need to be thought of such as EFL teachers being non-natives, who may often be not familiar enough with reduced forms either to teach them or to realise their importance (Ito 2006b: 21). As well as this, there is a lack of appropriate teaching materials available (Ito 2006b: 22).

1.8. Summary of the literature review

In this chapter we have established that connected speech processes exist and that they affect the phonetic shape of words, with consequences for spoken word recognition. They also operate differently in various languages and therefore both influence and are influenced by the native speakers' speech rhythm. Ways in which speech can be seen as having rhythm were also discussed, and ways to classify languages as having different rhythm types were outlined, since this thesis relies on using rhythmic speech as a tool for learning.

Connected speech comprehension seems to be a problem which has not been given much attention in EFL classroom and teaching, even though it is problematic for non-natives. While the native speaker does not struggle with comprehension of this kind of speech if given context, the foreign language learner often does, even when given context. There is some evidence that instruction on connected speech may help learners, however, research is not completely unanimous on that point. As well as this, there is evidence that having a good musical ability, and training one's musical ability, can have a positive influence on learners' L2 abilities.

A key feature of English rhythm is the alternation of weak and strong elements. This led to

discussing the perceptual evidence that entrainment to speech rhythm helps comprehension, as well as how Perceptual Grouping Hypothesis may be of importance to the present work.. As well as this, Dynamic Attending Theory has been brought into the picture to support the idea that our attention would entrain with an external stimulus, and would therefore facilitate learning through creating expectations of future events. This leads directly to the key idea that is tested in this thesis: that entrainment of attention to rhythmically-structured speech should help listeners learn to understand connected speech."the following section.

A question that may arise when relating entrainment to learning is the following: why should entrainment of attentional pulses aid learning? That is, apart from being simply exposed to native speech, or tasked with listening to and repeating the materials as happens in many studies investigating L2 learning, what does entrainment offer and how might it contribute to learning? A possible answer comes from Repp (2005: 978) who says "[w]ith repeated exposure to a sequence exhibiting a complex but fixed timing pattern, there is a gradual change from tracking to prediction, indicating learning." In other words, evidence of learning becomes apparent when expectations form, i.e. the participants correctly anticipate that an event will occur at a certain point in time in the future. This means they must have acquired the information needed for this expectation to form, i.e. they have gained some knowledge and are now able to predict what will happen next using that knowledge.

This thesis will present three experiments which test three different ways of employing rhythm as a tool to train connected speech comprehension in learners of English as a Foreign Language. Additional specialist literature for each of the experiments will be reviewed *in situ*. Chapter 2 presents the first experiment on one of these rhythmic methods, testing whether listening to highly rhythmic speech elicited using regular metronome beeps, compared to speech that is less rhythmic, can aid EFL learners' comprehension of connected speech in Glaswegian English.

Chapter II

Can listening to highly rhythmic speech aid L2 learners' comprehension of connected speech?

2.1. Experiment 1 overview

In addition to the benefits of musical training, or musical aptitude, there is also evidence that attention to speech and music is rhythmically guided, and that speech processing depends on language rhythm (e.g. Cutler 2012, Fuji & Wan 2014, Large & Jones 1999, Quené & Port 2005, Dilley & McAuley 2008). However, although some EFL teachers seek to develop rhythmic awareness through their pedagogy (see Section 1.7.3), little work has tested whether encouraging non-native learners to attend to rhythm might support their comprehension of casual speech.

The experiment reported in this chapter tests whether speech elicited using Cummins and Port's (1998) "speech cycling" paradigm develops listeners' awareness of English rhythm in a way that improves their comprehension of fast speech. Cycled speech is recorded as the speaker synchronises with a metronome, and is typically very regular in terms of both timing and metre. This means that speech events happen at predictable times, which should aid listeners' entrainment. The experiment contrasted training which used cycled-speech materials (with High-Rhythmicity) and materials not elicited through speech cycling, which had Low-Rhythmicity. The prediction was that High-Rhythmicity training would allow listeners to entrain better to the rhythm of the speech they heard, and would thus improve their comprehension of fast casual Glaswegian English speech. The participants of this study were all students of EFL living in Glasgow. Therefore, the pragmatic decision was made to use Glaswegian English as the variety for testing in this, as well as the following experiments.

2.2. The variety of the L1: Glaswegian English

Glasgow English shows several phonological differences from other varieties of English (e.g. Wells 1982, Stuart-Smith 2004). Hearing a Glaswegian English speaker means hearing speech which exists along a continuum from broad Glaswegian dialect to a Glaswegian form of Scottish Standard English (Stuart-Smith 2003, Corbett and Stuart-Smith 2010). Glaswegian vernacular continues a form of West Central Scots, influenced by Gaelic, Irish English, and Standard English, whilst Glaswegian Standard English is a form of Standard English, originally introduced into Scotland after the Union of the Crowns in 1603, which has developed its own distinct features. In practice, Glaswegian speakers shift up and down the continuum depending on style and context (Aitken 1984).

The most relevant features of Glaswegian English for the purposes of this thesis are those relating to rhythm and those that are known to affect comprehension by listeners of other native languages or varieties. The ones covered in this section will include the vowel system, including Scottish patterning of vowel duration (Abercrombie 1979), (Scobbie et al 1999) and some aspects of consonantal phonology, i.e. glottalling and rhoticity (or the possible lack thereof, e.g. Stuart-Smith 1999).

Glaswegian English has 9 monophthongs /i, ɪ, ɛ, e, a, ʌ, u, o, ɔ/ and 3 diphthongs /ae, ʌu, ʌi/ (Wells 1982: 399). It differs from other varieties, e.g. Standard Southern British English (SSBE) such that e.g. in Glaswegian the vowel in words such as e.g. *pam* and *palm* is the same, i.e. it is a single phoneme /a/, while in SSBE there are two: /a/ for *pam* and /ɑ/ for *palm*. Similarly, Glaswegian has the same phoneme, /ʉ/, in words such as *full* and *fool*, whereas SSBE has /u/ and /ʊ/ respectively. And in words such as *cot* and *caught*, Glaswegian also has one phoneme, /ɔ/, and SSBE has /ɒ/ for *cot* and /ɔ/ for *caught* (Abercrombie 1979).

Like other Scottish Englishes, Glaswegian also shows a distinctive pattern of vowel duration, known as the Scottish Vowel Length Rule (SVLR) (Aitken 1981). The SVLR observes that the duration of some vowels depends on phonological and morphological context. Before voiced fricatives, /r/ and a morpheme boundary, e.g. *breathe, beer, bee, bee#s,* vowels are long; otherwise they are short. This means that Glaswegian vowels are short after voiced stops, nasals and /l/, where the Voicing Effect would condition long vowels in most other varieties of English. Scobbie et al (1999) demonstrated that current evidence shows a restriction of the operation of the SVLR to /i/, /u/ and /ai/ for Glaswegian and Edinburgh dialects of English.

As a result of SVLR, the durational distinction between tense and lax vowels that is typical in other varieties of English is absent. This can best be seen in minimal pairs such as e.g. *bid* vs *bead,* where the durations of these vowels are very similar in Scottish English (Rathcke & Stuart-Smith 2015). The SVLR is one feature that contributes directly to the distinct rhythm of Glaswegian.

Another feature of Glaswegian is that traditionally it is a rhotic variety, just as Canadian and American English are, i.e. /r/ is pronounced after a vowel in words such as *car, farm* (Wells 1982). This feature has been undergoing some changes, such that working class speakers derhoticise their post-vocalic /r/-productions, while still maintaining a subtle

contrast between words with and without post-vocalic /r/, e.g. *hut* vs *hurt* (Lawson et al 2011, Lennon 2017, Lawson et al 2018). Listeners who speak other varieties of English do not find this subtle contrast easy to perceive (e.g. Smith & Rathcke 2017, Lennon 2017).

Glaswegian English intonation has been described as simply having a final rise on declarative statements (Cruttenden 1994).

2.3. Speech cycling

The High-rhythmicity materials for this experiment were elicited by asking the speakers to repeat the metrically regular sentences in time with a regularly speeding up metronome beat. First of all, the original speech cycling method will be looked into in more detail.

Cummins & Port (1998) conducted experiments in which subjects repeat a phrase, e.g. *big for a duck* together with a periodic stimulus. They looked at how the subjects align their speech to periodic stimuli in time – a high and a low tone (*beep – boop* (Port 1999)). The phase between the first high tone (called *a* in the formula below) and the low tone was always fixed so that the interval between those two stimuli was 0.7s. The interval between the second and third stimulus, or the low tone – high tone beep, is what they call the target phase. The phase value was selected at random and it varied from 0.3 – 0.7 (i.e. $0.3 <= \Phi <= 0.7$), this is called Φ in the formula below. The formula used for the random calculation was: $\Phi = a/(a+b)$. This resulted in the second interval being anything from 0.3s (for when the target phase value was the highest, i.e. 0.7) to 1.6333s (for when the target phase value was the lowest, i.e. 0.3). This meant the whole high to high tone cycle (phrase repetition cycle, which they call PRC) varied from 1s (0.7s+ 0.3s) to 2.333s (0.7s + 1.633s).

Cummins and Port found that subjects, even though they were given the opportunity to place the onset beats of the phrases anywhere, chose to place them in three specific points between the stimuli. This revealed that, at least for native speakers of English, these points clustered at 1/3, 1/2 and at 2/3 of the phrase repetition cycle. Cummins and Port (1998: 150) propose that what happens in such a speech circumstance is establishment of a harmonic relationship, or coupling, between the PRC and the metrical foot (the interval between the stress beats).

They then also conducted further experiments using the same design but controlling for the effect of musical training and speech rate. These showed that subjects without musical

training performed in the exact same way (1998: 159). In the experiment in which they looked for effects of speech rate, they reduced the first interval from being fixed at 0.7s to 0.45s, which resulted in PRC length varying from 0.638s to 1.5s. Results of this experiment were again, exactly the same as the results from the previous two experiments, i.e. no effect of speech rate was found (1998: 163).

2.4. Testing comprehension of reduced speech – background information

A specific kind of testing was used in the pre- and post-tests for the experiments in this thesis. They are based on the so-called Integrative Grammar Test (IGT). The origins of IGT are reviewed in more detail below.

The IGT (Bowen 1976) is a test designed to measure the learner's "competence in the structure of the language without involving the examinee in grammatical terminology and without using discrete-point test items of individual structural patterns" (1976:30). Discrete point tests check for the students' individual skills such as grammar, vocabulary, spelling, punctuation, listening comprehension, etc., whereas integrative tests attempts at combined measuring these skills.

As expected, native speakers of a language (English in this case), easily achieved high scores when given a task of understanding reduced forms. And the more L2 learners can do this, the more competent in the L2 they are (Bowen 1976 :30-31). In IGT the clarity of the message was worsened by use of speech reductions. Bowen reported that learners score between 0 and 80% and 90% only occasionally (Bowen 1976: 31), and suggested that these scores could "be regarded as a reliable index of overall competence in the language" (1976: 31).

An example of sentences Bowen used in the IGT:

- *Who'd he been to see?*
- *Who'd he wanna see?*
- *Who'd he like to see?*

The task was to listen to the sentence once and write down the second word. A group of native and non-native speakers were compared. Non-natives scored generally lower than natives, as expected. The range of scores was also much higher for the non-natives than for the natives. The non-native speakers' results of the IGT were then also compared with the American University in Cairo's (AUC) admission test. This comparison showed that IGT

and the admission test measured L2 competence consistently with each other (Bowen 1976: 34). Bowen also found that "the IGT scores seem to indicate that it is a grammar test rather than, say, a listening comprehension test, correlating with the AUC grammar test at a very high .871" (1976: 36).

Bowen concluded that if non-natives learn to identify meanings of the reduced forms of the English language "their mastery of English grammar will be commensurately improved" (1976: 37). These results altogether suggest that the non-native speakers' L2 competency, and especially the non-natives' grammar, can be successfully measured using the IGT.

2.5. Research Questions

The main research question for the present experiment is whether training with High-rhythmicity materials can improve EFL learners' connected speech comprehension better than materials of Low-rhythmicity.

In addition, the experiment explores what other factors influence whether such learning occurs. Participants were screened for their musical background and English language proficiency. A number of different speakers were heard, and materials featured several types of connected speech phenomenon, to explore the effects of these factors on learning.

2.6. Hypothesis

The hypothesis is that High-rhythmicity training will benefit EFL learners' English connected speech comprehension more than Low-rhythmicity training.

2.7. Experiment design

The experiment was designed so that two groups of EFL learners were tested in pre-test, training and post-test. Pre-test and post-test were identical for both groups, while the training phase was different. The experimental group received High-rhythmicity training and the control group received Low-rhythmicity training. The training was delivered through a specifically designed website.

2.7.1. Participants

Multiple speakers provided materials for this experiment, in order to give the learners High Variability Perceptual Training. Four native speakers of Glaswegian English were recorded, two females (aged 25 and 60), and two males (aged 28 and 52). The speakers varied socially as well as according to age and gender. The two female speakers were

middle-class, i.e. located at the Standard Scottish English end of the continuum and the two male speakers were working-class, i.e. speakers of a broader variety of Glaswegian English.

62 EFL learners took part in the experiment, living in Glasgow, aged 19-43 (mean = 27.5, SD = 6.3). They were of various language backgrounds, 29% Chinese, 13% Polish, 10% Thai and the remaining 48% were a few from many different languages each (Arabic, Burmese, Czech, French, German, Indonesian, Italian, Korean, Lithuanian, Mandarin, Portuguese, Romanian, Spanish, Urdu and Vietnamese) (for graphical representation of these see Appendix B). The rationale for including learners from all language backgrounds was of pragmatic nature, i.e. these were the students available at the time who were at a similar level of English and who were attending classes in the country where the language was spoken, so it could be assumed that they all had good motivation to learn English. They attended classes in EFL Units in Glasgow (such as University of Glasgow, University of Strathclyde, and Glasgow Clyde College). The English language level of the courses the participants attended ranged from pre-intermediate to upper intermediate/advanced according to their EFL Units.

2.7.2. Materials

Two sets of 36 sentence materials were designed, one for the High- and one for the Low-rhythmicity condition. They were similar in all respects except for the regularity of the sentences' metrical structure (explained below) and the elicitation procedure (explained in section 2.7.3.2. below). Out of the 36 sentences for each condition 24 were used for the training phase. The remaining 12 were used as pre-test and post-test materials, and divided between the pre-test and post-test so that each contained 6 of High-rhythmicity and 6 of Low-rhythmicity sentences.

For the High-rhythmicity condition, sentences with a regular metrical structure were designed, e.g. in *She was fast for a girl on a bike* there were weak and strong syllables alternating (*wwSwwSwwS* for the above example). In the High-rhythmicity materials, two different metrical structures were used, the one described above, i.e. *wwSwwSwwS* and *wSwwS*, as in e.g. *She can't sing her son*g.

For the Low-rhythmicity condition, the High-rhythmicity sentences were adapted to make their metrical structure less regular, e.g. *She was fast for a girl on a bike* was changed to

She is speedy for a girl on a bicycle in which the weak strong syllable alternation is less regular: *wwSwwwSwwSww*.

Seven different types of reduced speech phenomenon categories were chosen. These choices were based on the Hagen's (2000) account for EFL learners of connected speech phenomena in English, the reason for these choices was to provide a wide range of reduction types and explore which may cause most confusion among EFL learners. The chosen reduction types were:

- articles *a* and *the* (henceforth *article*); a total of 15 articles were used across pre-test and post-test, and 22 in the training phase (for High- and Low-rhythmicity training);
- the *is* and *was* forms of the verb to be (henceforth *be*); total of 8 *is/was* verbs were used across pre-test and post-test and 16 in the training phase (for High- and Low-rhythmicity training);
- contraction vs non-contraction in *can* vs *can't* and *should* vs *shouldn't* (henceforth *contraction*); total of 8 contractions were used across pre-test and post-test, and 8 in the training phase (for High- and Low-rhythmicity training);
- the personal pronoun *her* (henceforth *her*); total of 8 *her* pronouns were used across pre-test and post-test, and 10 in the training phase (for High- and Low-rhythmicity training);
- participle present ending *-ing* vs participle perfect ending *-en* in verbs in which the ending is the only feature that differentiates between the participles, e.g. *taken* vs *taking* (henceforth *participle*); total of 9 participle endings were used across pre-test and post-test, and 8 in the training phase (for High- and Low-rhythmicity training);
- a variety of prepositions (henceforth *preposition*); total of 15 prepositions were used across pre-test and post-test, and 16 in the training phase (for High- and Low-rhythmicity training);
- pronouns (henceforth *pronoun*); total of 12 pronouns were used across pre-test and post-test, and 24 in the training phase (for High- and Low-rhythmicity training);

Table 2.1 below gives examples of each category from high-rhythmicity and from low-rhythmicity materials. As can be seen from these examples, in most cases, more than one category could be tested using one sentence.

	High-Rhythmicity	Low-Rhythmicity
Article	He is fat for **a** cat in a box.	He is enormous for **a** cat in a box.
Be	She **was** great on the day of the race.	She **was** amazing on the day of the competition.
Contraction	I **shouldn't** leave her child.	We **shouldn't** allow her to fight.
Her	She can fight **her** cold.	She should take **her** medication.
Participle	He's **giving** them food.	He's **given** them macaroni.
Preposition	She was great **on** the day **of** the race.	She was amazing **on** the day **of** the competition.
Pronoun	**She** can fight her cold.	**She** should take her medication.

Table. 2.1. Example sentences per Reduction Type.

For the full materials see Appendix C.

2.7.3. Recording Procedure

All materials were recorded in a sound-attenuated booth at the University of Glasgow. An AKG SE 300B microphone was used together with a Rolls LiveMix MX34c microphone/headphone mixer and a Roland Duo Capture audio interface. In the High-Rhythmicity condition a set of Beyerdynamic headphones was used to play back the metronome beeps. The speech recordings were made using Audacity. The full set of materials was recorded from every speaker.

2.7.3.1. Low-Rhythmicity

First, speakers were asked to produce the Low-rhythmicity materials, i.e. the ones with a less regular metrical structure. These were recorded first, to make sure the speakers were not primed to produce them in as rhythmic a manner as they would be asked to do for the High-rhythmicity sentences. The sentences were embedded into a story in order to elicit speech with as much flow as possible. Participants were asked to read the story four times, at four different rates – very slow, normal, fast and fastest. Participants were allowed to set the rates themselves, without any external cue.

2.7.3.2. High-Rhythmicity

The High-rhythmicity materials were elicited by asking the speakers to repeat the metrically regular sentences in time with a regularly speeding up metronome beat. Their task was to align the stressed syllables of those sentences with the beeps. In order to capture connected speech processes each metronome beat sped up by 2bpm forcing the speakers to produce a variety of tempi from slow to as fast as possible. Because different metrical structures of the sentences were of two different types, with either two beeps per sentence (e.g. in *She's eaten the sweets*, for the structure *wSwwS*), or three (e.g. in *He's fat for a cat in a box*, for the structure *wwSwwSwwS*), two different metronome beat sequences had to be created. For the two-beat sentences the rate of the beeps started at 35 bpm and finished on 65bpm, which gave a total of 15 repetitions of the whole cycle; and for the three-beat sentences the rate of the beeps started at 20 bmp and finished on 65bpm which yielded a total of 23 repetitions of the whole cycle. Each time one cycle was finished, the next cycle sped up, i.e. no rate was repeated twice. They sped up until all repetitions were finished. The three-beat sentences started at a slower rate than the two-beat sentences because they were deemed to be too difficult to produce otherwise, and because slower rates were needed to later decide which sounded both slow and natural enough to be used in the stimuli.

The start and end of each repetition sequence was signalled by a longer (400ms) lower (440Hz) tone of peak amplitude 0.2 Pa and amplitude rise and decay time of 10 ms, which served as a warning beep. The first tone to synchronise to followed that beep immediately. The beats that the participants were to synchronise to alternated from High to Low, and there was a silent beat between each repetition. The high tone was at 400 Hz, peak amplitude 0.75 Pa*, onset/offset were not ramped and the low tone was at 800Hz, peak amplitude 0.25 Pa*, onset/offset were not ramped. Both had a duration of 40ms. The beeps alternated low-high for the two-beat sentences, and for low-high-low for the 3 beat sentences.

The sequences of beats were generated using Praat (see Appendix D) and were played back to the speakers using Audacity. The participants read the sentences one by one from a printed list. Their instruction was to start speaking with the beat as soon as they could. If they made a mistake, they were instructed to keep speaking until the metronome beeps for that sentence stopped. The experimenter would then ask for another recording of those sentences in which mistakes were made.

A quick demonstration by the experimenter was given to each participant on an example sentence, which showed them how to perform the task, but they were not given a practice session nor given feedback on their performance. In practice, the participants did not always starts speaking as soon as the first or even second cycle started, and since the beats themselves were not recorded during the materials recording, it was not possible to tell which cycle the speakers started with. This was likely due to the fact that there was no silence between the warning beep and the first beep to synchronise to, which, in retrospect, would possibly have given the participants a better chance to start speaking with the first beep. However, they all performed the task in the sense that once they started speaking, they spoke with the beats and were able to speed up gradually.

2.7.4. Stimulus Creation

To keep the High- and Low-rhythmicity conditions as comparable as possible, the stimuli in both conditions featured an equal number of sentence repetitions. For the Low-rhythmicity condition, the four recorded versions of each sentence were excised from the story and concatenated, ordered from slowest to fastest. This was done to gradually present the listeners with how speech changes when connected speech processes start to operate in faster speech. This is a form of Perceptual Fading technique (Terrace 1963, Jamieson and Morosan 1986), which is described in more detail in section 1.5.3.2.2 above. The silences between each repetition were not strictly controlled for, but were all made so that at least 100ms and at most 700 ms were silent in between the end of the preceding and start of the following sentence. For the High-rhythmicity condition, ideally four versions produced at the slowest and fastest tempi and two evenly spaced ones in between would have been selected. For example, these would be at 40 bpm, 50 bpm, 60 bpm, and 70 bpm (for the two beat sentences) and 60 bpm, 90bpm, 120bpm and 150bpm (for the three beat sentences). Furthermore, ideally those four tempi would be kept the same for each sentence of the same metrical type. However, as explained above, the speakers did not always start speaking straight away aligning their speech with the first beat they heard, and the beats they heard were not recorded, which meant it was not possible to judge whether the first sentence version they produced was at the intended starting tempo of 40 bpm, or a slightly faster tempo (e.g. 42 or 44). This rendered it impossible to consistently select the repetitions produced at the same four tempi for each of the sentences. Therefore, the decision was made to use the slowest repetition that a speaker produced that sounded sufficiently natural to the experimenter; and the fastest repetition that sounded comprehensible; and then to space the intervening two stimuli so that the distance between

each pair of stimuli was equal, e.g. every third repetition, or every second repetition of a sentence. This means that the difference in tempo between repetitions remained constant within each stimulus, but the differences varied from stimulus to stimulus.

An attempt was made to balance the number of sentences spoken by each of the four speakers, in each of the experimental phases, i.e. the pre-test, training and post-test. Ideally, 6 sentences from each speaker would have been used in the training phase, and 3 sentences from each speaker in each of pre-test and post-test. This was not possible in practice because some of the recordings were unusable as some of the repetitions did not sound fluent enough upon closer inspection, which would achieve a perfect balance between the speakers and between different sentence type or different metrical structure type in the training phase. Therefore in both the High- and Low-rhythmicity training phase, there are seven sentences produced by the older male, five by the younger male, and six by each of the females. In the pre-test and post-test it was possible to keep the balance, i.e. there were three sentences by each speaker. The details of which sentences were produced by which speaker can be viewed in Appendix E. As well as this, six of the 12 sentences in pre-test and post-test came from the high-recordings and six came from the Low-rhythmicity recordings. These sentences were then pseudo-randomised such that the same connected speech phenomenon type would not occur in two sentences in a row. This randomisation was the same for all participants. The same types of reduced speech phenomena were used in both pre-test and post-test, but used different sentences in order to test whether generalisation took place.

2.7.5. Website Creation

The materials were coded in Dreamweaver to html files and concatenated so that two websites resulted – one for the experimental group and one for the control group. The websites contained instructions for the participants on how to proceed to complete the task. These websites can be viewed here:

- experimental: http://www.micro-phonics.com/wip/research/expt-aug2015/page1-1.html;
- control: http://www.micro-phonics.com/wip/research/non-cycled-experiment-aug2015/pagenc1-1.html.

2.7.6. Procedure

The participants were tested at their own EFL units, in a group session for each of the groups. Each participant used a computer provided by their EFL unit and Beyerdynamic headphones provided by the Glasgow University Laboratory of Phonetics.

The experiment started with a short working memory test, intended to assess whether individual differences in working memory might affect comprehension. Participants were asked to memorise digits in the order they appeared on the screen and type them in. Due to a coding error, a large proportion of these data were not recorded and therefore this test could not be included in further analysis.

The second part of the experiment was an English language proficiency test consisting of 25 multiple-choice questions, obtained from http://www.cambridgeenglish.org/test-your-english/adult-learners/. An example of a question is:

Q: Shall we go to the gym now?

A. *I'm too tired.* (correct answer)
B. *It's very good.*
C. *Not at all.*

Participants then took a break before starting the main experiment. The main experiment had three phases: pre-test, exposure and post-test. The full materials for each group are in Appendix C and can be heard on the websites above.

In the pre-test, participants listened to sentences and were asked to fill in missing words on the computer screen. For example, they would hear the sentence *He was fat for a cat in a box* and would see: ____ fat ____ cat ____ box. Their task would be to fill in the blanks with the words they heard. They were told that each blank could contain more than one word.

In the exposure phase, the participants were asked to carefully listen to 24 sentences. While being played, each sentence was displayed for them on the screen, in order to make sure they understood as much as possible. The experimental group listened only to the High-rhythmicity sentences, while the control group listened to the Low-rhythmicity sentences. In order to maintain the participants' attention to the sentences, they were asked prior to

starting the exposure phase to count the number of times food was mentioned in these sentences. Upon completion of the exposure phase, participants were encouraged to take a short break if they wanted to and then to proceed to the post-test.

Next, the participants completed the post-test, for which the instructions were the same as the pre-test. Finally, they completed a questionnaire asking them for demographic information such as age, gender, their first language, length of studying English, any other languages spoken, and their musical training. The exact purpose of the experiment, i.e. that it was to train their comprehension of English connected speech, was disclosed to the participants via a debrief letter after the experiment.

2.7.7. Coding and cleaning of the data

The answers were coded as 1 for correct and 0 for incorrect. In cases where two words had to be typed into the answer field, each word was separately coded as correct or incorrect. For example, in a sentence such as *He was fat for a cat in a box*, if the participant typed in *in the,* they would score 1 for the first word and 0 for the second*;* if they typed *on a*, they would receive 0 for the first word and 1 for the second word.

In case of obvious misspellings, e.g. *abuve* instead of *above*, the answers were coded as correct. In case of any doubt as to the reason for the incorrect spelling of a word, the answer was scored 0.

After the experiment it was discovered that one of the participants in the control group knew one of the speakers used to record the materials and therefore had to be excluded from the analysis. Another participant in the control group did not record her answers to the multiple choice English proficiency test and therefore also had to be excluded from the analysis. Thus, 31 participants in the experimental group and 29 in the control group were analysed.

2.8. Results

A mixed-effects logistic regression analysis was performed using R (version 3.3.2). The dependent variable was Score. The model fitting procedure was to start with the simplest model possible containing only main effects, but no interactions. Then more variables and interactions between variables were added to the equation. A variable was retained if the model was significantly improved, based on a log likelihood test of the models.

The first model included two main predictors as fixed effects: Training Condition (levels: High-rhythmicity and Low-rhythmicity), Test (levels: Pre-test and Post-test), and their interaction, since the key prediction was that participants in the High-rhythmicity condition would improve more than those in the Low-rhythmicity condition. Other fixed predictors were English Language Proficiency (a score ranging from 0 to 25), the Reduction Type (levels: Article, Be, Contraction, Her, Participle, Preposition, Pronoun), the Speaker (levels: Older Female, Younger Female, Older Male, Younger Male) the participants heard in the pre- and post-test, and the Test Sentence Type i.e. whether the test sentence came from the High-rhythmicity or the Low-rhythmicity materials. It was planned to include Musical Training as a variable as well. However, a fairly large proportion of participants (15% of them, to be exact) did not provide an answer to that question (i.e. "Do you have any musical background? If so, what kind?") in the questionnaire, leaving it blank, and therefore this variable could not be used.

There were two random effects: Participant and Item. Random slopes were not included because the model would not converge if an attempt to include random slopes was made. Predictors that were non-significant in this first model, or that prevented the model from converging, were then removed. Finally, interactions of test with other factors were also tested to establish whether any other of the fixed effects affected participants' improvement at the task. The final model is shown in Table 2.2 below.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.30628	0.50167	-4.597	4.28e-06 ***
English Language Proficiency	0.13735	0.02533	5.422	5.90e-08 ***
Test Sentence (Type LR)	1.42849	0.43634	3.274	0.00106 **

Table 2.2. Final mixed effects model for Experiment 1 results.

Figure 2.1 below shows the results by Test and Condition. It reveals tendencies in the predicted directions: participants' scores were slightly higher in the Post-test than Pre-test, and this pattern was more pronounced in the High-rhythmicity than the Low-rhythmicity condition. Median scores were 55.56% in pre-test and 61.76% in post-test for the High-rhythmicity condition, an improvement of 6.2% on average; whereas for the Low-rhythmicity condition they were 58.33% in pre-test and 58.82% in post-test, an

improvement of 0.5% on average. However, neither main effect, nor their interaction, approached significance.

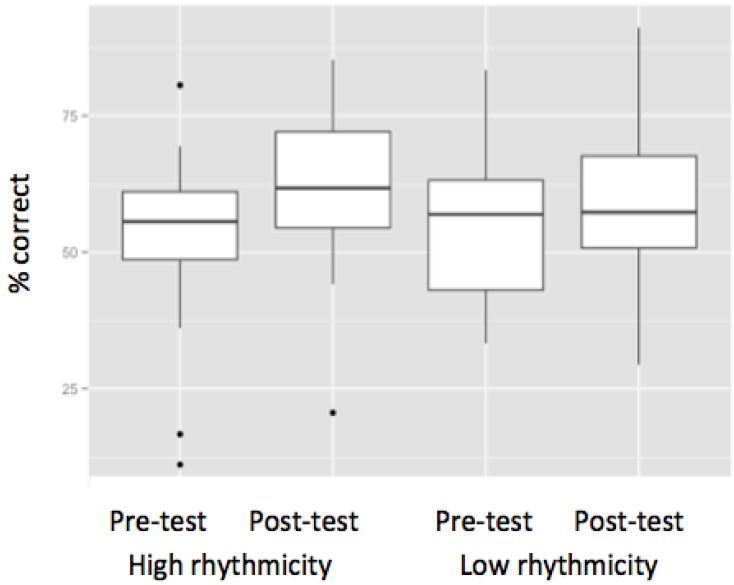


Fig 2.1. Boxplots illustrating the interaction of Test and Training Condition.

A highly significant effect of English language proficiency was found ($p < 0.0001$, $z = 5.4$), such that the higher the participant's proficiency, the higher the score. No significant interactions of English Language Proficiency with Test were found.

Another significant effect was Test Sentence Type, i.e. whether the sentence in the test was elicited using the High-rhythmicity or Low-rhythmicity method. Sentences that were taken from the Low-rhythmicity materials received significantly higher scores ($p < 0.05$, $z = 3.3$). Figure 2.2 graphs this difference, also including training condition. This shows that the Low-rhythmicity sentences were relatively easy for both groups to start with and no improvement was achieved in post-test. In contrast, both groups (especially the High-rhythmicity training group) appear to have improved on the High-rhythmicity sentences from pre- to post-test, but the interaction of Test with Test Sentence Type was not significant.

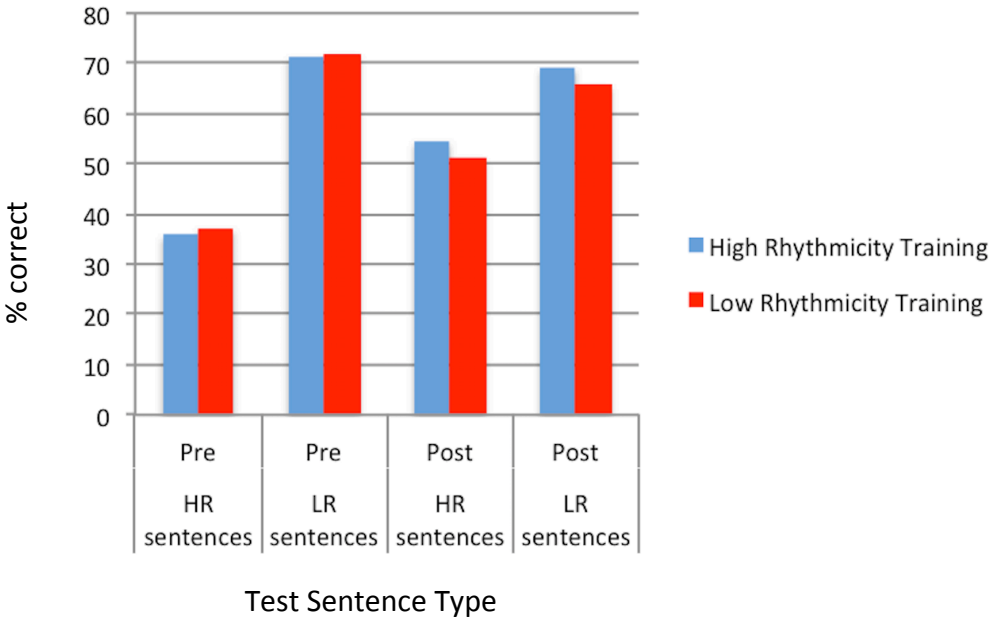


Fig. 2.2. Bar plots showing % correct responses in pre-test and post-test on High- and Low-rhythmicity test sentences. Colour indicates which training participants took part in: blue for High-rhythmicity and red for Low-rhythmicity.

The model did not converge when the Reduction Type variable was included in it, and therefore this variable is not in the final model. However, Figure 2.3 shows some trends as to which types of connected speech process were more difficult than others. The continuum of difficulty of perceiving these different reduction types therefore looks as follows:

Contraction < Pronoun < Preposition < Be < Participle < Article < Her

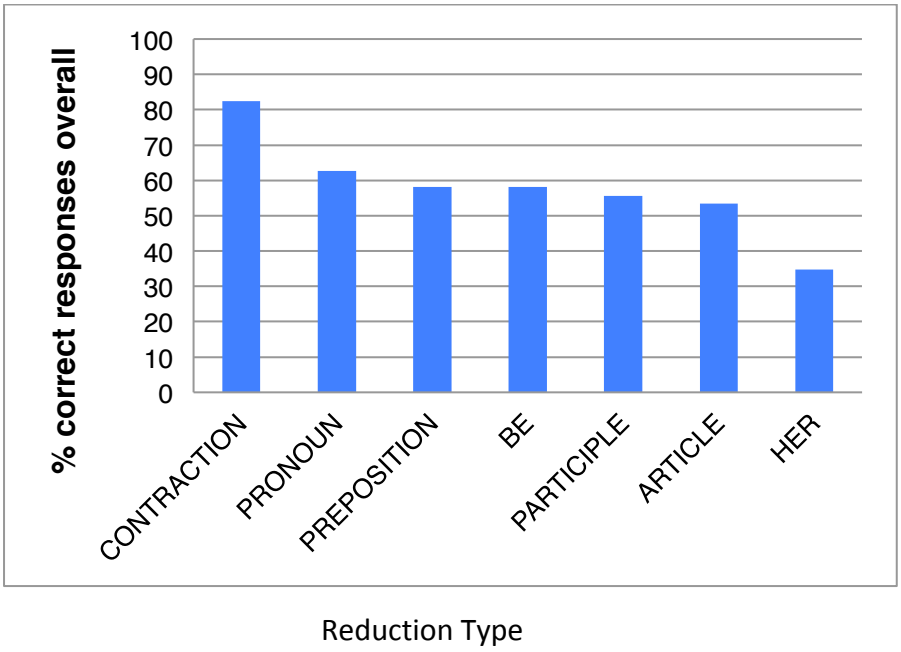


Fig 2.3. Overall % of correct responses on both Tests, by Reduction Type.

Figure 2.4 below further shows that there are numerical differences in how much improvement was seen for the different categories. Notably, only Participle (e.g. tak*en* - tak*ing*) and Article (e.g. *a*, *the*) improved by more than 10%; Preposition (e.g. *of*, *in*) improved by a smaller amount, while the other categories declined slightly.

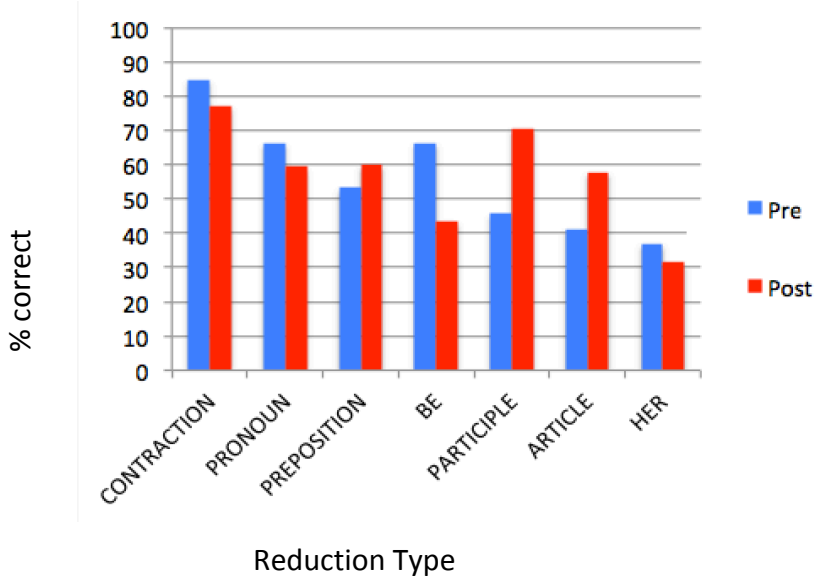


Fig. 2.4. Change from pre-test to post-test by Reduction Type.

In sum, the predicted result was not significant. Moreover, the stimuli varied in their intelligibility in two ways, of which the most important was that the High-rhythmicity materials were harder to understand than the Low-rhythmicity ones. Learners' proficiency also predicted their performance.

2.9. Discussion and directions for future experiments

The prediction for Experiment 1 was that the High-rhythmicity training would improve the listeners' comprehension more than the Low-rhythmicity training. This prediction was motivated by Dynamic Attending Theory (Jones 1976; Chapter 1, section 1.3.8) i.e. if our attention is rhythmical and synchronises, or couples, with a rhythm in the external world, then more learning would be expected to occur when that external rhythm is more regular and easier to synchronise our attention to. Prior support for this idea applied to speech comes from Quené and Port (2005), who found that regular timing of stressed and unstressed syllables significantly improved speech perception in a phoneme monitoring task, presumably because entrainment aided listeners' attention to the linguistic content of the speech they heard. In the case of Experiment 1 in the present study, however, although the numerical pattern was in the predicted direction, the result was not significant.

Why did this result did not reach significance? Firstly, this may be due to the fact that Low-rhythmicity materials were not different enough from the High-rhythmicity materials. Participants heard each sentence four times (with the rate speeding up) in both conditions, which means that on a higher level the Low-rhythmicity materials are also rhythmic in the sense of exhibiting a recurrent pattern in time. That is, even though the Low-rhythmicity training materials had an irregular metrical structure, four repetitions of the same sentence itself yields a rhythmic whole. Therefore, a better test of the hypothesis might be to use materials that are even less rhythmic than the Low-rhythmicity ones in the present experiment, in order to make the tasks for the experimental and the control group as different as possible. This could be done, for example by making the experimental group's task as rhythmic as in the present experiment, but making the control group listen to different sentences at different rates, i.e. not exposing them to repetition of the same sentence at different rates.

Moreover, it is possible that merely listening did not offer a strong enough means for entrainment to happen. That is, asking participants to entrain solely via listening rather than e.g. moving (a part of) their bodies to the beat was possibly not enough to aid the attentional oscillators create these expectations at stable points in time, or to entrain with the stimulus. This could relate to the fact that the participants were non-native listeners. Quené and Port's (2005) study was conducted on native listeners whereas the present study used listeners from a variety of L1 backgrounds, but no English natives. Perhaps this kind of entrainment was simply not enough for non-native listeners. As well as this, it is possible that non-native listeners would require more exposure than native listeners of a different variety.

The EFL learners in this experiment were of various L1 backgrounds. How this might have affected the results was explored in the modelling process. The L1 languages were coded according to whether they have been reported to be syllable or stress-timed, as well as whether they were tonal or non-tonal. However, the participants' L1 type (neither rhythm type or tonal vs non-tonal type) was not found to be a significant predictor.

The significant positive effect of English Language Proficiency is perhaps not surprising. Those participants whose general English language skills were higher, also performed better at the comprehension tests. An interesting point that can be made, however, is that the 25 multiple-choice questions, which designed to establish what level a learner is at are

based on simple reading comprehension (as shown by the example in section 2.6.6. above), but they seem to also be able to indicate how well participants comprehend speech and capture these individual differences. Individual differences may have also been due to the participants' musical background. Investigation of this effect on comprehension was planned, but was not possible because, as mentioned above a large proportion of the participants did not answer the question. This is possibly due to the question not being specific enough. More than one question may have been needed in order to establish to what extent a participant has been musically trained (as done by e.g. Lidji et al 2011), or indeed whether they have a more inherent and untrained musical ability (cf. Law & Zentner 2012, reviewed in more detail in Chapter 4 below).

Low-rhythmicity sentences were significantly easier than High-rhythmicity sentences for EFL learners to understand in both pre- and post-test. This was an unexpected result. Listening to the stimuli suggested that it could possibly be because the speech rate of the High-rhythmicity materials was higher. Two sets of measurements were taken in order to check this. The measurements were of the underlying syllables and feet, rather than the surface ones, i.e. as they would be in carefully produced utterances. In syllables per second, the result was similar for both, 6.3 syll/second for High-rhythmicity and 6.25 syll/second for Low-rhythmicity materials, but the difference was slightly bigger in feet per second, with 2.4 feet/s for High-rhythmicity vs. 2.05 feet/s for Low-rhythmicity test materials. Another possible reason why the High-rhythmicity materials may have been hard to understand could be the way it was produced in time to metronome beats. Impressionistically, this made the prominence contours in the speech sound slightly exaggerated: the weak syllables sounded more reduced than in speech produced without a metronome). In consequence, the cycled speech sounded slightly unnatural, which may go some way to accounting for the worse performance on these materials.

Interestingly, the High-rhythmicity speech also showed a trend for improvement larger than the Low-rhythmicity speech. The results for Low-rhythmicity sentences in pre-test were just as high as the results in post-test, which could mean that the participants reached their ceiling scores and could no longer improve. To establish whether this "ceiling" reflects a limit on the intelligibility of the materials or of the listeners' proficiency, we would need to test a group of native Scottish English listeners. In contrast, average performance on the High-rhythmicity materials improved by 18.8% from Pre- to Post-test in the group that had the High-rhythmicity training, and by 14% in the group that had the

Low-rhythmicity training, though – surprisingly – the improvement was not significant in either case. Perhaps further exposure to the fast rate and slightly unnatural prosody of these materials would have generated significant improvement.

There was a non-significant trend showing that some types of reduced speech phenomena are potentially harder than others. The three that showed a possible improvement were Preposition, Participle and Article. Article and Participle also are two of the three most difficult types in general and so is the category HER (for which the graph illustrating change shows worsening rather than improvement from pre- to post-test). A possible reason for Prepositions being easier to identify than participles or articles could lie in phonetic cues being more distinct in prepositions and hence more recognisable. For example, the /f/ and /r/ of the word *for* as opposed to the /fr/ and /m/ of *from* are possibly easier to distinguish from each other than the phonetic cues contained in the reduced *–en* vs *–ing* at the end of a participle are, or the phonetic cues contained in the reduced *her* vs *a*.

Finally, let us reflect on how well the speech elicitation method worked for the purpose of this experiment. In hindsight, it would have been useful to take rate measurements of both the speech and the training materials to ensure that the stimuli were comparable on these dimensions. As well as this, some more in depth acoustic measurements of the training materials, i.e. averages for all four rates from slowest to fastest in both conditions to investigate such question as, for example, whether the High-rhythmicity materials were generally faster or only at the fastest rate. In addition, measurements of how regular the speech was in the two conditions could be taken and compared by looking at intervals between stressed vowel onsets, or f0 peaks. This could inform us to what extent, if at all, the High-rhythmicity materials actually had more regular stressed beats than the LR materials.

A main drawback in this experiment was that the Tests (i.e. Pre-test and Post-test) were not counterbalanced. This means that we cannot be sure whether any change or lack of change was due to a difference in difficulty between the two tests. This could be addressed in the future by conducting an experiment with a group of participants who would answer all questions to both Pre-test and Post-test in an order that would be randomised for each participant. Results could be re-analysed in the light of that finding. Time constraints meant this was not conducted for this thesis.

In summary, the hypothesis that this experiment set out to test, was not supported. However, there are sufficient trends in the data to pursue the idea of regularity of the stimulus aiding entrainment, which in turn would aid processing of the material and therefore learning of it. These subtle indications prompted a further investigation of it with a different training set-up will be explored first. This new training set up, which is at the core of the next experiment, aims to increase the likelihood of entrainment to test whether this would result in improvement of comprehension.

Chapter III

Can tapping to the beat in speech aid L2 learner's comprehension of connected speech?

3.1. Introduction

Experiment 1, which investigated whether listening to highly rhythmic speech can benefit learners of English, did not show a significant change in the learners' comprehension of fast casual English. As discussed above, there could have been several possible reasons for this. To tackle some of these possible reasons, Experiment 2 was conducted. It builds on Experiment 1 by investigating whether performing a sensorimotor task (tapping to the beat in rhythmic speech) can help learners attend to rhythm, and thus improve their comprehension. Thus Experiment 2 seeks to establish whether stronger entrainment can be achieved through a tapping task, which would then lead to enhanced attention and therefore - through training - better comprehension of fast casual English speech.

Experiment 2 was conducted on three groups of listeners: Chinese learners of English as a Foreign Language, Canadian/US native English speakers and native Glaswegian English speakers. All of these groups were tested, via pre-test and post-test, and trained on comprehension of Glaswegian English. This allowed for a comparison of the effects of tapping training between listeners who were learners of English, listeners who were native speakers of a different variety of English and those who were native speakers of the variety they were hearing.

Because some additional background literature is needed for this experiment than was needed in Experiment 1, the introduction will cover relevant background in four main areas:

- Sensorimotor synchronisation, especially tapping to speech and how it could contribute to speech comprehension;
- P-centres, i.e. perceptual moment of occurrence of an event;
- Comprehension of accented speech by speakers of other varieties, and training effects;
- Some phonological features of Glaswegian English, with a brief comparison to General American and Canadian English.

3.1.1. Sensorimotor synchronisation (SMS)

The following section will give a brief overview of the existing literature on the topic of sensorimotor synchronisation (SMS hereafter). It will do so first in very general terms, and

then it will move on to the connection between SMS and speech and language. The end of this section will briefly look into ways in which information gained through the research on SMS can be applied practically.

3.1.1.1. General overview

SMS is most commonly investigated through a tapping paradigm, i.e. the participants tap their finger in an attempt to synchronise their movement with an external stimulus. Repp (2005) and Repp and Su (2013) provide comprehensive reviews of the existing tapping literature. They define SMS as rhythmic coordination of perception and action, or "coordination of rhythmic movement with an external rhythm" (Repp & Su 2013: 403). Repp & Su (2013) divide tapping tasks into different types:

- different forms of movement (tapping on a hard surface, finger/limb movement without contact with a surface);
- different modalities of stimulation (auditory, visual);
- different forms of coordination (in-phase, anti-phase; e.g. 1:1, 1:2 or 2:1).

The possible forms of phase coordination need further explanation here. The timing of the taps relative to the external stimulus, e.g. a metronome, can either be synchronised so that the taps occur at (or around) the same time as the stimulus sound, i.e. *in-phase*, or at (or around) the midpoint time between the stimulus sounds, i.e. *anti-phase*. 1:1 coordination means that taps occur on every beat of the metronome (and not at any other time); 1:2-coordination means that a tap occurs with every second beat of the metronome, while 2:1 coordination means that a tap occurs with each beat of the metronome and also at (around) the midpoint in time between each beat of the metronome. (Other relationships e.g. 1:3, 3:1 are also possible on the same principles.) These relationships are represented in Figure 3.1 below (taken from Repp 2005: 970). In Figure 3.1, IOI, or Inter-Onset Interval, refers to the interval between stimulus onsets; ITI, or Inter-Tap Interval, refers to the interval between the onsets of the taps performed by the participants.

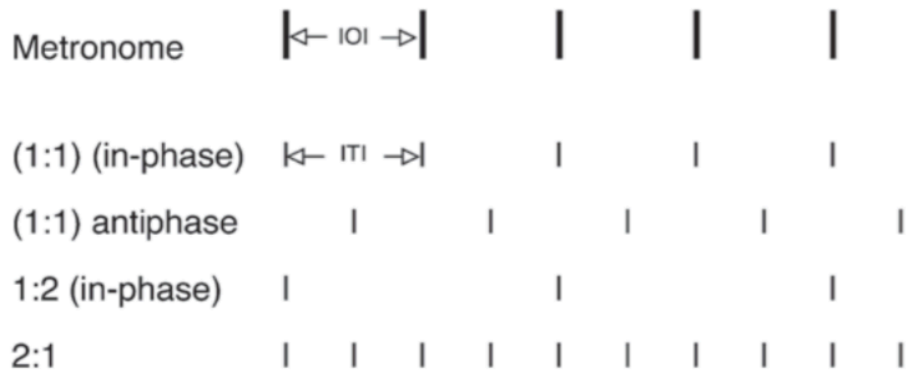


Figure 3.1. Common coordination modes. The bigger vertical bars represent metronome beats, the smaller vertical bars represent taps.

SMS occurs in time with what is commonly known as the *beat*. The beat could be defined as the most salient event of rhythm, which is called *tactus* (Repp & Su 2013: 416). However, as Repp & Su said, the beat "may not always be marked explicitly by the event onsets in the stimulus" (2013:416). It is not completely unambiguous where the beat falls in a sound pattern but is to be decided upon in some way or another (by the participant, experimenter, by context or notation). Repp & Su (2013) suggested that tapping would be coordinated with an external rhythm, but "synchronised with the internal periodic process that marks the beat" (2013: 416) and the synchronisation with the beat happens as a consequence of tapping to these external events. Martens (2011) conducted a study aiming at investigation of the tactus of music as it is tapped to. He showed that tactus depends on the tapper and categorized tappers into three groups: *surface tappers* (often non-musicians), who generally tap with the fastest beat in the music, and sometimes fail to synchronise; *variable tappers*, who choose beats of various rates; and *deep tappers*, who most often tap with a slow metrical level, which the author attributed to the, so called "subdivision benefit" (Repp 2005: 433), i.e. the notion that if a beat is subdivided by additional tones then the subjects' tapping variability decreases.

Repp (2005) discussed a range of different aspects of SMS, not all of which will be looked into detail here. These aspects are: the role of intention, i.e. whether the participants intend to move and to coordinate the movement with the external referent, since some movement may occur unintentionally (2005: 970); and tapping rate limits, i.e. what the maximum and minimum rates can be at which successful tapping can occur. Repp (2005: 973) also reviewed the phenomenon of negative mean asynchrony (NMA), which is an anticipation tendency, meaning that taps tend to precede the stimulus tones by tens of milliseconds,

rather than being *anywhere* around the tones. This is peculiar to non-musicians tapping in synchrony with a simple metronome.

Repp also covered the issue of individual variability in SMS. Generally, "musically trained and practiced individuals can achieve a standard deviation of asynchronies as small as 2% of the IOI or ITI duration in 1:1 in-phase tapping" (2005: 975). These asynchronies are at least twice as long for non-musicians as for musicians. The smaller the mean ITI, the smaller the variability of asynchronies and ITIs, i.e. ITI variability depends on ITI duration. The asynchronies were also found to depend on the mode of synchronisation, i.e. whether it is in-phase or anti-phase (smaller for anti-phase than in-phase), and on the modality of the metronome: auditory metronomes give results with smaller variability than visual metronomes (Repp 2005: 975). This is important for the present study because when investigating learning, information about individual differences is needed to explain any potential variability in the data.

Repp and Su (2013) review studies which support Repp's (2005) account, i.e. which found adult musicians to be able to synchronise their tapping much better than adult non-musicians, with much lower tapping variability, or standard deviations (Repp & Su 2013: 404). However, other studies found no differences between musicians and non-musicians on some aspects of SMS. For example Repp (2010) tested musicians and non-musicians on their synchronisation skills and perception of timing. He used three tests for this purpose: phase shift correction test, a period correction test and a tempo continuation test. Phase shift was a shortening or lengthening of only one IOI, so that the next stimulus occurs either early or late – and thereafter the period remains the same. The phase shift test aimed to measure the time needed for the participant to respond to phase shifts, by pressing a key at the end of sequence if they heard a timing perturbation and another key if they did not hear a timing perturbation. Period correction was a change in tempo of IOIs to faster or slower and the task was to continue tapping at the final tempo after the sequence had finished. The tempo continuation meant simply continuing to tap at the same tempo as the IOI sequence and the test asked the participants to continue tapping at the same tempo after the sequence of stimuli ended. The results showed that non-musicians performed better than musicians in the phase shift correction test, while musicians performed better on the other two tests. As well as this, Hove, Spivey, and Krumhansl (2010) conducted a study on tapping performance using "flashing visual metronomes with visual metronomes containing a spatial component, which was either compatible, incompatible, or orthogonal

to the tapping action" (2010: 1525). They found that the duration of musical training did not significantly correlate with tap timing or synchronization success. Repp, London, and Keller (2013) compared musicians' vs rhythm experts' (highly trained percussionists') performance on a synchronization-continuation task, which is a task in which participants synchronise their tapping with a rhythmic stimulus and are asked to continue tapping in the same way after the stimulus is switched off. They found that only some of the rhythm experts performed better than the non-percussionist musicians, and even then only at faster tempi.

In summary, the evidence on how musical and specifically rhythmic training affects SMS remains mixed. Therefore it will be important for the present study to remember that there may be individual differences in participants tapping performance, which may arise from e.g. musical training, or musical aptitude.

3.1.1.2. SMS and speech & language

Over the last decade, research has begun to address tapping to speech, and the various common aspects and applications that SMS research has with speech and language research. For example, several studies have found correlations between Speech and Language Impairments (SLI hereafter) as well as dyslexia and children's and adults' SMS ability. Some of these studies will be reviewed here. As well as this, links between SMS and memory and attention have been shown and these studies will be briefly outlined before. Finally, it has also been shown that brain responses are more consistent in those who are able to perform an SMS task in a more stable way. These areas are of interest to the present work because there is a potential for crossover for the findings from these research areas to the foreign language learning area.

Thomson et al (2006) investigated adults with dyslexia to test if specific rhythmic deficits in these subjects could be found. They used a variety of psychometric, phonological, psychoacoustic and motor tests. The results showed significant differences between adult dyslexic and control groups, and showed that rhythm perception measures were related to literacy (2006: 342). In the participants' rhythm production, the authors found that the within-subject variability of ITIs was the only measure that yielded a group difference result, such that the variability was higher in the dyslexic participants.

Corriveau & Goswami (2009) investigated tapping in synchrony with an auditory rhythm as well as self-paced tapping in children with SLI, compared to controls. The SLI children showed poorer performance on a metronome-paced tapping task compared to controls as well as greater inter-subject variability (2009: 127). As well as this, their taps at the slowest rate occurred significantly earlier, i.e. with a longer anticipation time, than those of the controls. No differences were found in the SLI and control children in the self-paced tapping task apart from the slowest rate, in which the SLI children exhibited the same result as in the metronome-paced task, i.e. higher inter-subject variability. The authors concluded with a recommendation that exploring the role of interventions which use rhythmic training could be worthwhile (2009: 129).

SMS also appears to have a link to memory. Miles, Nind, Henderson, and Macrae (2010) conducted a study in which they asked participants and a confederate to repeat words they heard over headphones, alternating the words between the confederate and the participant. They aimed to test whether the coordinated behaviour would induce similar memory effects to the so called self-reference effect (SRE), i.e. the better recall of information related to self than to others, which however disappears when there is an established relationship with the 'other', such as e.g. close friendship. During the experimental task the participant heard a regular metronome beat that they had been instructed to move their arm to. While the participant moved their arm, the confederate moved her arm either in phase or anti phase with the participant. This was synchronised with their word production, i.e. a word was interchangeably produced every 3 seconds by either the participant or the confederate. A subsequent surprise memory test showed that participants were able to recall self-produced words in the anti-phase coordination condition, while in the in-phase coordination condition they were able to recall the words equally well. In other words, the participants were able to better recall words they produced when the arm movement was more like turn-taking in a conversation than when it was synchronous. The authors interpret these results to be showing that the SRE, i.e. self-relevant information (such as the words produced by the participant in this experiment) is weakened when stable interpersonal synchrony is established, in a similar way as in long-standing relationships (2010: 9).

There is also a link between SMS and attention. Tierney & Kraus (2013a: 225) predicted that "since auditory-motor synchronization requires monitoring of the relationship between motor output and auditory input" those who are better at tapping to the beat would also

exhibit better performance on attention tests. They used a variety of tests to collect data on the participants reading, attention, IQ, and temporal processing as well as synchronised tapping. Tapping was measured using paced tapping, i.e. synchronising taps with an external auditory beat, and, what the authors call unpaced tapping, by which they mean that "the practice session was followed by a period of silence equivalent to 20 stimulus presentations, during which time the subject was asked to continue tapping as if the sound were still present" (2013a: 229).

Tierney & Kraus found that there was a relationship between the participants' tapping variability and three sets of behaviours. First, they found a negative correlation between their ability to read words and non-words and tapping variability, i.e. less variability correlated with higher reading score. Second, a correlation between tapping variability and their auditory (but not visual) attention, such that an auditory cue presented 0.5-1s before the target sound was detected faster than a visual cue presented 0.5-1s before a target picture was displayed. Third, the between tapping variability and the precision of the participants' auditory systems as observed in a tone detection test with backwards masking (where a noise burst follows the to-be-detected tone, and gives a measure of the participants' auditory system's fine temporal sensitivity) (2013a: 228). This correlation was such that the more sensitive the participants were the less variable their tapping was. Based on these correlations they speculate that training using synchronisation to the beat may lead to improvements in reading, some basic auditory functions as well as attention.

Tierney & Kraus (2013b) collected tapping and brainstem EEG data from participants. From the EEG data they calculated neural response consistency, by assessing the extent to which the brainstem's representation of sound varies from trial to trial, and inter-trial phase locking, which is a measure of the extent to which the phase of each frequency component of the brainstem response is consistent across trials (2013b: 14984). They then correlated these with the participants' tapping variability. They found that there was a link between the individual differences in the participants' auditory system function and their ability to synchronise with a beat, since they found that the less variable tappers exhibited a higher degree of consistency of the auditory brainstem response to sound (2013b: 14985).

One study which investigated what listeners perceived as the beat in speech (Lidji et al 2011). They also sought to relate the way the listener's synchronise their taps with to what their language background was found that listeners who spoke English (a stress-timed language) as their native language tapped more regularly than listeners who spoke French

as their native language (syllable-timed). As well as this, English natives tapped to a higher hierarchical level, i.e. the stressed syllable, rather than to every syllable, which was the case for the French native speakers. This study suggested that the events to which listeners entrain may also depend on their native language.

3.1.1.3. Beat Synchronisation Training

The benefits of training in beat synchronisation for speech and language have been most fully explored with regard to neurologically impaired participants. Though very different from the concerns of this thesis, this literature is reviewed briefly here because it offers the fullest account of how such training might actually work.

Many authors suggest there could be a benefit of training using beat synchronization, based on observed correlations between synchronization ability and other tasks. For example, Tierney & Kraus (2013b) conclude that training rhythmic abilities could lead to variable neural responses becoming more stable, more specifically that synchronisation of movement with a beat may improve these neural responses, and that this kind of training could be beneficial for those with auditory-based language impairments, since these kind of impairments are characterised by variable neural responses (2013b: 14981).

There is also some evidence that training in beat synchronization/musically cued training can lead to improvements on certain motor tasks. Schaefer (2014) reviews musically cued movement rehabilitation for post-stroke motor impairment, Parkinson's and Huntington's disease. She found that studies report mixed results for all these types of subjects, with the least evidence found for rhythmic auditory stimulation being useful in Huntington's disease, some evidence of benefits in rehabilitation in gait, but not upper limb movement, in stroke patients, and the highest amount of evidence for a specific type of Parkinson's disease patients (those who experience symptoms called 'freezing', i.e. motor blocks) benefitting from rhythmic auditory cues (2014: 4-5). She reports that many studies have shown that even though "clinical interventions using rhythmic auditory cueing do not show consistently positive results, it is argued that internal mechanisms of temporal prediction and tracking are crucial" (2014: 1). She identified four mechanisms through which motor learning that uses auditory cues can take place. The first one of them is *accelerated motor learning*, which means that repeated regular movement leads to faster learning and increased plasticity. The second one is *qualitatively different motor learning*, which occurs due to activating auditory as well as perception-driven motor learning areas which results in there being a "richer setting for motor learning and stimulating connectivity between

auditory and motor areas" (2014: 5). The third of these mechanisms is *acquiring temporal skills*, which occurs through music and rhythm training that relate to regularity detection and prediction which in turn impact motor control, cognition and cerebral processing (2014: 5). Finally, *motivation* is the fourth mechanism which can play a great role here, as it can lead to emotional engagement and improved rehabilitation experience. Schaefer (2014: 7) therefore suggests that these mechanisms arising from SMS need to be explored further in relation to the specific patient-type needs.

Dalla Bella et al (2015: 80) showed that musically cued gait training not only improved Parkinson's Disease patients' motor skills, but also their perceptual rhythmic skills. The training consisted of walking in time with a folk song with a superimposed beat in order to emphasise it and took place in three 30-minute sessions per week over a month. They showed that this training not only increased the patients' stride length and speed and their sensorimotor timing abilities, but also their perceptual timing which exhibited itself in improved discrimination of the durations of stimuli.

As Repp and Su (2013: 441) say, this research has clinical implications, however, entrainment research in this area may could also be relevant for training language and music acquisition. As well as this, there is some logic in considering reciprocity of perception and production of rhythm, i.e. that motor rhythms could also aid children's development of auditory rhythmic sensitivity (Corriveau and Goswami 2009: 129). And if these problems can be helped in children with SLI by using motor rhythms then it is also possible that motor rhythms can be used for foreign language learning. However, the evidence for speech is still very limited.

One recent development is *MusicSpeak* by Wang et al (2016) in the study of which Chinese EFL learners synchronised their speech with speech rhythm converted into musical notation. This study is reviewed in more detail in section 1.7.2 of Chapter 1 above.

3.1.1.4. Summary: SMS and speech and language

The SMS literature suggests that the ability to synchronise movement to a beat can have a positive impact on speech and language skills. More testing is needed to establish which specific linguistic skills benefit from SMS training. The present experiment will attempt to use SMS as a training method for improving listeners' comprehension of connected speech.

3.1.1.5. **The link between SMS and learning**

Just as in Chapter 1, section 1.9, the question why entrainment should help learning was asked, a parallel question could be asked about the link between sensorimotor synchronization and learning: why should SMS aid learning? Specifically, because the task in the experimental condition will be to tap to the *beat* the participants perceived in speech. And it is reasonable to expect that the *beat* would be the stressed syllable. Therefore, before the experimental condition is outlined, the rationale behind it should be addressed. Why should tapping to *stressed* syllables allow learners to improve their comprehension of *unstressed* syllables, i.e. the events in between the taps? As mentioned above, it may be that SMS will lead to stronger entrainment with the stimuli. It is possible that tapping to stressed syllables will allow participants to focus on the speech material in the intervals *between* the taps, and therefore to improve their comprehension of it. As well as this, the focusing on the intervals between the taps could be aided by yet another perceptual phenomenon, which may contribute to the learning process in Experiment 2, i.e. time-stretching. There is some research showing that perceptual "time-shrinking" (e.g. Sasaki et al 2002 & Nakajima et al 1991, 1992) as well as "time-stretching" (e.g. Sasaki et al 2010) can occur when participants are asked to pay attention to time. Under certain circumstances, i.e. that time intervals of equal physical durations can be perceived as intervals of shorter or longer durations, depending on what precedes or what follows them, or what environment they are found in. The perceptual time-stretching effect is of interest to the present thesis because learners were asked to tap to the beat in the speech they listened to while the beat itself was not the learning target, but the interval between the beat was.

To provide some examples of previously conducted studies, Fitzgibbons et al (1974) asked the question of whether the duration of a temporal gap between two auditory inputs is the only physical dimension which makes that gap perceivable. They used four tones of different frequencies, two of high (2,093 and 2,394 Hz) and two of low frequency (440 and 494 Hz) and the same duration for each tone. They concatenated these and placed gaps of different durations (20, 60 and 80ms) in different positions, such as between the two high tones (HGHLL, H=high tone, G=gap, L=low tone) between the high and the low tones (HHGLL), and between the two low tones (HHLGL). They found that participants were able to detect the gaps which were placed between tones of similar frequency much more easily than the gaps placed between the high and low tones. They interpreted this result as meaning that detection of temporal gaps is affected by the way we group auditory

information.

Later studies have shown that that perception of interval durations can be biased by a variety of properties of the stimuli. For example, it matters whether intervals are silent (Nakajima et al 1992, ten Hoopen et al 1993, Remijn et al 1999), or filled with sound (Sasaki et al 1993). Geiser & Gabrieli (2013: 1) found that intervals between perceptual groups are heard as longer than those within groups. Such perceptual elongation, Geiser & Gabrieli, could be in place in order to increase the perceived distinctiveness of the groups and therefore to strengthen group boundaries, which is useful in speech as well, e.g. when parts a stream of sounds into groups are related (Geiser & Gabrieli 2013: 4). As well as this, as Grondin (2010: 565) says "attending to the flow of time increases perceived duration and, somewhat similarly, that being distracted from time results in the shortening of perceived time."

The connection of these ideas to the experimental design is as follows: It is possible that making the participants attend to the timing of the stressed syllables and increasing their entrainment by asking them to tap to those events, will make the time in between the tapped-to events "stretch" perceptually. In other words, tapping to these events could bring participants' attention to the flow of time more than the control condition, or either condition of Experiment 1, and thus could increase the perceived duration of the intervals between the taps. This would also be in line with what Geiser & Gabrieli (2013) say about group boundaries, because in the present experiment the stressed syllables would help group the sound stream into feet and tapping would strengthen those group boundaries.

Additionally, such perceptual time interval stretching would be aided by repetition, i.e. by the fact that the participant becomes accustomed to the rate at which events occur, and to their tapping rate and therefore becomes able to form predictions which aid the tapping. Effectively, they might ultimately be able to perform the tapping nearly automatically which in turn would allow them to perceive the information in the intervals between the taps easier than those who did not perform any tapping. This information contained in the intervals between taps would be the cues to the unstressed syllables, which in this experiment were e.g. determiners (*a, the, her*), prepositions (*of, with, for, from* etc), or participle endings (*-ing, -en*).

3.1.2. P-centres

As discussed above, the fact that humans are able to synchronise their movement with a stimulus, with a certain degree of accuracy, seems to be well established. However, researchers have also been interested in finding out what it is in the signal that our SMS tasks are asking us to synchronise with. Or, what exactly do we tap to when we intend to produce taps synchronised to an event? We can probably assume quite safely, that we would tap at the moment in time at which we think the to-be-tapped-to event occurs, e.g. the beat in music. This is where the notion of P-centres, or perceptual centres, comes in. This is important for the present because the participants in this experiment were asked to tap at the moment at which they perceive the beat in speech to be. The review below reveals that there is some literature that suggests that P-centres can be different for speakers of L1s. If that is the case then, if learners of one can be directed to develop an awareness of how P-centres work in their target L2, through SMS training, then this could aid their comprehension of the L2.

Morton et al (1976: 405) first defined the P-centre of a word as the "psychological moment of occurrence" of a sound. The concept stemmed from attempting to record stimuli at regular intervals, specifically lists of numbers (i.e. the words *one*, *two*, *three*, etc). Examining the acoustics of such lists was how Morton et al realised that "producing items at regular intervals was not simply a question of having the onsets at regular intervals" (1976:405), and so the question arose: what did occur at regular intervals, if not the onsets?

As Scott says (1998: 5) "[t]here is no one signal for which the absolute P-centre is known. Thus any direct measure of P-centre location […] is impossible, since no signal could be used as the baseline against which other signals would be compared." In other words, the P-centres of sounds can only be measured as relative to one another. Therefore, the only viable way of measuring these is a rhythm-setting task, as used by Morton et al. Morton et al (1976) recorded spoken digits from *one* to *nine* and used a paradigm in which they presented listeners with a pair of these words in alternation. They had one of the words occur at regular intervals and the participants' task was to adjust the timing of when the other word would play using a knob, until they perceived these words as occurring at regular intervals. Observing these timings, the authors then concluded that timing of the P-centres would require some complex calculations, which remained to be discovered. This showed that indeed, acoustically the onsets were not what was aligned in temporally equal intervals, such that, e.g. the onset of the word *seven* needed to be 80ms before the onset of

the word *eight* would need to be, in order for the sequence to sound regular. The authors did not have a clear answer as to what exactly it was that was spaced at regular time intervals for the listeners, and conclude that this would be "a complex function, though the precise nature of the computation remains to be discovered" (1976: 406).

Morton et al (1976) also investigated how subjects time their productions if asked to produce two words in alternation, so that the intervals between the words remain constant. As they say: "If the P-center for one sound is time P_1 from the onset and that for the other is time P_2 from its onset, then onset-to-onset times will alternately be $(T + \tau)$ and $(T - \tau)$, where r, the relative P-center distance, is $(P_2\text{-}P_1)$" (1976: 406), as illustrated in Figure 3.2 below.

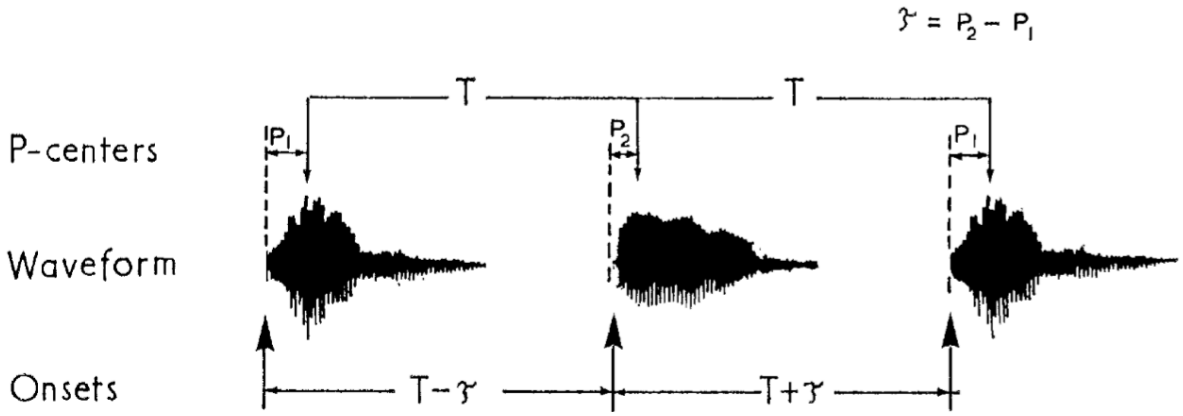


Figure 3.2. P-centre alignment of an alternating sequence of two stimuli (From Morton at al 1976: 407).

Morton et al (1976) conclude that the concept of a P-centre is important if trying to define what it is to perform an action "in time to the beat" (1976: 408) or in synchrony with another event. The key point is that, as Scott (1998: 4) says, the "perceptual regularity was found to not necessarily arise from evenly timed signal onsets."

Subsequent research has sought to calculate P-centre location in monosyllabic sequences more precisely using acoustic and phonological information about the syllable. Marcus (1981) found that both initial consonant duration and subsequent vowel and consonant duration had an effect on P-centre location, i.e. that over 90% of the variance in the P-centre placement could be accounted for by the following algorithm:

*P-centre = a * (acoustic onset to vowel onset duration) +*
*b * (vowel onset to acoustic offset duration) + c*

where *a* equals 0.65, *b* equals 0.25, and *c* "is an arbitrary constant representing the fact that we are only determining *relative* P-centre location of stimuli to one another" (1981: 253). As per this formula, the duration of the syllable onset consonant(s) has a larger influence on the P-centre location than the duration of the syllable rime, since the value of *a* is higher than the value of *b*.

Early work showed that P-centre location is relevant to speech perception on a number of tasks, and broadly confirmed Marcus' proposals. Buxton's (1983) aim was to test the hypothesis that P-centre-aligned rhythms are temporally highly predictive and allow more efficient speech perception than non-aligned rhythms. She conducted two experiments on P-centres, for which she used Marcus' (1976, 1981) P-centre algorithm.

In Buxton's first experiment the task was to detect a target digit in a randomised list of digits. The response was a key press and reaction time was measured. There were two sets of stimuli according to what was regular. In the first set, the stimuli were aligned so that the onsets of the words occurred at regular intervals, and in the other set what was aligned was the P-centres, as calculated using the above formula. The hypothesis was that the one that sounds more regular, i.e. the p-centre aligned one would yield faster reaction times because of the increased temporal predictability. And in fact this effect was found.

In the second study, Buxton investigated whether subjects would be able to trace excursions from a regular rhythm using a tapping task. She used sequences of monosyllabic words: either regular (P-centre aligned), or irregular (some words were displaced by 100ms from the regular timing). She found that in the regular condition, the standard deviation of these taps was 30ms, and in the irregular condition it was 55ms, and this difference was a significant one. She drew the conclusion that "sequences in the irregular condition did not have a rhythm that could be tapped out reliably, even when exactly repeated 10 times", and therefore that the "temporal pattern of irregular sequences simply cannot be learnt" (1983: 115). She reports that in a different study she had found that participants tapped equally as reliably (as the regular ones) to naturally-produced monosyllabic words, which lead her to the suggestion that "timing of natural speech has the same sort of inbuilt predictability" (1983: 115).

More recent research has added nuance to our understanding of the acoustic determinants of P-centres. Scott (1998: 5) conducted two experiments aiming to test what effect varying the onset and offset amplitude of the sound would have on the location of P-centres (Experiment 1), as compared to duration effects (Experiment 2). She asked two participants to adjust the timing of the intervals between the "ah" vowels in the same way as Morton et al (1976) had done. However, in Experiment 1 she varied either the vowel's onset rise time or offset decay time, and kept everything else, i.e. the duration, vowel formants, f0 and amplitude, constant. She found the subjects' adjustments were significantly different from absolute isochrony between the stimuli and concluded that perceptual regularity was influenced by the rise times but not by the decay times of the sounds. In other words, onset characteristics play a role in deciding where P-centres are located, but offset characteristics do not. The onset of a signal has the effect of moving the P-centre to a later point in time, the longer the onset rise (Scott 1998: 7).

Scott's second experiment examined the effect of duration on P-centres. Scott set out to test this, as the vowels in the numbers which were used by other studies (e.g. Morton et al 1976, discussed above), were of different spectral content and amplitude profile. This may have given rise to them being placed in at different times relative to each other, i.e. have affected when their P-centres occurred. The results showed a very weak effect such that the longer the duration, the later the P-centre occurred, but this was not statistically significant.

Research looking for acoustic correlates of P-centres found that they are often located *somewhere* close to the vowel onsets "within the transition between the onset of the syllable-initial consonant and the vowel" (Chow et al 2015: 56). While many studies find P-centres to be located *somewhere* near the vowel onset, neither acoustic nor articulatory correlates have been found so far which would demonstrate exactly what cues P-centres (cf Patel et al 1999, Hoequist 1983).

It is possible that there might be language differences, i.e. speakers of different languages would perceive or produce their P-centres at slightly different points. Chow et al (2015: 63) tested Cantonese speakers' synchronisation when producing various types of monosyllables in time with a metronome. They confirmed previous findings that manner of articulation of the initial consonant had an effect on the timing of production of the onsets e.g. syllable-initial fricative onsets were produced much earlier than other manners of articulation, relative to the metronome beat. They also found that the participants aligned

the syllable-initial consonants much closer to the metronome beat than their vowel onsets, from which they concluded that speakers of Cantonese "use the onset of syllables rather than the onset of vowels to align with a metronome", and suggested that this is possibly due to the fact that in Cantonese the syllable-initial consonants are shorter and less variable than in Germanic languages, which would allow syllable onsets to be "more reliable acoustic landmarks for P-center" (2015: 63).

To summarise, P-centres "have a complex relationship with syllable structure" (Hawkins 2014: 3), they can be influenced by segment or syllable duration, as well as by the properties of the end of the syllable (2014: 3), which results in the observed temporal variability of the word/syllable onsets relative to each other. As well as this, as Hawkins notes, the P-centre studies are often restricted to isolated words (or even isolated sounds, as in the case of the Scott 1998 study), which may result in them being produced in a more regular way than naturally occurring speech. Furthermore, the function that speech takes on, very much constrains the form it will have, i.e. the choice of words (morphology and phonology), as well as the grammatical choices (Hawkins 2014: 5). Therefore, as Hawkins says, one way in which a more comprehensive idea of what P-centres are could be one that "includes acoustic properties, but also prior knowledge about how the communicative medium is structured and functions" (2014: 5). Put simply, most research on P-centres focuses on word lists but not on connected speech. The question that arises then is the following, if the temporal location of stressed syllables in utterances was being considered would the same P-centre effects apply as are to be expected in monosyllabic sequences? Last but not least, different languages may permit different degrees to which regularity in terms of beats can be achieved (Hawkins 2014: 6) and speakers of different languages may have different "articulatory reference point[s]" in terms of speech timing (Chow et al 2015). If these points indeed are different, then would the way listeners tap to the beat in speech aid their comprehension of their target language speech in any way, by perhaps, indicating the timings that their target language uses?

3.1.3. Some phonological features of Chinese, General American and Canadian English

Experiment 2 works with Chinese EFL learners as well as native speakers of English of a different variety, i.e. US and Canadian English. In this section, some general features of Standard Mandarin Chinese and General American together with Canadian English and

will be briefly outlined. The English varieties are also briefly compared to Glaswegian English.

3.1.3.1. Chinese

Chinese is said to have seven main dialect families, each of which can have many dialects (Duanmu 2002). As Norman (1988: 3) says, for a variety or historical and socio-cultural reasons "it is very difficult to draw sharp boundaries between the different varieties of the language" spoken in China, and even the standard language is an amalgam of different varieties, such as Peking dialect, some literary language and other dialects. The largest of these dialects is Mandarin. Standard spoken Chinese was proposed and adopted in the middle of the 20th century (Duanmu 2002). This standard language is often called Standard Chinese, but as Duanmu says, it is also referred to as "Beijing Mandarin, Standard Mandarin, Mandarin Chinese, or simply Mandarin" (2002: 5). This standard variety is now used by the majority of speakers in China, often alongside their other local dialects. Therefore, for the purpose of this thesis, some of the main prosodic features of Standard Chinese (SC hereafter) will be briefly introduced. These features are syllable structure, vowel reduction, stress, tone and rhythmic classification.

A SC syllable corresponds to a morpheme, or a written graph (Duanmu 2014: 423). When a polysyllabic word occurs, it is mostly due to translation of foreign names and happens via concatenation of existing syllables (Duanmu 2014: 423). The maximal syllable, i.e. the longest possible syllable in SC, is one which contains CGVX, C being a consonant, G being a glide, VX being a diphthong, a short vowel plus a consonant, or a long vowel (Duanmu 2014: 423), e.g.:

[kʰwai] [tʰjan] [kwaŋ] [kwa:] [tswan]

"fast" "day" "light" "melon" "diamond" (examples from Duanmu 2014: 423).

The minimal syllable, i.e. the shortest possible one, in SC is normally CV, with an occasional V on its own. As well as this, SC has a syllabic C, i.e. consonants that constitute syllables on their own as they are words on their own and there is no opening of the mouth when pronouncing them, e.g.:

Shanghai [n] "fish"

Cantonese [m] "not" (examples from Duanmu 2014: 426).

There are two types of syllable in terms of weight: heavy (also known as full) and light (also known as weak). The heavy syllables are long in duration, carry lexical tone, cannot be reduced and can be stressed. In light syllables, the vowel undergoes reduction (often all the way to schwa), and can be shortened, and the coda is often deleted. Light syllables do not carry lexical tone and are not stressed. Duanmu reports (2014: 429) that the difference between a heavy and light syllable in Chinese is clear to native speakers, however judging the difference in weight between two heavy syllables, in e.g. a compound word, poses a problem to speakers of Chinese.

Another aspect of Chinese is that it is a tonal language, i.e. it uses pitch to differentiate between meanings of words which are otherwise pronounced the same (Zhang 2014: 443).

A study by Mok & Dellwo (2008) compared the rhythm of Beijing Mandarin and Cantonese with French and Italian, which are both classed as syllable-timed, and English and German, which are both classed as stress-timed. It showed clearly that both Beijing Mandarin and Cantonese place in the syllable-timed area, as measured by various acoustic rhythm metrics (VarcoC, %V and deltaC). Some slight differences between the two dialect families of Chinese were shown by these measures, which as the authors say, reflected the fact that there is no vowel reduction in Cantonese and every syllable carries a lexical tone, whereas there is vowel reduction in Mandarin and lexical tone is absent in light syllables (Mok & Dellwo 2008), which would lead to the conclusion that, canonically, Mandarin is less syllable-timed on these metrics than Cantonese is.

As Mandarin is typically classed as a language with a rhythm that is different to English though, i.e. syllable-timed, it is possible that training focused on rhythm would particularly benefit comprehension of connected speech for Mandarin learners of English.

3.1.3.2 Canadian, US and Glaswegian English – common features

All three varieties of English allow the same types of syllable structure, from a V only (e.g. *a*) to as many as three consonants in the syllable onset and in the coda (Setter 2006), i.e. CCCVCCC (e.g. *strengths*). As well as this, they all have weak and strong syllables. In strong syllables, the vowels tend to be longer, and in weak syllables vowels tend to be shorter. These shorter vowels are inherently shorter, rather than being shortened. These vowels in English are /ə ɪ ʊ /, or a syllabic consonant can occur instead (Setter 2006: 765). As well as this, all three varieties are normally considered rhotic, though there are certain exceptions. They differ slightly in their vowel inventory (some details of this also below).

As far as the rhythmic properties are concerned, the General American and Canadian and Glaswegian English varieties are generally classed as stress-timed (Mok & Dellwo 2008, Charles Boberg, personal communication), though different varieties can exhibit various timing differences, as shown by e.g. Rathcke & Smith (2011), who looked at the British English varieties of Cambridge, Leeds, Bradford and Glasgow. Some of the rhythmic differences that pertain to the varieties are important in the current research, i.e. Glaswegian, US/Canadian English are briefly outlined below.

3.1.3.3. Canadian/US English vs Glaswegian – the main differences

Canadian English and American English both have the following 10 monophthongs: /i, ɪ, ɛ, æ, ʊ, ʌ, u, o, ɑ, ɜ/ and the following 4 diphthongs: /eɪ, aɪ, ɔɪ, aʊ/ (Wells 1982: 492 and 472). As far as specific features Canadian English is characterised by is the so-called Canadian Raising. This is centralisation of the vowel onset in diphthongs [ai ~ əɪ] vs [au ~ əʊ] before voiceless consonants (Wells 1982), and therefore the vowels in e.g. *night* and *time*, are of different qualities.

As far as consonants are concerned, Glaswegian English has some in its inventory that Canadian/US Englishes do not. These are /x/ and /ʍ/, though /x/ is infrequent and has a very low functional load. All varieties are rhotic, but they differ in their realisations of rhotics, such that in Glaswegian, taps, approximants and pharyngealised variants are all possible for /r/ in Glaswegian, but fewer are possible in Canadian/US English. As well as this, taps and flaps are allophones of /t/, not of /r/ (Stuart-Smith et al 2015).

These varieties also differ in intonation. Glaswegian has default rise as a default nuclear accent on declarative sentences (e.g. Cruttenden 2007), while US/Canadian have a fall as default nuclear accent on declarative sentences, although rise is becoming more common.

3.1.3.4. Summary – prosodic differences in English and Chinese

To summarise, there are several key differences between the prosodic systems of English and Chinese. Whereas lexical tone is a feature of Chinese, English does not use pitch lexically, but rather "post-lexically", for intonation purposes. This means that in English one word can be pronounced with different tones, which may indicate various implications, e.g. in some dialects a rise may indicate incompleteness, or a fall definiteness, but the lexical meaning does not change (Wells 1982: 5).

The stress systems of the languages are not straightforwardly comparable in part because English has many polysyllabic words (each with a primary and sometimes with a secondary stressed syllable) whereas Chinese has few, except for compounds. In English compound words it is fairly easy to agree which stressed syllable is more prominent, e.g. *oil lamp, pancake, blackboard* (from Duanmu 2014: 429). In Chinese compound words, it is much more difficult to know which is more stressed than the other. Duanmu offers an explanation, saying that in Chinese, both words in such compounds, are heavy syllables and therefore each carry a lexical tone. In English, however, typically, only one syllable has a pitch accent (2014: 430).

As far as the rhythms of English and Chinese are concerned, according to the literature, they belong to two different rhythm classes, Chinese being classed as a syllable-timed language, and English as a stress-timed language. However, as said above, Chinese does exhibit some features that are typically associated with stress-timed languages, i.e. vowel reduction (cf Dauer 1983). Though, it lacks others, such as complex syllable structure.

3.1.4. Comprehension of an unfamiliar native accent and training

Given that Experiment 2 is primarily aimed to investigate the role of tapping in improving learner comprehension, the following section now briefly discusses some studies that deal with comprehension of unfamiliar native accents within the same language, as well as effects of training on comprehension of these accents.

3.1.4.1. Comprehension of an unfamiliar native accent

Speech in a different accent of the same language can be hard to understand. This is reflected in behaviour in various ways, such as lower general intelligibility (e.g. Adank et al 2009), lower processing speed (e.g. Adank et al 2009, Floccia et al 2009, Sumner & Samuel 2009), or lower intelligibility of specific features not existing in the listener's own variety (e.g. Sumner & Samuel 2009, Lennon 2017, Unpublished PhD thesis, Smith & Rathcke 2017). It has been shown for a range of accent pairs, such as Glaswegian and Leeds varieties (Smith & Rathcke 2017), and Glaswegian and Cambridge (Lennon 2017, Unpublished PhD thesis), Plymothian, and Irish English (Floccia et al 2009), General American and NYC (Sumner & Samuel 2009), and a range of phonetic features, such as for example vowel length or rhoticity.

The situation is similar to L2 comprehension in that differences in the phoneme systems of the accents can matter, differences in phonotactics (like rhoticity) or in phonetic realization

can also matter for the non-native listener. As well as this, the situation is also similar to L2 comprehension in that experience, or exposure to the accent most likely makes a difference. And it's also similar to L2 in that short-term training can be effective, at least up to a point.

The literature shows a generally clear picture on comprehension of a different variety of English, with respect to both segmental as well as non-segmental features of those varieties: being a speaker of one variety seems to comes with a processing cost when listening to a different variety.

3.1.4.2. Training for improving accent comprehension

There is some evidence that training can improve processing even using very short term familiarisation, e.g. Maye et al (2008) who showed American English listeners adapted to an artificial accent after only a 20 min exposure, as e.g. Smith et al (2014) and Adank et al (2009) showed, this kind of improvement may be limited, or incomplete in various ways. It is limited as e.g. Smith et al (2014) showed that short-term familiarisation with an unfamiliar accent only benefited the listeners in proportion to how good their comprehension was to begin with. It is incomplete as, e.g. Floccia et al (2009) showed that short-term training which involved listening to different accents improved how well the participants understood the message intended by the speaker but it did not improve comprehensibility, because participants' reaction times did not decrease, and therefore the cognitive effort did not decrease. As well as this, Lennon (2017, unpublished PhD thesis) showed that the English English listeners living in Cambridge were able to quickly adapt to the same level the English English listeners living in Glasgow, after the short-term training. However, the English listeners living in Glasgow did not improve even further after this training, suggesting that such training was possibly able to improve their comprehension only to a certain degree, and no further.

Adank et al (2010) also investigated the role of imitation in improving comprehension of an artificial variety. The artificial variety was created by switching all tense-lax vowel pairs (e.g. /ɛ/ was pronounced as /e:/, and vice versa), by changing all instances of the /u/-vowel to /Y/ and by all diphthongal vowels being realised as monophthongal vowels (e.g., /œy/ was pronounced as /y/) (2010: 1904-1905). The study showed that Dutch listeners' comprehension of this made-up variety of Dutch improved only after the participants vocally imitated the accent, as compared to when they listened to accented sentences, listened and repeated in their own accents, listened and transcribed the speech, or listened

and imitated the speech without being able to hear their own productions. This intriguing study is the main one that looked at motor involvement in speech comprehension training and while one type of motor involvement did not improve comprehension, the other type did. Therefore, Adank et al's (2010) study points to a possible role for sensorimotor involvement, though there are still many unknowns regarding what type of training, if any, would be the most effective for comprehension of native speech of an unfamiliar variety.

In sum, there is some slightly conflicting evidence on the effects of short-term training on processing of an unfamiliar variety of the listeners' native language. Literature on training learners of foreign languages, rather than L1 of a different variety was reviewed in Chapter 1, Section 1.5.3. L2 training. L2 training studies seem to concentrate mainly on finding the details of which techniques (e.g. High Variability Perceptual Training, or Perceptual Fading) that can be most successful at training, while L1 of a different variety mainly look to find effects of familiarisation (with the exception of Adank et al 2010 who compared the roles of different types of training).

3.1.5. Summary

To summarise, firstly, there are plenty of suggestions that rhythm training could beneficial for human cognition, learning and language in both healthy and impaired participants, as well as for movement rehabilitation. While many suggest these benefits for L2 learning, the evidence for specifically that purpose is very limited.

Secondly, the psychological moment of occurrence of an event (P-centre) is not yet very clearly defined. The P-centre occurs in our perception but since perception and production are very tightly linked, speaking with a metronome has been used to investigate P-centre location (e.g. Chow et al 2015), but no consensus has been reached across languages nd it may be the P-centre location is different for different languages. It is also entirely possible that getting as close to the P-centre as *somewhere around the vowel onset*, may have to suffice, as it may not be a moment that can be defined as precisely as many would like it to be. On the whole though humans are generally very good at synchronising action with an acoustic signal, or can be trained to do it.

Thirdly, comprehension of an unfamiliar accent in the listeners' L1 and comprehension of L2 speech both come with a processing cost and are usually difficult for the listener. Could it be that sensorimotor synchronisation with speech in the target language is useful as a

training method for listeners' comprehension of fast casual English, especially for non-native language speakers but also for listeners of a non-native variety of English? This is the core question of the present experiment. In addition, this experiment also attempts to answer the question of what listeners are synchronizing to, when performing SMS to speech.

3.2. Overview of Experiment 2

Experiment 2 set out to test the hypothesis that tapping to the perceived beat in speech can lead to a greater comprehension of fast casual English speech. It was designed to test subjects' comprehension of function words in a pre-test, after which all subjects received a short period of training. After the training phase all subjects' comprehension was tested once more in a post-test.

In all three phases the materials were short sentences, such as *He was fat for a cat in a box*, which had a very similar metrical structure (i.e. *wwSwwSwwS*, where *w*=weak syllable, *S*=strong syllable). The materials were recorded using metronome-cued speech, which was produced at slow and fast rates by a male Glaswegian native speaker. The fast rates resulted in highly reduced forms of the weak syllables, which were the learning target for test subjects.

Participants were divided into two groups, who received different types of training after the pre-test. One group received training, during which their task was to listen to the sentences, decide what the beat in the sentences was, and tap to that perceived beat on a silent electronic keyboard which recorded their responses. The other group received training during which they listened to the same sentences as the other group, but their task was to tap the keyboard when they heard a randomly placed click sound that occurred during some of the sentences, rather than to tap to the beat.

There were three groups of participants. The first group were non-native learners of English, i.e. Chinese EFL learners living in Glasgow. The second group were native speakers of a different variety, i.e. Canadian/US speakers who were unfamiliar with the Glaswegian variety. The third group was a control group of 40 Glaswegian native speakers.

3.2.1. Research Questions

The main research question for Experiment 2 was whether training involving tapping to the beat perceived in speech would improve the EFL learners' comprehension of

English connected speech more than a control form of training which is not beat-based.

As well as this, Experiment 2 attempts to answer the question of what other factors can have an effect on such learning such as participants' musical background and English language proficiency.

3.2.2. Hypothesis

The hypothesis is that training involving tapping to the beat perceived in speech will benefit connected speech comprehension of Glaswegian English in both the EFL learners and the native listeners of a different variety, more than control training. The Glaswegian native group serves as a sort of control group. In other words, two controls are used in this experiment, one for training (Tapping vs Click identification) and one for accent (Chinese EFL, Canadian/US, and native Glaswegian). Since previous research has shown that Glaswegian English can also be generally more difficult for listeners of Glaswegian English background (Smith et al 2014), a small improvement was also expected in the Glaswegian native listeners.

3.2.3. Participants

There were 120 participants, in three different groups (n=40 in each). The first group, and the main focus of the experiment, were learners of English as a Foreign Language. Chinese EFL learners living in Glasgow were chosen, for two main reasons. First, the rhythm of the native language (all Mandarin, with a few additionally speaking Cantonese) of these participants is very different to English and is traditionally classed as syllable-timed. As advised by two experienced EFL teachers (Gemma Archer and Clare Thomson, personal communication), Chinese learners are also those who struggle with comprehension of casual English the most. There is also a large community of Chinese students at the University of Glasgow and they were therefore more available than speakers of other languages. Participants in this group were 40 Chinese EFL learners (12 male, 28 female) aged 17-33 (mean = 23.25, standard deviation = 3.09), who had been living in Glasgow for between 3 months and 27 months prior to the experiment (henceforth called "Chinese group"). This was controlled for so that no complete newcomers were tested, meaning they had at least a small amount of time to get accustomed to Glaswegian. The reason for excluding participants who have lived in Glasgow for longer than 3 years was to ensure the task was not going to be too easy. 31 of the participants were native speakers of Mandarin, 7 specified their native language

simply as “Chinese”), one was a native speaker of Hakka and one of Shanghainese.

The second participant group were 40 native English speakers of an accent of English other than Glaswegian, here, either Canadian or US English (henceforth Canadian/US group). Variability in whether these participants were US or Canadian English speakers (10 male, 30 female) was assumed to be unlikely to matter in that Glaswegian English is likely to be difficult for speakers of both varieties (Charles Boberg, personal communication)). These participants were resident in Montreal, were aged 18-35 (mean = 22.525, standard deviation = 3.25) and had had no previous interactions with anyone from Scotland.

Finally, 40 native speakers of Glaswegian English (11 male, 29 female) were tested as a control group (henceforth “Glaswegian group”). They were aged 17-33, apart from one participant aged 43 (mean = 22.75, standard deviation = 5.93), were resident in Glasgow, and had not lived in a different country for longer than 3 months. They were all monolingual speakers of English and came from various parts of the Greater Glasgow area.

3.2.4. Materials

There were three types of materials for the three different phases of the experiment: Pre-test materials, Training phase materials, and Post-test materials. The materials were designed so that all sentences in all three phases followed the same regular metrical structure of alternating weak and strong syllables: *wwSwwSwwS*, e.g. But it sat on the path of the duck (w= weak syllable, S=strong syllable).

As in Experiment 1, the weak syllables were the learning targets. They were carefully selected to instantiate certain types of word or morpheme which were expected to undergo reduction in connected speech. The types of words and morphemes chosen were:

- Determiners (*the*, *a* or *her*)
- Participle endings (*-ing*, *-en*) for verbs which are identical in present and past participle apart from the *–ing* and *–en* ending, e.g. *taken – taking, beaten – beating* and *eaten – eating*
- Prepositions (*for*, *from*, *in*, *on*, *with*, *of*).

The decision to choose these three categories was based on the results of Experiment 1, as they were the three categories that showed largest numerical improvement from pre- to post-test. (Note that Experiment 1's category of Articles included only *a* and *the; her* was a distinct category, which was the most difficult overall for participants, and therefore was included here within the Determiner category.)

3.2.4.1. Pre- and Post-test Materials

Two sets of sentences were designed for the pre-test and post-test (A and B). Each set contained 57 "learning targets" (i.e. gaps for participants to complete), though the number of sentences was uneven between the tests, i.e. there were 19 in Test A and 21 in Test B. All participants heard the same sentences, but the assignment of the set (A or B) to pre- or post-test was counterbalanced across participants within each group

The design of the pre- and post-test materials sought to avoid the possibility that the target words could be comprehended due to the context they were in. That is, the pre- and post-test sentences were designed so that they were as ambiguous as possible with respect to the target words: there were as few semantic cues as possible to which of the prepositions/determiners/participle endings were used. This was intended to ensure as far as possible that listeners would have to rely on phonetic cues only. While obeying these constraints slightly reduced the naturalness of the sentences, this was judged to be worthwhile to avoid the function words being predictable from context. Examples are in the following subsections.

3.2.4.1. Determiners

The sentences were also designed so that any one of the three determiners could be used in place of another, i.e. *a*, *the* and *her* could be used interchangeably, e.g. *And he's eaten her/a/the cake on the/a/her seat*. In total there were 33 determiners in each of the tests (set A and set B) i.e. each preposition occurred 11 times in each set.

3.2.4.2. Participle endings

Three verbs were used which are identical in their present and past participle, apart from the ending (-*en* vs –*ing*). This gave 3 verbs ending in –*en* and 3 verbs ending in -*ing*. The sentences were also designed so that they were ambiguous with respect to which participle could be used, e.g. *Or she's eating/eaten a meal for her son.*

In total there were 6 verb forms in each of the tests (set A and set B), i.e. each verb

form occurred once in each set.

3.2.4.3. Prepositions

Six different prepositions were used: *for, from, on, in, of, with.* The sentences were designed so that the use of pairs of them was ambiguous:

- *for* could be used where *from* was used and vice versa:
  e.g. *So I came for/from a show of a friend.*
- *on* could be used where *in* was used and vice versa:
  e.g. *And she's beaten a man in/on the chair.*
- *of* could be used where *with* was used and vice versa:
  e.g. *But we stayed on the road with/of the tour.*

In total there were 18 prepositions in each of the tests (set A and set B), i.e. each preposition occurred three times in each set. For the full materials, see Appendix F.

3.2.4.2. Training materials

38 sentences with the same metrical structure as the pre-/post-test sentences were designed for the training phase. Both the experimental, i.e. the tapping group, and the control, i.e. the click identification group heard the same sentences.

These sentences contained the same types of reduced words/morphemes as the test sentences, but were novel sentences. It was not necessary for the reduced words/morphemes to be ambiguous in the training sentences. Rather, they were designed for the target words to be contextually supported, in order for the participants to understand the sentence well and be able to learn from the context of the sentence.

Example training sentences are:
And he's taking her pup in a box.
And he spoke of a girl from her team.

The full set of training materials is in Appendix F.

3.2.5. Recording Procedure

One male native speaker of standard Glaswegian English, aged 26, was recorded producing the sentences. Recording took place in a sound-attenuated booth at the

University of Glasgow. An AKG SE 300B stand microphone was used to record the participant's speech, together with a Rolls LiveMix MX34c mixer and a Roland Duo Capture audio interface. During recording, the participant wore a set of Beyerdynamic headphones. The speech recordings were made using Audacity.

The speaker produced these stimuli at four different speech rates, from slow to very fast. In order to obtain stable rates, all speech was cued by a metronome beat. This means that the speaker heard a metronome over headphones at a certain speed for approximately 5 seconds before beginning to read the sentences aloud. He was asked to read them at the tempo provided by the metronome after the metronome was switched off. He read 5 sentences at a time, then stopped, listened to the metronome for approximately 5 seconds again and proceeded to read the next 5 sentences. This procedure resulted in considerably more natural-sounding recordings than in Experiment 1, where the speaker spoke simultaneously with the metronome beeps. All materials were recorded at four different metronome rates (80bpm, 120bpm, 160bpm & 180bpm).

3.2.6. Stimulus presentation in pre-test and post-test

All pre-test and post-test materials were sentences produced by the speaker at the second fastest rate, i.e. cued by the metronome at 160bpm. This rate was chosen because the fastest rate, i.e. 180bpm sounded too fast, slightly out of breath, and unnatural.

3.2.7. Stimulus presentation in training phase

For the training materials, four repetitions of each sentence were concatenated. The first repetition was at the 80bpm metronome rate and the remaining three were at 160bpm. As in Lidji et al (2011), there was a 350ms high-pitched tone (500Hz) at the beginning followed by a 700ms silence. After that the first repetition played which was at the slow rate. After that there was 1000ms silence followed by another 350ms high-pitched tone plus 700ms silence. Then the first slow repetition played followed by a 1000ms silence and a 300ms low tone (200Hz). The two repetitions after that were again the fast ones (cued by a 160bmp metronome rate), these were preceded by a low tone of 300ms duration and a 700ms silence and followed by 1000ms silence.

There were 38 different sentences for the training phase. For presentation to the

participants, these were concatenated into 6 blocks. This meant that four blocks consisted of six sentences (four repetitions of each) and two blocks consisted of seven sentences.

The same training materials were used for both the experimental and the control group. The only difference in the materials for the control group was that there was a 160ms long click sound which was added to about 20% of the sentence repetitions in a pseudo-random fashion. The purpose of adding this click sound was to give the control participants a motor task that was similar to the experimental task without being beat-based, i.e. the participants would be asked to tap the keyboard with their index finger when they heard the click sound. 34 of the 38 sentences had a click sound presented during one of their repetitions; the other 4 sentences had no click sound to reduce predictability. Clicks were randomly assigned to one of the three fast repetitions of a sentence, and were never presented on slow repetitions. Clicks were presented during one of the stressed syllables, so as not to acoustically mask the unstressed syllables which were the learning targets, and also to parallel the experimental condition, where participants were expected to tap on the stressed syllables.

3.2.8. Data collection

Data collection for the Canadian and US English listeners was conducted at McGill University in Montreal, Canada. For the Chinese learners of English and Glaswegian English native speakers, the data were collected at the University of Glasgow, UK.

3.2.9. Testing procedure

Participants were tested individually. The stimuli were presented to Glaswegian and Chinese participants over Sennheiser headphones and a MacBook Pro. Tapping responses iwere made using Casio CDP120BK keyboard using Cubase 8.5 software. The stimuli were presented to the Canadian/US participants over AKG K271 headphones and a Windows PC. Tapping responses were made using Roland RD700 keyboard using Cubase SX software. All tapping responses were recorded as MIDI files. The experiment was presented as a series of webpages.

The experiment started with the Edinburgh handedness questionnaire (Oldfield 1971), which was needed to establish the index finder of which hand the participants would use to tap on the keyboard during the training phase. For the EFL learners, the next part

of the experiment was 25 multiple-choice questions in order to test the participant's English language proficiency. As in Experiment 1, this was the official test used on the Cambridge English website (http://www.cambridgeenglish.org/test-your-english/adult-learners/).

After that the participants were offered a break before starting the main experiment. The main experiment had three phases: pre-test, exposure and post-test. These materials can be read in full in Appendix F.

The pre-test and post-test involved listening to sentences and filling in missing words on the computer screen. As in Experiment 1, participants saw each sentence on a computer screen with gaps, e.g.:
So I came _____ show _____ friend.

Their task was to fill out the blanks with the words they heard. They were made aware that each blank could contain more than one word. The tests were counterbalanced, so that half of the participants in each group heard Test A as pre-test and Test B as post-test, and the other half of the participants heard Test B as pre-test and Test A as post-test.

In the training phase all participants listened to the same 38 sentences, each repeated four times (1x slow and 3x fast), and divided into four blocks of six sentences and two of seven sentences, as described above. After each block the participants had a short break (of around 1 minute, which was the time needed for the experimenter to save the current file and open the next file).

The experimental group were told that during the first two repetitions of each sentence, i.e. the first slow and first fast repetition, they should simply listen to the speech (these repetitions were preceded by a high tone to identify them). The aim was for them to first hear the sentence at the slow rate in order to understand it as well as possible, and then to hear it once at the fast rate to decide at which points they thought they heard the "beat". They were then told that during the second two fast repetitions, identified by a preceding low tone, they should tap the index finger of their dominant hand on a key on the keyboard to what they perceived as the "beat" in those sentences.

The control group were asked to listen to the sentences as well as listen out for a randomly placed click sound and tap the index finger of their dominant hand on the keyboard as soon as they heard that click sound. Upon completion of the training phase, participants were offered the chance to take a short break if they wished to, and then to proceed to the post-test.

The task for the post-test was the same as pre-test. After the post-test, there was a questionnaire asking participants for demographic information relating to their age, gender, their first language, length of studying English (if a learner), any other languages spoken, and their musical training (years of training, hours of practice, self-rated ability to dance in time with music, self-rated ability to tap in time with music, etc). The full questionnaires are in Appendix G. Once they completed the questionnaire, the purpose of the experiment was explained to them, i.e. that it was to test whether rhythmic training could aid their comprehension of English connected speech. They were then given a chance to ask questions and comment, and were paid a small fee for their participation.

3.2.10. Coding the Pre-test & Post-test data

The answers were coded as 1 for correct and 0 for incorrect. In the case where two words had to be typed into the answer field, they were split into two, i.e. each word could be coded as correct or incorrect separately. That is, if the sentence contained two missing words in a row, e.g. *in a* in a sentence such as *He was fat for a cat in a box*, if the participant typed the correct answer to the first or second word only, e. g. *in the* or *on a*, they would receive one out of two possible points.

In case of obvious misspellings, e.g. *eatting* instead of *eating*, the answers were coded as correct. In case of any doubt as to the reason for the incorrect spelling of a word, the answer was scored 0.

3.2.11. Coding the tapping data

The MIDI files containing the participants' taps were converted using MatLab into DAT files. The DAT files were then compared in R against the file containing the segmented speech stimuli data extracted from Praat into CSV files. For each tap, the R script identified the syllable during which it occurred, and the timing of the tap. This allowed for further analysis of the tapping data, e.g. whether or not the participant

tapped three times per repetition as expected; calculation of the durations of the inter-tap intervals (ITIs), the means and standard deviations of these ITIs, and the timing of the taps with reference to other events in the speech signal.

3.3. Experiment 2 Results

In the results section, first, the comprehension data for all three groups, i.e. the Chinese learners of English, the Canadian/US native speakers and the Glaswegian native speakers, were considered together. Subsequently, each listener group is considered separately and a separate model is created for each group. This is because in the Chinese group there are some extra variables, such as length of stay in an English-speaking country, general level of English etc. After that, the tapping data is also considered.

3.3.1. Comprehension data from Pre-test and Post-test

The procedure for modelling was always as follows: starting with a model containing only the variables which would answer the research question (i.e. Training condition, Group and Test) and adding main effects and interactions one by one, checking every time whether the more complex model was a better fit (i.e. $p < 0.05$) to the data using a log likelihood test.

3.3.2. Analysis

<u>The dependent variable was:</u>

1) Score - 1 or 0 score for each correctly or incorrectly typed in word in pre and post-test

<u>There were three manipulated variables:</u>

1) Group (Glaswegian, Canadian, Chinese)
2) Condition (Tapping or Click Identification)
3) Test (Pre-test or Post-test)

<u>The following variables were also included:</u>

1) TypeLangKnown (if participants knew any other language than English they were coded as “StrT” (if it was stress-timed only) or “both” (if it was a syllable-timed; this is because all participants knew a stress-timed language, namely English)
2) Reduction Type (whether the missing word was a Participle, Determiner or a

Preposition)

3) Order (AB if Set A was heard in Pre-test and Set B in Post-test; BA if vice versa)

4) SelfTap (self-rated ability to tap in time with music; levels: None, VeryLittle, Moderate, Good, Very Good)

5) Years of musical training

There were also two random factors, i.e.:

1) Item

2) Participant.

3.3.3. The SelfTap variable

SelfTap is a variable that requires further explanation first of all. It was selected to represent the musical ability of participants. They were asked to rate their ability to tap in time with music. This variable was positively correlated with reported years of musical training and hours per week spent practicing; as seen in Panels A and B in Fig. 3.3 below, these correlations are positive. Exploratory modelling suggested that of the musical variables, SelfTap was the most predictive of participants' responses. Therefore, it was chosen to represent the participants' musical ability.

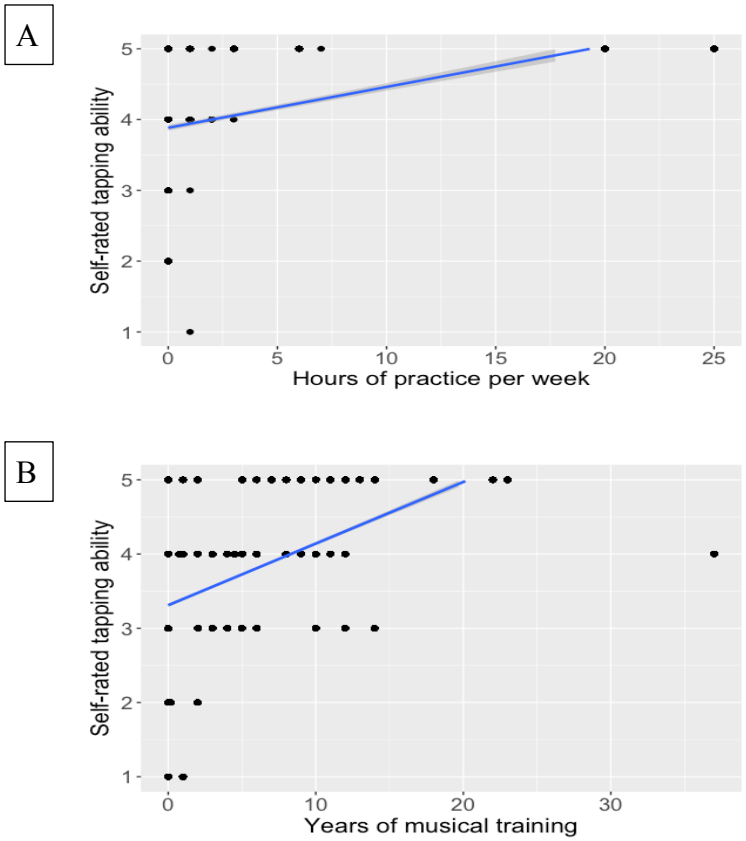


Fig. 3.3. Correlations between SelfTap variable and musical training variables. The x-axis, for A: hours

of musical training per week, for B: Years of musical training. The y-axis shows categories of SelfTap: 1 = None, 2 = Very Little, 3 = Moderate, 4 = Good, 5 = Very Good).

3.3.4. Model results

The final model contained the following predictors:

- Test (levels: Pre or Post)
- Group (levels: Chinese, Canadian/US, Glaswegian)
- Self-rated tapping ability (called 'SelfTap', levels: None, Very Little, Moderate, Good, Very Good)

The final model is given in Table 3.1.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.11616	0.18346	6.084	1.17e-09 ***
Test (Pre)	-0.22518	0.04694	-4.797	1.61e-06 ***
Group (Chinese)	-1.35718	0.09429	-14.394	< 2e-16 ***
Group (Glaswegian)	0.71629	0.09449	7.58	3.45e-14 ***
SelfTap (Moderate)	-0.14132	0.10078	-1.402	0.16084
SelfTap (None)	-0.88884	0.21066	-4.219	2.45e-05 ***
SelfTap (VeryGood)	0.03357	0.09493	0.354	0.72362
SelfTap (VeryLittle)	-0.40855	0.13595	-3.005	0.00266 **

Table 3.1. Final model table for all listener groups.

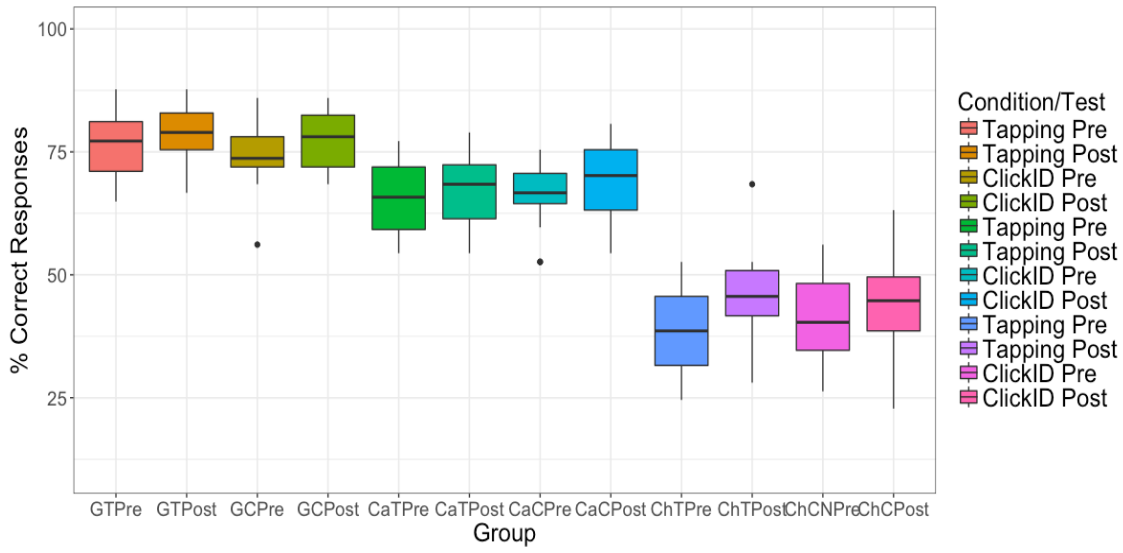


Fig 3.4. Mean % of correct responses (i.e. Score) on y axis. Group on x axis: The leftmost four boxes represent Glaswegian data, the middle four represent Canadian/US data and the rightmost four represent Chinese data. Within each group, the first two box plots are the scores of pre-test and post-test in the condition in which the participants were asked to tap to the beat they heard in the speech

Fig. 3.3 and Table 3.1 show that the Glaswegian participants had significantly higher scores than the Canadian/US participants ($z = 7.6$, $p < 0.0001$), who in turn had significantly higher scores than the Chinese participants ($z = 14.4$, $p < 0.0001$). The Glaswegians also had higher scores than the Chinese listeners ($z = 20.3$, $p < 0.0001$). For re-levelled tables, see Appendix H. Note that though a few of the Glaswegians participants come close to it, no one scored 100% correct in either of the tests, this means that overall the test is difficult, even for native speakers of the same variety of English. Figure 3.3 and Table 3.1 also show that there was a main effect of Test, such that overall scores were significantly lower on the Pre-test than they were on the Post-test ($z = -4.8$, $p < 0.005$).

No statistically significant differences of the interactions between Test and condition (Tapping vs Click ID), nor between Test and Group (Glaswegian, Canadian/US, Chinese) were found in this overall model.

Interestingly, a significant main effect of the SelfTap variable on comprehension was found. Those participants who self-rated as having a Good ability to tap in time with music did not score differently than those who self-rated as having Very Good and Moderate ability to tap in time with music. However, the participants who rated themselves as having Very Little ability to tap in time with music scored significantly lower than those who rated themselves as Good ($z = -3.2$, $p < 0.005$), as did those who rated themselves as having No such ability at all ($z = -4.2$, $p < 0.0005$). As well as this, those who rated themselves as having Very Little ability to tap in time with music scored significantly lower those who rated themselves as having Very Good ability ($z = -2.6$, $p < 0.05$), and lower than those who rated their ability to tap in time with music as Moderate ($z = -2.3$, $p < 0.05$). See re-levelled tables in Appendix H. This relationship can also be seen in Fig. 3.5 below.

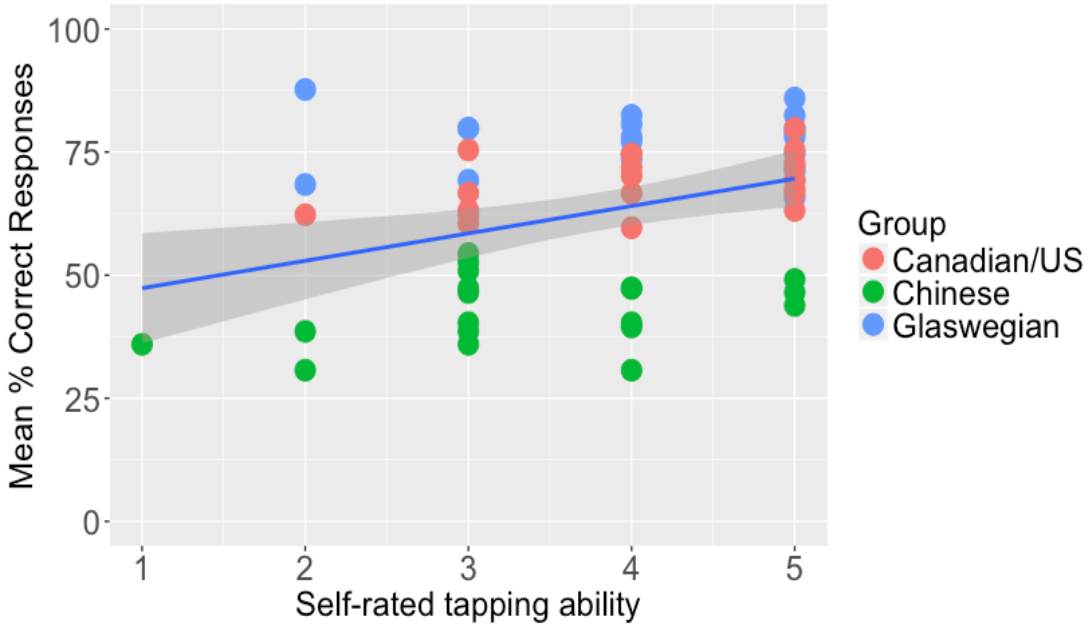


Fig. 3.5. Relationship between SelfTap and mean % correct responses across all participants. SelfTap scale 1=None, 2=Very Little, 3 =Moderate, 4 = Good, 5 = Very Good.

To sum up these results, there is an improvement from Pre-test to Post-test for all participants together; there are overall differences in score between the groups, such that Glaswegians score highest, then Canadians slightly lower and the Chinese score the lowest. The participants' self-rated tapping ability is also significant such that the higher the participants rate their own tapping ability, the better their score.

As mentioned above, there were a number of questions in the final questionnaire which were different for each of the listener groups, e.g. the Chinese EFL learners were asked about how long they had lived in an English speaking country, how long they had been studying English, etc. Therefore, separate modelling was performed for each listener group in order to explore in more detail what, if any, the relationships between these variables were.

3.3.5 Chinese Data analysis

The same procedure was employed for modelling the Chinese learners' of English data, i.e. starting with a model containing only Test, Condition and the random effects. Random slopes were not included because the model would not converge if an attempt was made to include random slopes. Then variables and their interactions were added one by one, checking which model fit the data better using log likelihood test (*anova(model1, model2)*). Interactions were of the significant variables were tested, plus interactions which were key to the hypotheses were tested even if their main effects were not significant. In

the Chinese group, some additional variables were included, because of these participants being non-native English speakers:

- CamEng - general level of English as measured by the Cambridge English "Test your English" test (http://www.cambridgeenglish.org/test-your-english/general-english/)
- StartEngl - age when they started learning English
- YearsLivEnglSpCtry - years lived in an English speaking country

The final model is given in Table 3.2.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.98382	0.40761	-4.867	0.00000113***
Condition (Control)	-0.03124	0.12557	-0.249	0.8036
Test (Pre)	-0.45287	0.11011	-4.113	0.0000391***
CamEng	0.09146	0.02277	4.016	0.0000592***
SelfTap (Moderate)	0.07299	0.13522	0.54	0.5893
SelfTap (None)	-0.69178	0.17028	-4.063	0.0000485***
SelfTap (VeryGood)	0.26385	0.15515	1.701	0.089.
SelfTap (VeryLittle)	-0.22974	0.14019	-1.639	0.1013
YearsLivEnglSpCtry	0.13234	0.05851	2.262	0.0237*
Condition (Control):Test(Pre)	0.27413	0.15382	1.782	0.0747.

Table 3.2. Final model results for the Chinese EFL learner group. The reference levels are 'tapping' for Condition, and 'Post' for Test.

In this model, the interaction between Condition and Test is marginally significant ($z = 1.8$, $p=0.07$). The criterion for treating a result as marginal is $0.05 < p < 0.1$. From this, as well as from the re-levelled tables (Appendix I), it can be seen that for the tapping group, Pre-test scores were significantly lower than Post-test scores ($z = -4.1$, $p < 0.005$), while for the click identification group, i.e. the control group, Pre-test scores were lower than the Post-test scores but not significantly so ($z = -1.7$, $p = 0.1$). Note also that in Pre-test, the Tapping group has marginally lower scores than the Click ID group ($z= -1.9$, $p=0.0534$, see re-levelled table in Appendix I).

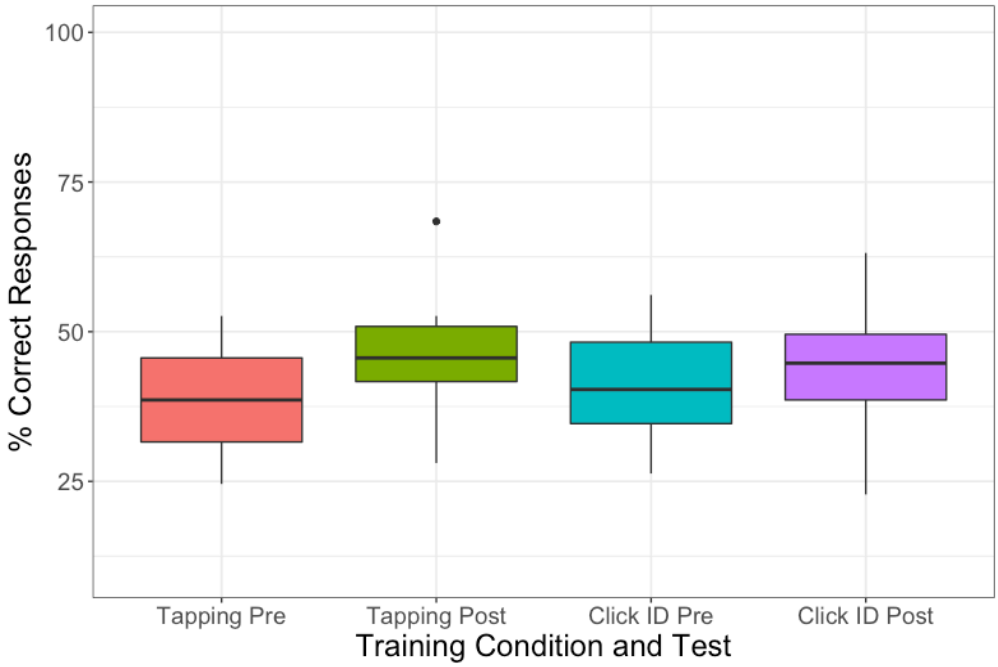


Fig 3.6. Boxplots representing mean % of correct responses on y axis, for Chinese participants only. Leftmost two boxplots show Pre-test and Post-test for the tapping participants; rightmost two for the click identification participants.

Figure 3.6 above represents these differences, and Fig 3.7 shows improvement from Pre-test to Post-test for individual participants. Note the presence of an outlier, i.e. one participant in the tapping condition who has a very high improvement score in Figure 6. However, if this outlier is removed, the improvement from Pre- to Post-test remains significant ($z = -3.3$, $p < 0.005$), and the interaction of Condition and Test remains marginally significant.

As can be seen in Figure 3.7 below, a smaller number of participants got worse in the tapping group than in the click ID group: in fact only one scored worse in Post-test than Pre-test in the tapping condition, whereas 8 scored worse in Post-test in the click ID condition.

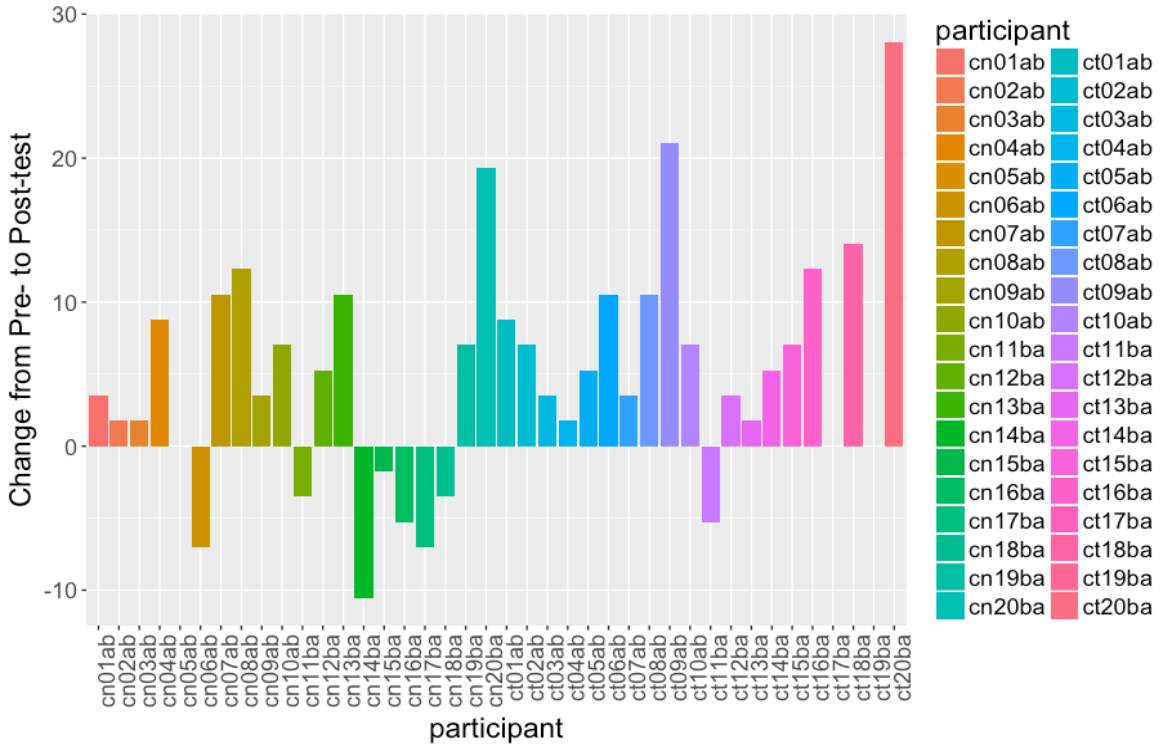


Fig 3.7. Change from Pre-test to Post-test per participant for Chinese participants only. Leftmost 20 participants (with codes beginning *cn*) did Click ID; rightmost 20 (with codes beginning *ct*) did Tapping training.

A significant effect was found of the participants' general level of English, such that the higher their level of English, the better their score ($z = 4$, $p < 0.005$ for the CamEng variable in the model output). There was also a significant positive effect of the amount of time spent in an English speaking country ($z = 2.3$, $p < 0.05$).

As in the model for all three participant groups, here the SelfTap variable also had a significant effect, represented graphically in Figure 3.8. Those who self-rated as having No ability to tap in time with music at all scored significantly worse than those who rated themselves as having a Good ability to tap in time with music ($z = -4.1$, $p < 0.005$). Those who rated themselves as having a Very Good ability to tap in time with music scored marginally better than those who rated themselves as Good ($z = 1.7$, $p = 0.09$); those who said they had Moderate or Very little such ability were not different than those who said they had a Good ability ($z = 0,05$, $p = 0.5$ and $z = -1.6$, $p = 0.1$, respectively, see also re-levelled tables in Appendix I).

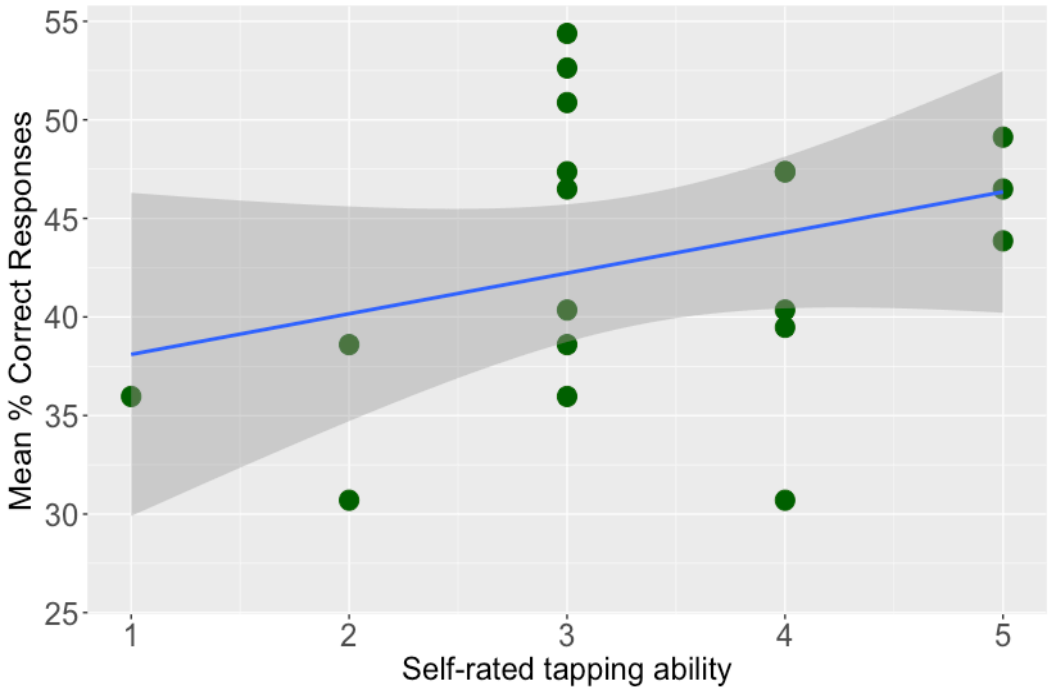


Fig 3.8. For Chinese participants only: correlation between mean % of correct responses on y axis, Self-rated tapping ability on x axis. Self-rated tapping scores from 1 to 5 correspond to the levels of SelfTap in order from None to Very Good, in the modelling above. When treated as a linear variable the regression line shows a weak positive correlation. Note that re-running models with SelfTap as a linear variable did not change the overall pattern of significances. Participant code names: the first letter in the name of each of these participants, *c*, stands for "Chinese", the second is either *t*, for those who took part in the tapping training, or *n*, for those who took part in the the control training condition, after that each participants is assigned a unique two digit number (i.e. 01, 02, 03 etc). The following two characters are *ab* or *ba* for each participant, which represents the order test A and test B were administered to the participant, i.e. as a pre-test or as a post-test.

Fig. 3.7 is a graphic representation of the correlation between self-rated tapping ability and the Mean % of correct score (overall across both pre-test and post-test). It shows a positive ($r = 0.54$, $p < 0.0005$) correlation between mean % of correct score and self-rated tapping ability, i.e. the better the participant self-rated on their musical ability, the higher their scores.

3.3.6. Canadian/US data analysis

The modelling procedure was the same as previously, i.e. starting with a model containing only Test, Condition and the random effects. Random slopes were not included because the model would not converge if an attempt to include random slopes was made. Then variables and their interactions were added one by one, checking which model fit the data better using log likelihood test (*anova(model1, model2)*). Interactions were of the significant variables were tested, plus interactions which were key to the hypotheses were tested even if their main effects were not significant.

Final model is given in Table 3.3.

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.0205	0.2919	3.496	0.000472***
Test (Pre)	-0.1407	0.146	-0.964	0.335271
SelfTap (Moderate)	-0.3435	0.2011	-1.708	0.087657.
SelfTap (VeryGood)	0.2854	0.1956	1.459	0.144597
SelfTap (VeryLittle)	-0.8972	0.3635	-2.468	0.013585*
Reduction Type (Preposition)	1.5901	0.4449	3.574	0.000352***
Reduction Type (Participle)	-0.3853	0.656	-0.587	0.55699
Test (Pre): SelfTap(Moderate)	-0.1314	0.2181	-0.603	0.546721
Test (Pre): SelfTap (VeryGood)	-0.2294	0.2121	-1.082	0.279353
Test (Pre): SelfTap (VeryLittle)	0.9796	0.396	2.473	0.013381*

Table 3.3. Final model results for the Canadian/US group. The reference levels are 'Post' for Test, 'Good' for SelfTap, and 'Determiner' for Reduction Type.

For the Canadian listeners, there was no effect of Condition, and no interaction of Condition with Test: that is, results were not affected by whether the participants performed tapping or click identification. The very slight numerical improvement in comprehension scores in participants across both groups, shown in Figure 3.4 was not significant, and Fig 3.11 shows that improvement and worsening each occurred for individual participants in both groups.

Improvement was, however, significantly affected by the participants' musical ability, as reflected in a significant interaction of the SelfTap variable and Test. This can be seen in Figures 3.9 and 3.10 below.

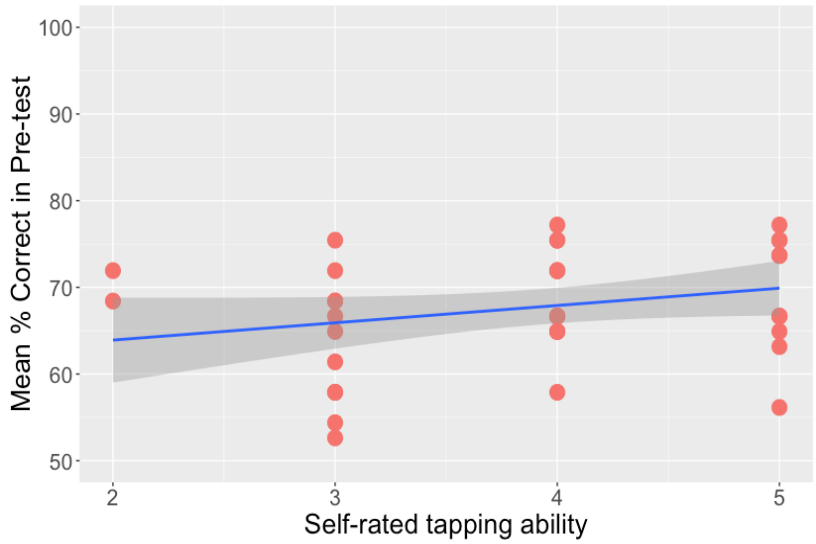


Fig. 3.9. Correlation between mean % correct responses in Pre-test.

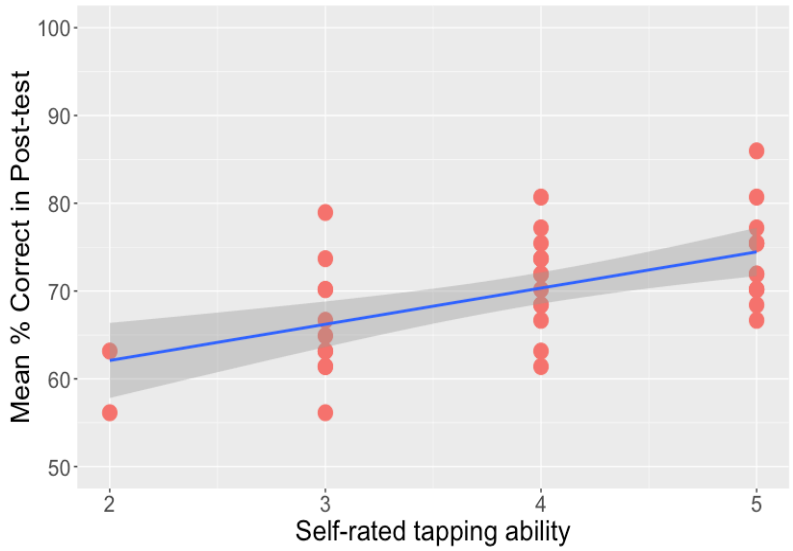


Fig. 3.10. Correlation between mean % correct responses in Post-test.

No participants rated themselves as having no ability at all. Those participants who rated their tapping ability as Very Good scored significantly higher on the Post-test than the Pre-test ($z = 2.421$, $p < 0.05$), those who rated themselves as Good improved numerically but not significantly ($z = -1$, $p = 0.33$); while those who rated themselves as Moderate actually did marginally worse on Pre-test than on Post-test ($z = -1.718$, $p = 0.086$) and those who rated themselves as Very Little, showed significant worsening ($z = 2.281$, $p < 0.05$). For re-levelled tables see Appendix J.

There was also a significant main effect of Trial Type Category, not found for the Chinese group. Scores on Prepositions were significantly higher than scores on Determiners ($z = 3.6$, $p < 0.005$), and on Verbs ($z = 2.8$, $p < 0.005$), while scores on Verbs were not significantly different than scores on Determiners (for re-levelled tables see Appendix J).

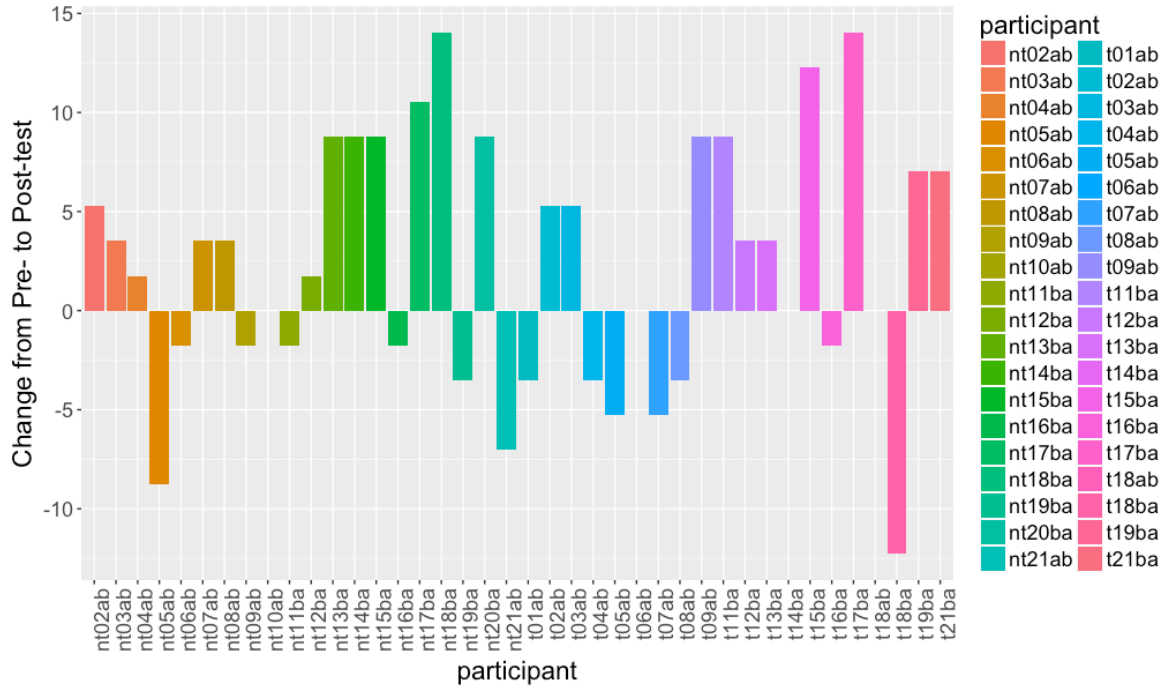


Fig 3.11. Change, i.e. difference from Pre-test to Post-test per individual participant for Canadian/US participants only. The first 20 are participants who did Click ID and the last 20 are the tapping participants (for this, also see participant code name). Canadian and US participants code names: the first character in the code name is either a *t*, for those who took part in the tapping training, or *n*, for those who took part in the non-tapping training, after that each participants is assigned a unique two digit number (i.e. 01, 02, 03 etc). The following two characters are *ab* or *ba* for each participant, which represent the order test A and test B were administered to the participant, i.e. as a pre-test or as a post-test.

A closer look at a graphic representation of the relationship between the SelfTap variable and Mean % correct per participant in the tapping participants (Figure 3.12 below) shows that the higher the participants' musical ability (as represented by the self-rated ability to tap in time with music, x axis), the higher the % score (y axis). This correlation was here: $z = 0.48$, $p < 0.05$.

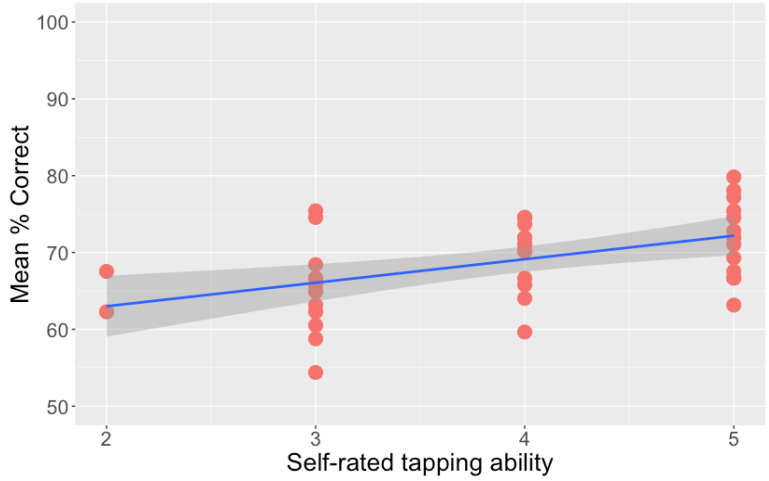


Fig 3.12. Mean % of correct responses on y axis, Self-rated tapping ability on x axis.

3.3.7. Glaswegian data analysis

The modelling procedure employed was again the same. The procedure was to start with the model containing only the key variables for the hypotheses (i.e. Test and Condition), adding variables and interactions one by one, while comparing models each time using a log likelihood test (*anova(model1, model2)*) and removing non-significant variables.

The final model looked as follows:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.18408	0.24984	8.742	< 2e-16***
Test (Pre)	-0.2757	0.09567	-2.882	0.00396**
SelfTap (Moderate)	-0.18823	0.22583	-0.834	0.40455
SelfTap (VeryGood)	-0.15454	0.16453	-0.939	0.34759
SelfTap (VeryLittle)	-0.96388	0.31748	-3.036	0.0024**

Table 3.4. Final model results for Glaswegian native listeners. Reference level for Test is 'Post' and for SelfTap is 'Good'.

In this model, there is a significant effect of Test, i.e. improvement overall from Pre-test to Post-test ($z = 2.9$, $p < 0.005$), but no effect of Condition (tapping vs click identification) and no interaction between Condition and Test. As well as this, a main effect of SelfTap was found. Those who rated themselves as having Very Good, Good and Moderate ability to tap to a beat all had significantly higher comprehension scores than those who rated themselves as not having any ability to tap in time with music ($z = 2.5$, $p < 0.05$, $z = 3$, $p < 0.005$, $z = 2.2$, $p < 0.05$ respectively; for re-levelled tables, see Appendix K). There were no significant differences in scores between those who rated themselves as Very Good, Good and Moderate (Appendix K).

There was no significant effect of Trial Type Category, nor any significant interactions.

Tables showing scores per participant in Pre-test and Post-test and a % difference per participant in the Click Identification Group are in the Appendix L.

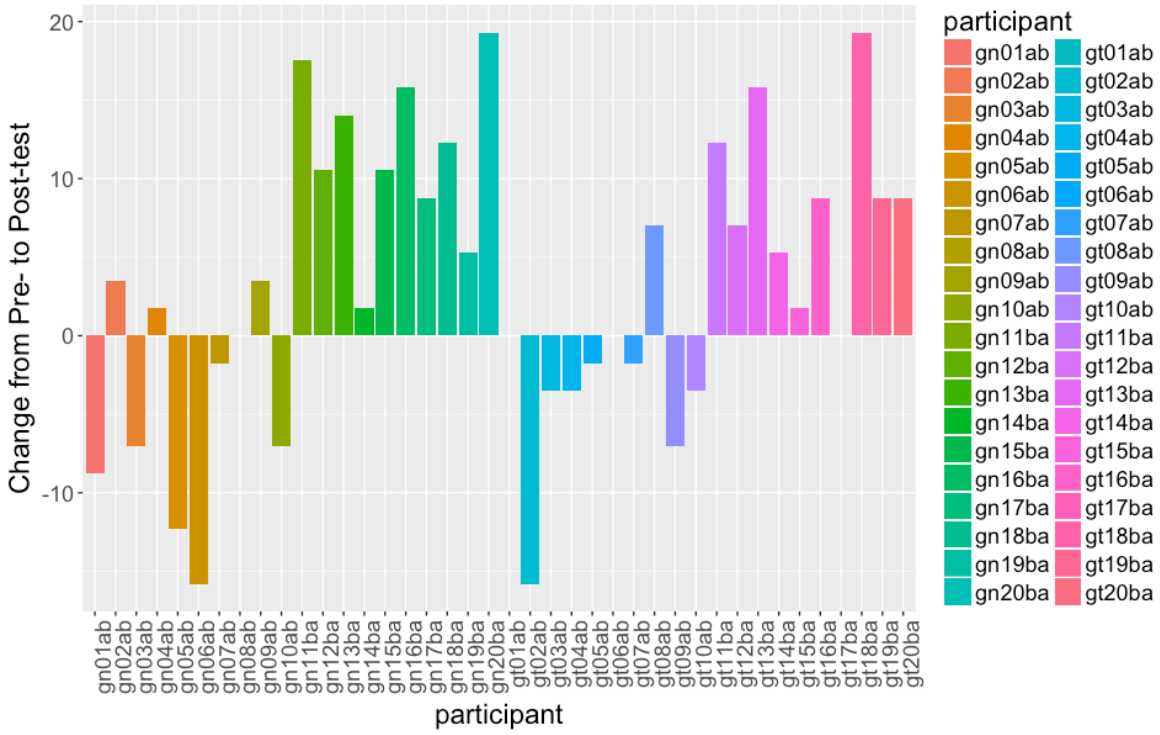


Fig. 3.13. Change, i.e. difference from Pre-test to Post-test per individual participant, for native listeners only. The first 20 are participants who did Click ID and the last 20 are the tapping participants (for this, also see participant code name).

In the Glaswegian speakers the pattern found in the Chinese and Canadian/US speakers' data is not found (r = -0.078, p = 0.7432) (see Fig. 3.14 below).

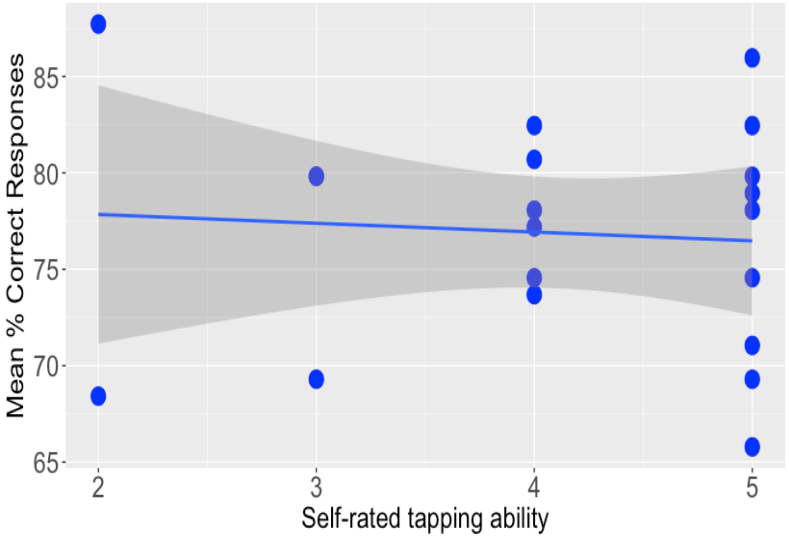


Fig 3.14. Mean % of correct responses (i.e. Score) on y axis, Self-rated tapping ability on x axis.

3.3.8. Summary of comprehension results

To sum up, the hypothesis that tapping training would improve the EFL learners' comprehension of Glaswegian English was supported. The other two groups, i.e. native

Glaswegian and Canadian/US English listeners' comprehension did not improve significantly. However, the Canadian/US group showed an interaction of Test and self-rated tapping ability, such that those participants who rated themselves higher benefitted from training in both conditions (i.e. tapping and click identification), and those who rated themselves as having very little tapping ability did not benefit from the training.

Apart from the comprehension data, the experimental group's tapping data was also collected, in order to look for the relationship between the tapping timing and the stimuli. These data are considered next.

3.3.9. Analysis of tapping data

The aim of the side question of this experiment was to investigate the relationship of the participants' taps to the speech events, and also of their tapping to their comprehension scores. In order to achieve this aim, both the participants tapping timings as well as the rhythmic structure of the speech they heard was analysed.

As mentioned above, the recording procedure of the speech materials which the participants tapped to was such that the speaker heard a metronome at a rate of 160 beats per minute (bpm) first, and was asked to produce these sentences as if he was aligning the stressed syllables with the metronome, after the metronome was switched off. 160bpm means that a beat occurs every 0.375s (60 seconds / 160 beats = 0.375s per beat). This metronome-defined beat interval is what will be referred to when examining the relationship of beat in speech and taps. However, there are a number of assumptions that were made about the speech as well as about the tapping performed to the speech. Firstly, it was assumed that the speaker was able to produce a regular beat in his speech by aligning some aspect of his speech to the beat and that he managed to maintain the beat rate defined by the metronome after it was switched off, i.e. that he did not speed up or slow down, or become otherwise variable. Secondly, even if the first assumptions about the speaker, or the materials provided by him, are valid the listeners' task was to find those beats, i.e. they did not have a metronome beat to rely on. This introduces more potential sources of variability, i.e. the location of the perceptual beat could be affected by acoustics, or by phonology, and it could also vary across listeners due to it being perceptual. As well as this, it needs to be noted that because participants were asked to tap to speech, beats are more difficult to define and their placement may depend on many factors, e.g. the type of consonant preceding or following a vowel; or

it could be different for different speaker/listeners and is therefore more subjective than a beat produced by a metronome. The acoustic or articulatory landmark speakers align with beats when they are asked to do so has not been defined with absolute precision (see section 3.1.2 on P-centres). Therefore, a variety of different candidates for the exact timing of the beat need to be considered. In other words, the speech signal the participants tapped to was coded for a variety of possible points at which beats could have occurred. How this was done, is described in the following section.

3.3.10. Acoustic analysis of materials

In order for the tapping data to be correlated with the speech signals which the participants listened and tapped to during the experiment, the speech materials were labelled in Praat. This was not an easy task: the materials were very difficult to segment due to the very fast speech rate. Basic segmentation and labelling were carried out by the author and were checked for consistency by the main supervisor.

All segmentation decisions were based on acoustic data (waveform, spectrograms, and f0 trajectories) in conjunction with auditory impression. Segmentation was carried out of two different kinds of syllable, of stressed vowel onsets, and of f0 peaks.

Two ways of determining syllable boundaries phonologically were employed. One was guided by the maximal onset principle, which says that all consonant sequences are assigned to syllable onsets as long as they do not form phonotactically illegal clusters (Selkirk 1981). The other way of segmenting syllables was based on the maximal onset principle but also respecting any morphological and lexical boundaries. In most cases the two approaches give the same result. However, an example of where they differed would be a sentence such as *And she's eaten a roll on the train*, the /z/ after *she* would belong to the following syllable and form /zi/ in the maximal onset principle, but in the morpho-lexical syllable, the /z/ would belong to the preceding syllable and form the syllable /ʃiz/.

While stressed syllable onset seems an obvious choice for the occurrence of beat, literature on P-centre location suggested that this location could also coincide with stressed vowel onset, or be very near *after* the vowel onset (section 3.1.2). Therefore stressed vowel onsets were also labeled separately. As well as this, f0 peaks, or pikes, were also considered. These are the f0 peaks of accented syllables, which, for example,

have been shown to be what speakers synchronise with in conversation (e.g. Loehr 2007, Ogden & Hawkins 2015).

For the acoustic segmentation of vocalic and consonantal intervals, the guidelines discussed in Peterson and Lehiste (1960) and refined by White and colleagues (White and Mattys 2007, White et al. 2009) were used.

The general segmentation rules were therefore as follows. Vowel onsets were marked as starting at the onset of voicing for the vowel. If voicing was continuous through the previous segment, the presence of vocalic formant structures was the criterion. Voice onset time was included in consonantal intervals. For nasal-vowel and vowel-nasal boundaries, discontinuities in waveform amplitude and waveshape (especially sinusoidality) were used. The start and end of fricatives were marked using the start and end of aperiodicity. In cases of fricatives, followed by a vowel, the start of the fricative was marked at start of major frication noise, however, auditory criteria were also employed. Due to the speech being fast, it was often unclear where to put a boundary between a fricative and a vowel, e.g. in such words as *she's* there can be strong frication all the way through the word, even in the most vowel-like portion. Here, the boundaries for the start of the fricative sound were marked as from where frication started to dominate and ended where the vowel portion started to dominate; as well as this auditory criteria were employed.

Approximant /r/-sounds were labeled as vowels. Where /r/ occurred in a cluster with a voiceless obstruent (e.g. /fr/) and was partially devoiced, only the voiced part of /r/ was included in the vowel segment (as in the procedure employed by Arvaniti 2012:357). Tap instances of /r/ were labeled as consonants. For the start of an /m/- sound, drop in amplitude and complexity of waveform as well as some loss of energy spectrogram were considered.

Where strong glottalisation or glottal stops occurred they were labeled as separate segments. However cases of weak glottalisation and nasalization, i.e. where there was weak or no cue in the spectrogram for them and they were only perceivable auditorily, were treated as secondary articulations and therefore not segmented as separate sounds. For each syllable, its prominence was marked as weak or strong.

F0 peaks were identified by looking for prominences in the f0 track curve in Praat and marked accordingly. An example picture of these can be seen in Fig 3.15 below.

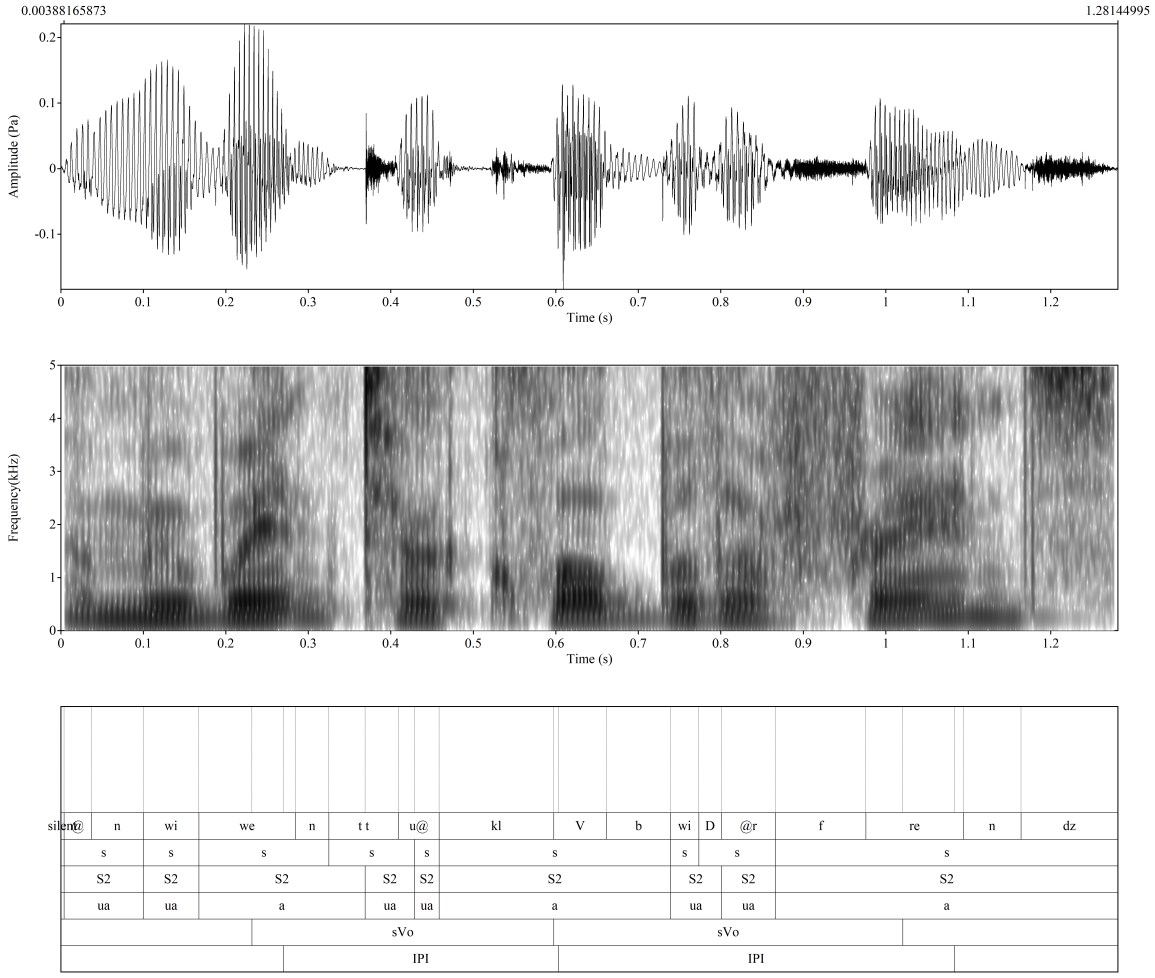


Fig. 3.15. Material segmentation procedure for the phrase *And we went to a club with her friends*. Uppermost panel shows the waveform, middle panel the spectrogram and bottom panel the textgrid. In the textgrid, six tiers can be seen. From the top to the bottom tier, these represent: segments, maximal onset syllable (symbolized by the character *s*), morpho-lexical syllable (S2), accented or unaccented syllable (a, ua), stressed vowel onset (sVo), and pikes (IPI, for inter-pike interval).

3.3.11. Results of acoustic analyses of materials

For each type of data as described above (i.e. onsets of stressed syllables of two different types - maximal onset and morpho-lexical), stressed vowel onsets and pikes, i.e. f0 peaks), the intervals between successive units of each type were measured. Based on these intervals mean, standard deviation, and coefficient of variation (i.e. standard deviation divided by the mean) were calculated. Coefficient of variation is a useful measure of variability in the data, because the standard deviation tends to co-vary with the mean, whereas the coefficient of variation is independent of the mean. It allows us to compare which of the four different measurements, i.e. pikes, maximal stressed

syllable onset, morpho-lexical stressed syllable onset or stressed vowel onset is the least variable, i.e. the most regularly occurring one.

Table 3.5 below shows coefficient of variation was the smallest for pikes and highest for stressed vowel onsets, meaning that pikes were the most regular inter-stress intervals in these materials. The mean duration of the inter-pike intervals is also close to 0.375s which corresponds to the 160bpm of the metronome the speaker heard before producing the speech.

	Pikes	Stressed syllable onset (Max-onset)	Stressed syllable onset (Morpho-lexical)	Stressed vowel onset
Mean (s)	0.378	0.376	0.373	0.4
Standard deviation (s)	0.041	0.044	0.046	0.062
Coefficient of variation (s)	0.109	0.117	0.125	0.156

Table 3.5. Calculations of intervals between various events in speech.

3.3.12. Tapping Data

Each of the sentences the subjects heard during the experiment had three stressed syllables. Therefore, if participants were able to identify the stressed syllables as the "beat" of the sentence, they would tap three times to each sentence.

The recorded tap times were analysed to establish the following, for each participant across all sentences:

- on average, how many taps were recorded per sentence (mean taps per repetition)
- the average duration of the ITIs
- the standard deviation of the ITIs
- the coefficient of variation of the inter-tap intervals was.

This section contains three graphs which represent the durations per participant, of the following measurements: mean inter-tap interval in light blue, coefficient of variation of inter-tap interval in red, and mean number of taps per repetition of each sentence in dark blue.

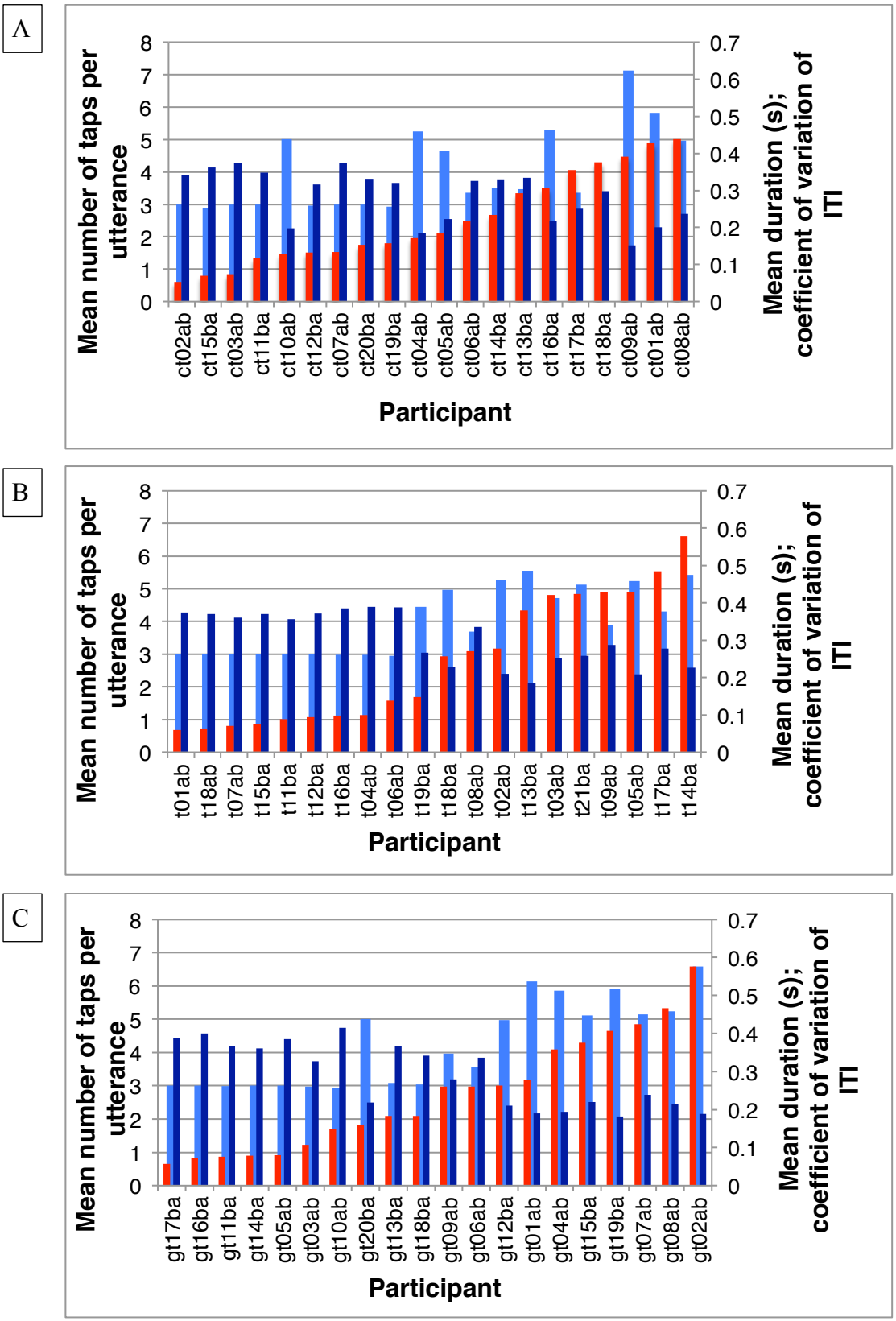


Fig. 3.16. Tapping data for all three listener groups: Panel A: Chinese participants. Panel B: Canadian participants. Panel C: Glaswegian participants. The y axis has three different measures: the light blue bars represent the mean number of taps per utterance (scale on the left hand side of the graph); the dark blue bars represent the mean duration of inter tap intervals (ITI) and the red bars the coefficient of variation of the ITIs (scale on the right hand side of the graph). All ordered from the smallest to highest coefficient of variation.

From graphs in Figure 3.16 above we can see that, generally, almost half of the participants in all three groups tapped three times per utterance, i.e. as expected. These participants also tended to have a mean ITI that was close to the metronome beat interval of 0.375 sec, which were in turn close to the mean interval between stressed syllable onsets or f0 pikes (see Table 3.5 above). They also tended to have the lowest coefficient of variation. The remaining participants tapped more times per utterance – up to twice as often in some cases – which decreases the mean ITI durations; these participants generally tapped at less regular intervals, i.e. their ITIs had higher coefficient of variation.

3.3.13. Correlating tapping data with other variables

The measurement of taps mentioned above and tapping regularity were then correlated with the dependent variable in the pre-test and post-test, i.e. comprehension score, in order to explore possible relationships. It was thought that success in tapping to the stressed syllable, i.e. three times per utterance would lead to greater entrainment and therefore greater comprehension scores. I was advised (Caroline Palmer, personal communication) not to use the participants' tapping data in the modelling procedure as one of the predictors of the participants' comprehension, i.e. that one dependent variable should not be used as a predictor of another.

The dependent variable, i.e. Score reflecting comprehension, was represented by the overall percentage of correct responses per participant (i.e. averaged across Pre-test and Post-test) was correlated with the mean number of taps per utterance per participant (Figure 3.17 below) and the coefficient of variation of the inter-tap intervals (Figure 3.18 below). Correlations between Pre-test scores and Post-test scores with these variables were investigated separately as well, but since those patterns did not reveal a different result than the overall percentage of corrected responses across the two tests, only the overall data are presented below.

Finally, the last correlation performed in this section was between the participants' average number of taps per sentence and their self-rated ability to tap in time with music.

3.3.13.1. Relationship between number of taps and comprehension score

First of all, the three panels of Figure 3.17 below show the relationship between the number of taps per utterance and the mean score, per participant.

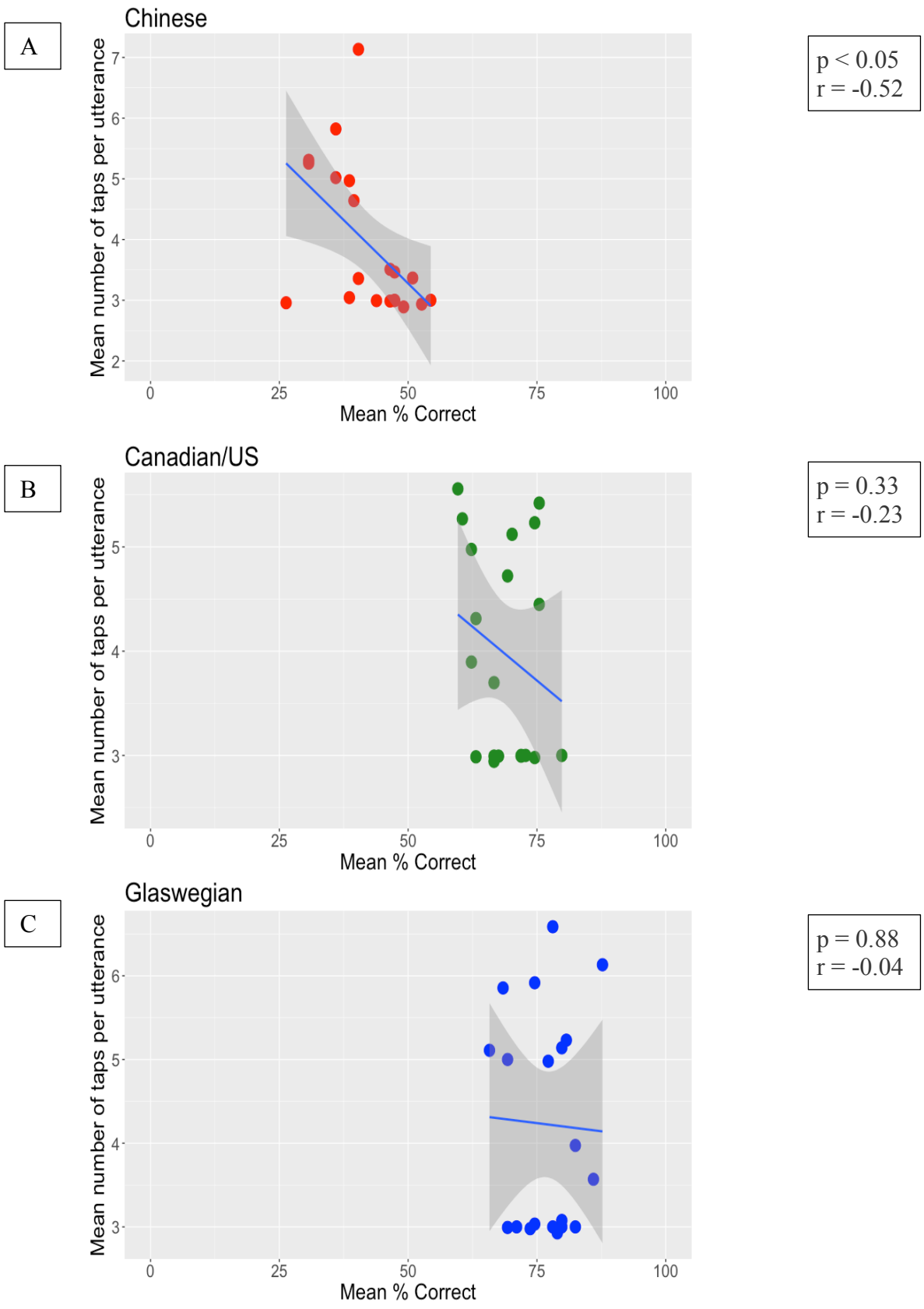


Fig. 3.17. Correlations of mean number of taps per utterance per participant with mean % correct responses on the comprehension test. Panel A: Chinese listeners, Panel B: Canadian/US listeners, Panel C: Glaswegian listeners. Pearson's correlation outputs are presented next to each graph.

The correlation between Chinese participants' comprehension scores and their number of taps was found to be significant. Canadian data exhibits the same tendency, however without significance. They both show the same direction of the relationship, i.e. as expected, the closer the participants' number of taps was to 3, the higher their comprehension scores were. For the Glaswegian speakers, however, there is no evidence of a relationship of these variables.

3.3.13.2. Relationship between coefficient of variation of the ITIs and comprehension score

Figure 3.18 below shows two different patterns of the coefficient of variation correlations with score. For the Chinese and Canadian/US participants, there are weak negative correlations between the two variables (Chinese: $r = -0.23$, $p = 0.32$, Canadian/US: $r = -0.23$, $p = 0.34$), meaning that the more regular the participants' tapping, the higher their comprehension score. In the Glaswegian data, there is no correlation of coefficient of variation and score at all ($r = 0.06$, $p = 0.81$).

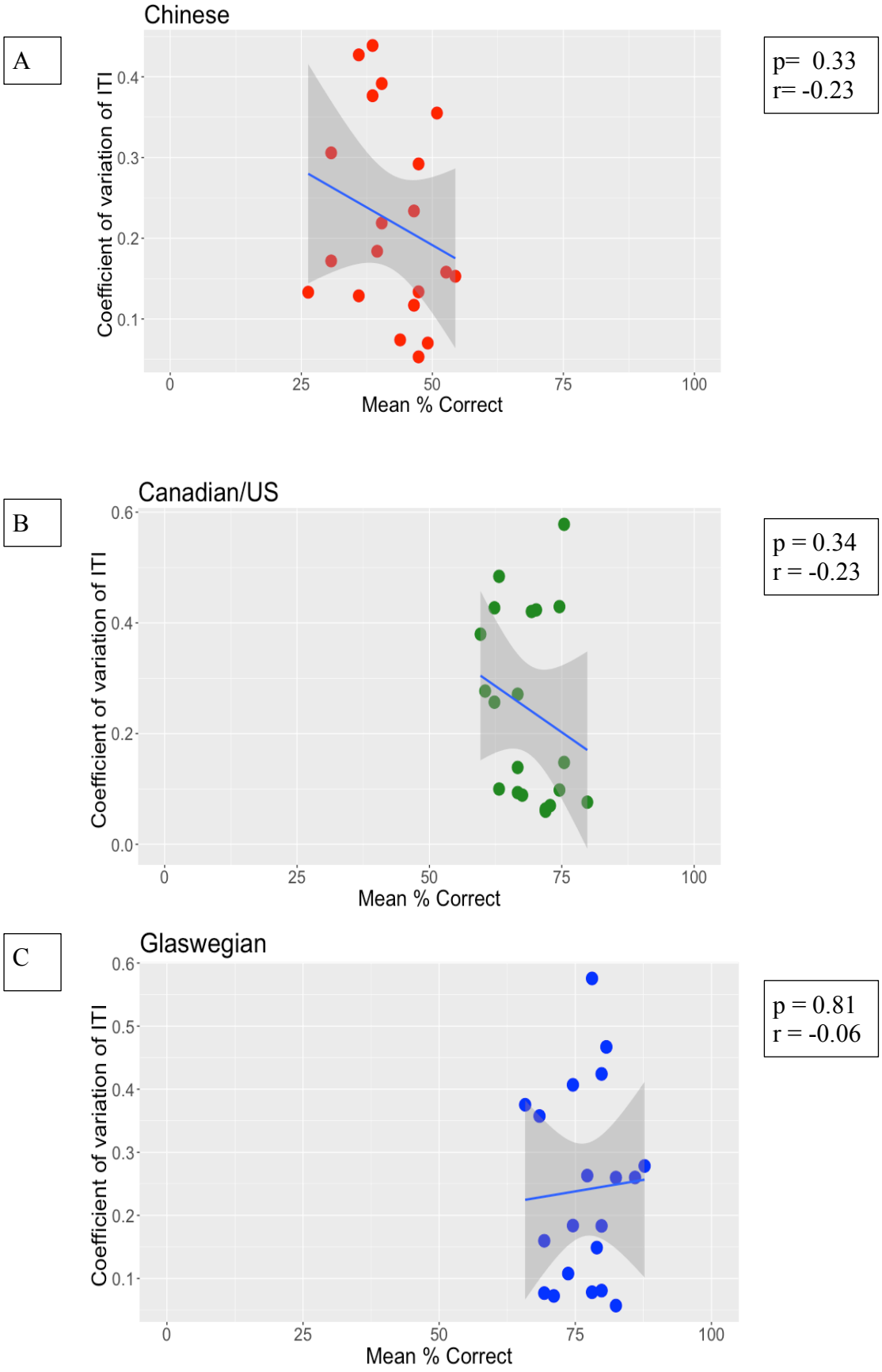


Fig. 3.18. Correlations of participants' coefficient of variation of their ITIs with mean % correct responses on the comprehension test. Panel A: Chinese listeners, Panel B: Canadian/US listeners, Panel C: Glaswegian listeners. Pearson's correlation outputs are presented next to each graph.

3.3.14. SelfTap correlations with tapping data

Figure 3.19 below represents the correlations between the SelfTap variable, i.e. how participants rated their own ability to tap in time with music, and the actual number of taps per utterance. These are negatively correlated—though not significantly so—in all three speaker groups, meaning that the closer to 3 times per utterance these participants tapped, the better they rated their own tapping ability. In this case the strongest correlation was in the Glaswegian participants ($r = -0.36$, $p = 0.10$), followed by that in the Canadian/US participants ($r = -0.26$, $p = 0.26$), and the weakest correlation was found for the Chinese participants ($r = -0.22$, $p = 0.34$).

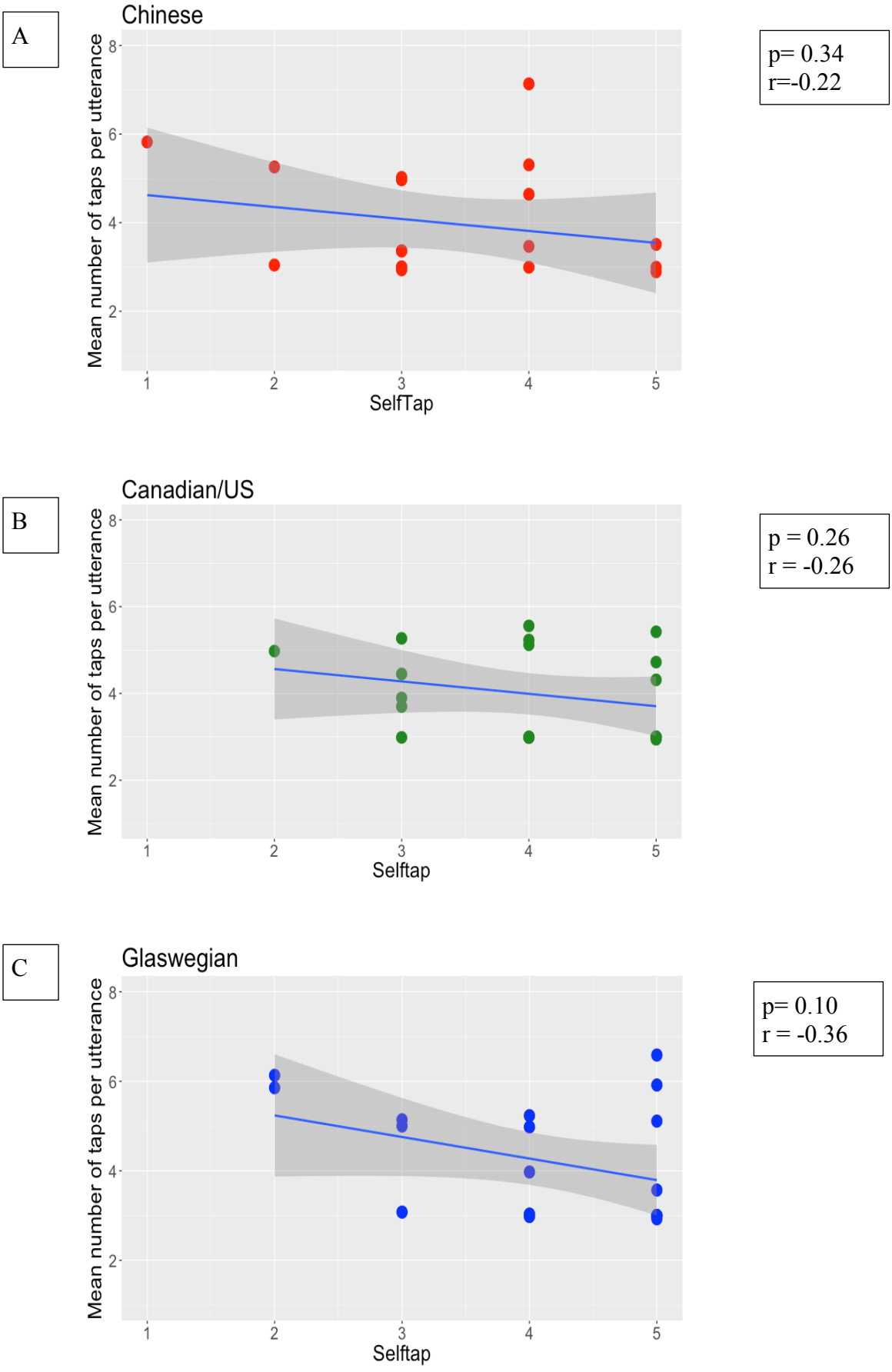


Fig. 3.19. Self-rated tapping ability (on a scale from 1-5, 1=None, 2=Very little, 3=Moderate, 4=Good, 5=Very Good) on x axis; Mean number of taps per utterance on y axis. Panel A: Chinese, Panel B: Canadian/US, Panel C: Glaswegian participants.

3.3.15. Summary of tapping and comprehension/self-tapping report

To sum up, the overall pattern is such that the closer the number of participants' taps was to 3 the better their score on comprehension tests was, though this is significant only for the non-native listeners, the Chinese participants. This pattern was the same but not significant for native but unfamiliar participants, Canadian/US participants. No such relationship was found for the native Glaswegian participants. As well as this, a correlation in the Chinese and Canadian groups was found between the coefficient of variation of their ITIs and overall comprehension score, but not in the Glaswegian group. In all three listener groups, SelfTap was weakly correlated with the number of taps per utterance, such that the closer it gets to 3 taps, the higher the participants rated their tapping ability.

3.4. Discussion of Experiment 2

The following section will discuss the results of this experiment from the point of view of the following questions: 1) Does tapping to the beat in speech improve non-native listeners' comprehension of English connected speech more than a non-beat based training? Can it also improve native listeners' connected speech comprehension? 2) What other factors can influence a listener's connected speech comprehension ability? As an answer to this latter question, musical ability emerged as a particularly interesting factor, and therefore a further question was asked: is musical ability linked to the participants' actual tapping performance? A side question, which is addressed briefly in section 3.4.4. is: what event in the speech signal do listeners synchronise to when performing sensorimotor synchronisation (SMS) with speech?

3.4.1. Does entrainment with speech improve connected speech comprehension?

The main research question that Experiment 2 aimed to answer was whether rhythmic training, based on entrainment achieved via SMS to the beat in speech, could be useful as a method for improving comprehension of fast casual English speech in non-native listeners.

Chinese EFL learners were the only group out of the three tested whose comprehension of fast spoken English improved after short-term training, as reflected in a marginally significant interaction of Condition and Test, and a significant difference between Pre- and Post-test scores for the tapping condition. The other two groups showed a small amount of improvement numerically, but it was not significant. This would suggest that entrainment achieved through SMS has at least some potential to be used as a training method that could lead to better comprehension of fast spoken English. The training received by the

tapping and the control groups was quite similar, but the click identification group did not improve, which suggests the experimental group may have improved due to the tapping training.

However, the difference in the task between the experimental and control conditions should be considered further. It could be that the improvement in the experimental condition in the Chinese EFL learners was simply due to the fact that their task entailed attending to speech, whereas the control group's task was to listen out for a non-speech click sound. This non-speech sound that the participants listened out for could have acted as a distractor, or a contributor to the listener's perceptual load that is independent of the speech materials. Click detection task was used in order to make the experimental and control conditions as similar to each other as possible, i.e. the participants had to perform a motor action using their finger in both groups in response to a sound (advised by Caroline Palmer, personal communication). Nevertheless, to investigate the possibility that the extra perceptual load caused the difference in results, it would be advisable to re-run this part of Experiment 2 with a different control task that did not involve attention to non-speech, perhaps one similar to the control condition in Experiment 1, i.e. counting how many times a certain type of word occurred, or to tap when they heard that certain word, or a certain speech sound.

However, considering the possibility that the SMS training did aid the non-native listeners' comprehension of connected speech further, this improvement was achieved over a single session, the training phase itself lasting no longer than 30 minutes, and without any feedback being offered. Therefore, this improvement over this short timescale and with no feedback seems promising.

In this experiment, the native speaker groups, i.e. Canadian/US and Glaswegian did not benefit from the training in general. The Canadian/US speakers benefitted from both the control and the tapping training if they rated their tapping ability as higher, rather than lower, as reflected by the interaction of the SelfTap variable and Test. Even for the Chinese EFL learners the interaction of Test and training condition was rather weak. Here, learning of the target morphemes occurred, even though the learning targets were not the tapped-to events. An alternative approach could have involved tapping to every syllable, but that could have caused confusion as to what clearly is and is not the beat. Yet another approach could have involved asking participants to tap

strongly to strong beats and weakly to weak beats, however, this was deemed to be a rather difficult task because of the fast speech rates they were asked to tap to. Therefore, the decision to ask participants to tap to stressed syllables only was expected to bring the participants' attention to the rhythm of these sentences as a whole, as this would help the participants to notice the contrast between the weak and strong syllables.

Future research could seek to vary the learning target and the to-be-tapped-to event. If the strong syllables were the learning targets: would an even larger effect be observed? If weak syllables remained the learning targets, would the effect be larger if participants tapped along with them? It is possible that learning could be enhanced in these ways, but it also seems plausible that discovering the general structure, i.e. the hierarchical relationship between strong and weak beats, is what aids comprehension of the weak syllables in between the strong syllables.

The fact that there was no significant improvement in the native speaker groups (Glaswegian or Canadian/US) could be due to there not being as much room for improvement in the first place—although this seems unlikely to be the whole story, as even the highest performing listeners in these groups achieved only 88% and 85% correct responses respectively. More plausible perhaps is that these listeners were already familiar with the rhythm of English and therefore the training focusing on rhythm was not useful. One could still expect that the Canadians might benefit from exposure to the specifics of Glaswegian rhythm; but Smith et al (2014) also comment that "short-term familiarisation with an accent […] does not compensate for long-term familiarity" (2014: 604), and point to the fact that some studies actually failed to show improvement after short-term familiarisation (e.g. Floccia et al 2009). Since the Chinese EFL learners are less familiar with the rhythm of English, it is possible that they would benefit from such training more than listeners who are familiar with it. In other words, the Chinese EFL listeners need to learn the fundamental aspects of another language's rhythm, whereas North American English listeners, even though rhythmically different from Glaswegian, already share the language-specific rhythmical structure.

Another possibility was that, as shown by the results in the Canadian/US listeners, prior rhythmic ability positively affected the usefulness of the training, i.e. those who rated themselves as good at tapping benefitted from this training more than those who rated

themselves as less good at tapping. In other words, it is possible that, as native English speakers, they had sufficient knowledge of English rhythm already, e.g. they knew the stress patterns, however that having better rhythmic skills leads to greater benefit from training which uses rhythmically structured repeated sentences (recall that both the control and tapping training used the same materials and only the participants' task differed). As well as this, the test phases turned out to be quite difficult generally, which exhibited itself in there not being a single native Glaswegian speaker who would score 100% correct answers, and the average of correct scores being at around 80%.

Therefore, a speech-beat synchronisation task, which acts as a means for entrainment, improves speech comprehension for foreign language learners, i.e. SMS training can be useful in improving comprehension for foreign language learners. Comparing these results to Experiment 1, in which participants only listened to metrically regular speech recorded using a metronome, in which however participants did not show significant improvement, one possible explanation is that stronger entrainment was achieved in Experiment 2 than in Experiment 1, due to SMS. It was discussed in Chapter 1, that since entrainment leads to expectation about what happens next and therefore to quicker processing of information (Jones and Boltz 1989, Kohler 2009), more regular speech timing would aid the listener in more accurate and faster speech comprehension (Quené & Port 2005). In Experiment 2, participants were asked to entrain by synchronising their finger movement with the speech beat. It was assumed that SMS would reinforce any entrainment which may have been occurring. Therefore, the present study added more support to the existing literature which suggests benefits of training that uses SMS, by showing that entrainment achieved through synchronising movement with speech beat, improved the participants' speech comprehension.

As Tierney and Kraus suspected (2013a, reviewed in section 3.1.1.2), based on the relationships they observed between reading, attention, temporal processing skills and synchronised tapping behaviour: training using beat synchronisation may indeed lead to improvement in basic auditory function and cognitive flexibility. In the present study's case this manifested itself as an improvement of L2 speech comprehension. As well as this, in their other study Tierney & Kraus (2013b: 14985) concluded that "it is possible that training in rhythmic abilities including beat synchronization practice could lead to a more stable neural representation of sound, in addition to improving linguistic skills, such as phonological awareness and reading". The present study's contribution to this idea is

therefore that another linguistic skill this type of training can improve is speech comprehension in L2, and in particular connected speech comprehension of Glaswegian English by Chinese EFL learners.

Last but not least, in the Canadian group there was a significant effect of Reduction Type variable, which was whether the reduced morpheme was a determiner, preposition or a participle ending, such that scores were highest on prepositions than on determiners and participle endings, but scores on participle endings and determiners did not differ significantly. This pattern was the same for all three listener groups, though not significant in the Glaswegian and Chinese. This is the same pattern that emerged in the previous experiment for the Chinese listeners, which however, did not reach significance in Experiment 1. The explanation was that there were more cues in the phonetic shape of Prepositions than there were in Determiners and Participle ending. While this was not a strong enough effect in the non-native listeners in the previous experiment, in the native unfamiliar listeners in this experiment, it did. This may be due to the fact that the Canadian/US listeners, already have sufficient knowledge of English rhythm in general, and recalibration for comprehension of fine detail in other varieties can happen very quickly, as shown by some previous experiments on exposure to other varieties (e.g. Smith et al 2014, Lennon 2017).

3.4.2. On musical ability

Experiment 2 revealed correlations between the participants' musical ability, as measured by the questionnaire, and their comprehension of fast casual speech. Rhythmic ability, as represented by the self-reported tapping ability variable, had a positive effect on all listeners' comprehension of reduced words produced by a Glaswegian English speaker. This held true regardless of training and across both pre-test and post-test. The literature so far, as reviewed in section 1.6, describes benefits of musical ability on a variety of other skills, such as L2 learners' sound production (e.g. Slevc & Miyake 2006, Posedel et al 2011, Milovanov 2008, Milovanov 2010), learners' L2 tone processing (e.g. Gottfried 2007, Marques et al 2007, Delogu 2008, Marie et al 2010), listeners' processing of speech timing characteristics (e.g. Marie at al 2011, Sadakata & Sekiyama 2011, François 2012, Herrera et al 2011), listeners' L2 comprehension (e.g. Swaminathan and Gopinath 2013) and L2 reading ability (Herrera et al 2011) as well as L1 reading ability (Strait et al 2011). The new contribution from the present study is that rhythmic ability is also correlated with

how well listeners comprehend fast casual speech, whether the speech is the listeners' native speech and native accent, native speech but an unfamiliar accent, or L2 speech.

In the Glaswegian and the Chinese groups, the general trend was that the lower the participants' self-rated tapping ability, the lower their speech comprehension test score; their tapping ability did not relate specifically to their ability to improve at the task. In the Canadian/US group, in contrast, the pattern was that the higher the participants' self-rated tapping ability, the greater their improvement on the comprehension test. As mentioned above, short-term training has been shown to benefit native listeners of a different variety in relation to how good their comprehension was to begin with (Smith et al 2014). It is possible that those who were better at comprehension of Glaswegian connected speech to begin with, were also those who rated themselves higher on tapping ability, and that these are those participants who improved more.

The results from the present study taken together indicate that listeners' rhythmic ability, as represented by the self-rated tapping ability, may be linked to speech comprehension. As well as this, the results of the Canadian group suggest that improvement after a training that uses rhythmically organised, predictable speech may be more useful to those subjects who are more musically able to begin with. A reasonable conclusion is that training one's general rhythmic skills — as distinct from training the specific skill of tapping to speech — might lead to better speech comprehension. However, further research would be needed to establish whether those who already have a higher rhythmic aptitude or skill benefit from this type of training more, and/or whether tapping training could benefit those that lack this aptitude to begin with.

3.4.3. The relationship between tapping performance and other variables

The correlations of the tapping performance with some of the variables were also investigated in order to explore these relationships. The number of taps per utterance was expected to be 3 because each sentence contained three stressed syllables. The number of taps the participants performed varied however. These numbers of taps were significantly correlated with the Chinese participants' with how well they performed on the comprehension tests. The pattern was the same for Canadian/US participants, though not significant. As well as this, the participants' tapping regularity (coefficient of variation) showed patterns suggesting that the more regular tappers performed higher on the comprehension tests.

Lidji et al (2011) found that native speakers of English, i.e. a traditionally stress-timed language, tap to the higher hierarchical level in speech than speakers of French, i.e. a traditionally syllable-timed language. This meant that those who spoke English as a first language tapped to stressed syllables only, while those who spoke French only as a first language tapped to a lower hierarchical level, i.e. more often than on stressed syllables only. If this is true, then the native English speakers in this experiment should have tapped three times only, because they are speakers of a stress-timed language and there were only three stressed syllables in the materials. The Chinese EFL learners would also be expected to tap more than 3 times, i.e. perhaps attempt to tap on every syllable, due to Chinese being classed as a syllable-timed language. However, about half of the participants in each group (Chinese, Canadian/US, Glaswegian) tapped more than three times.

Potentially, it is due to the fact that the instructions for the tapping groups did not specify exactly *how* to tap to the beat. These instructions were kept the same as the ones Lidji et al (2011) gave to their participants, i.e. they were asked tap to what they perceived to be the beat in the sentences they heard. Therefore it was left to the participant to decide how exactly to perform the tapping task, rather than explicitly instructing them to tap to a specific event in the speech, which gave some mixed results, but which helped find other potential correlations between SMS to speech and musical training. It would, however, be good to know if giving more explicit instructions on how many times to tap to each sentence would yield an even better result.

What the above-mentioned correlations did show was that, generally, the closer the participants' taps were to three per utterance, i.e. as expected, the better their scores were on the comprehension tests. Their coefficient of variation, i.e. the measure of regularity of their taps, was not significantly correlated with comprehension, though numerically the trend was such that the smaller the ITI coefficient of variation, the better the comprehension. Since these were also (weakly) correlated with the self-rated ability to tap in time with music, the conclusion that rhythmic ability is linked to speech comprehension.

As mentioned above, previous work has suggested that listeners' choice of tapping level relates to the rhythm class of their native language, but the present study does not bear this out. Contrastively to Lidji et al (2011), the Chinese native speakers in this study as well as

the English native speaker groups were all divided into roughly half and half, i.e. around half of the participants in each speaker group tapped on the higher hierarchical level and the other half tapped on a lower level. Since Chinese languages are classed as syllable-timed and English as stress-timed, but there was no difference in their tapping responses, on the one hand it seems that native language rhythm class is not a very strong determinant of the level at which participants tapped. On the other hand, that the listeners' comprehension of these sentences tended to be better the closer their number of taps came to three. This means that we can say that there may be something about these English sentences that makes them more comprehensible if paying attention to the higher hierarchical level, which could be seen as supportive of Lidji et al's (2011) conclusions.

As outlined in Section 3.3.3 above, the SelfTap variable was selected to represent the participants' general rhythmic ability, as it positively correlated with other musicality variables such as years of musical training and hours of practicing a musical instrument. It is possible that due to the more introspective nature of the question, i.e. "How do you rate your own ability to tap in time with music?" it was also able to pick up on the more intrinsic ability, or natural aptitude, that participants had. As well as this, since it asked about how well they could tap, it is possible that it is picking up on the participants' rhythm production abilities. This was demonstrated in the correlation between the self-rated tapping ability and the participants' actual ability to find the beat, i.e. getting the right amount of taps. Therefore, the variable not only reflected how well the participants thought they could tap, but how well they actually tapped.

3.4.4. Which event(s) did participants tap to?

An interesting question arising from Experiment 2 concerns what event(s) in the speech signal participants tapped to. Relatedly, when the speaker who produced the materials spoke in time with the metronome beat he had heard, we can ask which events in his speech he synchronised with the beat. Due to space and time constraints, these questions have only been looked into superficially in this thesis i.e. via some basic calculations on the speech and the tapping data separately. Preliminary indications are that a more in-depth analysis would be valuable at a later stage.

The acoustic analyses revealed that the most regular event in the materials was the pike (cf Loehr 2007), which is the lowest or highest f0 point. This had a lower coefficient of variation compared to the stressed syllable onset and vowel onset, i.e. it occurred most

regularly out of these possible events, though the difference in regularity between pikes and stressed syllable onsets was small. Further analyses of e.g. the timing of participants' taps with respect to these specific events could potentially add to our understanding both of sensorimotor synchronisation to speech, and of P-centres. For example, if it turned out that the most regular taps, i.e. the ones with the lowest coefficient of variation were timed consistently with respect to pikes then we could possibly infer that pikes are a good proxy for P-centres. A future in-depth analysis of these data will also allow for a comparison between the way the Glaswegian, Canadian/US and Chinese tappers perceived these beats, and could shed light on the types of question raised by Chow et al (2016) regarding cross-linguistic differences in P-centre location.

3.4.5. Future directions and unanswered questions

There are some questions that remain unanswered after this experiment. One of the main drawbacks of this study was that in all three phases (pre-test, training and post-test) the same type of metrical structure was used for all the sentences. Whether participants would be able to perform just as well on differently metrically-structured sentences, i.e. to what extent generalisation would take place, is not known. Tests of generalisation were not done in order to keep the experiment as short as possible, and to allow for maximal entrainment by keeping the rhythm the same for all sentences. It would be useful to know if these participants improved their connected speech comprehension in other sentence structures.

Moreover, it would be good to know whether this improvement was a lasting one. No delayed post-test was conducted in order to test whether the improvement was retained days or even weeks after the training took place. This was not done due to time restrictions, i.e. there was a limited time within each experiment had to be conducted in order to complete the thesis in a timely manner.

The SelfTap variable showed significant effects on the non-native listeners' comprehension. While an explanation of this variable as well as some possible reasons for its significance have been offered above, it is also quite an unusual variable in itself, because the question asked of the participants was "Please rate your own ability to tap in time with music." It would be valuable to employ a more objective test of participants' rhythmic ability in future experiments.

Perhaps the broadest question that Experiment 3 raises is whether this type of training

could be implemented in an EFL class. This experimental training was conducted in laboratory conditions, rather than in a more natural classroom setting. Whether it can be useful for teachers and learners in groups is therefore unknown. Chapter 5 presents Experiment 3 which addresses this question directly. First, Chapter 4 delves deeper into objective ways to measure participants' rhythmic ability.

Chapter IV

Musical skills and how to test them

4.1. Introduction

Several questions surfaced during and after conducting the previous experiments, as evidence emerged that participants' pre-existing rhythmic ability might affect their speech comprehension and/or the benefit they could derive from the training that was provided. These were such questions as: "What is musical ability?" and "How is musical ability different to rhythmic ability?" as well as "How can these abilities be measured in a systematic way?" This chapter gives a brief overview of the research that deals with these questions, in order to inform the approach taken in the final experiment.

4.2. What is musical ability?

First of all, it seems unlikely that musical ability is a single ability that a person can either have or not have. In other words, it is important to understand three issues: 1) that an individual can possess some abilities without being trained (in the form of aptitude), 2) that even despite not having a great deal of aptitude, an individual can learn some abilities; and 3) that the term musical ability may actually be an umbrella term for a variety of different abilities, which may or may not be closely related to one another.

In other words, musical ability is an ability that is composed of a variety of elements, such as perception and production skills in rhythm, meter, pitch and loudness. While these are most likely interlinked with one another, there is some evidence showing that the relationships between these elements are not always straightforward. Furthermore, these abilities can be either seen as aptitudes, i.e. existing before any training has been undertaken by an individual, or as abilities that have been developed mainly as a consequence of training (Law & Zentner 2012).

4.3. Tests of Musical Abilities

A range of tests have been devised throughout many years for various purposes with the aim of measuring the various elements of musical ability as well as their interdependencies. One of the oldest tests, which some of the most recent tests still validate themselves against, is Gordon's Musical Aptitude Test (MAP; Gordon 1965). This test battery asked students for their judgments of musical excerpts in terms of whether they were the same, similar or different on various aspects, such as Melody and Harmony which were grouped together as "Tonal Imagery", Tempo and Meter, which were grouped as "Rhythm Imagery" and Phrasing, Balance and Style, which were grouped as "Musical Sensitivity". The students were asked to decide which of the two excerpts of in the

"Musical Sensitivity" group was "indicative of a more musical performance" (Gordon 1965: 14). As well as this, students' own musical performance was judged by professionals.

Similarly to this old test, the newer test batteries also consist of a variety of subtests whose aim is to give a full picture of a person's abilities through investigating various components of what would be generally understood by the term musical ability. They recognise that there may be different elements such as skills in rhythm, meter, pitch, loudness, etc.

It is now common to divide the concept of musical ability into receptive and production skills, which have been shown to be correlated with each other, however, possibly not in all possible directions (see section 4.4.1 below). It is also becoming more common for tests to attempt to distinguish between aptitudes and learned skills.

For the next experiment in the thesis, a more systematic and reliable way of testing participants' musical, or to be more precise, rhythmic skills was needed. A range of recent test batteries were considered for this purpose, and are reviewed briefly below. First, tests are discussed which aim to test general musical skills, as MAP did, and then those that aim to test rhythmic skills only.

While for the purposes of the present research, it is reasonable to think that the most important skills would be the ones to do with rhythm and timing, a few general musical aptitude test batteries are reviewed as well. This is because these batteries use various methods of measuring rhythm and timing skills, and such a review allows us to compare what kinds of tests have been used. As well as this, some of these tests show how the various other musical skills, as measured by their various subtests, can be intercorrelated.

First of all, two batteries that aim to test general musical skills will be considered in detail: the Montreal Battery of Evaluation of Amusia (MBEA, Peretz et al 2003), and the Profile of Music Perception Skills (PROMS, Law & Zentner 2012).

4.3.1. The Montreal Battery of Evaluation of Amusia (MBEA)

Amusia is a term used for loss of musical abilities. It is possible for disorders specific to music alone to occur, e.g. as a result of damage to a particular part of the brain, responsible

for processing a particular component of musical ability (Peretz et al 2003: 59). The goal for the design of the MBEA was to present a tool which could be used "as a standardised evaluation of musical abilities" (2003: 59). The authors proposed a model in which musical input has a melodic and a temporal component, which Peretz et al refer to as the 'what' and the 'when' respectively. The melodic component consists "of melodic contour and the tonal functions of successive pitches" (2003: 60). The temporal organisation can have two different mechanisms: rhythm and meter. Rhythm is the "tendency to group events according to temporal proximity without regard to periodicity" and meter is "concerned with the extraction of an underlying temporal regularity or beat" (2003: 62). Their model assumes "the operation of separable perceptual subsystems for analysing pitch and temporal variations" (Peretz et al 2003: 61). The MBEA consists of six tests, which are used to assess these components: contour, interval, scale, rhythm, meter and memory tests.

For the tests of melodic organisation, three types of manipulations were applied to the same tone in 15 sequences, which were presented in pairs asking the participants to judge whether they were the same or different:

1) changing the scale while maintaining the original melodic contour, i.e. pitch was modified so that it would be out of scale and would sound out of tune;
2) changing of the pitch direction of the surrounding intervals while maintaining the original key, i.e. playing a note that e.g. goes down in pitch rather than up compared to the original, but remains in the same key;
3) changing one note by one semi-tone but preserving the contour and scale, i.e. changing the pitch interval.

For the temporal organisation tests, rhythmic and metric components were assessed. Manipulations were the following:

1) changing the durations of two adjacent tones to change rhythmic grouping, while retaining the same meter and total number of sounds
2) presenting melodies so that their binary or ternary structure were accentuated, half of the sequences being written in duple and half in triple meter; participants were asked to categorise them as such (i.e. either waltz or march)

As well as these, the MBEA had a memory recognition component, which presented participants with sequences and asked them whether or not they had heard these earlier during the session. The authors tested the battery for sensitivity, getting a result that

showed that there was no single subtest on which over 80% of participants scored 100%. They also tested it for test-retest reliability, and showed that participants who were asked to re-take the test after 4 months improved slightly on the second test, but none obtained a perfect score. The authors took this to mean that the test is reliable on re-test.

Finally, they also tested the battery against Gordon's commonly-used Musical Aptitude Profile (1965), which they found it correlated highly with ($r = 0.53$, $p < 0.001$). Based on these results, Peretz et al (2003) concluded that MBEA is "suitable for detecting the presence of a musical disorder in perception and memory", as well as a good test of musical abilities in the general population.

The authors also looked at the diagnostic value of the MBEA, i.e. whether it is able to diagnose a disorder when there is no known evidence of brain damage (2003: 68), e.g. to detect amusia. They found that the test was able to confirm amusia in 89% of participants who self-declared as amusic. Here they also discovered that for these individuals the results of the three melodic tests correlated with each other, but not with the temporal element of MBEA.

One of the advantages of MBEA is that, as the authors say, it was able to pinpoint "which music recognition-processing component is likely at fault in a given individual" (2003: 72), compared to other tests that existed at the time. Another advantage the authors listed was that the test is not only able to capture the most musically talented individuals, but also those who have underlying deficits (2003: 72). However, Law and Zentner (2012) criticise MBEA for being "developed to assess amusia" (2012: 3). In short, while MBEA was originally developed to test and evaluate amusic individuals and other researchers still criticize it as such nearly a decade later (Law and Zentner 2012), the authors argue that it can also be used for the general population (Law and Zentner 2012).

4.3.2. Profile of Music Perception Skills (PROMS)

PROMS is a test designed to "identify musically untrained individuals with musical skill, as well as those who, despite extensive musical training, may not be as skilled" (Law & Zentner 2012: 1). It was designed to measure perceptual musical skills, "across multiple domains: tonal (melody, pitch), qualitative (timbre, tuning), temporal (rhythm, rhythm-to-melody, accent, tempo), and dynamic (loudness)" (2012: 1). The authors aimed to create a test that would be useful for the general population, that would include more musical

perceptual components than any previous test, that would be able to test each component very specifically and that would be valid and reliable. The components which the authors say had previously been neglected are timbre, tuning, tempo, and accent (2012: 11). What the authors used as a tool for distinguishing training from aptitude is an aggregate index of information on the amount of musical training and qualifications their participants had.

The participants' task was to decide whether two sequences were the same or different, and they were also asked to rate their confidence in their judgment. The authors tested PROMS for internal consistency and test-retest reliability, and found promising results. As well as this, they validated their test for convergence with other, older musical ability tests (Gordon's MAP (1965) & Gordon's Advanced Measures of Music Audiation (1990), Musical Ear Test (Wallentin 2010)). They also correlated their results with the participants' musical training, which confirmed the validity of the idea of "musical sleepers", i.e. those who perform well without musical training, and "sleeping musicians" who did not perform well, despite being trained. This can be seen in Figure 4.1 below, along with the general correlation between the musical training and total PROMS score.

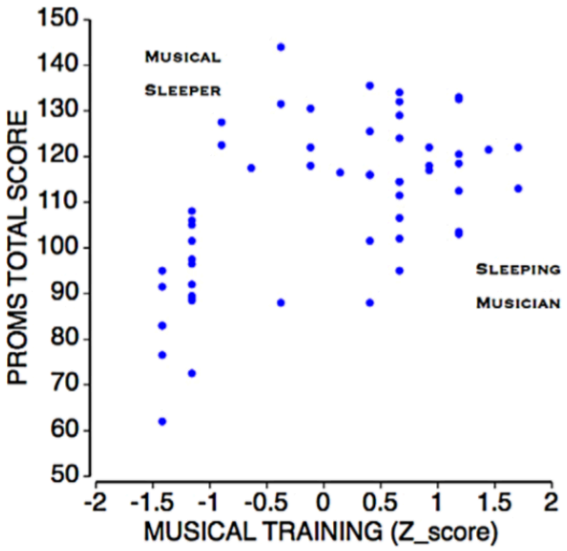


Figure 4.1. Scattergram plotting PROMS performance against an aggregate index of musical training. Training includes years of musical training, music degrees and qualifications, critical listening activities, and musicianship status. Extent of training predicts PROMS performance substantially but imperfectly ($r=0.57$, $p<0.01$). Upper left corner: Example of a ''musical sleeper'' performing well despite minimal musical training. Lower right corner: Example of a ''sleeping musician'' posting a lesser performance despite extensive musical training. Reproduced from Law & Zentner (2012: 10).

The various subtests of PROMS also showed significant correlations with each other, but not with a gap detection task (the smaller the gaps the listeners are able to detect, the better

their auditory discrimination skills (2012: 11)). This is because the auditory discrimination skill test was a non-musical one, i.e. gap detection (or silence detection), in white noise which did not correlate significantly with any of the component tests. The PROMS battery does not include any measurement of musical production skills. The reason why the authors decided to omit this completely was that that "production tasks would confer an advantage to those with experience in handling a musical instrument, including the human voice" (2012: 13), which would be more likely to be a result of practice than of aptitude, which the test's main goal was to measure. However, as the authors say, PROMS can be used in conjunction with other measures to study the links between musical perception and production (2012: 13).

4.4. Rhythm Skills Tests

Moving on to batteries that test for rhythmic skills specifically, three tests will be discussed below: the Beat Alignment Test (BAT, Iversen & Patel 2008), Harvard Beat Assessment Test (HBAT, Fuji & Schlaug 2013), and Battery for the Assessment of Auditory Sensorimotor and Timing Abilities (BAASTA, Dalla Bella et al 2017).

4.4.1. Beat Alignment Test (BAT)

The Beat Alignment Test (BAT) was designed to "complement existing tests of rhythm processing by directly examining beat perception in isolation from beat synchronisation" (Iversen & Patel 2008: 465), with a twofold aim: to be able to find what the patterns of beat-based processing abilities are in the general population and to be able to find individuals who exhibit difficulties in the beat processing domain, but who are not tone deaf.

It aims to answer the question of the extent to which synchronisation skills are linked to beat perception ability, i.e. whether it is possible for an individual to be able to synchronise their movement with a beat without being able to perceive it well, or vice-versa. At the time, the beat perception part of the test was an innovative idea, i.e. no previous tests had combined beat production with beat perception.

It consists of the following subtests:

1. spontaneous tapping, i.e. regular tapping at a preferred tempo
2. synchronisation of taps to a metronome at various tempi
3. synchronisation to musical passages, i.e. tapping to what the participants perceive as the

beat in music

4. beat perception in music, i.e. a test in which participants listen to excerpts of music without tapping, while there is a superimposed beep which falls either on or off beat; the off-beat version occurs either with beeps that were too fast or too slow, or out of phase (early or late) with the actual beat, The test measures reaction time, i.e. the time which it takes the participants to decide whether or not the beeps were on or off beat, and they are also asked to rate how confident they were of their judgment on a scale from 1 to 3.

In the synchronisation component two measures were taken: the ability to match the tapping rate to the tempo of the musical excerpt and the variability of tapping. Iversen & Patel (2008) found that generally participants were able to synchronise their taps with the musical excerpt well, with some individual variation.

In the beat perception test, they found that most participants were able to correctly identify the on-beat excerpts in most cases, but that generally participants performed much worse on off-beat ones: they frequently incorrectly judged an excerpt which contained off-beat clicks as on-beat, i.e. they were biased to answer "on beat".

Iversen & Patel (2008) found that participants' performance on the beat perception task were only weakly correlated with one measure of their synchronisation performance (i.e. correlation of the participants' tapping tempo with music tempo). In contrast, the coefficients of variation of their participants' inter-tap intervals were strongly correlated with their score on the beat perception test. They conclude that participants' perceptual judgements were correlated with how accurately they were able to synchronise their tapping with the auditory stimuli, though this link would require further investigation (2008: 468).

4.4.2. Harvard Beat Assessment Test (HBAT)

The HBAT (Fuji & Schlaug 2013) was developed to test both beat perception and beat production. Sowinski and Dalla Bella (2013) found individuals who were able to perceive rhythm but were unable to synchronise to music. Fuji & Schlaug were interested in finding out whether poor rhythm perception can exist without any impairment of production, i.e. whether there is dissociation between rhythm production and perception. As well as this, they were interested in establishing objective measures and cut-off scores in order to have a tool which would identify beat-deaf individuals (2013).

Fuji and Schlaug criticised MBEA because its rhythm and meter subtests also require the use of pitch-processing abilities. They claimed that the same confound can be found in BAT, because it uses musical excerpts and therefore includes perception of pitch, melody, harmony, timbre and emotion and memory (2013: 2). To reduce the potential for confounds of rhythm and pitch processing, they proposed the use of "psychophysics with controlled auditory stimuli" (2013: 2), specifically monotonic or woodblock sounds (2013: 2).

HBAT has four subtests:

1. *music tapping test* (MTT), i.e. a test of how well participants can synchronise their taps with the beat of music. For this they used 3 of the BAT excerpts which they had first rated on the clarity of their pulse, and manipulated each of them to have three different tempi (slow, medium and fast). The researchers demonstrated to the participants how they expected them to tap to the underlying beat, before recording their taps and corrected them until the participants understood the way they were expected to tap.
2. *beat saliency test* (BST), *beat interval test* (BIT), *beat finding and interval test* (BFIT), which all test perception and production thresholds via an adaptive psychophysical stair-case method (2013:1). In this method, the starting point for the tests is to present two stimuli that differ by an initial amount (20dB for intensity differences between beats in the BST and 20ms in the BIT and BFIT). When a participant responds correctly twice in a row, these parameters are divided by two (i.e. become, 10, 5, etc), while when incorrect responses are made twice, they get doubled. The points at which the change, or the "turnaround," happens are recorded until such points occur 6 times. The average of these values is then said to be the threshold for perception and production.

The BST used a woodblock tone whose intensity was manipulated to create either a triple or a duple meter. The task for perception was to discriminate between the two meters by a key press on keyboard. The task for production used a drum pad and asked the participants to synchronise their finger tapping with a stimulus and to reflect the intensity in the stimulus by changing the amplitudes of the taps. The BIT used a woodblock tone in which the tempo either gradually decreased or increased. In the perceptual part of this test participants had to decide whether the sequence was slowing down or speeding up. In the production part participants were asked to tap while adjusting their tempo to the stimulus as well as possible. In the beat finding and interval test they used a series of repeating notes

and created either a speeding up or a slowing down beat. Again, there were two tasks: perceptual discrimination between a slowing down or speeding up beat which asked specifically to determine the beat of the underlying quarter-note, and a production task where participants had to synchronise with that underlying quarter-note. The difference between BIT and BFIT was that in BIT they were asked to perform the tasks on woodblock clicks, while in BFIT they were asked to perform the task on musical beat. In all three of these tests the adaptive two-alternative forced choice discrimination paradigm was used, which adjusted the parameters of these tests depending on the participants' responses in order to measure their beat/meter production and perception thresholds.

Fuji and Schlaug found that a lower degree of synchronization in the music tapping test was associated with higher perception and production thresholds in the beat saliency, beat interval and the beat finding and interval test (2013: 14). As well as this, there was a significant correlation between production and perception thresholds in the beat saliency test, but no such correlation in the other beat tests. Furthermore, the authors found significant positive correlation between the duration and amount of participants' musical training with their synchronisation to the musical beat, which suggests that the Music Tapping Test might be sensitive to the effect of musical training, but not with the other tests.

Fuji and Schlaug (2013: 14) conclude that the music tapping might be a useful measure to differ from group mean performance, and that the other measures, i.e. the BST, BIT and BFIT might be useful measure for further investigations of ways in which perception and production are dissociated.

4.4.3. Battery for the Assessment of Auditory Sensorimotor and Timing Abilities (BAASTA)

BAASTA (Dalla Bella et al 2017), the final test discussed here, was developed to test a broad range of timing skills while being sensitive enough to study individual differences and impairments in populations of both healthy and patient populations. Dalla Bella et al measured perceptual thresholds using two different procedures (in order to validate their results). The tasks used are briefly described below.

The perceptual tasks were as follows

1) duration discrimination, which was a same-different task using two pure tones of different durations. Participants were asked whether the second tone was longer or the same as the first task;
2) anisochrony detection with tones, in which participants were asked to judge whether a sequence of 5 tones was regular or irregular. In the irregular sequences the fourth tones occurred earlier relative to the inter-onset intervals (IOIs) of the previous tones;
3) anisochrony detection with music, which was a similar task to the previous one, but used short musical excerpts. In the irregular versions one of the IOIs between the musical beats was manipulated. The task was to detect the excerpt with a deviant beat;
4) music beat detection, which was an adapted version of BAT. They used "four computer-generated musical excerpts with a salient beat structure" (2017: 1132), on which they superimposed a triangle-timbered beat which was either on or off beat. The off-beat versions were created by aligning the beat so that either the relative phase or the period were misaligned. These excerpts were presented at four different tempi.

The production tasks were as follows:

1) unpaced finger tapping, in which the task was to tap regularly at any rate, as well as at the highest possible and the lowest possible rate; this was also repeated after the participants completed all the production tasks;
2) paced finger tapping, in which participants were asked to synchronise their finger taps to an isochronous sequence of piano tones, at various tempi;
3) paced tapping to music, in which participants were asked to tap to the beat that they found in the same classical pieces as in the music beat detection task
4) synchronisation continuation, in which participants tapped with a regular piano tone stimulus first and after 10 taps the stimulus stopped and the task was to keep tapping at the same rate. This task was performed at 3 different rates.
5) adaptive tapping to a tempo changing sequence, in which 10 tones were presented to participants, 6 of which were at the same tempo and four were either at a faster or a slower tempo. The task was to synchronise with the first tempo, then adapt the tapping to the changed tempo and then to continue tapping at the changed tempo after the stimulus was no longer presented.

Within these various BAASTA tasks, some correlations were found, however they were not consistent. A positive correlation was found between participants' results on the modified BAT and the paced tapping task, which the authors interpret as a suggestion that these tasks "may engage common beat-based processes" (2017: 1139, also similar to Iversen & Patel's 2008 findings). However, correlations between some of the other tasks did not emerge, e.g. no other results correlated with either the duration discrimination task (a perception task), nor the adaptive tapping task (a production task). The fact that only some of the other tasks showed correlations was interpreted as showing that BAASTA is able to assess a variety of different timing skills.

4.5. Summary, and usefulness for the purpose of the present work

An overview of these tests and their main purposes can be seen in Table 4.1 below.

Test	**General Population or Disorder Testing**	**Rhythm-specific or Generally-musical**	**Perception only or Perception & Production**	**Unique selling point/main purpose**	**Earlier test validated against**
MBEA	Disorder Testing	Generally musical	Perception only	suitable for screening of amusic cases in brain damaged individuals	Gordon's *Musical Aptitude Profile* (1965)
PROMS	General Population	Generally musical	Perception only	to identify musically untrained individuals with musical skill, as well as those who, despite extensive musical training, may not be as skilled	Gordon's *Musical Aptitude Profile* (1965) & Gordon's *Advanced Measures of Music Audiation* (1990), Musical Ear Test (Wallentin 2010)
BAT	General Population	Rhythm	Perception & Production	a more direct test of beat perception created by asking participants to detect asynchronies rather than merely discriminating pairs	N/A (mentioned Seashore 1960 but doesn't say it was tested against it)
HBAT	General Population	Rhythm	Perception & Production	to test for dissociation between beat perception and production	N/A
BAASTA	Both	Rhythm	Perception & Production	to test a broad range of timing skills while being sensitive enough to study individual differences and impairments in populations of both healthy and patient populations	N/A

Table 4.1. Overview of rhythmic and musical abilities tests.

In sum, some of these tests intend to test the various components of what we would call a musical ability and test whether correlations between these components exist, while others concentrate on one of those components, specifically rhythmic ability. There is some evidence showing that rhythmic ability is not necessarily one skill, but rather a more diverse set of components of timing skill. However, what is of interest for the purpose for the present work, both BAASTA and BAT showed correlations between perception and production using the tapping task, while HBAT suggested a link between production and perception in the task which required meter discrimination. Fuji and Schlaug (2013) showed that tapping performance was correlated with the amount and duration of musical training, which is perhaps not surprising because musical training usually involves synchronisation training in form of synchronising with a metronome, for example.

The results of Experiment 2 indicate that musical ability, specifically rhythmic ability, plays a role in casual speech comprehension and in the ability of listeners to benefit from rhythmic training. For the next experiment, it is desirable to examine this further by testing participants' ability. The next experiment was to be carried out in a setting that is quite different to the usual lab conditions, as it aimed at EFL learners in classroom. Therefore, a widely-used test was required which would be able to identify the participants' rhythmic ability, but which would not make the test phases of the experiment long. The perceptual part of BAT was chosen, because the perceptual part of BAT was used by many of these studies as one of their subtests, and the original study showed a correlation between BAT's perception and production subtests. As well as this, the most recent battery, BAASTA, also used BAT as a subtest and showed that its perceptual task was correlated with participants' tapping ability. For these reasons, the perceptual part of BAT was incorporated into the test phases of the next experiment.

Chapter V

Can in-class group training using drumming aid L2 learners' comprehension of connected speech?

5.1. Introduction

The previous experiment showed that training involving tapping aids comprehension of connected speech in non-native learners of English, and that rhythmic ability correlates with comprehension. However, these results were obtained in controlled lab conditions. In order for them to also be able to be applied practically, for example, in a classroom or as part of a CALL app, there was a need to test them in classroom learning conditions.

5.1.1. Review of a choice of textbooks

A small selection of EFL textbooks was reviewed in search for what, if any, methods have been used for teaching connected speech or rhythm. The resources looked at here are those which were found at the EFL library of the University of Glasgow. Some of them have no mention of teaching rhythm e.g. Roberts 2009, others cover the very basics of stress (e.g. Roach 2000). However, a few have been found to cover more than just basics, as well as some suggested ways of teaching rhythm through games, poetry, songs and music, using movements of limbs (hands beating the rhythms or feet walking), e.g. Kelly (2000), Kenworthy (2000), O'Connor (1980), Bowen & Marks (1992). These are mostly recommendations from experienced EFL teachers who have been able to observe what kind of activities at least seem to bring some results in class.

5.1.2. EFL Teacher Survey

In order to investigate current practices in EFL classroom, a brief survey was conducted which was sent out to EFL teachers around the world. The general aims of the survey were the following: to become informed about what type of teaching they were doing (EFL, ESOL, EAP etc), what kind of populations they taught (which countries, what learner populations etc), whether the teachers consider connected speech and connected speech comprehension an obstacle to their students' communication, and whether the consider it important. As well as this, it contained question on whether they cover connected speech and rhythm in class and how much training they have to teach these, as well as what resources they use etc. The questions the survey asked can be seen in Appendix A. The survey participants were recruited by sending emails to existing contacts within the EFL teacher community, and requesting they forward the emails on, as well as via some mailing lists, e.g. the Education Scotland ESOL (English for Speakers of Other Languages) group, and the PronSig (Pronunciation Special Interest Group) of the IATEFL (International Association for Teachers of English as a Foreign Language) group.

From the EFL teacher survey, 59 responses were collected. 60% of them were from the UK, 10% from Argentina, 3.4% from France, and 1.7% from Japan, Vietnam, Taiwan, Iraq, Russia, Germany, Greece, Netherlands, Malta, Czech Republic, Poland, Ireland, Colombia and Peru each. 86 % of them teach adults, and 14% teach children (in primary or secondary school). Around 20% of those who teach adults also teach children, and around 9% of the overall number also teach adults with disabilities.

The results showed 89% thought their students' comprehension is a barrier to communication (example quotation: "Students often do not understand different accents or fast speech, particularly when faced with multiple locals talking or elderly locals talking"). Some questions asked the teachers on their training to teach connected speech and rhythm. These results showed that: 81% cover connected speech, and 80% cover rhythm in their classroom, but only 49% say they have been trained on how to teach it. Out of these 38% commented that they had done their training while studying at the university, or for a degree, or CELTA/DELTA/TEFL qualification. As well as this, 17% say they learned it through self-study & developed their own methods and 31% say that it was either a short training or that the training was a long time ago, 10% did not comment at all/did not comment with a relevant comment (an example quotation from the survey: "I have never been trained and am not aware of effective resources").

86% of the teachers reported that they cover connected speech and rhythm in their own teaching via drills, listening activities, chunking, songs/poetry, gestures, beating/clapping rhythms, kazoos, specialised Pronunciation materials (e.g. Hancock 1996, Gilbert 2005, Vaughan-Rees 2010, Underhill 1998). However, 93% would use new resources for rhythm and connected speech if they were available.

The teachers were also asked whether they thought that students would use materials to learn connected speech and rhythm, if they were more readily available. Only 5% thought the students wouldn't use them. 64% thought they would use them as an app and 80% thought they would like to use them in class (choosing both options as answers was possible).

To summarise, first of all, these results show that there may be a general scarcity of materials for teachers to use, as well as a scarcity of training provided on appropriate ways of teaching rhythm and connected speech to students, despite these areas being considered

important, or even crucial, for everyday communication. As well as this, these results show that some methods which use rhythmic and sensorimotor training, such as e.g. clapping or beating, or performing other gestures are already being used in classroom. However, the effectiveness of these methods may have not been investigated, but rather have been recommended by various textbook authors, or possibly even made up by the teachers themselves.

Therefore, Experiment 3 aimed to investigate the effects of rhythmic training as they could potentially be employed in a more natural language-learning setting, i.e. in EFL classroom. There were several major differences between the third and the second experiment. In the third experiment, training was given over multiple sessions spread over a month, pre-test and post-test included a test of the participants' rhythmic ability rather than just a questionnaire, and the training itself was of a different kind. Therefore, first of all some literature on one-off vs multiple training sessions is reviewed in the next section, below.

5.2. Duration and amount of training for perceptual learning

Many perceptual language learning studies use a one-off training session (e.g. Maye et al 2008, Floccia et al 2009, Adank et al 2010, Smith et al 2014, Morano et al 2015, Jügler & Möbius 2015, Lennon 2017). Many other studies addressing various training techniques used longer-term training ranging from 4 or 15 sessions (e.g. Logan et al 1991, Lively et al 1993, Iverson et al 2005, Iverson and Evans 2009, Saloranta et al 2015) to 45 sessions (e.g. Bradlow et al 1997). These studies are reviewed in more detail in the literature reviews for Chapter 1, or Chapter 3 above.

The way that teaching and learning a language have worked in practice is based on the assumption that repetition and reinforcement aid the learner. This is consistent with the incremental theory which says that in learning, associations are created gradually, rather than instantly, i.e. in one trial (Underwood & Keppel 1962). Moreover, as Wright and Sabin (2007: 727) say, "[i]mprovements in performance on many perceptual tasks continue over multiple days of practice." Citing McGaugh (2000), they also argue that to transfer learning from short- to long-term memory, consolidation must take place (2007: 727). In search for the answer to the question of how much daily training is enough, Wright & Sabin (2007) compared different amounts of auditory discrimination training over 6 days. The types of perceptual task they trained the participants on were frequency discrimination and temporal-interval discrimination, and they used either 360 or 900 trials per day for

each of these conditions. They found that less training (i.e. fewer trials) were needed to gain the temporal-interval discrimination skill and additional practice did not improve the skill further, whereas more training was needed for the frequency discrimination skill to improve, concluding that the amount of training per day depends, among other things, upon the type of target skill.

However, one-trial learning can occur according to some (i.e. Estes et al 1960, Rock 1957). Estes et al (1960) conducted an experiment which showed that subjects were able to form associations between a syllable and a number after being exposed to these associations only once. However, these experiments also showed that half of these associations were not learned after one trial exposure. Estes et al (1960) then asked the question whether this meant that the probability of an item being learned was 0.5, or whether it was subject-dependent, such that for half of the subjects the probability of one-trial learning was 1 and for the other half it was 0. Therefore, they repeated the trial. This time it showed that 71% of the previously correct items were also correct, and only 9% of the items that were not correct on the first trial were correct. From this Estes et al (1960) concluded that the one-trial learning theory is supported.

Underwood & Keppel (1962) point out a few issues with this approach. One important one is that according to the one-trial theory, it is assumed that no degree of strength of association whatsoever has been formed for the items that were incorrectly responded to, whereas in the incremental theory, some associative strength would have developed that could have explained these results just as well, if not better, than the one-trial theory. As well as this, Rock (1957: 193) says that "repetition after the association is formed is effective in strengthening it", and even Estes et al admit that repeated reinforcements give "repeated opportunities for the formation of an association between a stimulus pattern and the reinforced response" (1960: 338).

Despite the consensus that learning tends to be incremental, many studies in L2 learning have successfully used a single, relatively short training session. One study that tried to establish the optimal amount of training for auditory tasks was conducted by Molloy et al (2012). They focused on perceptual learning of tone discrimination (a frequency discrimination task) in both training and testing. More precisely, they aimed to answer two questions: what is the best amount of training provided in a single session, and how are training sessions best distributed over time to achieve most efficient learning? They

conducted two experiments. In the first, they compared four training groups, varying both the total amount of training, and the extent to which the training was concentrated into a short period of time vs distributed over a longer period (from 8 min to 1 hour per session). They found that in the early training stage, the group with the shortest sessions learned faster compared to the group with longer sessions. In a second experiment they compared a group who received training over 4 sessions with a group who received training over 1 session (4 hours vs 1 hour) and found no difference in these groups' performance. Additionally, they found that those who received a single training session continued to improve after the training stopped, while those who received more training sessions did not. It is therefore certainly possible that a single training session can lead to meaningful improvement.

From the point of view of second language speech learning, when considering the amount of training needed for learners to achieve a learning goal, it is also important to consider other factors, such as the learners' L1, and what they are being trained on. For example, Iverson and Evans (2009) compared the effects of high variability perceptual training (HVPT) on German and Spanish learners of English. German and Spanish differ in the extent of their vowel inventories, Spanish with 5 vowels and German with 18. Iverson and Evans (2009) trained their participants on identification of English vowels over 5 sessions, and tested them in a pre-test and post-test. The results showed that German listeners improved more than the Spanish listeners in the initial 5 sessions. However the Spanish listeners improved after another 10 sessions to the same level as the German listeners. This is probably dependent on the relationship between the participants' prior knowledge and what they are being trained on.

Thus, various approaches to the amount of training have shown various effects. It seems that the amount of training needed is dependent on such factors as what the trained skill is, as well as (in linguistic studies) what the learners' L1 is. The present work does not seek to establish exactly how many sessions are optimal to improve the comprehension of fast casual speech. Rather, since there is some evidence for multiple sessions aiding learning, and since that is what happens in language classrooms in practice, in Experiment 3 multiple sessions will be given to the participants.

5.3. Overview of Experiment 3

Experiment 3 set out to test whether in-class, group training using drumming to the beat in speech can lead to greater comprehension of fast casual English speech. It was designed in collaboration with a community musician, Jane Bentley, who specialises in music in health, social and communicative wellbeing. Her PhD (Bentley 2011) examined the enabling and outcomes of musical interaction among non-specialists. Furthermore, two EFL teachers from Glasgow were involved in the creation of materials. They advised on the areas of speech comprehension which their students tend to struggle with most.

The participants were learners of English who lived in Glasgow at the time. The experiment first tested the subjects' comprehension of function words in a pre-test, after which all subjects received training during their English classes. For the Experimental group, part of their normal class time was assigned to 4 rhythmic training sessions each lasting approximately 25 minutes, spread over a month. The Control group instead followed their normal curriculum for learning English, except that their listening comprehension tasks were modified so that they received exposure to the voice heard in the pre- and post-test. After the training phase all subjects' comprehension was tested once more in a post-test.

While the test phases were the same as in Experiment 2, the training phase was quite different. This was not only because of the training taking place in class, with a group rather than individually. The second main difference was that drums were being used as instruments rather than finger tapping. Finally, the stimuli were not presented with 4 repetitions of each sentence concatenated into one file with carefully controlled beep sounds signalling the start of each repetition, as there were in Experiment 2. This approach was to simply make the participants aware of the rhythmic structure of English connected speech.

5.3.2. Research Questions

The main research question for this experiment is whether multiple sessions of training involving drumming to the beat perceived in speech can improve EFL learners' comprehension of English connected speech more than a control form of training which does not use the drumming method.

As well as this, Experiment 3 attempts to establish what other factors can have an effect on such learning (e.g. participants' musical background, English language proficiency, and so on), with additional emphasis placed on musical skills in particular, in light of the results of Experiment 2.

5.3.3. Hypothesis

The hypothesis is that training involving drumming to the beat in speech will benefit EFL learners' English connected speech comprehension more than the control training.

5.3.4. How the study was developed

Two EFL teachers from Glasgow were asked to identify any connected speech phenomena their students struggled with. They provided a list of these, together with examples. The training materials were developed based on these suggestions. The general points the EFL teachers made were that the students mostly struggled with the following:

1) perceiving word boundaries in connected speech, e.g. when hearing a phrase such as *two till four* would be perceived as something like *tootil four,* or *two eggs* would be perceived as *two weggs,* or *this week* as *the sweek*
2) perceiving word boundaries between words that end and start with the same or similar sounds, e.g. *at the, best student, first three, bit tired, good day, want to*
3) comprehending contractions in e.g. *don't, should've, people'll,*
4) identifying words where /h/ has been elided, in e.g. *him*
5) comprehending word sequences where coalescence has occurred, e.g. *would you* produced as /wudʒu/, *do you* as /dʒu/
6) dealing with reduction processes in function words such as *a* to [ə], or *and* to [n]
7) comprehending polysyllabic words when syllables have been elided, e.g. *library, naturally*
8) detecting participle endings in syllable codas, e.g. /t/ in talked, or in weak syllables, e.g. /n/ in broken.

As well as this, the community musician provided some ideas on the types of tasks that might be suitable for the class, based on her experience teaching musical rhythm in various community groups.

5.3.5. Participants

One male native speaker of standard Glaswegian, aged 26, was recorded to provide all materials. This was the same speaker that provided all materials for Experiment 2.

All EFL learners in this experiment were Chinese learners living in Glasgow. There were 30 EFL learners (18 male, 12 female) aged 17-33 (mean = 24.3, standard deviation = 3.98), who had been living in Glasgow for between 1 month to 1 year prior to the experiment. They all spoke Mandarin as their first language; some also spoke other varieties of Chinese (Ankang, Yangzhou and Shanghainese). All were studying at the KAPLAN centre, which is attached to the University of Glasgow and teaches the students English in order for them to undertake studying at the university, i.e. these are the pre-sessional courses which the students must pass in order to continue their studies.

There were 12 participants in the experimental group, and 18 in the control group. These groups had already been formed as classes in the school, i.e. one class was assigned to be the experimental group and another to be the control group. It was not possible to intervene in the composition of the groups to balance numbers, since the groups were already formed and attending EFL classes as a whole.

All participants were also attending their regular classes at the time of the experiment. Each group had 8 hours of classes per week. According to the school's system, the experimental group was at a pre-intermediate level and the control group was at an intermediate level. It was unfortunately not possible to access two groups who were more closely matched for level. All of the 8 hours of classes of the experimental group were English language classes, while the control group were attending 4 hours of English classes and 4 hours of 'study skills' classes. In those classes, which are delivered entirely in English, they learned about research methods, referencing, plagiarism, giving presentations, making posters etc. The lessons were not language-focused but the students were learning how to put into practice the language they have learned in reading/writing/listening/speaking classes.

To sum up, it was not possible to assign the participants to the control and experimental groups randomly because in order to access the participants at all, the training had to be incorporated into their usual EFL class rather than separate sessions. Therefore, there were

some underlying differences between the groups which could not be controlled for, especially how long they had lived in the UK for, and their level of English.

5.3.6. Materials

As in Experiment 1 and 2, there were three types of materials for the three different phases of the experiment: Pre- test materials, Training phase materials, and Post-test materials. The Materials for Pre-test and Post-test were the same as in Experiment 2. However, the training materials were different to the ones used in Experiment 2. All materials for all three phases were recorded using the same male Glaswegian English voice.

5.3.6.1. Pre- and Post-test Materials

In both test phases the materials were the same as those used in Experiment 2, i.e. recordings of the native male Glaswegian speaker producing short, metrically-structured sentences such as *He was fat for a cat in a box.* These sentences contained highly reduced forms of weak syllables, which were the learning target. Since the materials were the same as in Experiment 2 they are only briefly summarised below (see section 3.2.4.1 above and Appendix F for more detail).

Two test sets containing 57 items each were used (sets A and B). Both the control and the experimental groups heard the same sentences, but the assignment of the set (A or B) to Pre- and Post-test was counterbalanced across participants. The test materials were designed so that all sentences in both test phases followed the same regular metrical structure of alternating weak and strong syllables: *wwSwwSwwS*, e.g. *But it sat on the path of the duck.*

As in the previous experiments, the weak syllables were the learning targets. The types of words and morphemes chosen were:

- Prepositions (*for*, *from*, *in*, *on*, *with*, *of*)
- Determiners (*the*, *a* or *her*)
- Participle endings (*-ing*, *-en*) for those verbs whose present and past participle are identical apart from the *–ing* and *–en* ending, e.g. *taken – taking, beaten – beating* and *eaten – eating*.

The context in which these learning targets appeared was ambiguous, so that as few semantic cues as possible could be used, e.g.: *for* could be used where *from* was used and vice versa: e.g. *And they came from/for a break at the beach*.

In total there were 18 prepositions in each of the tests (set A and set B), i.e. each preposition occurred three times in each set; 33 determiners i.e. each of the three determiners occurred 11 times in each set; and 6 verb forms in each of the tests (set A and set B), i.e. each verb form occurred once in each set. For the full materials see Appendix M and N. Further details of Pre- and Post-test stimulus creation are in Chapter 3, section 3.2.4.1.

5.3.6.2. Experimental training

The training materials for the experimental group involved a variety of different tasks designed so that participants would first learn basics about rhythm in general, and later concentrate on finding stressed and unstressed elements in pre-recorded speech. They were delivered to the group in four separate sessions. Sessions 1 and 2 were delivered by the community musician, Jane Bentley, and sessions 3 and 4 were delivered by the present researcher. Each session took approximately 30 minutes and participants performed a variety of tasks with materials of increasing difficulty. At the start of each session, there was always a brief reminder of what had been covered in the previous session.

5.3.6.2.1. The drums

A special type of drums were used to carry out most of the training tasks. These are called Soundshapes and are a facilitators kit constructed with a Skyndeep™ Graphic drumhead framed in colour coded Sound Shape Circles (SoundShapes Facilitators Kit 2018). An example drum can be seen in Fig. 5.1 below. These are of different colours depending on their size. The sizes used for this experiment were 8.25”, 10.5” and 12.75”. Sticks, which are a part of the kit, were used together with the drums in order to perform the drumming.

Fig. 5.1. Sound shape circle.

5.3.6.2.2. Overview of the training sessions

The 4 training sessions were designed so that first of all the participants were introduced to the basics of musical rhythm as afterwards to the idea of language having rhythm. This was done in a progressive way, i.e. starting with the easiest, i.e. least complex tasks and rhythms and finishing with the most difficult, or most complex tasks and rhythms. Participants did a variety of different tasks, from drumming out simple musical rhythms on their own, to being shown a simple link between different syllables having different stresses (weak or strong), through repeating rhythms modelled by the musician (where musician and participants produced the rhythms either with instruments, or verbally, or both). There were some multiple choice tasks, in which participants matched rhythms of phrases and sentences to other phrases or sentences (speech to speech tasks), or matched drummed rhythms with spoken utterances (rhythm to speech tasks).

The language-focused tasks always involved either perception or production of rhythm of language, starting with short phases, such as: *Hot and cold* and progressing to more complex ones, such as: *They exercise every day apart from Sundays*, and short dialogues. In the rhythm production tasks with language stimuli, participants drummed to the strong syllables (or beats) they heard. For the more complex sentences, the option for them to drum to every syllable was considered the option for them to drum to every syllable was considered, but rejected as much of the speech was too fast for it to be possible, and also because it was desirable to draw attention to the difference between strong and weak beats. This is consistent with the training provided in Experiment 2, i.e. training to engage

entrainment for the strong beat. Each session lasted approximately 35-40 minutes.

For the experimental training phase, the speaker was asked to produce the stimuli at a fast yet natural rate. For the control training, he was asked to read the materials as naturally as possible, as if for an educational purpose, for students of EFL.

Session 1

Session 1 started with a short presentation by the researcher on the reason why we decided to give the participants this training, i.e. previous research. The purpose was to make participants aware of the idea that language has rhythm. The presentation included results of the Experiment 2, showing that tapping to the beat in the speech was beneficial for comprehension of fast casual English speech.

After that, the community musician started out with a simple call and response task using the instruments. She played simple rhythms and asked the participants to repeat them after her. She used the middle of the drum to represent one sound, called "boom", and the wooden side of the drum to represent another sound, called "pa". The "boom" represented strong beats/syllables and the "pa" represented weak beats/syllables. The next task involved each subject playing one rhythm and the rest of the group repeating that rhythm together on their drums. The following task involved saying the sound the drum was making out loud at the same time as playing the different rhythms. This was done to reinforce the idea of there being a link between the different syllables as uttered by participants, i.e. "boom" or "pa" and the sounds the drum made. Next was a speech-to-drum task, in which the musician spoke the syllables without playing them on her drum and the participants drummed them back to her. The task was to drum the rhythm back in synchrony with each other, as a group, rather than as an individual. If she felt the group was not able to drum the pattern back to her, she asked them to repeat it. However, if they still could not replicate the patterns, the musician would not ask for a third repetition. This was done in order not to discourage participants from taking part in the task. After this task, the musician reversed the speech-to-drumming order: she drummed patterns to the group and their task was to produce the pattern verbally, using the "boom" and "pa" syllables. She explained to the group the aim of the task was to get them to 'translate' words into rhythm and rhythm into words. The task after this was for everyone to think of a fruit or a vegetable in English, and to think of how they would drum the word on their

drum, and then to drum it and for the rest of the group to drum it back. Each participant had a turn at this and the group repeated the fruit.

After this short practical introduction to musical rhythm and the connection between musical rhythm and speech, the group moved on to tasks specific to speech rhythm. These started with various short phrases and short sentences, e.g. *Hot and cold* or *I'm hungry*. The musician first produced the sentence while simultaneously drumming it, in the middle of the drum for the strong syllables (i.e. for the "boom" sound), and on the side of the drum for the weak syllables (i.e. for the "pa" sound). She then indicated with her hand that it was the group's turn, and they produced the rhythm back to her (both verbally and by drumming simultaneously). In session 1 there were 16 sentences in total, starting with a number of simple short phrases, such as the examples above, and progressively becoming longer and more complicated, with the last and longest sentence being *They exercise every day apart from Sundays*. For the full list of tasks and materials in each session, see Appendix M and Appendix N. Once the group reached the end of these sentences, the task was repeated, but this time with the musician producing some of the sentences verbally only and the group drumming them back.

The last part of session 1 was to listen to and drum to the same set of 16 sentences, but this time these were pre-recorded, produced by the Glaswegian male, and were therefore at a faster and more natural rate than the musician had produced. Using the pre-recorded sentences also meant that it was not possible to adjust them in any way, e.g. by changing their rate, for the listeners. Because the rate was faster than during the practice with the musician, and more natural, the participants were asked to drum on the "boom" syllables, i.e. in the middle of the drum on the strong syllables only. It was originally envisaged that they would drum in the middle of the drum for strong syllables ("boom") and on the side for weak syllables ("pa"), but during the preparation of this task, it proved it be a very difficult task to switch between drumming in the middle of the drum and the edge of the drum fast enough to keep up with the rate of natural speech, so the alternative of drumming just to strong syllables was preferred. The participants were asked to first only listen to the sentences and then say the sentences out loud at the same time as drumming the sentences; and the group did the task simultaneously, i.e. in synchrony. Each utterance was played twice, and after each repetition they produced it verbally and by simultaneously drumming. The musician did the task together with the group on her drum. If she did not feel the participants performed the drumming well enough, the sentence was played a third time.

To end the session, the musician led a fun and short drumming game in which participants were put into small groups colour coded by the drums they were using. She asked one group to drum out the phrase *Hot and cold*, i.e. where the drum would correspond to the strong-weak-strong syllables ("boom pa boom"), the next group to drum out the sentence *It's a lot of fun*, i.e. where the drum would correspond to the weak-weak-weak-weak-strong sounds ("pa pa pa pa boom"), and the last group to drum *My phone's very old*, i.e. weak-strong-weak-weak-strong ("pa boom pa pa boom"). The musician then started by asking only one of the groups to play their rhythm in synchrony. After they had established good synchronisation with each other, she asked another group to come in and play their rhythm simultaneously, and eventually also the third group. This is one of the tasks the musician uses in her everyday work with community groups with the aim of group integration and ice-breaking/team building.

Session 2

The second session started with an exercise to remind the participants about the "translation" between the syllables "boom" and "pa" and the way they would drum these. The musician started out by producing the syllables simultaneously with drumming them and the participants' task was to repeat the drumming after her. After the initial few trials, the musician proceeded to drumming only, i.e. without verbal production of the syllables, and subsequently to producing syllables verbally only which the students drummed back, and finally to drumming rhythms which the students produced in "boom-pa" form.
They then repeated the game in which the participants were divided into three groups and each group drummed a different rhythm simultaneously with the whole group.

The next part of the session consisted of repeating some of the sentences drummed in the first session, and adding new more complicated ones. First, these sentences were practised together with the musician at a slower, more adjustable rate and the musician judged how well the participants were performing each time and when needed repeated each sentence up to three times. Then, the participants drummed after hearing the recording, as in the previous session. Each time the recording was played twice, or three times if needed.

Session 3

Session 3 consisted first of a reminder of the sentences used in the previous session, where participants listened to the sentences and then drummed them back (drumming on stressed syllables only, i.e. the "boom" syllable on the drum) while simultaneously saying them out loud.

After that they performed two new perceptual tasks. The first was a multiple choice task in which they listened to one sentence followed by three others, A, B and C. Their task was to identify which of A, B or C best matched the question sentence in terms of rhythm. For example, the question sentence has the same structure as B in question 1 below:

1. Tell him.

A. He stopped.
B. Help her.
C. Talk to them.

There were 5 questions of this type and the group were asked to answer as a group rather than as individuals, i.e. they could call out their own answers together and were told whether their answer was correct or not.

In the second multiple-choice task, participants listened to a pre-recorded drummed rhythm. They then decided which of 5 pre-recorded sentences (spoken by the same Glaswegian male speaker) best matched the rhythm in the question. The answer sentences were all the same sentence, but each time a different word carried an emphatic accent (see example below, accented words are marked in bold):

1. pa pa pa pa pa pa boom pa

*A. I **said** I wanted to go there.*
*B. I said **I** wanted to go there.*
*C. I said I **wanted** to go there.*
*D. I said I wanted to **go** there.*
*E. I said I wanted to go **there**.*

Here, the answer was D.

Session 4

In the final session there were a variety of perceptual and production tasks. First, pre-recorded rhythms were played and the task was to choose which of three pre-recorded sentences best matched each one, as in example 1:

1. *pa pa pa boom boom*

A. It's for you.
B. This took a long time.
C. This essay is great! (Correct answer is B)

There were three items of this kind. After that there were three items where a sentence was played and the participants chose the pre-recorded rhythm that best matched it, as in example 2:

2. This cat is too big for this box!

A. boom boom pa pa pa pa boom pa
B. pa pa pa pa pa boom boom boom
C. pa boom pa pa boom pa pa boom (Correct answer is C)

After this the teacher played a non-verbal rhythm game called "Don't drum this one back" with the participants. She played various rhythms and the group listened and repeated each one back in synchrony with one another. However, there was one particular rhythm, which they were shown at the start of the game, that they were not to repeat after she played it. This forces the participants to pay attention to each rhythm and actively make a decision to not repeat the one particular rhythm. The rhythm is what would correspond to the phrase *Don't drum this one back* (rhythm: *boom boom pa pa boom*) This gave the participants a fun short break from performing speech tasks.

The final part of this session was drumming short dialogues. The dialogues were pre-recorded and played to the participants, who also received a printed version of them. They were first asked to listen to the dialogues and underline the stressed syllables in each sentence. The group was then divided into two and each of the two groups took on one of the roles in the dialogues (A or B). They listened to each sentence once first, and were then asked to drum their part of the dialogue out while simultaneously saying it out loud.

5.3.6.3. Control training

The control group continued their usual curriculum covering speaking, listening and writing and reading. However, since the experimental group's training and test materials were produced by the same speaker, steps had to be taken to eliminate a potential confound

of rhythm training with voice/speaker familiarity. To address this, some of the listening tasks from the control group's normal curriculum were re-recorded using the same speaker who provided materials for both the test phases and the experimental groups' training and played to them instead of the recordings provided in their usual materials.
Therefore the control group's training consisted of listening to recordings of stories which would have been a part of their original curriculum, but these were presented in the same voice that the experimental group listened to. The recording was approximately 5 minutes long.

Links to these sessions can be found in Appendix N.

5.3.7. Recording Procedure

All pre-recorded materials for the training phase were recorded in a sound-attenuated booth at the University of Glasgow. An AKG SE 300B stand microphone was used to record the participant's speech, together with a Rolls LiveMix MX34c mixer and a Roland Duo Capture audio interface. During recording, the participant wore a set of Beyerdynamic headphones. The speech recordings were made using Audacity.

For the test phases, the same recordings were used as for Experiment 2; refer to the Chapter 3 (section 3.2.4.1) for the details of their recording procedure.

5.4. Data collection

Pre-test and post-test were conducted at the University of Glasgow's teaching lab. There were 22 days between the pre-test and post-test. Participants were tested individually at the University of Glasgow's STELLA teaching lab, and submitted their responses via webpages. Training sessions, as described above, were conducted at the Kaplan centre, in groups.

5.5. Procedure for pre- and post-test

The pre-test started with a slightly modified version of the perceptual part of Beat Alignment Test (BAT). The modification was such that the item randomization was such that it didn't vary from participant to participant, and participants were not asked to rate the confidence of their judgment (where 1 = guessing, 2 = somewhat sure, 3 = completely certain). The experimenter explained the tasks, and there were then 6 practice examples, followed by the main test which had 30 items.

The next part of the experiment was 25 multiple-choice questions in order to test the participant's English language proficiency. As in Experiments 1 and 2, the test on the Cambridge English website was used (http:// www.cambridgeenglish.org/test-your-english/adult-learners/).

The participants were offered a break, after which they began the pre-test. Just as in Experiment 2, they listened to sentences and filled in missing words on the computer screen. After this, the participants were thanked for their time and received no payment. The training sessions began 7-8 days later.

The post-test took place 3-4 days after the end of the training. It contained the same types of reduced speech phenomena as the pre-test, but used different sentences in order to test whether generalisation had taken place. The procedure was the same as for the pre-test, but without the Cambridge English test. After the post-test, participants completed a demographic questionnaire (Appendix O) which elicited information on age, gender, first language, number of years studying English, other languages spoken, and musical training (e.g. years of training, hours of practice, self-rated ability to dance in time with music, self-rated ability to tap in time with music).

New questions for Experiment 3 concerned participants' motivation for learning English. As the literature suggests, motivation to learn a language can be a factor which contributes to success or failure at foreign speech and language learning (Flege 2005). To examine this, an adapted version of the questionnaire used by Saheeb (2015) was used. 20 statements were chosen which participants were asked whether they agreed or disagreed with, on a scale from 1 to 5. The statements asked about a variety of motivational aspects, both intrinsic (i.e. whether they enjoyed the act of studying English), and extrinsic (e.g. whether the participants' parents wanted them to know English). The full questionnaire can be found in Appendix O.

Finally, the experimental group were also asked additional questions about their experience of the 4 training sessions, e.g. whether they found them useful or enjoyable.

5.6. Coding the data

As in Experiment 2, pre- and post-test responses were coded as 1 for correct and 0 for incorrect. In the case where two words had to be typed into the answer field, they were split into two, i.e. each word could be coded as correct or incorrect separately. That is, if the sentence contained two missing words in a row, e.g. *in a* in a sentence such as *He was*

fat for a cat in a box, if the participant typed the correct answer to the first or second word only, e. g. *in the* or *on a*, they would receive one out of two possible points.

In case of obvious misspellings, e.g. *eatting* instead of *eating*, the answers were coded as correct. In case of any doubt as to the reason for the incorrect spelling of a word, the answer was scored 0. Participles were coded as correct if they contained the correct participle ending, i.e. *-ing* when the correct answer was *-ing*, and *-en*, when the correct answer was *-en*, even if the verb itself was not the correct one. E.g. *putting* was coded as correct, even if the verb was *eating*.

5.7. Analysis

The results from the Pre- and Post-test data were analysed in R (version 3.3.2) as follows.

The dependent variable was:

1) Score (1 or 0 for each correctly or incorrectly typed word)

There were two manipulated variables:

1) Condition (Experimental or Control)
2) Test (Pre-test or Post-test)

The following variables were also included:

1) CamEng (participants' general level of English, as in previous experiments)
2) Reduction Type (whether the missing word was a Verb, Determiner or a Preposition)
3) AB_BA (order in which Pre-test and Post-test were performed, which was counterbalanced across subjects)
4) UKLive (how many years the participant had lived in the UK at the time of testing)
5) SelfTap (self-rated ability to tap in time with music; levels: None, Very Little, Moderate, Good, Very Good)
6) SelfRhythm (self-rated general rhythmic ability; levels: None, VeryLittle, Moderate, Good, Very Good)
7) Years (number of years of musical training)
8) Hours (number of hours of practising a musical instrument per week)
9) BAT (Beat Alignment Test, score on a scale from 0-30)
10) Motivation (collected through a set of 20 questions on participants' motivation to learn English, shown in Appendix P; an average score was calculated for each participant

based on their answers, which were numbers from 1 to 5, 1 meaning strongly disagreeing and 5 meaning strongly agreeing with a statement; the ends of the scale always corresponded to the same level of agreement, such that a simple average of the numerical scores could be calculated)

There were also two random factors, i.e.:

1) Item
2) Participant.

Random slopes were not included because the model would not converge if an attempt to include random slopes was made.

As well as this, data was collected on how enjoyable the participants found the rhythmic training. These data could not be used in the modelling procedure due to the fact that these could only be collected from the experimental group, who were the ones who took part in the rhythmic training.

5.7.1. Descriptive statistics relevant to participants' learning of English

As mentioned above, participants were not assigned to groups randomly, but rather had already been grouped into classes by the school, according to their level of English. As Figure 5.2 shows, this meant that the experimental group had lived in the UK for a shorter time on average than the control group, and were much more homogeneous than the control group with respect to length of residence. The control group also averaged higher CamEng scores than the experimental group, as Table 5.1 shows; their motivation scores were higher too, but only slightly,

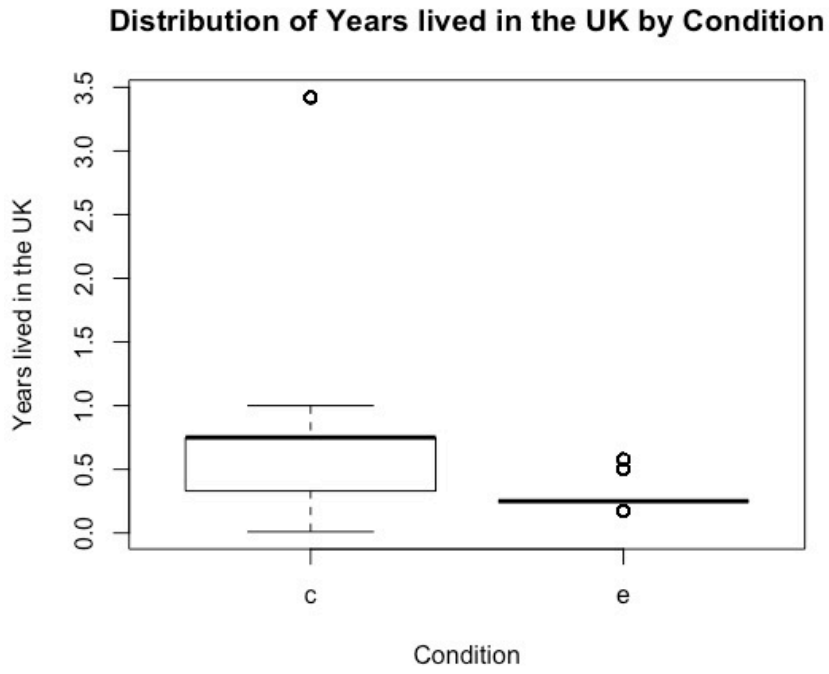


Fig. 5.2. Distribution of years lived in the UK by experimental (e) or control (c) condition

	Motivation Score		CamEng Score	
	Mean	SD	Mean	SD
Experimental	3.75	0.97	9.58	2.39
Control	3.83	0.9	11	3.2

Table 5.1. Summary statistics for Motivation and CamEng Scores and Time Spent in the UK by group.

5.7.2. Modelling

The procedure for modelling was as follows: starting with a model containing Test and condition, since these were the variables that would answer the main research question. After that main effects were added one by one, checking every time whether the more complex model was a better fit to the data using the log likelihood test, while keeping Test and condition so that interactions could be checked. After checking all main effects, the two-way interaction between Test and condition was tested first, since this was the interaction most important to the research question.

5.7.3. Results

The final model contained the following predictors:

- Condition (experimental, control)
- UKLive
- Reduction Type (Determiner, Preposition, Participle Ending)

The output table for this model was:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-1.7539	0.2885	-6.079	1.21e-9***
Condition (Experimental)	-0.7144	0.1861	-3.839	0.000123***
UKLive	0.3588	0.1429	2.51	0.012075*
Reduction Type (Preposition)	0.8283	0.4065	2.038	0.041582*
Reduction Type (Participle)	-0.8788	0.6429	-1.367	0.171682

Table. 5.2. Final model for Experiment 3. Reference levels were: 'Control' for Condition and 'Determiner for Reduction Type.

In Fig. 5.3 below, percentage of correct responses are shown. Boxplots from left to right, the first two represent the Experimental Group, with overall lower scores, the last two are for the Control Group with overall higher scores.

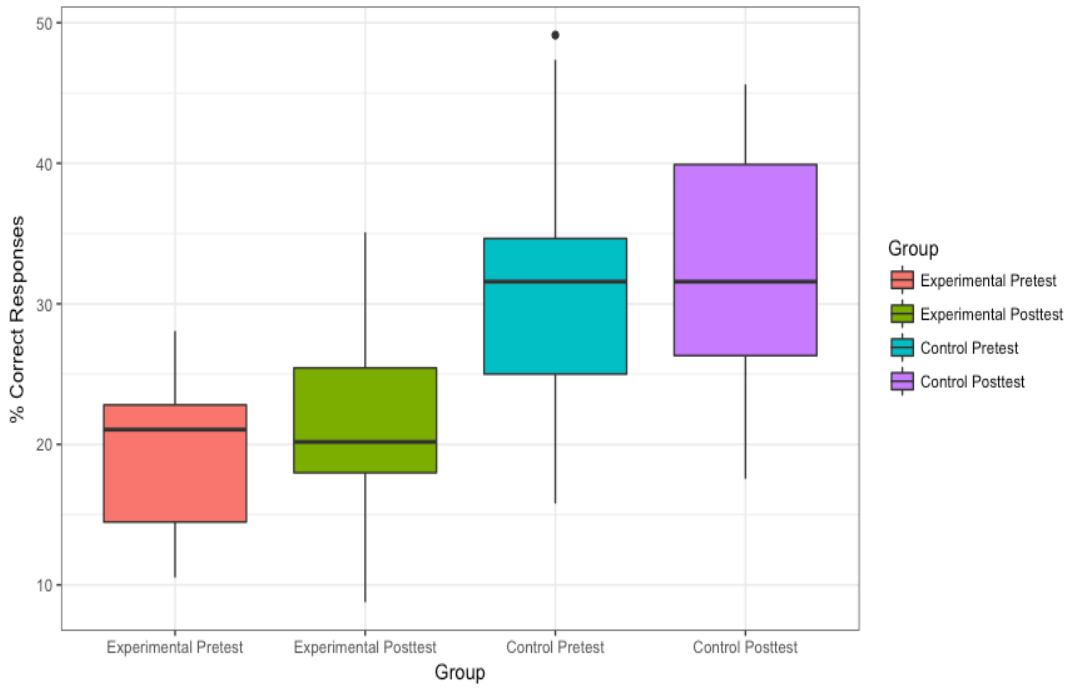


Fig. 5.3. % correct responses by condition in Pre-test and Post-test.

In the model, the main effect of condition was that Experimental Group scored significantly lower than the control group ($z = -3.839$, $p < 0.005$). There was no effect of Test, i.e. no significant improvement from Pre- to Post-test; the predicted interaction of Condition with Test was not significant.

Time spent living in the UK was a significant predictor, such that the longer the participants had lived in the UK, the better their comprehension score ($z = 2.510$, $p < 0.05$). This relationship can be seen in Fig 5.4 below.

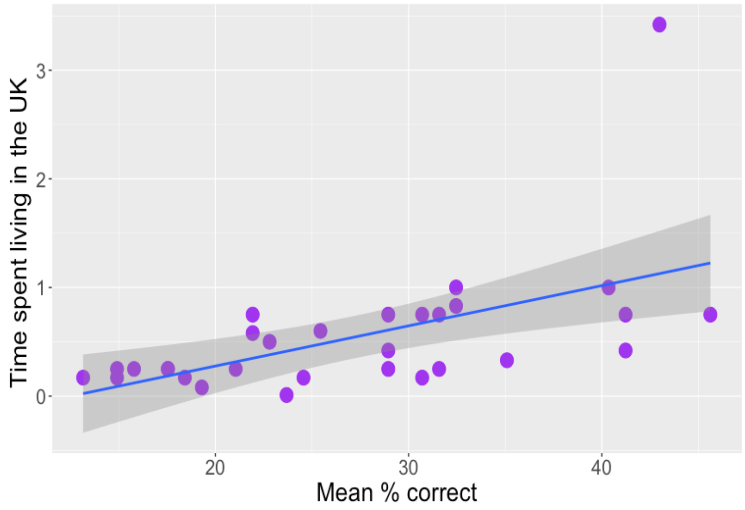


Fig. 5.4. The relationship between time spent living in the UK (in years) and overall mean % of correct responses on the comprehension tests (Pre- and Post-test pooled).

Reduction Type also significantly affected scores (for re-levelled tables see Appendix Q). Participants scored significantly higher on Prepositions than on Determiners ($z = 2.038$, $p < 0.05$), and significantly higher on Prepositions than Participle Endings ($z = 2.520$, $p < 0.05$); the participants' scores on Participle Endings did not differ significantly from their scores on Determiners.

5.7.4. Variables relating to rhythmic skills

The self-rated tapping ability (the variable called SelfTap), had no significant effect in this experiment, and neither did the BAT score. However, since SelfTap had an effect in the previous experiment, and since the BAT data was collected for this experiment, it was still of interest to explore the relationships between BAT scores, SelfTap and other variables relating to rhythmic ability, such as Years and Hours of practising a musical instrument, as well as the participants' self-rating of their general rhythmic skills.

5.7.5. Beat Alignment Test (BAT) score

To explore the relationship of BAT scores to the other musical variables, a separate regression model (glm in R) was conducted. This model had BAT as the dependent variable and Self-rated tapping ability, Self-rated rhythmic ability, Hours of practising a musical instrument, and Years of practising a musical instrument as predictors. The modelling procedure was to first include all the independent variables and then remove the non-significant ones one by one.

Table 5.3 shows the final model:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.6772	2.5442	5.376	9.94e-06 ***
SelfTap	2.1555	0.8155	2.643	0.0133 *

Table 5.3. Final model for BAT score.

The only significant variable was SelfTap. The positive relationship between BAT and the SelfTap variable can be seen in Figure 5.5 below.

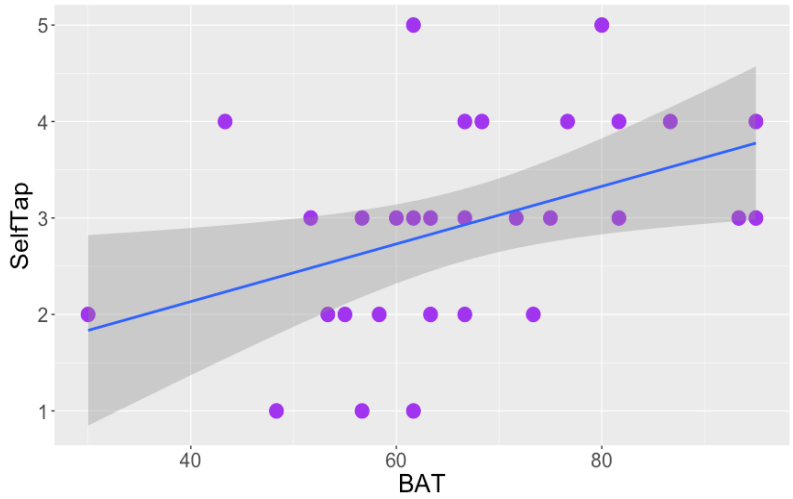


Fig. 5.5. Relationship between BAT score and SelfTap.

Neither Years of musical training, nor Hours of practicing a musical instrument, or Self-rated rhythmic ability were significant in this model.

5.7.6. Enjoyment and usefulness of training according to participants

The variables which measured how much participants enjoyed the training and to what extent they judged the training to be useful were not significant in the modelling. However, the results of these are reported here briefly. Both questions asked participants to rate the enjoyableness and usefulness of the training on a scale from 1-10 (1 being least enjoyable or useful and 10 being most enjoyable or useful). The mean enjoyableness score was 7.5, with a median of 8 (standard deviation of 1.68), and the mean usefulness score was 6.42, with a median of 7.5 (standard deviation of 2.15).

5.8. Discussion

Experiment 3 set out to investigate whether multiple sessions of training involving drumming to the beat perceived in speech could improve EFL learners' comprehension of English connected speech more than a control training which does not use the drumming method. It was predicted that participants who did the course of rhythm training would improve in speech comprehension more than participants who did not. It was additionally predicted that participants with stronger rhythmic skills would have better comprehension scores and/or would improve more after the training.

These predictions were not confirmed. The experimental group did not improve more than the control group, indeed neither group showed detectable improvement at the task; and the difference between experimental and control groups that persisted from Pre-test to Post-

test seems most likely due to the baseline difference in level of English across the two groups.

There are several possible reasons for no effect of training. First, even though there were more training sessions and a greater overall training time in Experiment 3 than Experiment 2, the training differed from that in Experiment 2 in a variety of ways: 1) it was conducted in a group rather than individually, which likely means that the auditory feedback the participants received was different; 2) it offered a different kind of tactile/kinaesthetic feedback than a finger tapping task does; and 3) there was no regularity of repetitions of each of the training sentences, which may have affected the extent to which entrainment could occur.

Let us now address these reasons in turn. First, we need to consider in some detail the role of feedback—both auditory and tactile/kinaesthetic—in sensorimotor synchronisation. Aschersleben & Prinz (1995) investigated Negative Mean Asynchrony (NMA) in three experiments. NMA, as explained in Chapter 3, is the tapper's tendency to tap their finger very slightly before the auditory stimulus occurs. NMA is usually between 20-50ms (Aschersleben & Prinz 1995: 306). An early and still widely-accepted explanation of the NMA phenomenon was based on the idea that the brain uses two sensory codes - one auditory, i.e. the sound of a click, and one tactile/kinaesthetic, i.e. the feeling of a tap (1995: 306). This explanation is called the Paillard-Fraisse hypothesis after Fraisse proposed it in 1980, and credited Paillard for it. The negative asynchrony stems from the idea that if the brain is to perceive the tap and the click as happening simultaneously, then they cannot be happening simultaneously "in the real world" because the processing times involved in the generation of these two codes are different. Sensory information from the tip of a finger takes longer to travel to the brain than auditory information from the ear to the brain, i.e. the times of nerve conduction are different for these. Therefore, taps lead over the clicks in the real world in order to coincide in the brain of the tapper.

In Aschersleben and Prinz's first experiment, participants used hands and feet as effectors for tapping, so that different times would be needed for the different sensory codes (auditory or tactile/kinaesthetic), depending on the effector. They found asynchrony in both hands and feet, confirming the results of Fraisse (1980). As well as this, the asynchronies were larger in the foot than in the hand condition, which is in line with the idea that it takes information longer to travel from the foot than from the hand to the brain.

This experiment was conducted with no auditory or visual feedback on the participants' taps.

In Aschersleben and Prinz's second experiment, participants were given extrinsic auditory feedback in addition to the intrinsic tactile feedback in an attempt to shift the onsets of the codes generated by the taps to an earlier point in time. Asynchronies were found, as in the previous experiment, and were again larger for feet than for hands, even with the auditory feedback. However, these were slightly but significantly smaller than in the condition with no auditory feedback. As well as this, they found a significant interaction of the effector and feedback, suggesting that both kinds of feedback "determine the generation of the common event code" (1995: 312), i.e. both auditory and tactile/kinaesthetic feedback are useful in minimising the NMA, and result in more accurate tapping.

For the purpose of the current experiment, if we assume that the Paillard-Fraisse hypothesis stands, an explanation of the reasons why Experiment 3 did not show a significant improvement in the rhythm-trained group can be offered. That is, if both tactile/kinaesthetic and auditory feedback are necessary for the coincidence of events to occur in a person's perception, then neither of these conditions were fulfilled in Experiment 3. First of all, the participants performed all the training tasks in a group, which therefore means hearing feedback from everyone else's drumming. This may have been very distracting and could have interfered highly with each participant's auditory perception of their own drumming. Furthermore, using drumsticks and the Soundshape drums is quite a different motor action to tapping one finger of one hand on a surface. First of all, it involves more than one limb since one holds the drum and the other holds the stick. This in turn would offer more complex tactile/kinaesthetic feedback to the participant than simple finger tapping.

Aschersleben & Prinz later also conducted experiments testing whether delayed auditory feedback had any effect on the participants' asynchronies. They tested various delays, from 0-70ms, and found "a linear relationship between the size of the delay and the asynchrony between tap and click" (1997: 35).

The results of Aschersleben & Prinz's studies (1995 & 1997) together can serve as possible reasons for Experiment 3 not having brought a significant change in the participants' English speech comprehension. The auditory feedback the participants were receiving was

perturbed due to the training phase being conducted in a group: their ability to relate the tactile/kinaesthetic feedback they were receiving to the sound they were hearing. These cover the possible reasons 1 (auditory disruption) and 2 (tactile/kinaesthetic disruption) mentioned above.

Let us now consider the third reason for the lack of improvement suggested in the opening of this section – namely that there was no regularity of repetitions of each of the training sentences. The results of Experiment 3 could suggest that entrainment of a person's motor action with the external stimulus is necessary for the type of learning to occur that leads to comprehension improvement. Recall that the sentences in the training materials in this experiment were presented only once, i.e. the teacher or the recording said the sentence out loud first and then the students performed the drumming, as in the typical repeat-after-me classroom setup. While this task might still have brought the participants' attention to the different strengths of syllables in English, it may have lacked the entrainment element, i.e. the repeated, or sustained, patterning in time that allows entrainment to occur. This will be discussed further in the general discussion.

Finally, apart from, or in addition to, these three reasons, there may be another overall important reason why training in Experiment 3 did not bring the expected result, while training in Experiment 2 did. A quick look at the Chinese EFL learners' overall comprehension results across the two experiments, participants in Experiment 2 scored generally much higher (around 35%-50%) on the comprehension tests, but participants in Experiment 3 scored generally just above 20% correct. It is possible that this type of task is too difficult for participants at a generally lower level. If we assume that Experiment 3 learners have not yet had the chance to acquire at least some of the rhythmic properties of English that the learners in Experiment 2 had, then this would be in line with some other research suggesting that various prosodic L2 skills develop at different stages of acquisition (e.g. Li & Post 2014).

One or more of these reasons may explain why the experimental training did not yield a significant improvement. However, more research would be needed to investigate these.

The significant effects that were found are those of the condition, i.e. experimental or control, and the amount of time spent in the UK. The CamEng variable, i.e. the participants' score on the Cambridge English test, which places learners in "levels" of their

English language skills, did not significantly predict comprehension scores. This result would perhaps not be very surprising because the test is a 25-question multiple-choice test which does not test any listening skills. However, it was a significant predictor in the previous experiment, and therefore a different explanation is likely. It is possible that in this experiment the variation among participants in their general level of English, which has been assessed by their teachers and used to assign them to different classes meant that those classes they were assigned to predicted the score leaving CamEng without enough variance left to explain. In other words, the fact that these participants were assigned to classes based on their level of English could have given rise to the significant difference in Condition.

However, time spent in the UK was significant, meaning that those who have spent more time in the UK, are better at comprehending Glaswegian than those who have spent less time in the UK, regardless of their scores on the Cambridge English test. .

As mentioned in the general literature review (Chapter 1, Section 1.5), the sole amount of time a person has spent in a country in which the target language is spoken is not usually a significant predictor of the degree of accentedness in their speech production (Piske et al 2001, Strange 2007). The fact that here it does seem to matter may have two reasons.

First of all, the conclusions drawn by these authors are based on measures of speech production, whereas the present study measured the participants' speech perception.

Secondly, it is possible that the amount of time spent in the target language country plays a more important role for highly motivated students. Piske et al (2001) and Strange (2007) found that motivation does affect accentedness. There was no direct significant effect of motivation in Experiment 3, but the students had all moved to Glasgow from abroad in order to learn English, and to study at university, and mostly had a high level of motivation, reflected in a median overall score of 4.2 and mean of 3.8 on the 20 questions asked, on a scale where 1 reflected the most extrinsic motivation and 5 the most intrinsic motivation. Perhaps given these high levels of motivation, they were able to show a direct link between the amount of time in the UK and thus opportunity to speak English they had had, and their comprehension.

Another difference between this and previous experiment is such that the self-rated tapping

ability was not significant here and neither was the more objective measure of participants' rhythmic ability which was included in this experiment, the Beat Alignment Test score.

However, the variables looking into musical and rhythmic training and ability were explored further for their interactions with each other. The results showed was that BAT scores were not significantly predicted by either Years of musical training or Hours spent practising a musical instrument, but were predicted by participants' scores on the SelfTap variable. This finding may support the idea of musical sleepers and sleeping musicians (Law & Zentner 2012). In other words, it is not necessarily surprising that the amount of musical training and practice do not correlate with the participant's beat perception performance, because those who have training may be less inherently skilled and those who have more inherent ability may lack training. However, what is interesting is that the participants' score on BAT *is* correlated with self-rated tapping ability. Even though the participants' BAT score did not show a significant effect in the model in this experiment, and neither did the self-rated tapping ability, these two variables were correlated. Therefore, Experiment 3 shows that self-rated tapping ability does relate to more accepted measures. This is important because it is an extremely quick and easy way to assess rhythmic skill, compared to having to do a more complex test. This may be a possible useful methodological contribution of this experiment.

Finally, the results also showed that scores were higher on Prepositions than on Determiners and that scores on Determiners and Participle endings did not differ from each other. This indicates that Prepositions easier than Determiners and Participle endings. This pattern is again, the same as in the previous two experiments, and therefore the same explanation can be offered for these. The explanation was the following. The phonetic cues are likely easier to distinguish between the different prepositions, e.g. /f/ and /r/ in *for* are easier to distinguish from the /fr/ and /m/ in *from*, than the phonetic cues in reduced Participle endings, *-en* vs *–ing*, or in the reduced *a* vs *her*.

To summarise, this experiment showed an unexpected result, i.e. did not support the prediction that the tested approach would work well in classroom, despite a carefully constructed set of tasks, which were spread over several sessions. Furthermore, the close relationship between task performance and rhythmic skill, which was found in Experiment 2, was not found here. Some possible reasons why the training might have helped comprehension less than expected have been given. What conclusions can be drawn from

these results? It is possible that a classroom setting is not a good context to train these skills using sensorimotor synchronisation as a means for entrainment.

However, there may still be some potential for developing classroom tasks that would use this method. For example, the issue of the auditory and tactile feedback being disrupted due to the drumming task may resolved if the group tasks were supplemented with some individual practice in a setting that encouraged entrainment. This would possibly not involve any instruments but rather silent finger tapping on a hard surface to avoid interference caused by the noise from other participants' drumming.

To briefly reflect on the methods, this experiment was conducted in school rather than lab conditions, which was for two reasons, one being to see whether the method could work in more natural language learning conditions as well as for time-restriction reasons. This, however, meant non-optimal conditions for conducting an experiment, and meant a mis-match in the groups' level of English, which may have played a role in the lack of success of the experiment. A future experiment of this kind would require a more careful matching of participants' general level of English across the experimental group and the control group, in order to for the experiment to be better-controlled.

Finally, even if the participants' enjoyment of judgement of the usefulness of the training that involved drumming was not a significant predictor of their comprehension, this type of training was generally judged to be enjoyable and useful. Therefore, it would be interesting to know whether any aspects of their ongoing learning, or performance on the assessments carried out by the school at later stages would have been affected by taking part in these activities in any way. These data were not collected, however, they possibly could be in a future study of this kind.

Chapter VI

General discussion and conclusion

This chapter discusses the results of all three experiments together, comparing them and identifying their contribution in the context of the various research areas that this thesis touches upon. First of all, given that only one of the three experiments showed a positive result for rhythm training, and even this rested on a marginally significant interaction, the possibility will be considered that rhythmic, or tapping, training does not have a positive effect on the learners' comprehension of English connected speech. After that, the results will be considered in the light of the various previous literature findings and theories. To this end, the results will be discussed first in the light of Dynamic Attending Theory (section 6.1), and then in the context of the literature on Sensorimotor Synchronisation (SMS, section 6.2). Links between entrainment and learning (section 6.3) as well as SMS and learning (section 6.4) will then be discussed, and the results will be viewed in the light of research on perceptual learning (section 6.5). Then, a reflection is made on the role of the learner's L1 linguistic rhythm (section 6.6) and the role of rhythmic abilities in language learning (section 6.7) as well as on P-centres (section 6.8) in the light of these results. Finally, a conclusion and some future research directions and applications are discussed.

6.1. The possibility that rhythmic training does not help comprehension

The interaction of Test and Condition in Experiment 2 was only marginally significant, even if the experimental group improved significantly and the control group did not. This could be partly a function of lack of power (as could the other experiments' negative results). Nevertheless, taken together with the negative results from Experiments 1 and 3, the rather weak nature of Experiment 2's finding means that we need to give serious consideration to the possibility that rhythmic training simply does not work.

There could be several reasons why rhythm training might not aid comprehension of connected speech. First of all, it is possible that training that focuses on using speech that is in a sense artificially made to be isochronous does not aid comprehension of connected speech, because real connected speech is rarely isochronous itself. Therefore the materials heard in the experiments were likely to be fairly new to many of the participants in the sense that they were isochronous. This is a reasonable objection, but as argued in Chapter I, speech does have rhythms, and, while isochrony does not usually exist in natural speech, there seem to be more complex rhythmic patterns in speech (which include not only timing, but also other dimensions, such as pitch and loudness), which native listeners are attuned to, which is possibly what helps their more efficient comprehension of speech that

is native to them, as opposed to non-native speech.

Second, it is possible that comprehension of reduced forms could not be improved by the rhythmic methodology used in the experiments in this thesis because the rhythmicity of the speech foregrounded the strong, or stressed, syllables, whereas the targets were the weak/unstressed syllables. However, this does not seem to accurately reflect the nature of attention. There are many studies that show that speech perception is rhythmically guided which are reviewed in section 1.3.9 above, which suggest that it is the overall rhythm that facilitates speech comprehension, rather than merely the prominence that certain syllables have. How listeners would have possibly attended to the unstressed syllables in these materials is discussed in more detail in section 6.2 below. To get more clarity on this issue in future work, testing of the comprehension of the strong syllables as well as the weak ones in future experiments would be helpful, to show whether there was a difference in improvement in comprehension of strong as opposed to weak syllables. Alternatively, a training method which entailed foregrounding of the weak syllables in some way could also be helpful, e.g. if the participants listened to the reduced sections of the utterances in isolation, as well as in their original context.

Furthermore, it is possible that rhythmic training does not aid connected speech comprehension because it is not explicit about what the learning targets are. It is possible that an additional explanation of what connected speech processes are, and how they work combined with the training would have brought a greater effect. This explicit explanation might entail for example the participants being explicitly taught about vowel reduction to schwa, and how segments become elided, or deleted. The lack of explicit instruction does not seem likely to be the whole story: there are plenty of studies which use training based on implicit rather than explicit learning, which show improvement in participants' speech comprehension (e.g. Adank et al 2009, Floccia et al 2009, Sumner & Samuel 2009, Adank et al 2010, Mitterer & Tuinman 2012, Morano et al 2015, Lennon 2017, Smith & Rathcke 2017, to name a few of the studies looked at in Chapter I). However, it might still be that what learners need is to have their attention brought to connected speech processes, rather than to rhythm. If a training approach that foregrounded connected speech processes succeeded where rhythmic training failed, it would suggest that it is connected speech processes themselves that learners need to be taught about, and that their connection to rhythm is either not very strong, or not very important from the perspective of learning.

Given the question mark over the strength of Experiment 2's positive finding, a replication study would be the logical next step in order to confirm whether or not tapping training does significantly improve connected speech comprehension. In addition to a straightforward replication, it might be a good idea to compare a different, more widely known, or more widely used, training method as a control. This could be, for example, training using feedback where students perform tests and are offered immediate feedback on their responses. It could also be training their knowledge of connected speech processes in English. Only then could we be sure whether or not SMS training significantly improves the learners' comprehension of connected speech relative to other training techniques.

Due to time and space constraints, no replications or extensions could be attempted in the present thesis. Therefore, given that Experiment 2's finding *was* positive (even if weakly so), the approach taken in the sections that follow is to try to interpret the combined pattern of results across the three experiments in the light first of Dynamic Attending Theory and sensorimotor synchronisation literature, and then of theories of perceptual learning. The focus is on accounting for why Experiment 2, but not the other experiments, yielded support for the hypothesis that rhythm training would benefit connected speech comprehension. This discussion will, in future work, be used to inform the design of replication and extension studies; if those, too, turn out to produce negative results, then the logical conclusion will have to be that rhythmic attention does not help connected speech comprehension.

6.2. Results in the light of Dynamic Attending Theory

Dynamic Attending Theory says that human attention entrains to the stimuli coming from the outside world (Jones 1976). In the training phase of Experiment 1, the aim was to achieve entrainment through listening to highly rhythmic stimuli. In the training phase of Experiment 2, a higher degree of entrainment was encouraged through additional sensorimotor synchronisation of a finger movement with the rhythmic stimuli. In the training phase of Experiment 3, the task was to use drumming to entrain the movement with the external stimulus. As well as this, the training phases of experiments 1 and 2 offered the participants four concatenated repetitions of each sentence, while the training phase of Experiment 3 offered only a single presentation of each sentence. Crucially, Experiment 2 showed a significant improvement in the EFL learners' comprehension from pre-test to post-test as compared to the control group. Therefore, we need to consider the differences in the design of these three experiments in terms of how entrainment was to be

achieved.

Experiment 2 was different in two important ways from both Experiment 1 and 3. Firstly, the training materials in Experiment 1 and 3 were not as regular as the materials in Experiment 2. Experiment 1 did have four repetitions of the same sentence concatenated into a long string but the pauses between the repetitions were not of the same duration and the start of each repetition was not signalled by a tone, as was done in Experiment 2. Therefore, while participants had the same sentence repeating four times in both experiments, they may have been able to rely more on the regularity in Experiment 2 than Experiment 1. Similarly, in Experiment 3, each sentence was presented to the participants only once. Participants could listen to it multiple times but there was no regularity between the repetitions: the sound files were played by the teacher (i.e. the musician or the present researcher), who would warn the participants when the stimulus was about to play, but not as systematically as in Experiment 2, i.e. with a tone of a certain frequency, duration and amplitude followed by a silence of a certain duration.

This may be important because, as Dynamic Attending Theory says, an expectation is created when synchronising, or entraining, attention to a regular stimulus. Large and Jones (1999) propose that the two elements which entrain, and are therefore coordinated with each other, are the external rhythms (in our case the speech stimuli) and the internal attending rhythms. The external rhythms can be isochronous or non-isochronous. In other words, an external rhythm is seen as a sequence of regular pattern of onsets, but periodicity does not always need to be constant: the period of the components of such a rhythm can systematically change over time. The internal attending rhythm is seen as a self-sustaining oscillation which generates a periodic activity which is called *expectation*. This expectation actively anticipates events within the external rhythm, i.e. couples with it. When such coupling, or synchronisation, is stable, it can deal very well with perturbations, or fluctuations, that may happen in the external rhythm. As well as this, the more stable the rhythm in the stimulus, the stronger, i.e. more focused and narrower the attentional pulse, as outlined in more detail in section 1.3.8.

In terms of the current work, Experiment 2 gave the participants the most temporally regular training. They could form expectancies as to when each repetition of a sentence would start, due to the regular tones followed by regular silences. As well as this, after the first repetition which allowed the participants to listen to the slow and least reduced

version of each sentence, the next three repetitions were exactly the same, therefore, temporal expectancies could form as to when the various elements of these sentences would occur. If, for now, we think of only the stressed syllables, which the participants tapped to, these occurred at three stable points, a certain amount of time after the initial warning beep sound. The attentional energy pulse for these could then be illustrated as in Fig 6.2 below.

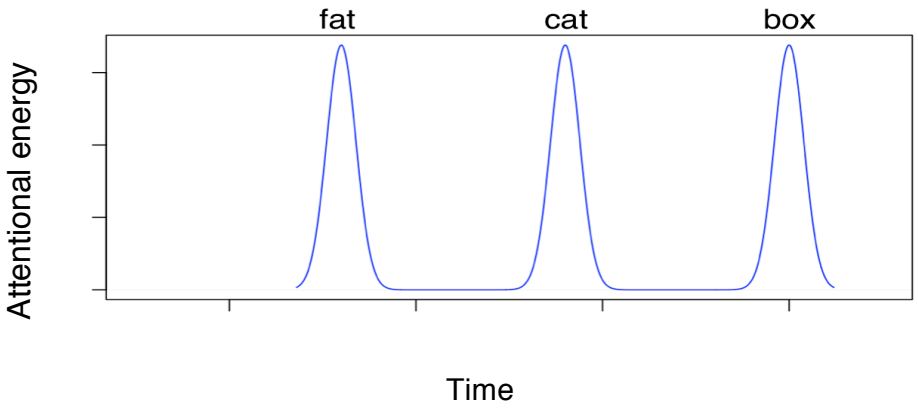
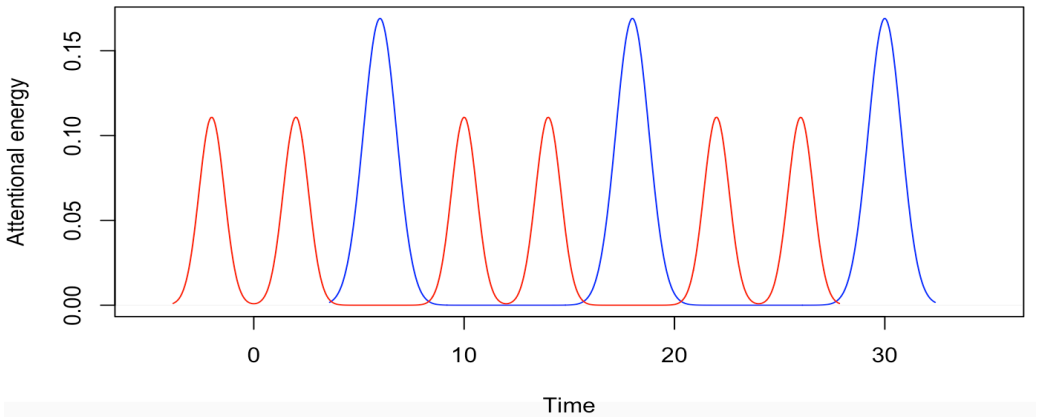


Fig. 6.1. Hypothetical attentional energy pulse for the stressed syllables in the sentence *He is fat for a cat in a box* over time. Units are arbitrary.

However, as Large & Jones (1999:132) say, in the real world we experience complex rhythms which contain multiple pseudo-periodicities, and thus we will use multiple attending rhythms which will couple with them. Therefore, it is not only the rhythm of the strong beats that our attentional pulse is coupled to, but also to that of the weak beats. Furthermore, Large and Jones (1999) state that when multiple internal rhythms occur, they are also coupled with one another so that their phase and period relationships can be preserved. The same idea would work for the external rhythm that participants in this Experiment 2 synchronised to, i.e. speech. Here the elements of each utterance are also in phase relationships with one another. Thus, as in our example, the stressed syllables *fat*, *cat* and *box* are not only in a relationship with the internal attentional pulse, but also with one another; in this case, that relationship is a simple one, i.e. they occur pseudo-periodically every third syllable in the case *He is fat for a cat in a box.* This is illustrated in Figure 6.3 below. As well as this, in this example there are two unstressed syllables in between each stressed syllable, i.e. *he is*, *for a* and *in a* and these are also in relationship with each other as well as with the stressed syllables, and with the attentional pulse.

A



B

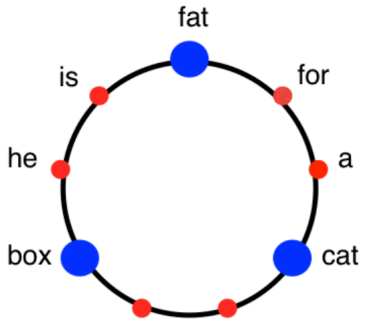


Fig. 6.2. Both pictures in A and B represent the hypothetical attentional pulses for both the stressed and unstressed syllables in the sentence *He is fat for a cat in a box*. Units are arbitrary. A represents these relationships in a linear manner, while B represents them in a circular manner.

Compared to Experiment 2, the training sentences in Experiment 1 may have led to poorer entrainment due to their regularity being perturbed because of the change in rate. That is, in both Experiments 1 and 2 participants listened to four repetitions of each sentence. However, in Experiment 1 all four repetitions were different productions of the sentence at a different rate, while in Experiment 2 only the first repetition was at a different rate than the remaining repetitions. While the phase relationships between the elements within each repetition cycle were likely similar with every repetition (though we need to remember about the reduction processes which take place more the faster the speech), the four repetitions of a sentence from Experiment 1 compared to the four repetitions in Experiment 2 taken as a whole, could have given a less stable attentional pulse. In other words, one reason why Experiment 1 was less successful at training the participants' comprehension of Glaswegian speech may relate to the stability and strength of the attentional pulse. The attentional pulses would have been less focused and therefore broader in Experiment 1 than in Experiment 2 because each production of a repetition in Experiment 1 was new, at a different rate, rather than a copy of the same one as was the case for the last three

repetitions of each sentence in Experiment 2. The stimuli for Experiment 2 also had clearly timed warning beeps followed by a silence of the same duration, which may also have made the attentional pulses narrower and more focused than those in Experiment 1. Altogether, the expectancies that participants could form as to the precise moment of occurrence of the next event was less predictable in Experiment 1 than in Experiment 2, even if the material was the same and the elements within each repetition cycle remained in very similar phase relationships with each other. For example *fat* would remain in a very similar phase relationship with the next syllable, *for* whether it was the first, or the second repetition, but the timing of these elements from cycle to cycle would change, as illustrated schematically in Fig. 6.4 below.

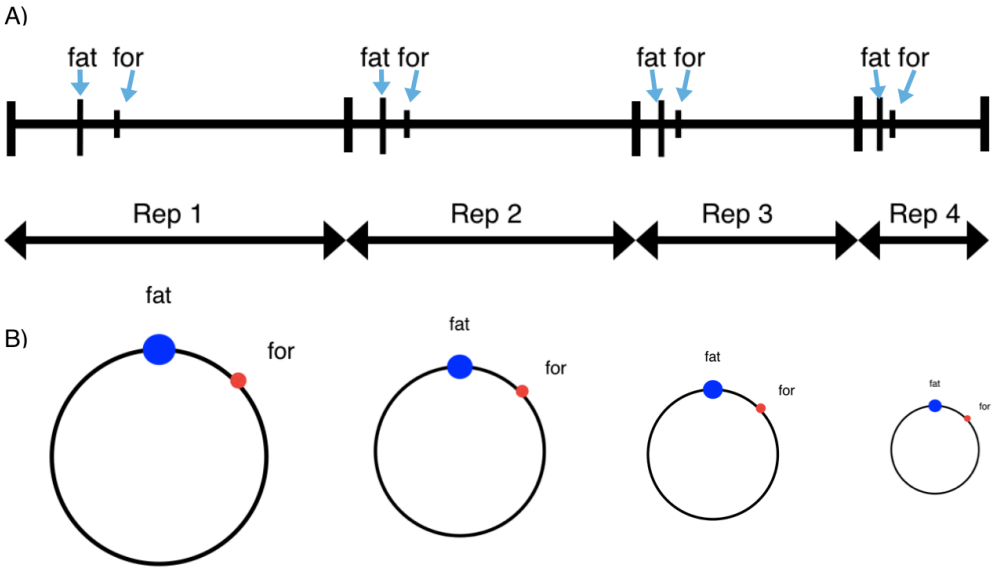


Fig. 6.3 A) A schematic representation of the utterance repetition cycle of the experimental condition in Experiment 1 and the relationships between the elements in the utterance repetition cycle. B) A schematic representation of the relationships between two elements from the utterance *He is fat for a cat in a box,* in four circles of different sizes; the size of the circle depends on the speech rate at which the utterance was produced.

Both panels A and B in Fig 6.4 above help to illustrate that if time was measured from the start of the utterance to the end separately for each utterance, the relative timing of each element within the utterance repetition cycle would likely remain constant, i.e. in B *for* occurs after *fat* at the same moment relative to fat in each repetition cycle, regardless of rate (cf Port 2007), i.e. with the same phase. The faster the rate, the smaller the circle, to represent the diminishing amount of time the whole cycle takes. However, taken as a whole, the timing of each *fat* and each *for* varies, i.e. speeds up relative to the previous one.

Having to dynamically adjust the phase every repetition would then, according to Dynamic Attending Theory, cause the attentional pulse to broaden and focus to be less concentrated, as was shown by the curve K=1 in Fig. 6.1 above.

Comparing this to Experiment 3, there were no repetitions concatenated, but rather each presentation of a sentence was at the same rate and its onset was controlled by the teacher, or the experimenter. Therefore, the participants' attention can be expected to have been perturbed to an even larger extent by the irregularity of the breaks between the repetitions.

6.3. Results in the light of sensorimotor synchronisation (SMS)

It is possible that these differences between the presentation of training stimuli in the three experiments alone are the reason for the key result that only Experiment 2 showed significant improvement in the participants' connected speech comprehension. While this may seem like quite a plausible explanation, there was also another factor which may have contributed to the pattern of results. In Experiment 1 participants were asked to merely listen to the sentences, while in Experiment 2 and 3 they performed a SMS task. Of these, only Experiment 2 showed significant improvement. On the surface, this result may seem confusing: how could performing an SMS task have made a difference for the participants in Experiment 2 but not in Experiment 3? However, these SMS tasks themselves were quite different: in Experiment 2 the participants engaged in individual finger tapping to the stimuli, in Experiment 3, the participants drummed to the stimuli in a group. As reflected upon in the discussion section for Experiment 3, there may be some tactile/kinaesthetic issues there as well as the fact that the auditory feedback was disrupted due to group drumming.

An interesting future direction would be to conduct Experiment 1 again, and to create an experimental condition in which participants would perform the tapping task to materials from Experiment 1. This would make it possible to test whether tapping improved learners' comprehension of connected speech. To improve their regularity, the materials from Experiment 1 could be adjusted such that there would be warning beeps of constant durations between each repetition followed by a silence of constant duration as well. Corresponding adjustments could be made to the control condition. The conditions would be the following: tapping vs listening only, at a gradually speeding up vs a more stable tempo. This would allow a partial replication of Experiment 2, and a more direct comparison between Experiment 1 and Experiment 2. There is evidence showing that

participants are able to adjust their tap timings to a speeding up or slowing down metronome click (e.g. Repp 2001, Schulze et al 2005), which suggests that participants may be able to perform this task.

The results of this proposed experiment would help to answer the question of whether it is the fact that participants performed a SMS task as a part of their training that contributed to the significant improvement of their comprehension of connected speech, or whether that effect derived more from the fact that they could form more focused and narrower attentional pulses in Experiment 2, and their expectancies formed as to the timing of the following speech event were upheld.

6.4. Results in the light of perceptual learning

Another way to think about the differences in improvement observed in the experiments is to consider the opportunities for perceptual learning that each experiment afforded.

Training in Experiment 1 can be described as High Variability Perceptual Training (HVPT) in that it used multiple speakers, two male and two female (Iverson et al 2005). It also used a modified version of Perceptual Fading (PF), i.e. it started out with the slowest repetition containing the least amount of reduction and through the remaining repetitions which gradually increased in rate, it demonstrated the various degrees of reduction that can occur when speech becomes faster.

Training in Experiment 2 involved no HVPT, as only one speaker was used. It did involve PF, but arguably to a lesser extent than Experiment 1, because one slow version of each sentence was followed by three repetitions at a faster rate, which were all identical to one another.

Training in Experiment 3 did not use HVPT, as only one speaker was used. It included recordings of various phrases from simple ones such as *Hot and cold* to more complex ones such as *Fructose is a naturally occurring sugar*, but no PF in the sense that slower/more careful as well as faster/more reduced versions of the same phrase would be presented to the participants. It did, however, contain a variety of tasks which directly demonstrated to the learners where the strong as opposed to weak elements in the speech occurred.

Training for Experiments 2 and 3 both involved a motor activity, i.e. drumming or tapping the beats out. Participants' perception of the speech materials in Experiment 3 was possibly hindered by the general noise generated by the drumming of other learners in the group, whereas in Experiment 2 training was conducted on a one-to-one basis and even then, they could not hear their own keyboard presses, due to the keyboard being silent. In that way, in Experiment 2 participants were given more chance to hear the materials while drumming to their beat. For any possible future experiments which would like to test usefulness of drumming/tapping in class, a solution to this issue would need to be found, to ensure that the motor task would not interfere with the perception of the auditory stimuli.

In summary, the modified version of PF with no HPVT, i.e. one in which the learner is presented with one unreduced version of speech repeated once and one reduced version but repeated three times (Experiment 2) brought a better result than the HPVT combined with PF on a larger scale, i.e. 4 different gradually more reduced versions of the same utterance (Experiment 1). However, we do not know if the lack of significant improvement in Experiment 1 was because HPVT is not useful for this task, or because of some other factors. This could be an area for future research. On one hand, it is possible HPVT is not suitable for tasks which rely on entrainment to speech, because different talkers can have different speech timing, rhythm and tempo patterns, or habits, and therefore switching the talker might disrupt entrainment. On the other hand, HPVT does seem to have a large amount of research behind it showing it supports perceptual learning. This could be looked into by setting up an experiment in which favourable conditions for entrainment were created, but which would also use variability in voices. For example, ensuring all speakers used the same tempo which could be achieved by eliciting speech in a way similar to the way it was elicited in Experiment 2, i.e. asking the speaker to produce speech after the metronome stops.

6.5. Linguistic rhythm

While this thesis did not set out to test how learners from rhythmically different L1 backgrounds respond to rhythmic training, a few suggestions come to mind when considering the results of the experiments conducted for this thesis. In all three experiments, most (in Experiment 1) or all (in Experiments 2 and 3) of the EFL learners were from a Chinese background, reflecting demographics of EFL learners in Glasgow currently. Chinese is rhythmically different from English, in that it is usually classed as a syllable-timed language. Previous studies (e.g. White & Mattys 2007, Astruc et al 2013, Li

& Post 2014, reviewed in section 1.7.1) have shown L2 rhythm is harder to acquire if the learner is of a language background with a different rhythm class.

As well as this, Canadian/US listeners' comprehension did not significantly improve, though compared to native Glaswegians, there was a small amount of room for improvement. A speculation comes to mind, that this may be not only because the Canadian/US listeners were also native speakers, albeit of a different variety, but possibly also because this particular method would be more useful for listeners from a language background of a different rhythmic class.

To sum up, it is possible that the tapping-to-the-beat method is more useful for learners of a rhythmically different language, such as in these experiments Chinese natives learning English. This question could be addressed if native English-speaking listeners of a contact-variety, e.g. Singapore English, which is thought to be syllable-timed (Deterding 2001), were to perform the same task: their comprehension might improve more than the Canadians'.

6.6. The role of rhythmic abilities in language learning

In Experiment 1 data on participants' musical ability were not available to use in the analysis. In Experiment 2 a more detailed questionnaire about the participants' musical abilities revealed that the measure that predicted their comprehension ability to some degree was their self-rating of their own ability to tap in time with music (called SelfTap). In Experiment 2, for the Glaswegian and the Chinese participants, the lower the participants' self-rated tapping ability, the lower their speech comprehension test score was. For the Canadian participants, the higher their self-rated tapping ability, the higher their improvement on the comprehension test after training.

Because of these results, in Experiment 3, Beat Alignment Test (BAT) was used with the intention of including a more objective way of measuring participants' rhythmic skills alongside the same questions as in Experiment 2. Here, surprisingly neither the BAT score nor the SelfTap rating significantly predicted comprehension score. However, a separate modelling of BAT showed that SelfTap was – as for Experiment 2 - the only significant predictor of BAT score. Reported Years of musical training, Hours per week practising, and even Self-rated rhythmic ability (which would seem very close to the Self-rated tapping ability variable) did not predict the BAT score.

What is it exactly that the SelfTap variable captures? It is possible that participants were extremely accurate at rating their own ability to tap in time with music, and that it is simply the answer to the question asked which was *Please rate your ability to tap in time with music*. This in turn reflected their ability to tap to the beat in speech, as shown by Experiment 2, but not by Experiment 3. Experiment 3, however, showed that the participants' perceptual beat alignment ability (as measured by BAT), is indeed predicted by the SelfTap variable. This could be interpreted such that participants had the awareness of how well they were able to tap to the beat in music and that this was correlated to their finger tapping ability (as in Experiment 2).

That said, it is important to remember that the questionnaire was administered after the experiment and therefore the participants' responses to the SelfTap question could have been based on their actual experience of tapping to the beat in the speech task. This would mean participants were remarkably good at noticing how well they were actually able to synchronise their taps with the beat in the speech they were listening to. Moreover, this could explain why SelfTap did not predict comprehension score in Experiment 3, because participants got less informative feedback about how well they were synchronising their drumming to the beat.

Since SelfTap was a significant predictor of the participants' general comprehension of connected speech in Experiment 2, it is an indication that having a good beat-finding ability is linked to better comprehension of connected speech in an L2, but also in L1 and in L1 of a different variety. Additionally, the fact SelfTap was also linked to improvement in the Canadian/US English native speakers in Experiment 2 suggests that training using rhythmic stimuli could be beneficial for training connected speech comprehension.

6.7. P-centres

The tapping data collected in Experiment 2 could also be analysed in order to add to the research on P-centres: that is, to establish with which event in connected speech participants tend to time their taps, be this related to vowel onsets, syllable onsets, salient f0 events, or some combination of these As mentioned above, this was not attempted in depth in this thesis, due to time and space restrictions. What these data could potentially contribute to is understanding the differences in how P-centres are perceived by speakers of different languages or varieties. If we assume that the tapped-to event is the P-centre, or

the moment at which an event is perceived to occur, then we can ask the following questions. Do Chinese listeners perceive the moment of occurrence of the beat in EFL in a different way to native speakers? Is that moment the same for Glaswegian vs Canadian/US speakers? As well as this, how close were the taps of non-natives and speakers of a different variety to the timings of taps of the native speakers? Was it the case that those who tapped to the same event as the native speakers were better at comprehension?

6.8. Summary, or how to train your entrainment

To summarise the above reflection, it is clear from the results, especially those of Experiment 2, that there is a relationship between speech comprehension, rhythmic ability and entrainment, or tapping ability, which could be illustrated graphically, as in Figure 6.5 below.

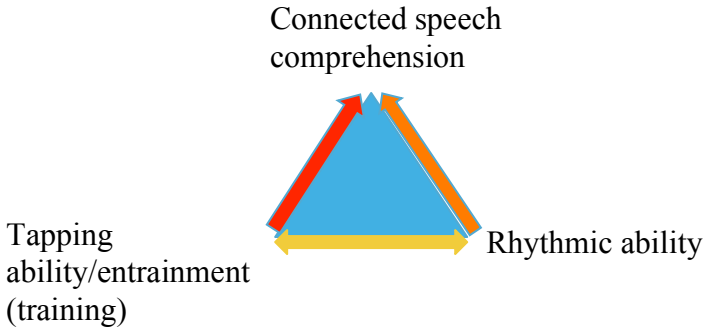


Figure 6.5. An illustration of the relationship between rhythmic ability, tapping ability or ability to entrain and connected speech comprehension.

All three components are related to each other. Having a rhythmic ability may mean having a natural aptitude, acquiring skill through training, or both (cf. Law & Zentner 2012). However it arises, this rhythmic ability will exhibit itself in being able to tap to the beat, whether this is in music or speech. Tapping to a steady beat, in turn, is a means to entraining one's behaviour with the external stimulus. As Experiment 2 revealed, entrainment achieved via tapping, i.e. synchronising one's movement with a beat, seems to be a valid method for improving listeners' connected speech comprehension. As well as that, the higher the individual's rhythmic ability, the better their connected speech comprehension. It seems plausible that training one's rhythmic skills can lead to both enhanced tapping ability (yellow arrow in Fig. 6.5 above) as well as enhanced connected speech comprehension (red arrow in Fig. 6.5 above).

Rhythmic ability is therefore linked with connected speech comprehension in two ways.

The first one is direct, meaning that better rhythmic ability is linked to better comprehension of fast casual speech, which due to connected speech processes undergoes a variety of temporal changes compared to careful speech (orange arrow in Fig. 6.5 above). The second one is less direct: rhythmic ability is also linked to the ability to entrain to the beat, which in turn is linked to the ability to improve via exposure to speech, so entrainment can be used as a training method for connected speech comprehension. The key point is that the relationships appear to be reciprocal.

6.9. Unanswered questions and future directions

This section discusses areas for future research. Some of these have been mentioned in the discussion above, such as whether the effects of this type of training would be lasting, and whether generalisation would take place (i.e. whether the training would improve the learners' comprehension of sentences with different metrical structures, sentences produced by unfamiliar speakers, or sentences containing different types of connected speech phenomena).

There are, however, more questions. One of them is to investigate the effects of tapping on speech production, specifically the production of connected speech and intelligible rhythm in L2. This type of data could be added quite easily to the existing experiment design, by asking the participants to produce sentences of similar structure before and after the training. Rhythm metrics could be used to compare the pre-test and post-test speech with native speaker productions. Alternatively, or additionally, the pre-test and post-test recordings could be rated by native speakers on how well connected speech was produced, or how native-like the rhythm sounded.

It is possible that training using SMS with the beat in speech in form of tapping might help improve temporal aspects of production, better than speaking/imitating, or at least would complement speaking. This is because it is possible that SMS training provides a clearer, or more explicit and easy to observe, kind of feedback as to the timing of events, because it can be seen as well as felt, i.e. contain tactile feedback through involving the motor system and through the feedback from the finger's touch. Therefore, the question arises whether it could be that SMS using finger tapping provides better training as to *when* events occur (e.g. P-centres of syllables) in a foreign language than other kinds of training with synchronisation, e.g. synchronous speech, or a simple vocal imitation? In other words, would the "articulatory reference point[s]" (Chow et al 2015) in terms of speech timing

become clearer to the learner if it was tapped to, rather than when repeating of the material was attempted? Relatedly, would those timings of the non-native and different-variety speakers become closer to the native ones with SMS training?

Finally, a question arises whether the difference results between Experiment 2 and 3 was due to the difference in training method or due to the difference in learner proficiency? Experiments 2 and 3 used a different kind of synchronisation method, one using finger tapping and one using drumming. As outlined in detail in discussion for Experiment 3, the fact that improvement was achieved only in Experiment 2 in which finger tapping method was used but not in the Experiment 3 that used drumming, could be to do with the tactile/kinaesthetic as well as auditory feedback disruptions. However, as also mentioned in the discussion to Experiment 3, it is possible that the drumming method in Experiment 3 did not bring a significant result because participants in Experiment 3 were at a lower level of English and therefore did not benefit from the drumming method. This also points to the need for more investigation. For example, how and at what stage of L2 learning can SMS be helpful?

6.10. CLLAPP - a practical application of the research findings

From the outset of the thesis research, there has been an aim to put the fundamental findings/methods into something practical for learners and the project has been is collaborative with Micro-phonics Ltd. As reviewed in Chapter 1, Computer Assisted Language Learning has a variety of advantages, the main one perhaps being that the student is able to use these kinds of applications on their own, for self-study.

As far as rhythm-training games are concerned, Bégel et al (2017) recently reviewed the readily available games that train rhythmic skills. Using technology for re-education, such as rehabilitation of movement for example, is called *Exergaming*. The games they reviewed had to fulfil two conditions: they had to focus on rhythmic skills and they had to record the temporal precision of the players' responses. They then categorised the games in terms of: "(a) the peripheral used to capture and record the response; (b) the type of response that is recorded; and (c) the output" (2017: 3). They found games that involve full body movement (e.g. ones for Kinect or Wii), ones that involve finger tapping on a touch screen (e.g. Beat Sneak Bandit), or finger tapping on a key (e.g. Guitar Hero), and games involving singing, i.e. karaoke-type games.

Bégel et al (2018) have been working on developing a music-based game whose aim would be to train rhythm skills. In this game, they included tasks of progressive difficulty, which they based on consistency and accuracy of synchronisation in their first experiment. Their second experiment was a proof of concept study, which trained participants to either synchronise with the stimuli, i.e. tap to them, or trained their perception of rhythm. These results were compared with the results of a control group who received no training. After two weeks' training on either synchronisation or perception, a significant improvement of beat perception (but not production) occurred compared to the control group.

If this type of game can be successful at training the learner at musical rhythm, and if musical and speech rhythm are linked, as discussed in section 6.9 above, then making a rhythm game for improving speech comprehension in L2 would be a natural way to apply the findings of this thesis. The main finding of this thesis was that Chinese EFL learners' comprehension of fast connected speech as produced by a Glaswegian native speaker was improved after the training that involved tapping their finger to the beat in that speech and entraining to that beat. Since this result was obtained in laboratory conditions, with only one participant at a time, what resulted was the idea to create an app, which would be used by learners for self-study. In-class application of rhythmic training techniques appears to be more difficult, possibly due to the fact that the auditory as well as tactile feedback was disrupted by other attendees' noise. Therefore, it seems more appropriate, at least initially, to develop a way of implementing the findings for individual training, rather than group training.

Such an app would aim to provide the learner with SMS training, using tapping to the beat in speech similarly to the way participants tapped to the beat in speech in Experiment 2 of this thesis. As well as this, it would allow this research to grow by collecting the participants' speech comprehension and tapping data, asking them for demographic information as well as their musical ability, motivation to learn etc, as in the experiments in this thesis.

The app would include short videos which would take the user through basics of musical rhythm and demonstrate how to synchronise tapping with the speech materials. It would also make the connection between music and speech by demonstrating tapping to the beat of both music and speech. The materials would start with easy, i.e. shorter and slower, musical as well as speech materials and they would take the learner from easiest to more

complex rhythms. It would use speech materials that listeners would tap along to, while receiving feedback on their tapping accuracy, and having the opportunity to practice under conditions that which would be set up to optimize entrainment.

The data collected through this app would be the participants' tapping timings, speech comprehension responses collected throughout the course as well as their demographic background. An app which is freely available would allow to collect data from a wide range of learners. This in turn may enable research of such issues as the following: How well listeners' motor action needs to be entrained with the speech beat in order for a comprehension benefit to be seen? Do speakers of different types of languages perceive beat in speech in the same way, or more precisely, at the same time, i.e. are P-centres the same for L1 speakers of different languages? As well as this, the musical ability variable could be explored in more depth using this app, to establish whether those with more musical training or musical aptitude benefit more, or whether those with little training in music could catch up through practice, and if so, whether their improvement would generalise to musical tasks. The data could also help answer other questions such as whether participants' L1 has an effect on their comprehension of English connected speech, i.e. if their learning is affected by the rhythm class of their L1? And if so, is it the case that L1 speakers of a syllable-timed language would benefit from this training more than L1 speakers of a stress-timed language, or vice versa? Answering these questions and more could potentially be tackled by means of a freely-available app.

6.11. Conclusion

The experiments reported in this thesis set out to test the role of attending to the rhythm in speech in perceptual learning of English as a Foreign Language. Experiment 2 in this thesis provided some evidence that entraining to the rhythm in speech in the target language can be a useful learning tool improving the learners' comprehension. In order to be sure that this is the case, more research is required. If a replication of the study showed positive results, we would be able to say that, supporting Dynamic Attending Theory, such learning occurs when the listener forms expectations as to what event occurs next. The results of the three experiments taken together suggest that for the learner to be able to entrain to speech well enough, it may not be sufficient to merely listen to rhythmic speech in the target language, but that engaging in a simple sensorimotor synchronisation task, achieved through finger tapping, and thus entraining that movement to that speech is more helpful. As well as this, there is reason to believe that the tactile (from the motor part of the task) as

well as auditory (from the listening part of the task) feedback, which the learner receives during such training, seems important. In other words, the learner requires to be able to perceive auditorily when the exact moment occurred to which they entrained their attention and movement, or to which they performed the sensorimotor synchronisation. Put simply, it seems likely that the learner needs to hear the sound and feel the moment when their finger touched on a surface at the same time.

Interestingly, the results also show some support for the role of musical training, or musical ability, in determining the extent to which listeners can benefit from training with auditory and motor entrainment. This suggests that having, or gaining, musical abilities is beneficial for speech comprehension, as has been found in other areas of speech and language research from dyslexia to speech and language impairments.

Bibliography:

Abercrombie, D. (1967). *Elements of general phonetic*s. Edinburgh: Edinburgh University Press.

Abercrombie, D. (1979). The accents of Standard English in Scotland, In Aitken, A. J. and McArthur, T. (eds) *Languages of Scotland.* Edinburgh: W. and R. Chambers, 68–84.

Adank, P., Evans, B., Stuart-Smith, J., Scott, S. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology Human Perception Performance. 2009 Apr;35*(2):520-9. http://dx.doi.org/10.1037/a0013552.

Adank, P., Hagoort, P., Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science, 21*(12), 1903-1909.

Adank, P., & McQueen, J. M. (2007). The effect of an unfamiliar regional accent on the speed of word processing. In J. Trouvain and W. J. Barry (eds), *Proceedings of the XVIth International Congress of Phonetic Sciences, Saarbrücken, Germany,* 1925–1928.

Aitken, A. J. (1981). The Scottish Vowel Length Rule. In Benskin, M., & Samuels, M. L. (eds), *So meny people longages and tonges: Philological essays in Scots and Mediaeval English presented to Angus McIntosh* (131-157). Edinburgh: Middle English Dialect Project.

Aitken, A. J. (1984) Scottish accents and dialects. In Trudgill, P. (ed) *Language in the British Isles.* Cambridge University Press.

Alexander, J and Wang, G. (2016). The perceptual assimilation model for suprasegmentals and cross-language lexical-tone identification. *The Journal of the Acoustical Society of America 139*, 2161 (2016). https://doi-org.ezproxy.lib.gla.ac.uk/10.1121/1.4950403.

Altmann, H. (2006). *The perception and production of second language stress: A cross-linguistic experimental study* (Unpublished doctoral dissertation). University of Delaware, Newark.

Anderson-Hsieh, J., Riney, T., & Koehler, K. (1994). Connected speech modifications in the English of Japanese ESL learners, *Issues and Developments in English and Applied Linguistics, 7,* 31-52.

Anvari S. H., Trainor L. J., Woodside J., Levy B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *J.*

Exp. Child Psychol. 83 111–130. 10.1016/S0022-0965(02)00124-8.

Archer, G. (2018). Pronunciation models in regional environments: a comparison and assessment of RP and SEE. Unpublished MRes dissertation.

Archibald, J. (1997). The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics*, *35*, 167–181.

Arvaniti, A. (2009). Rhythm, Timing and the Timing of Rhythm. *Phonetica*, *66* (1-2), 46-63.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. doi:10.1016/j.wocn.2012.02.003.

Arvaniti, A., & Ross, T. (2010). Rhythm classes and speech perception. *Proceedings of Speech Prosody 2010*, Chicago, 11-14 May 2010.

Aschersleben, G., & Prinz, W. (1995). Synchronizing actions wtih events: The role of sensory information. *Perception and Psychophysics*, *57*(3), 305–317.

Aschersleben, G., & Prinz, W. (1997). Delayed Auditory Feedback in Synchronization. *Journal of Motor Behavior*, *29*(1), 35–46. http://doi.org/10.1080/00222899709603468

Astruc, L., Payne, E., Post, B., Vanrell Bosch, M., & Prieto, P. (2013). Tonal Targets in Early Child English, Spanish, and Catalan. Language and speech. 56. 229-53. doi:10.1177/0023830912460494.

Baker, R. (2008). The production and perception of morphologically and grammatically conditioned phonetic detail. Unpublished PhD dissertation, University of Cambridge.

Beatty, K. (2003). *Teaching and Researching Computer Assisted Language Learning*, New York: Longman.

Bégel V, Di Loreto I, Seilles A and Dalla Bella S. (2017). Music Games: Potential Application and Considerations for Rhythmic Training. *Front. Hum. Neurosci.* 11:273. doi: 10.3389/fnhum.2017.00273.

Bégel, V., Seilles, A., & Dalla Bella, S. (2018). Rhythm Workers: A music-based serious game for training rhythm skills. To appear in *Music and Science,* 2018. Retrieved from: https://www.researchgate.net/publication/327535584_Rhythm_Workers_A_music-based_serious_game_for_training_rhythm_skills.

Bella, S. D., Benoit, C.-E., Farrugia, N., Schwartze, M., & Kotz, S. A. (2015). Effects of

musically cued gait training in Parkinson's disease: beyond a motor benefit. *Annals of the New York Academy of Sciences, 1337*(1), 77–85. http://doi.org/10.1111/nyas.12651.

Bentley, J.E. (2011). Tuning in: Towards a grounded theory of integrative musical interaction. PhD thesis, University of Strathclyde. Available from: http://digitool.lib.strath.ac.uk/webclient/StreamGate?folder_id=0&dvs=1537990697617~268&usePid1=true&usePid2=true. Accessed on 26.09.18.

Best, C. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (ed), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, York, Timonium, MD, 167–200.

Best CT, McRoberts GW, Goodell E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J Acoust Soc Am. 109*:775–794.

Best, Catherine T. & Michael D. Tyler. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M.J. Munro & O.-S. Bohn (eds), *Second language speech learning: The role of language experience in speech perception and production*. 13-34. Amsterdam: John Benjamins.

Best CT, Hallé PA. (2010). Perception of initial obstruent voicing is influenced by gestural organization. *J Phonet. 38*:109–126.

Bloom, B. S. (1984). The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring. *Educational Researcher 13*. 4-16.

Bohn, O-S & Munro MJ. (2007). Language Experience in Second Language Speech Learning. Amsterdam: John Benjamins.

Bowen, J.D. (1976). Current research on an integrative test of English grammar. *RELC Journal, 7,* 30-37.

Bowen, T., & Marks, J. (1992). The Pronunciation book: Student-centered activities for pronunciation work. NY: Longman.

Bradlow, A.R., Pisoni, D.B., Akahane-Yamada, R., Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *J. Acoust. Soc. Am*. 101, 2299–2310.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & psychophysics* 61, 977–985.

Brookhart, S.M. (1997). Effects of the classroom assessment environment on

mathematics and science achievement. *The Journal of Educational Research, 90*:6, 323-330.

Brown, J. D., & Hilferty, A. G. (1986). The effectiveness of teaching reduced forms for listening comprehension. *RELC Journal*, 17(2), 59-70.

Brown, J. D. & Kondo-Brown, K. (2006). Introducing connected speech. In J. D. Brown & K. Kondo-Brown (eds), *Perspectives on teaching connected speech to second language speakers,* 1-15. Honolulu, HI: University of Hawaii Press.

Buxton H. (1983). Temporal predictability in the perception of English speech. In A Cutler, DR Ladd (eds) *Prosody: models and measurements,* 111 – 121. Berlin, Germany: Springer.

Carreira, J.M. (2008). Effects of teaching reduced forms in a university preparatory course. In K. Bradford Watts, T. Muller, & M. Swanson (eds), *JALT2007 Conference Proceedings*. Tokyo: JALT.

Catford, J.C., Pisoni, D.B. (1970). Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal* 54(7), 477–481.

Celce-Murcia, M., Brinton, D.M. & Goodwin, J.M. (1996). *Teaching pronunciation. A reference for teachers of English to speakers of other languages.* Cambridge, UK: Cambridge University Press.

Chow, I., Belyk, M., Tran, V., & Brown, S. (2015). Syllable synchronization and the P-center in Cantonese. *Journal of Phonetics, 49*, 55–66. http://doi.org/10.1016/j.wocn.2014.10.006.

Chung, Y., & Arvaniti, A. (2013). Speech rhythm in Korean: Experiments in speech cycling, *Proc. Meet. Acoust.* 19, 060216. doi:10.1121/1.4801062.

Clifford, R. (1987). The Status of Computer-Assisted Language Instruction. *CALICO Journal 4*, 4.

Corbett, J. and Stuart-Smith, J. (2010). *Standard English in Scotland, Standards of English: Codified Varieties Around the World.* doi: 10.1017/CBO9781139023832.005.

Corriveau, K. H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, 45, 119–130.

Cruttenden, A. (2001). *Gimson's Pronunciation of English*. London: Edward Arnold.

Crystal, D. (1980). *A first dictionary of linguistics and phonetics*. London: Deutsch.

Crystal, D. (2003). *English as a Global Language*. Cambridge University Press. 2nd ed.

Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26*(2), 145– 171.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Boston: MIT Press.

Dalla Bella, S., Farrugia, N., Benoit, C.E., Begel, V., Verga, L., Harding, E., Kotz, S.A. (2017). BAASTA: battery for the assessment of auditory sensorimotor and timing abilities. *Behav. Res. Methods*, 49:1128–1145, doi:10.3758/s13428-016-0773-6.

Dauer, R. (1983). Stress timing and syllable timing reanalyzed. *Journal of Phonetics. 11*, 51-62.

Delogu F., Lampis, G. & Olivetti Belardinelli, M. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception, European Journal of Cognitive Psychology, 22:1, 46-61, doi: 10.1080/09541440802708136.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for deltaC. In P. Karnowski, & I. Szigeti (eds), *Language and language processing: Proceedings of the 38th linguistic colloquium* (231–241). Piliscsaba 2003. Frankfurt: Peter Lang.

Dellwo, V., & Wagner, P. (2003). Relations between language rhythm and speech rate. In *Proceedings of the 15th international congress of phonetics sciences* (471–474). Barcelona.

Deterding, D. (2001). The measurement of rhythm: A comparison of Singapore and British English. *Journal of Phonetics*, 29, 217–230.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language,* 59, 294-311.

Duanmu, S. (2002). *The phonology of Standard Chinese*. New York: Oxford University Press.

Duanmu, S. (2014). Syllable structure and stress. In James Huang, C.-T., Audrey Li Y.-H., & Simpson A. (eds), *The Handbook of Chinese Linguistics.* John Wiley & Sons, Inc.

Dubois, B. L. (1980). English speech rhythm and the foreign learner by Corrine Adams. *TESOL Quarterly*, Vol. 14, No. 3 (Sep., 1980). 375-378.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, *106*, 682–706.

Ernestus, M. & Warner, N. (2011). Editorial. *J. Phon., 39*, 253-260.

Fitzgibbons PJ, Pollatsek A, Thomas IB. (1974). Detection of Temporal Gaps within and between Perceptual Tonal Groups. *Percept Psychophys 16*(3): 522– 528.

Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics 15*, 47-65.

Flege, J. & Eefting, W. (1986). Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43, 155-171.

Flege, J. & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, 83, 729-740.

Flege, J., Munro, M. & MacKay, I. (1995). The effect of age of second language learning on the production of English consonants. *Speech Communication*, 16, 1-26.

Flege, J. (2005). The origins and development of the Speech Learning Model. *Keynote lecture at the 1st Acoustical Society of America Workshop on L2 Speech Learning, Simon Fraser University, Vancouver, CA* (April 14-15, 2005).

Flege, J. and MacKay, I. (2011). What accounts for "age" effects on overall degree of foreign accent? In M. Wrembel, M. Kul and Dziubalska-Kołaczyk, K. (eds) Achievements and perspectives in the acquisition of second *language speech: New Sounds 2010*, Vol. 2, Bern, Switzerland: Peter Lang. 65-82.

Flege, J. & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26, 1-34.

Fletcher, J. (2010). The Prosody of Speech: Timing and Rhythm. In , W. J. Hardcastle, J. Laver and F. E. Gibbon (eds). *The Handbook of Phonetic Sciences, Second Edition*, Blackwell Publishing Ltd., Oxford, UK.

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: can listeners adapt? *Journal of psycholinguistic research*, *38*(4), 379–412.

François, C., Chobert, J., Besson, M., & Schön, D. (2012). Music training for the development of speech segmentation. *Cerebral Cortex*. doi:http://dx.doi.org/10.1093/cercor/bhs180.

Fujii, S., Schlaug, G., (2013). The Harvard beat assessment test (H-bat): a battery for assessing beat perception and production and their dissociation. *Front. Hum. Neurosci. 7*, 771.

Geiser E, Gabrieli JDE (2013) Influence of Rhythmic Grouping on Duration Perception: A Novel Auditory Illusion. *PLoS ONE 8(1)*: e54273. https://doi.org/10.1371/journal.pone.0054273.

Gick, B. Bernhardt, B.M., Bacsfalvi, P. and Wilson, I. (2008) Ultrasound imaging applications in second language acquisition. In J. Hansen & M. Zampini (eds.) *Phonology and Second Language Acquisition. Amsterdam*: John Benjamins, 309-322.

Gilbert, (2005). Clear Speech Teacher's Resource Book: Pronunciation and Listening Comprehension in American English. Cambridge University Press.

Gordon, E. (1965). The Musical Aptitude Profile: A New and Unique Musical Aptitude Test Battery. *Bulletin of the Council for Research in Music Education,* (6), 12-16.

Gordon, E. (1990). *Predictive validity study of AMMA: A one-year longitudinal predictive validity study of the Advanced Measures of Music Audiation.* Chicago, IL: GIA Publications.

Grabe, E. & Low, E.L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. *Laboratory Phonology 7*. Berlin, Boston: Mouton de Gruyter. 515-546.

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attent. Percept. Psychophys.* 72, 561–582. doi: 10.3758/APP.72.3.561

Hagen, SA. (2000). Sound Advice: A basis for listening. New York: Pearson Education.

Hallé PA, Best CT. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/clusters. *J Acoust Soc Am 121*:2899–2914.

Hallé PA, Segui J, Frauenfelder U, Meunier C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *J Exp Psychol Hum Percept Perform. 24*:592–608.

Hancock, M. (1996). *Pronunciation games.* Cambridge University Press.

Harnsberger JD. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *J Acoust Soc Am. 110:*489–503.

Hawkins S, Smith R H. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian J. Linguistics-Rivista di Linguistica 13:* 99–188. http://kiri.ling.cam.ac.uk/sarah/TIPS/hawkinssmith-101.pdf

Henrichsen, L. (1984). Sandhi-variation: A filter of input for learners of ESL. *Lang. Learn., 34,* 103-126.

Herrera, L., Lorenzo, O., Defior, S., Fernandez-Smith, G., & Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading readiness of native- and foreign-Spanish-speaking children. *Psychology of Music*, *39*(1), 68−81. doi: http://dx.doi. org/10.1177/0305735610361995.

Hoequist, C. E. (1983). The perceptual centre and rhythm categories. *Language and speech, Vol. 26*(4), 367–376.

Hornickel, J., & Kraus, N. (2013). Unstable representation of sound: A biological marker of dyslexia. *The Journal of Neuroscience, 33*(8), 3500-3504.

Hove, M. J., & Schwartze, M. (2014). Deconstructing the Ability to Move to a Beat. *The Journal of Neuroscience*, *34*(7), 2403–2405. doi: http://doi.org/10.1523/JNEUROSCI.4880-13.2014.

Hove, M. J., Spivey, M. J., & Krumhansl, C. L. (2010). Compatibility of motion facilitates visuomotor synchronization. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1525–1534. doi:10.1037/a0019059.

Hubbard, P. (2009). *Computer Assisted Language Learning: Critical Concepts in Linguistics, Volumes I-IV*. London & New York: Routledge.

Ito, Y. (2006a). Comprehension of English reduced forms by second language learners and its effect on input-intake process. In J. D. Brown & K. Kondo-Brown (eds), *Perspectives on teaching connected speech to second language speakers*, 67-81. Honolulu, HI: University of Hawaii Press.

Ito, Y. (2006b). The significance of reduced forms in L2 pedagogy. In J. D. Brown & K. Kondo-Brown (eds), *Perspectives on teaching connected speech to second language speakers*, 17-25. Honolulu, HI: University of Hawaii Press.

Iversen, J. R., and Patel, A. D. (2008). The beat alignment test (BAT): Surveying beat processing abilities in the general population. In M. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, and M. Tsuzaki (eds), *The 10th International Conference on Music Perception, and Cognition (ICMPC 10).* Sapporo.

Iverson, P., Hazan, V., Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *J. Acoust. Soc. Am. 118,* 3267–3278. doi:10.1121/1.2062307.

Iverson, P., and Evans, B. G. (2009). Learning English vowels with different first-language

vowel systems II: Auditory training for native Spanish and German speakers. *J. Acoust. Soc. Am.* 126, 866–877. doi:10.1121/1.3148196

Jenkins, J. (2000). *The phonology of English as an international language: new models, new norms, new goals*. Oxford: Oxford University Press.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5).

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), Jul 1989, 459-491.

Jügler, J and Möbius, B. (2015). Auditory feedback methods to improve the pronunciation of stops by German learners of French. In The Scottish Consortium for ICPhS 2015, *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. Paper number 0184. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0184.pdf.

Kohler, K. J. (2009). *Rhythm in Speech and Language From Theory to Data Special Issue: Phonetica 2009, Vol. 66, No. 1-2*. S Karger Pub.

Ladd, D.R., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment in rising pitch accents in Dutch. *Journal of the Acoustical Society of America, 107*, 2685–2696.

Ladd, D. R., Schepman, A., White, L., Quarmby, L. M., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, *37*(2), 145–161. doi:http://doi.org/https://doi.org/10.1016/j.wocn.2008.11.001.

Large, E. W. (2008). Resonating to musical rhythm: Theory and experiment. In S. Grondin (ed) *The Psychology of Time.* West Yorkshire: Emerald.

Large, E. W., and Jones, M. R. (1999). The dynamics of attending: How we track time varying events. *Psychological Review*, 106 (1), 119-159.

Laver, J. (1994). *Principles of Phonetics*. Cambridge: Cambridge University Press.

Larsen D.P., Butler A.C., Roediger H.L. 3rd. (2008). Test-enhanced learning in medical education. *Medical Education 42*. 955-66.

Law LNC, Zentner M (2012) Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PLoS ONE 7*(12): e52508. doi:10.1371/journal.pone.0052508.

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011b). The social stratification of tongue shape for postvocalic/r/in Scottish English. *Journal of Sociolinguistics*, 256–268.

Lawson, E., Stuart-Smith, J. and Scobbie, J. M. (2018) 'The role of gesture delay in coda /r/ weakening: An articulatory, auditory and acoustic study', Journal of the Acoustical Society of America, 143(3). doi: 10.1121/1.5027833.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics 5*: 253–263.

Lennon, R. (2017). Experience and learning in cross-dialect perception: derhoticised /r/ in Glasgow (Unpublished PhD thesis).

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39,* 3.

Li, A, & Post, B. (2014). L2 Acquisition of Prosodic Properties of Speech Rhythm: Evidence from L1 Mandarin and German Learners of English. *Studies in Second Language Acquisition* 36(2).

Lidji, P., Palmer, C., Peretz, I., and Morningstar, M. (2011). Listeners feel the beat: entrainment to English and French speech rhythms. *Psychon. Bull. Rev.* 18, 1035–1041. doi: 10.3758/s13423-011-0163-0.

Loehr, D. (2007). Aspects of rhythm in gesture and speech. *Gesture* 7: 179-214.

London, J. (2001). Pulse. In Deane Root (ed), *Grove Music Online.* Accessed 18 February, 2018. Retrieved from: http://www.oxfordmusiconline.com.

Manuel, S. (1995). Speakers nasalize /∂/ after /n/, but listeners still hear /∂/. *Journal of Phonetics,* 23. 453-476. 10.1006/jpho.1995.0033.

Marcus, S. M. (1981). Acoustic determinants of perceptual-center (P-center) location. *Perception & Psychophysics*, *30*, 247–256.

Marie, C., Delogu, F., Lampis, G., Belardinelli, M., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*(10), 2701–2715. doi: http://dx.doi.org/10.1162/jocn.2010.21585.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463. doi: http://dx.doi. org/10.1162/jocn.2007.19.9.1453.

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: lexical adaptation to a novel accent. *Cognitive science*, *32*(3), 543–562.

MacLean, M., Bryant. P. E. and Bradley, L. (1987). Rhymes, nursery rhymes, and reading in early childhood. *Merrill-Palmer Quarterly,* 33, 255-281.

Mennen, I. & de Leeuw, E. (2014). Beyond segments. *Studies in Second Language Acquisition, 36*, 183–194. doi: 10.1017/S0272263114000138.

Miles, L. K., Nind, L. K., Henderson, Z., & Macrae, C. N. (2010b). Moving memories: Behavioral synchrony and memory for self and others. *Journal of Experimental Social Psychology, 46*, 457–460.

Morton, J., Marcus, S.M. and Frankish, C.R. 1976: Perceptual centers (P-centers). *Psych. Rev. 83,* 405-8.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research*, *1194*, 81–89. doi: http://dx.doi.org/10.1016/j. brainres.2007.11.042.

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, *20*(1), 56–60. doi: http://dx.doi.org/10.1016/j. lindif.2009.11.003.

Mitterer, H., & Tuinman, A. (2012). The role of native-language knowledge in the perception of casual speech in a second language. *Frontiers in Psychology*, *3*(July), 249. doi:10.3389/fpsyg.2012.00249.

Mok, P. & Dellwo, V. (2008) Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. In *Proceedings of the 4th Speech Prosody (2008)*. 423-426. Campinas, Brazil.

Molloy, K., Moore, D. R., Sohoglu, E., & Amitay, S. (2012). Less is more: Latent learning is maximized by shorter training sessions in auditory perceptual learning. *PLoS ONE, 7*(5). http://doi.org/10.1371/journal.pone.0036929.

Morano, L., Ernestus, M. & ten Bosch, L. (2015). Schwa reduction in low-proficiency L2 speakers: learning and generalization. In The Scottish Consortium for ICPhS 2015, *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0457. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0457.pdf.

Morrill, T., Dilley, L., McAuley, J. D., & Pitt, M. (2014). Distal rhythm influences

whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition*, *131*, 69–74. doi:10.1016/j.cognition.2013.12.006

Morrill, T., McAuley, J. D., Dilley, L. C., Zdziarska, P. A., Jones, K. B., & Sanders, L. D. (2014b). Distal prosody affects learning of novel words in an artificial language. *Psychonomic Bulletin Review*, 1–9. doi:10.3758/s13423-014-0733-z.

Mora, J., Keidel, J. and Flege, J. (2011). Why are the Catalan contrasts between /e/-/eh/ and /o/-/oh/ so difficult for even early Spanish-Catalan bilinguals to perceive? In M. Wrembel, M. Kul and Dziubalska-Kołaczyk, K. (eds), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010, Vol. 2,* Bern, Switzerland: Peter Lang. 183-193.

Mora, J.C., Keidel, J.L. & Flege, J.E. (2015). Effects of Spanish use on the production of Catalan vowels by early Spanish-Catalan bilinguals. In J. Romero & M. Riero (eds), *The Phonetics-Phonology Interface, Representations and methodologies*. Amsterdam: John Benjamins.

Nagle, C. (2018). Examining the temporal structure of the perception-production link in SLA: A longitudinal study. *Language Learning* (Vol. 68). doi:http://doi.org/10.1111/lang.12275

Nakajima Y., ten Hoopen G., van der Wilk RGH. (1991). A new illusion of time perception. *Music Percept 8*: 431–448.

Nakajima, Y., ten Hoopen, G., Hilkhuysen, G., & Sasaki, T. (1992). Time-shrinking: A discontinuity in the perception of auditory temporal patterns. *Perception & Psychophysics*, 51, 504-507.

Nicolson, R. I., & Fawcett, A. J. (2011). Dyslexia, dysgraphia, proce- dural learning and the cerebellum. Cortex, 47, 117–127. doi:10.1016/j.cortex.2009.08.016

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Nolan, F., & Jeon, H. (2014). Speech rhythm: a metaphor? *Phil. Trans. R. Soc. B* 369: 20130396. doi:http://dx.doi.org/10.1098/rstb.2013.0396.

Nouveau, D. (2012). Limites perceptives de l'e caduc chez des apprenants néerlandophones. *Revue Canadienne de linguistiquea Appliquée, 15,* 60–78.

O'Connor, J. D. (1980). Better English Pronunciation. 2nd edition. Cambridge:Cambridge University Press.

Ogden, R. and Hawkins, S. (2015). Entrainment as a basis for co-ordinated actions in speech. *Proceedings of the 18th International Congress of Phonetic Sciences.* Glasgow: University of Glasgow. Paper number 0599.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97-113.

Ordin, M. & Polyanskaya, L. (2014). Development of timing patterns in first and second languages. *System. 42.* 244–257. 10.1016/j.system.2013.12.004.

Ordin, M. & Polyanskaya, L. (2015). Acquisition of English speech rhythm by monolingual children. *Interspeech 16th Annual Conference of the International Speech Communication Association, Dresden, September 6-10.*

Overy, K. (2000). Dyslexia, temporal processing and music: The potential of music as an early learning aid for dyslexic children. *Psychology of Music. 28*(2), 218-229.

Patel, AD. (1999). The acoustics and kinematics of regularly timed speech: a database and method for the study of the P-center problem. *Proc. 14th International Congress of Phonetic Sciences, (August), 1–4.* Retrieved from http://web.haskins.yale.edu/Reprints/HL1137.pdf

Patel, A.D. (2008). *Music*, *language*, and the *brain*. Oxford: Oxford University Press.

Peretz, I., A.S. Champod & K.L. Hyde. (2003). Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Ann. N.Y. Acad. Sci.* 999: 58–75.

Peterson, G. E. & Lehiste, I. (1960) Duration of syllable nuclei in English. *JASA* 32: 693-703.

Pike, K. L. (1945). *The intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29* (2), 191.

Polka L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *J Acoust Soc Am. 89*:2961–2977.

Polka L. (1992). Characterizing the influence of native experience on adult speech perception. *Percept Psychophys. 52*:37–52.

Polka L. (1995) Linguistic influences in adult perception of non-native vowel contrasts. *J Acoust Soc Am.* 97:1286–1296.

Polka L, Bohn OS. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *J Acoust Soc Am.100*:577–592

Polka L, Colantonio C, Sundara M. (2001). A cross-language comparison of/d/-/ð/perception: Evidence for a new developmental pattern. *J Acoust Soc Am. 109*:2190–2201.

Port, Robert. (1999). Speech cycling by Fred Cummins. *The Trustees of Indiana University.* http://www.cs.indiana.edu/rhythmsp/cycling.html. Accessed on 24/03/14.

Posedel, J., Emery, L., Souza, B., & Fountain, C. (2011). Pitch perception, working memory, and second-language phonological production. *Psychology of Music*, *40*(4), 508–517. doi: http://dx.doi.org/ 10.1177/0305735611415145.

Przybylski, L., Bedoin, N., Krifi-Papoz, S., Herbillon, V., Roch, D., Leculier, L., Kotz, S.A., & Tillmann, B. (2013). Rhythmic auditory stimulation influences syntactic processing in children with developmental language disorders. *Neuropsychology, 27* (1), 121–131.

Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*(1), 1–13. doi:10.1159/000087222.

Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*(3), 265–292. doi:10.1016/S0010-0277(99)00058-X.

Rathcke, Tamara V. and Smith, Rachel H. (2015) Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies. *The Journal of the Acoustical Society of America, 137*, 2834-2845. doi:http://dx.doi.org/10.1121/1.4919322.

Remijn, G., van der Meulen, G., ten Hoopen, G., Nakajima, Y., Komori, Y., & Sasaki, T. (1999). On the robustness of time- shrinking. *Journal of the Acoustical Society of Japan (E), 20*, 365–373.

Repp, BH. (2001). Processes underlying adaptation to tempo changes in sensorimotor synchronization. *Human Movement Science*. 20(3) 277-312. doi:10.1016/s0167-9457(01)00049-5.

Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin & Review, 12,* 969–992.

Repp, B. H. (2010). Sensorimotor synchronization and perception of timing: Effects of music training and task experience. *Human Movement Science, 29*, 200–213.

Repp, B. H., & Su, Y.-H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review, 20*(3), 403–452. http://doi.org/10.3758/s13423-012-0371-2

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In Crystal, D. (ed). *Linguistic Controversies: Essays in Honour of F.R. Palmer*. London: Arnold, 73-79.

Roach, P. (2010). *English Phonetics and Phonology* (3rd Edition), Cambridge University Press.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *1*. 382-395.

Saber Khaghaninezhad, M., & Jafarzadeh, G. (2013). Investigating the effect of reduced forms instruction on EFL learners' listening and speaking abilities. *English Language Teaching*, *7*(1), 159–171. doi: http://doi.org/10.5539/elt.v7n1p159

Sadakata, M., & Sekiyama, K. (2011). Enhanced perception of various linguistic features by musicians: A cross-linguistic study. *Acta Psychologica*, *138*(1), 1–10. doi: http://dx.doi.org/10.1016/j.actpsy.2011.03.007.

Saheb, V. (2015). Motivation in English as a Foreign Language Learning : A study of motivation toward English language learning in Stockholm´s upper secondary schools for adults (KOMVUX). (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-27544.

Saloranta, A, Tamminen, H, Alku, P, Peltola, MS. (2015). Learning of a non-native vowel through instructed production training. In The Scottish Consortium for ICPhS 2015, *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4. Paper number 0235. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0235.pdf.

Sasaki, T., Nakajima, Y., & ten Hoopen, G. (1993). The effect of a preceding neighboring tone on the perception of filled durations. *Proceedings of the Acoustical Society of Japan, spring meeting*, 347-348. Tokyo: Acoustical Society of Japan.

Sasaki T., Suetomi D., Nakajima Y., ten Hoopen G. (2002). Time-shrinking, its propagation, and Gestalt principles. *Percept Psychophys 64*: 919–931.

Sasaki T., Nakajima Y., ten Hoopen G, van Buuringen E, Massier B, et al. (2010) Time stretching: Illusory lengthening of filled auditory durations. *Atten Percept Psychophys 72*(5): 1404–1421.

Schaefer RS. (2014). Auditory rhythmic cueing in movement rehabilitation: findings and possible mechanisms. *Philos Trans R Soc B Biol Sci.* doi: 10.1098/rstb.2013.0402.

Schittek, M., Mattheos, N., Lyon H. C., Attström, R. (2001). Computer assisted learning. A review. *European Journal of Dental Education 5.* 93–100.

Schulze, H.-H., Cordes, A., & Vorberg, D. (2005). Keeping synchrony while tempo changes: Accelerando and ritardando. *Music Perception*, 22, 461-477.

Scobbie, J. M., Hewlett, N., & Turk, A. (1999). Standard English in Edinburgh and Glasgow: the Scottish Vowel Length Rule revisited. In Foulkes, P. & Docherty, G. (eds), *Urban voices: Accent studies in the British Isles*, 230-245. London: Hodder.

Scott, S. K. (1998). The point of P-centres. *Psychol Res.* 61: 4-11.

Selkirk, E.O. 1981. *English Compounding and the Theory of Word-structure.* In M. Moortgat, H. Van der Hulst & T. Hoestra (eds), The Scope of Lexical Rules, Foris, Dordrecht.

Setter, J. (2006). Speech rhythm in world Englishes: the case of Hong Kong. *TESOL Quarterly 40*, 763-782.

Shinohara, Yasuaki. (2015). Effects of English /r/-/l/ perceptual training on Japanese children's production. *Proceedings of the 18th International Congress of Phonetic Sciences*. Paper number 0540. Retrieved from: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0540.pdf.

Simpson, Adrian P. (2013). Spontaneous speech. In Mark J. Jones & Rachael-Anne Knight (eds), *The Bloomsbury companion to phonetics*. London: Bloomsbury, 155-170.

Slevc, L. R. & Miyake, A. (2006). Individual differences in second language proficiency: Does musical ability matter? *Psychological Science*, 17(8), 675-681.

Smith, R., Holmes-Elliott, S., Pettinato, M., and Knight, R.-A. (2014). Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization. *Quarterly Journal of Experimental Psychology*, 67(3), 590-608.

Smith, R. and Rathcke, T. (2016). Glasgow *gloom* or Leeds *glue*? Dialect-specific vowel duration constrains lexical segmentation and access. *Phonetica*, 74(1), 1-24. doi:10.1159/000444857.

SoundShapes, Facilitators Kit. (2018). https://remo.com/products/product/sound-shape-facilitators-kit/, accessed on 26/09/2018.

Sowinski, J., and Dalla Bella, S. (2013). Poor synchronization to the beat may result from deficient auditory-motor mapping. *Neuropsychologia* 51, 1952–1963. doi: 10.1016/j.neuropsychologia.2013.06.027.

Strait, D. L., Hornickel, J., & Kraus, N. (2011). Sub-cortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions*, *7*(44), 1–11. doi: http://dx.doi. org/10.1186/1744-9081-7-44.

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In Foulkes, P. and Docherty, G. J. (eds), *Urban voices: Accent Studies in the British Isles.* Arnold, 203–222.

Stuart-Smith, J. (2003). The Phonology of Modern Urban Scots, In Corbett, J., McClure, D. J., and Stuart-Smith, J. (eds), *The Edinburgh Companion to Scots*. Edinburgh University Press, 110–137.

Stuart-Smith, J. (2004) Scottish English: Phonology, a handbook of varieties of English: A multimedia reference tool two volumes plus CD-ROM. doi: 10.1515/9783110197181.

Stuart-Smith, J. (2007). A sociophonetic investigation of postvocalic /r/ in Glaswegian adolescents. *Proceedings of the 16th International Congress of Phonetic Science*, 1449–1452. Retrieved from http://www.icphs2007.de/conference/ Papers/1307/1307.pdf

Stuart-Smith, J., Timmins, C., & Tweedie, F. (2007). "Talkin" Jockney'? Variation and change in Glaswegian accent. *Journal of Sociolinguistics*, *11*(2), 221–260. doi:10.1111/j.1467-9841.2007.00319.x.

Stuart-Smith, J., & Smith, R. (2013). *Technical Report on Encompass Award: Effectiveness of Micro-phonics Pronunciation Coach Software.*

Swaminathan, S., & Gopinath, J. K. (2013). Music training and second-language English comprehension and vocabulary skills in Indian children. *Psychological Studies*, 1–7. http://dx.doi.org/10.1007/ s12646-013-0180-3.

ten Bosch, L., Giezenaar, G., Boves, L., & Ernestus, M. (2016). Modeling language-learners'errors in understanding casual speech. In G. Adda, V. Barbu Mititelu, J. Mariani, D. Tufis, & I. Vasilescu (eds), *Errors by humans and machines in multimedia, multimodal, multilingual data processing, Proceedings of Errare* 2015, 7–121. Bucharest: Editura Academiei Române.

ten Hoopen, G. Hilkhuysen, G, Vis, G. Nakajima, Y, Yamamuchi, F. & Sasaki, T. (1993) A new illusion of time perception: II. *Music Perception, 11*, 15-38.

Tierney, A., & Kraus, N. (2013a). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain Lang. 2013 Mar;124*(3):225-31. doi: 10.1016/j.bandl.2012.12.014. Epub 2013 Feb 9.

Tierney, A., & Kraus, N. (2013b). The ability to move to a beat is linked to the consistency of neural responses to sound. *The Journal of Neuroscience, 33*(38), 14981 LP-14988.

Tilsen, S. & Arvaniti, A. (2013) Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America,* 134, 1: 628-639. doi:10.1121/1.4807565.

Thomson, M. (1993). Teaching the dyslexic child: some evaluation studies. In: *Meeting Points in Dyslexia.*

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29*, 553–584.

Tuinman, A., Mitterer, H., & Cutler, A. (2012). Resolving ambiguity in familiar and unfamiliar casual speech. *Journal of Memory and Language*, *66*(4), 530–544. doi:10.1016/j.jml.2012.02.001.

Turk, A., & Shattuck-Hufnagel, S. (2013). What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology*, *4*(1), 93–118. doi:10.1515/lp-2013-0005.

Underhill, A. (1998). *Sound foundation: Living phonology*. Oxford: Heinemann.

Vaughan-Rees, M. (2010). *Rhymes and Rhythm: A Poem-Based Course for English Pronunciation Study.* Garnet Education.

Walker, R. (2010). *Teaching the pronunciation of English as a Lingua Franca.* Oxford: Oxford University Press.

Wang, H., Mok, P., & Meng, H. (2016). Capitalizing on musical rhythm for prosodic training in computer-aided language learning. C*omputer Speech & Language, 37*, 67–81. http://doi.org/10.1016/J.CSL.2015.10.002.

Wells, J. C. (1982). *Accents of English*. Cambridge: Cambridge University Press.

Wells, J. C. (1982b). *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: first language and second language studies. *Journal of Phonetics*, *35*(4), 501–522. doi:10.1016/j.wocn.2007.02.003.

White, L., Payne, E., and Mattys, S. L. (2009). Rhythmic and prosodic contrast in Venetian and Sicilian Italian. In M. Vigario, S. Frota, and M. J. Freitas (eds), *Phonetics and Phonology: Interactions and Interrelations*. John Benjamins, Amsterdam, 137–158.

Wilson, I. and Gick, B. (2006). Ultrasound technology and Second Language Acquisition research. In Mary Grantham O'Brien, Christine Shea, and John Archibald (eds), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*, Somerville, MA: Cascadilla Proceedings Project, 148-152.

Wong, S. W. L., Mok, P. P. K., Chung, K. K.-H., Leung, V. W. H., Bishop, D. V. M., &

Chow, B. W.-Y. (2017). Perception of native English reduced forms in Chinese learners: its role in listening omprehension and its phonological correlates. *TESOL Quarterly, 51*(1), 7–31. http://doi.org/10.1002/tesq.273.

Zeromskaite, I. (2014). The potential role of music in second language learning: a review article. *Journal of European Psychology Students, 5*(3), 78-88, doi: http://dx.doi.org/10.5334/jeps.ci.

Zhang, J. (2014). Tones, Tonal Phonology, and Tone Sandhi. In James Huang, C.-T., Audrey Li Y.-H., & Simpson A. (eds), *The Handbook of Chinese Linguistics*. John Wiley & Sons, Inc.

Appendices
Appendix A – Rhythm Matters EFL teacher survey

Overview of the Project

Comprehension of speech in a foreign language is often difficult for the learner. Words often sound different in a foreign language class to what they sound like in every-day, real-life situations. This leads to misunderstandings, which in turn can have consequences ranging from lighthearted and funny, to catastrophic.

In native English speech specifically, there are processes by which some syllables become "weak", and some remain "strong", or "stressed". E.g. in the sentence "He was fat for a cat in a box", the three words that will most likely be stressed are fat, cat, and box. The remaining words will be left unstressed, weak. This kind of process occurs when words are linked to one another. I will call this 'connected speech'. Connected speech is the most natural thing to happen in native speech. It is normally fast and/or casual, where the speaker is not carefully articulating every sound.

These processes give the English language its specific rhythm. As in the example sentence above, the rhythm would be:
"- - O - - O - - O"
i.e. two weak syllables (signified by a "-") and one strong syllable (signified by an "O"), alternating.

Although it is natural, connected speech can be problematic for EFL learners' comprehension of every-day English. It can be difficult to know how to help learners improve these skills. I would like to invite you to answer some questions which may help us design a workshop for EFL teachers and/or learners, with the objective of aiding both learning and teaching of this difficult and important matter.

Information for Participants
CONSENT TO THE USE OF DATA

I understand that Ewa Wanat is collecting data in the form of completed questionnaires for use in an academic research project at the University of Glasgow.

Participants will be asked to answer 11 questions concerning their opinions on teaching rhythm in EFL.

I give my consent to the use of data for this purpose on the understanding that:

- All names and other material likely to identify individuals will be anonymised.
- The material will be treated as confidential and kept in secure storage at all times, including after the project is complete.

At any point during the questionnaire you may refrain from answering a question. Should you wish to withdraw at any point during the completion of the questionnaire, exit the window and no data from any previous answers will be stored. Should you wish to withdraw your answers at any point after the completion of the questionnaire, please contact:

Researcher's name and email contact: Ewa Wanat - e.wanat.1@research.gla.ac.uk

Supervisor's name and email contact: Dr Rachel Smith - Rachel.Smith@glasgow.ac.uk

Department address: 12 University Gardens, University of Glasgow, Glasgow, G20 8QQ

By clicking 'next' you are agreeing to the above conditions.

About you

Which of the following types of English Teaching are you engaged in? (tick all that apply)
□ English as a Foreign Language (EFL)
□ English for Speakers of Other Languages (ESOL)
□ English as an Additional Language (EAL)
□ English for Academic Purposes (EAP)
□ Business English
□ English as a Second Language (ESL)
□ English for Special Purposes (ESP)
□ Other:

Which of the following age groups do you teach? (tick all that apply)
□ Pre-school children
□ Primary-age children
□ Secondary-age children
□ Adults
□ Adults with disabilities
Other:

Where is your teaching carried out? (Specify city/region and country, or online)

Rhythm Matters survey

1) Do you consider your students' pronunciation to be an obstacle to their successful communication? If so, please explain how.

2) Do you consider your students' comprehension to be an obstacle to their successful communication? If so, please explain how.

3) Do you cover connected speech in your pronunciation teaching?
□ Yes
□ No

If you answered 'yes' to question 3, can you describe how you do this, what resources you use etc?

If you answered 'no' to question 3, why not? Would you consider it important to implement this in the future?

4) Do you cover English rhythm in your pronunciation teaching?
□ Yes

□ No

If you answered 'yes' to question 4, can you describe how you do this, what resources you use etc?

If you answered 'no' to question 4, why not? Would you consider it important to implement this in the future?

5) Do you know of any resources for teaching rhythm and connected speech that you do not currently use? If so, please provide details of what they are.

6) Have you been trained on how to teach connected speech and rhythm of English?

□ Yes
□ No

If you answered 'yes' to question 6, how much training did you receive? Did you feel the training was beneficial to your understanding and ability to teach connected speech and/or rhythm?

7) How confident do you feel in your understanding of how speech rhythm affects pronunciation?

Not at all confident

□ 1
□ 2
□ 3
□ 4
□ 5

Very confident

8) How could better understanding of connected speech and rhythm help your students?

9) If materials to teach connected speech and rhythm were more readily available, would you use them? Why/Why not?

10) Do you think your students would be interested in using such materials?

□ Yes, in class
□ Yes, for self study e.g. via an app
□ No

Based on your response to question 10, why/why not?

_____

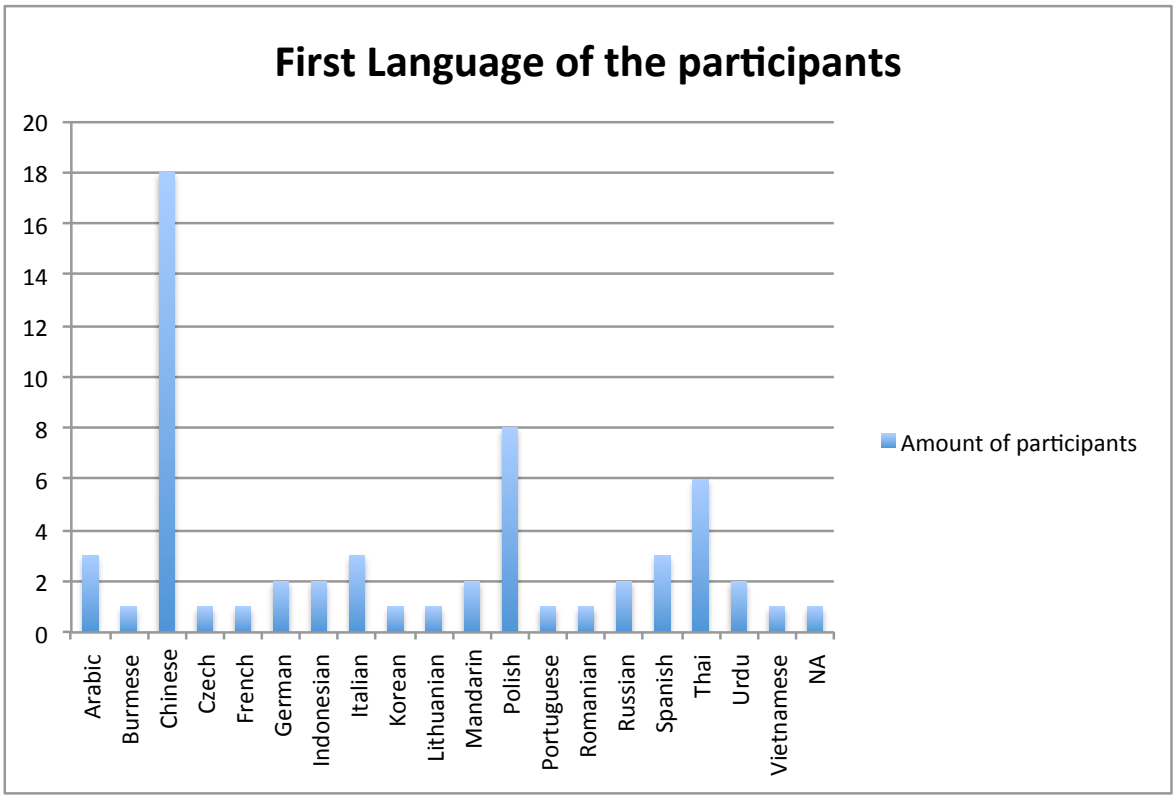
11) Would you attend a workshop/CPD event on teaching and learning connected speech and rhythm? Do you think it would be useful to have a joint event for both students and teachers, or a separate one?

12) Would you be willing to collaborate with researchers on developing materials and teaching methods for rhythm and connected speech?

□ Yes
□ No
□ Don't know

If you answered yes to question 12, please provide a contact e-mail (optional) so that we can get in touch.

Appendix B - First Language of participants in Experiment 1



Appendix C - Experiment 1 materials

Pre-test:

1. It is big even for a chicken in a farm.
2. It was loud for a bird on a roof.
3. She is late for the concert in the theatre.
4. He was nice to the man at the shop.
5. We can smell her sandwich.
6. They can't hear her thing!
7. They should fix her machine.
8. They shouldn't break her chair.
9. He's beaten the poor dog.
10. It's giving them food.
11. It's taking quite a while.
12. He's eaten a meal.

Training high rhythmicity:

1. He is fat for a cat in a box.
2. She was small for a girl in the class.
3. It is cold for a day in the spring.
4. It was made for the man at the church.
5. He is short for a boy in the team.
6. He was kind to the girl at the dance.
7. She is fast for a girl on a bike.
8. She was great on the day of the race.
9. I can't sing her song.
10. She can win her game.
11. She should take her pill.
12. We shouldn't let her fight.
13. She can fight her cold.
14. He can't smell her feet.
15. We should paint her desk.
16. I shouldn't leave her child.
17. He's eating the bread.
18. She's taken her stuff.
19. He's given them chips.

20. She’s beaten the score.
21. He’s taking a nap.
22. She’s giving a talk.
23. She’s beating her time.
24. She’s eaten the sweets.

Training low rhythmicity:

25. He is enormous for a kitten in a box.
26. She can win her contest.
27. He is eating the desert.
28. It is cold for a morning in the summer.
29. I shouldn’t be leaving her child.
30. He’s beaten the result.
31. He can’t smell her very well.
32. He is so little for a player in the team.
33. He’s taking a comfortable nap.
34. He was kind to the young lady at the party.
35. We should decorate her furniture.
36. She is speedy for a girl on a bicycle.
37. She’s giving a presentation.
38. She was amazing on the day of the competition.
39. She’s eaten all the sweeties.
40. I can’t sing her melody.
41. She is very tiny even for a girl in the class.
42. She’s beating her own record.
43. She should take her medication.
44. It was prepared for the people at the church.
45. We shouldn’t allow her to fight.
46. She’s taken her sandwiches.
47. She can get rid of her cold.
48. He’s given them macaroni.

Post test

1. He was decent to the dog in a cage.
2. She is loud for a child on a bus.

3. It was read by a person in a t-shirt.
4. It is stuck to the seat in the car.
5. He can bring her special knife.
6. They can't have her pen!
7. I shouldn't be touching her book.
8. He should steal her bike.
9. She's beating the gentlemen.
10. It's given her joy.
11. She's eating the marmalade.
12. It's taken some time.

Appendix D – Praat Script for creating metronome beats

```
# this is a three-beat metronome
# for use in Exp.1
# Hugo Quené, hquene@indiana.edu, Nov 2001
# modified by Rachel Smith and Ewa Wanat, April 2015

form Three-Beat Metronome
  comment This metronome generates a 3-beat pattern, with
  comment a low tone on the first beat and a high tone on the
  comment second beat. The third beat is unmarked.
  comment The rate (speed) refers to the whole `measure` cycle.
  comment The metronome speeds up gradually, from the initial Rate to the Endrate,
  comment increasing each cycle by the amount specified in the Increment.
  positive Rate_(bpm) 20
  positive Endrate_(bpm) 65
  positive Increment_(bpm) 2
endform

# prepare something special for last cycle
Read from file... /Volumes/PHILIPS UFD/pilot_from_labs_pc/praat_scripts/beep.wav
mywhooshid = selected("Sound")

rate='Rate'
while rate < 'Endrate'
        cycletime = 60/'rate'
        thirdcycletime = 20/'rate'
        tonetime = 0.040
        pausetime = 'thirdcycletime'-'tonetime'

        # for debugging
        printline Cycle time is 'cycletime'
        printline Third-cycle time is 'thirdcycletime'
        printline Pausetime per third-cycle is 'pausetime'
```

```
        Create Sound... Ltone 0 'tonetime' 44100  3/4 * sin(2*pi*400*x)
        Create Sound... Htone 0 'tonetime' 44100 1/4 * sin(2*pi*800*x)
        Create Sound... stil 0 'pausetime' 44100 0
        Create Sound... Notone 0 'tonetime' 44100 0

        select Sound Ltone
        plus Sound stil
        Concatenate
        Rename... part1

        select Sound Htone
        plus Sound stil
        Concatenate
        Rename... part2

        select Sound Notone
        plus Sound stil
        Concatenate
        Rename... part3

        select Sound part1
        plus Sound part2
        plus Sound part3
        Concatenate
        Rename... chain'rate'
        rate = rate+'Increment'
endwhile

rate = 'Rate'
select Sound chain'rate'
#loop to select all the copies of chain
while rate < 'Endrate'
  plus Sound chain'rate'
  rate = rate+'Increment'
endwhile

#add the final sound to the selection
plus 'mywhooshid'

#wait for user input and then concatenate the files and save the output
pause
        Concatenate
        longchain = selected("Sound")
        Save as WAV file... 3beatchain_'Rate'bpm_'Endrate'bpm_'Increment'.wav

select all
Remove
```

Appendix E - Experiment 1 materials by speaker

The speakers for each sentence have been colour coded as the following:

blue - younger male

green - older male

orange - younger female

red - older female

Pre-test:

1. It was loud for a bird on a roof.
2. They shouldn't break her chair.
3. He's eaten a meal.
4. It's giving them food
5. They can't hear her thing!
6. He's beaten the poor dog.
7. It is big even for a chicken in a farm.
8. It's taking quite a while.
9. He was nice to the man at the shop.
10. She is late for the concert in the theatre.
11. They should fix her machine.
12. We can smell her sandwich.

Low-Rhythmicity Training:

1. He is enormous for a cat in a box.
2. She can win her contest.
3. He's eating the dessert.
4. It is cold for a morning in the summer.
5. I shouldn't be leaving her child.
6. She's beaten the result.
7. He can't smell her very well.
8. He is little for a player in the team.
9. He's taking a comfortable nap.
10. He was kind to the young lady at the party.
11. We should decorate her furniture.
12. She is speedy for a girl on a bicycle.

13. She’s giving a presentation.
14. She was amazing on the day of the competition.
15. She’s eaten all the sweeties.
16. I can’t sing her melody.
17. She is very tiny even for a girl in the class.
18. She’s beating her own record.
19. She should take her medication.
20. It was prepared for the people at the church.
21. We shouldn’t allow her to fight.
22. She’s taken her sandwiches.
23. She can get rid of her cold.
24. He’s given them macaroni.

High-Rhythmicity Training:

1. He is fat for a cat in a box.
2. She can win her game.
3. He’s eating the bread.
4. It is cold for a day in the spring.
5. I shouldn’t leave her child.
6. She’s beaten the score.
7. He can’t smell her feet.
8. He is short for a boy in the team.
9. He’s taking a nap.
10. He was kind to the girl at the dance.
11. We should paint her desk.
12. She is fast for a girl on a bike.
13. She’s giving a talk.
14. She was great on the day of the race.
15. She’s eaten the sweets.
16. I can’t sing her song.
17. She was small for a girl in the class.
18. She’s beating her time.
19. She should take her pill.
20. It was made for the man at the church.

21. We shouldn't let her fight.
22. She's taken her stuff.
23. She can fight her cold.
24. He's given them chips.

Post-test:

1. She is loud for a child on a bus.
2. It's given her joy.
3. They can't have her pen!
4. He can bring her special knife.
5. He was decent to the dog in a cage.
6. I shouldn't be touching her book.
7. It was read by a person in a t-shirt.
8. It is stuck to the seat in the car.
9. She's beating the gentlemen.
10. She's eating the marmalade.
11. He should steal her bike.
12. It's taken some time.

Appendix F – Materials for Experiment 2

Test A

1. But it sat on the path of the duck.
2. And she's beaten a man in her chair.
3. And it stayed on the bed for a while.
4. Or she's eating a meal for her pal.
5. So I came for the show at the club.
6. And we're baking the cake of a bride.
7. And we spoke of her friend in the pub.
8. So it stayed in a tree by her house.
9. And she's taken a fruit from a boy.
10. So we sang in the church with her group.
11. And he's beating her boy at the game.
12 So he walks on a street to her car.
13. And she left from her place at the lake.

14 So they went from the line on the ground.

15. But we stayed on a field by the hut.

16. And they came for a walk at the beach.

17. So I played to a beat with her drum.

18. Or he's taking a turn with her ball.

19. And he's eaten her cake in the seat.

Training

1. And he spoke of a girl from her team.

2. So you stayed on the plane for a while.

3. So I came for a hike by the shore.

4. And you spoke of the girl on the beach.

5. So we work for a place in the state.

6. So I sit on a stool with her kid.

7. So they talked of the child with the judge.

8. And we stayed on the road with the tour.

9. So he jumped in the pool from a board.

10. And I'll speak of a guy from her group.

11. So we go for the trip on her coach.

12. And she works on a book with the child.

13. And they spoke from a car with a crowd.

14. And we're writing a list of her points.

15. But she thought of the words of her mum.

16. And I stared at the skill of her art.

17. And we went to a club with her friends.

18. And we went for a pint in her pub.

19. And he's beaten her dog for a start.

20. Or she's beating her boy on the bus.

21. And he's taken her book from the boy.

22. And he's taking her pup in a box.

23. And she's eaten a cake from a store.

24. But she's eating her steak at the bar.

25. And they came from a break at the beach.

26. And she's giving a talk in her course.

27. But she's given her speech on the news.

28. And she’s beaten a score with her team.
29. Or he’s beating her crew in a game.
30. And he’s taken a job with the chap.
31. But she’s taking her notes for a friend.
32. And she’s eaten her roll in the train.
33. And she’s eating her soup at the bar.
34. And he’s given a gift for her band.
35. And she’s giving her task to the staff.
36. But we ran from the bar in her town.
37. But we went to a house by the lake.
38. So she talked to a girl and her dog.

Test B:

1.And we came in her boat for some sun.
2. And he’s beaten a dog in a cage.
3. But we came for a game at a pool.
4. And she stayed in the team for some time.
5. But she’s beating a boy for her town.
6. And we played in the team with the boss.
7. Or he’s taking a bag to her shop.
8. But we look at her cat on the box.
9. And she’s eaten her slice of the bread.
10.And I sat on a train to the beach.
11. But we brought them the gift from a group.
12. Or she’s taking her duck for the pond.
13. So she went in a ship at the shore.
14. But they stayed at the bar of the guy.
15. But she’ll speak with her pal at her desk.
16 Or she’s eating a wrap at her shop.
17. And she came from a test at her school.
18. And she stayed at the house of a lord.
19. So he ran from the back of the room.
20 It was stuck on the head of a girl.
21. And we’re dancing her dance with the troupe.

Appendix G - Questionnaires for Experiment 2

For Chinese learners of English speakers living in Glasgow:

1. What gender are you?
2. How old are you?
3. Do you have any speech, hearing or reading problems? If yes, please specify.
4. What language(s) and dialects did you learn first when you were a child?
5. What language(s) did your mother and/or father or other care-giver speak to you when you were a child?
6. Do you know any other foreign languages (apart from English)? If not, go to question 10.
7. What age were you when you started learning this language (or languages)?
8. Have you ever learned a foreign language in school (apart from English? If yes, what language and how many years of schooling (primary school through university) did you have in this language (or languages)?
9. Do you consider yourself proficient in your second language(s)? If so, what age did you start considering yourself proficient at it?
10. Are you left or right handed?
11. Where were you born and raised (area of the city)?
12. Have you lived in a different country (apart from the place you were born and raised in or the one where you live currently)? If so, where and how long for? Please specify which city & country.
13. At your home, are there/were there any other languages spoken by the members of your family? If so, which languages and which members of your family?
14. How long have you lived in Scotland?
15. Have you ever lived in another English-speaking country? If yes, how long for and where?
16. About how many hours per week do you spend listening to music?
17. Do you have any musical training/background? If yes, please specify what instrument(s), or whether you sing or dance, and how long for. This includes playing an instrument without being able to read music.
18. If you have musical training, how many hours per week do you currently play or practice?
19. Do you currently take or have you ever taken music classes? If yes, how long for, how many hours per week and what instruments (or if singing, please specify).
20. Can you read music?
21. Do you think you can tap along on time with music? Please rate your ability tap on the beat. Type in one of the following: no ability at all / very little / moderate / good / very good
22. Please rate your ability to dance in time with music. Type in one of the following: no ability at all / very little / moderate / good / very good
23. What age were you when you started learning English?

For Canadian/US speakers living in Montreal:

1. Your subject ID (experimenter will type this in)
2. What gender are you?
3. How old are you?
4. Do you have any speech, hearing or reading problems? If yes, please specify.
5. What language did you learn first when you were a child?
6. What language(s) did your mother and/or father or other care-giver speak to you when

you were a child?
7. Do you know any other languages (apart from English)? If not, go to question 11.
8. What age were you when you started learning this language (or languages)?
9. Have you ever learned a foreign language in school? If yes, what language and how many years of schooling (primary school through university) did you have in this language (or languages)?
10. Do you consider yourself proficient in your second language(s)? If so, what age did you start considering yourself proficient at it?
11. Are you left or right handed?
12. Where were you born and raised?
13. Have you lived in a different country (apart from the place you were born and raised in or the one where you live currently)? If so, where and how long for? Please specify which city and state.
14. At your home, are there/were there any other languages spoken by the members of your family? If so, which languages and which members of your family?
15. If you were not born in Canada, how long have you lived here?
16. How long have you lived in the Quebec area?
17. About how many hours per week do you spend listening to music?
18. Do you have any musical training/background? If yes, please specify what instrument, or whether you sing or dance, and how long for. This includes playing an instrument without being able to read music.
19. If you have musical training, how many hours per week do you currently play or practice?
20. Do you currently take or have you ever taken music classes? If yes, how long for, how many hours per week and what instruments (or if singing, please specify).
21. Can you read music?
22. Do you think you can tap along on time with music? Please rate your ability tap on the beat. Type in one of the following: no ability at all / very little / moderate / good / very good
23. Please rate your ability to dance in time with music. Type in one of the following: no ability at all / very little / moderate / good / very good
24. Do you have any Scottish friends you interact with regularly?
25. Do you have any Scottish relatives you interact with regularly?
26. Have you ever lived in or visited Scotland? If so, how long for?
27. Have you ever lived in or visited the UK? If so, how long for?

For Glaswegian participants:

The questionnaire was the same as for the Chinese participants, apart from questions asking about English as a foreign Language. Instead, it asked if participants knew any other foreign languages.

Appendix H – Experiment 2: all groups' re-levelled tables

Re-levelled Group to Chinese:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.24102	0.18608	-1.295	1.95E-01
TestPre	-0.22518	0.04695	-4.797	0.00000161***
Group (CanadianUS)	1.35718	0.09429	14.394	< 2e-16***
Group (Glaswegian)	2.07347	0.10232	20.265	< 2e-16***
SelfTap (Moderate)	-0.14132	0.10077	-1.402	0.16082
SelfTap (None)	-0.88884	0.21069	-4.219	0.0000246***
SelfTap (VeryGood)	0.03357	0.09493	0.354	0.72359
SelfTap (VeryLittle)	-0.40855	0.13595	-3.005	0.00265**

Re-levelled Group to Glaswegian:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.83246	0.18567	9.87	<2e-16***
Test (Pre)	-0.22518	0.04694	-4.797	0.00000161***
Group (Chinese)	-2.07348	0.10232	-20.264	< 2e-16***
Group (CanadianUS)	-0.71629	0.09449	-7.581	3.44e-14***
SelfTap (Moderate)	-0.14131	0.10078	-1.402	0.16083
SelfTap (None)	-0.88884	0.21069	-4.219	0.0000246***
SelfTap (VeryGood)	0.03357	0.09493	0.354	0.7236
SelfTap (VeryLittle)	-0.40855	0.13596	-3.005	0.00266**

Re-levelled SelfTap to “None”

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.94362	0.27092	3.483	0.000496***
Test (Pre)	-0.22518	0.04695	-4.797	0.00000161***
Group (Chinese)	-2.07348	0.10231	-20.266	< 2e-16***
Group (CanadianUS)	-0.71629	0.09449	-7.581	3.44e-14***
SelfTapGood	0.88883	0.21058	4.221	0.0000243***
SelfTap (Moderate)	0.74752	0.21204	3.525	0.000423***
SelfTap (VeryGood)	0.92241	0.2144	4.302	0.0000169***
SelfTap (VeryLittle)	0.48028	0.22545	2.13	0.033145*

Re-levelled SelfTap to “Moderate”

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.69114	0.1946	8.691	< 2e-16***
Test (Pre)	-0.22518	0.04694	-4.797	0.00000161***
Group (Chinese)	-2.07347	0.10232	-20.265	< 2e-16***
Group (CanadianUS)	-0.71629	0.09449	-7.581	3.44e-14***
SelfTap (None)	-0.74753	0.21212	-3.524	0.000425***
SelfTap (Good)	0.14131	0.10077	1.402	0.160809
SelfTap (VeryGood)	0.17489	0.10522	1.662	0.09649.
SelfTap (VeryLittle)	-0.26724	0.1404	-1.903	0.056993.

Re-levelled SelfTap to “Very Good”

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.86603	0.18614	10.025	< 2e-16***
Test (Pre)	-0.22518	0.04695	-4.797	0.00000161***
Group (Chinese)	-2.07347	0.10232	-20.264	< 2e-16***
Group (CanadianUS)	-0.71629	0.09449	-7.58	3.45e-14***
SelfTap (Moderate)	-0.17489	0.10523	-1.662	0.09651.
SelfTap (None)	-0.92242	0.21455	-4.299	0.0000171***
SelfTap (Good)	-0.03357	0.09493	-0.354	0.72358
SelfTap (VeryLittle)	-0.44213	0.14025	-3.152	0.00162**

Re-levelled SelfTap to “Very Little”

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.4239	0.21477	6.63	3.36e-11***
Test (Pre)	-0.22518	0.04694	-4.797	0.00000161***
Group (Chinese	-2.07347	0.10232	-20.265	< 2e-16***
Group (CanadianUS)	-0.71629	0.09449	-7.58	3.44e-14***
SelfTap (VeryGood)	0.44213	0.14022	3.153	0.00162**
SelfTap (Moderate)	0.26724	0.14039	1.904	0.05696.
SelfTap (None)	-0.48029	0.22551	-2.13	0.03319*
SelfTap (Good)	0.40855	0.13593	3.006	0.00265**

Appendix I – Experiment 2: Chinese group re-levelled tables

- Re-levelled condition to non-tapping:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.01503	0.42025	-4.795	0.00000163***
Condition (Tapping)	0.03124	0.12557	0.249	0.8035
Test (Pre)	-0.17874	0.10746	-1.663	0.0963
CamEng	0.09146	0.02277	4.016	0.0000592***
SelfTap (Moderate)	0.07299	0.13522	0.54	0.5893
SelfTap (None)	-0.69178	0.17027	-4.063	0.0000485***
SelfTap (VeryGood)	0.26385	0.15514	1.701	0.089.
SelfTap (VeryLittle)	-0.22974	0.14019	-1.639	0.1012
YearsLivEnglSpCtry	0.13234	0.05851	2.262	0.0237*
Condition (tapping):Test (Pre)	-0.27414	0.15381	-1.782	0.0747.

- Re-levelled Test to Pre:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.19379	0.42095	-5.212	1.87e-07 ***
Condition (tapping)	-0.2429	0.12576	-1.931	0.0534
Test (Post)	0.17874	0.10746	1.663	0.0963 .
CamEng	0.09146	0.02277	4.016	5.92e-05 ***
SelfTap (Moderate)	0.07299	0.13522	0.54	0.5893
SelfTap (None)	-0.69178	0.17027	-4.063	4.85e-05 ***
SelfTap (VeryGood)	0.26385	0.15514	1.701	0.0890 .
SelfTap (VeryLittle)	-0.22974	0.14018	-1.639	0.1012
YearsLivEnglSpCtry	0.13233	0.05851	2.262	0.0237 *
Condition (tapping):Test (Post)	0.27413	0.15381	1.782	0.0747 .

- Re-levelled SelfTap to "Very Good"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.17285	0.4445	-4.888	0.00000102***
Condition (Click ID)	0.2429	0.12576	1.931	0.05343.
Test (Post)	0.45288	0.11011	4.113	0.0000391***
CamEng	0.09146	0.02277	4.016	0.0000592***
SelfTap (Good)	-0.26385	0.15514	-1.701	0.089.
SelfTap (Moderate)	-0.19085	0.15392	-1.24	0.21499
SelfTap (None)	-0.95562	0.1923	-4.97	0.000000671***
SelfTap (VeryLittle)	-0.49358	0.16315	-3.025	0.00248**
YearsLivEnglSpCtry	0.13233	0.05851	2.262	0.02371*
Condition (Click ID) :Test (Post)	-0.27414	0.15381	-1.782	0.07471.

-Re-levelled SelfTap to "Very Little"

	Estimate	Std. Error	z value P	Pr(>|z|)
(Intercept)	-2.66646	0.41739	-6.388	1.68e-10***
Condition (Click ID)	0.2429	0.12576	1.931	0.05344.
Test (Post)	0.45288	0.11011	4.113	0.0000391***
CamEng	0.09146	0.02278	4.016	0.0000593***
SelfTap (VeryGood)	0.49359	0.16316	3.025	0.00248**
SelfTap (Good)	0.22974	0.14019	1.639	0.10126
SelfTap (Moderate)	0.30273	0.15209	1.99	0.04655*
SelfTap (None)	-0.46204	0.17754	-2.602	0.00926**
YearsLivEnglSpCtry	0.13234	0.05851	2.262	0.02371*
Condition (Click ID):Test (Post)	-0.27413	0.15382	-1.782	0.07472.

- Re-levelled SelfTap to "None"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-3.12848	0.41933	-7.461	8.61e-14***
Condition (Click ID)	0.2429	0.12576	1.931	0.05343,
Test (Post)	0.45288	0.11011	4.113	0.0000391***
CamEng	0.09146	0.02277	4.016	0.0000592***
SelfTap (VeryLittle)	0.46204	0.17754	2.602	0.00926**
SelfTap (VeryGood)	0.95563	0.1923	4.97	0.000000671***
SelfTap (Good)	0.69178	0.17027	4.063	0.0000485***
SelfTap (Moderate)	0.76477	0.18391	4.158	0.0000321***
YearsLivEnglSpCtry	0.13234	0.05851	2.262	0.02371*
Condition (Click ID):Test (Post)	-0.27414	0.15381	-1.782	0.0747.

- Re-levelled SelfTap to "Moderate"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.36371	0.4474	-5.283	0.000000127***
Condition (Click ID)	0.24289	0.12576	1.931	0.0534.
Test (Post)	0.45287	0.11011	4.113	0.0000391***
CamEng	0.09146	0.02277	4.016	0.0000592***
SelfTap (None)	-0.76477	0.18391	-4.158	0.0000321***
SelfTap (VeryLittle)	-0.30273	0.15209	-1.99	0.0465*
SelfTap (VeryGood)	0.19086	0.15392	1.24	0.215
SelfTap (Good)	-0.07299	0.13522	-0.54	0.5893
YearsLivEnglSpCtry	0.13234	0.05851	2.262	0.0237*
Condition (Click ID):Test (Post)	-0.27413	0.15381	-1.782	0.0747.

Appendix J - Experiment 2: Canadian/US group re-levelled tables

- Re-levelled SelfTap to "Moderate"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.67702	0.29857	2.268	0.023359*
Test (Pre)	-0.27213	0.15836	-1.718	0.085729.
SelfTap (Good)	0.34348	0.2011	1.708	0.087636.
SelfTap (VeryGood)	0.6289	0.20555	3.06	0.002216**
SelfTap (VeryLittle)	-0.55371	0.36877	-1.501	0.133229
Reduction Type (Preposition)	1.59007	0.44489	3.574	0.000351***
Reduction Type (Participle)	-0.38531	0.65599	-0.587	0.556956
Test (Pre):SelfTap (Good)	0.13142	0.21805	0.603	0.54669
Test (Pre):SelfTap (VeryGood)	-0.09802	0.21916	-0.447	0.654703
Test (Pre):SelfTap VeryLittle)	1.11099	0.40115	2.769	0.005615**

- Re-levelled SelfTap to "Very Good"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.30592	0.2955	4.419	0.0000099***
Test (Pre)	-0.37015	0.15288	-2.421	0.015472*
SelfTap (Moderate)	-0.6289	0.20555	-3.06	0.002216**
SelfTap (Good)	-0.28542	0.19562	-1.459	0.144545
SelfTap (VeryLittle)	-1.1826	0.3667	-3.225	0.00126**
Reduction Type (Preposition)	1.59007	0.4449	3.574	0.000352***
Reduction Type (Participle)	-0.38531	0.65567	-0.588	0.556762
Test (Pre):SelfTap(Moderate)	0.09802	0.21916	0.447	0.654688
Test (Pre):SelfTap (Good)	0.22944	0.21207	1.082	0.279284
Test (Pre):SelfTap (VeryLittle)	1.209	0.39888	3.031	0.002437**

- Re-levelled SelfTap to "Very Little"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.1233	0.4249	0.29	0.771642
Test (Pre)	0.8389	0.3677	2.281	0.02254*
SelfTap (VeryGood)	1.1826	0.3665	3.227	0.001251**
SelfTap (Moderate)	0.5537	0.3685	1.503	0.13293
SelfTap (Good)	0.8972	0.3632	2.47	0.013513*
Reduction Type (Preposition)	1.5901	0.4448	3.574	0.000351**
Reduction Type (Participle)	-0.3853	0.6557	-0.588	0.556789
Test (Pre) :SelfTap (VeryGood)	-1.209	0.3986	-3.033	0.002421**
Test (Pre):SelfTap (Moderate)	-1.111	0.4008	-2.772	0.005578**
Test (Pre):SelfTap (Good)	-0.9796	0.3957	-2.476	0.013295*

- Re-levelled Test to "Pre" and SelfTap to "Very Good"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.93577	0.2936	3.187	0.001437**
Test (Post)	0.37015	0.1529	2.421	0.015483*
SelfTap (VeryLittle)	0.0264	0.37888	0.07	0.944456
SelfTap (Moderate)	-0.53088	0.20128	-2.637	0.008352**
SelfTap (Good)	-0.05598	0.19201	-0.292	0.770636
Reduction Type (Preposition)	1.59008	0.44495	3.574	0.000352***
Reduction Type (Participle)	-0.38528	0.65599	-0.587	0.556988
Test (Post):SelfTap (VeryLittle)	-1.209	0.39894	-3.031	0.002441**
Test (Post):SelfTap (Moderate)	-0.09802	0.21918	-0.447	0.65472
Test (Post) :SelfTap(Good)	-0.22944	0.21209	-1.082	0.279325

- Re-levelled SelfTap to "Good"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.87979	0.29124	3.021	0.002521**
Test (Post)	0.1407	0.146	0.964	0.3352
SelfTap (VeryGood)	0.05598	0.19199	0.292	0.770632
SelfTap (VeryLittle)	0.08238	0.37697	0.219	0.827013
SelfTap (Moderate)	-0.47491	0.19877	-2.389	0.016885*
Reduction Type (Preposition)	1.59008	0.44486	3.574	0.000351***
Reduction Type (Participle)	-0.38527	0.65588	-0.587	0.556932
Test (Post):SelfTap (VeryGood)	0.22944	0.21207	1.082	0.279274
Test (Post):SelfTap (VeryLittle)	-0.97956	0.39594	-2.474	0.013361*
Test (Post):SelfTap(Moderate)	0.13143	0.21803	0.603	0.546651

- Re-levelled SelfTap to "Moderate"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.40489	0.29737	1.362	0.173326
Test (Post)	0.27213	0.1584	1.718	0.085791.
SelfTap (Good)	0.4749	0.19881	2.389	0.016909*
SelfTap (VeryGood)	0.53088	0.2013	2.637	0.008358**
SelfTap (VeryLittle)	0.55729	0.38229	1.458	0.144907
Reduction Type (Preposition)	1.59007	0.44505	3.573	0.000353***
Reduction Type (Participle)	-0.38529	0.65622	-0.587	0.55711
Test (Post):SelfTap (Good)	-0.13142	0.21809	-0.603	0.546763
Test (Post):SelfTap (VeryGood)	0.09802	0.2192	0.447	0.654762
Test (Post):SelfTap (VeryLittle)	-1.11099	0.40133	-2.768	0.005636**

- Re-levelled Reduction Type to "Verb"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.01956	0.61782	0.032	0.97474
Test (Post)	0.27213	0.15838	1.718	0.08576.
SelfTap (Good)	0.4749	0.1988	2.389	0.0169*
SelfTap (VeryGood)	0.53088	0.20129	2.637	0.00835**
SelfTap (VeryLittle)	0.55728	0.38218	1.458	0.1448
Reduction Type (Determiner)	0.38533	0.65618	0.587	0.55705
Reduction Type (Preposition)	1.97541	0.70077	2.819	0.00482**
Test (Post):SelfTap (Good)	-0.13142	0.21807	-0.603	0.54673
Test (Post):SelfTap (VeryGood)	0.09802	0.21919	0.447	0.65475
Test (Post):SelfTap (VeryLittle)	-1.11099	0.4012	-2.769	0.00562**

- Re-levelled Reduction Type to "Preposition"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.99497	0.38521	5.179	0.000000223***
Test (Post)	0.27213	0.15836	1.718	0.085728.
SelfTap (Good)	0.4749	0.19878	2.389	0.016892*
SelfTap (VeryGood)	0.53088	0.20127	2.638	0.008347**
SelfTap (VeryLittle)	0.55728	0.38214	1.458	0.144751
Reduction Type (Participle)	-1.97536	0.70044	-2.82	0.0048**
Reduction Type (Determiner)	-1.59008	0.44486	-3.574	0.000351***
Test (Post):SelfTap (Good)	-0.13142	0.21805	-0.603	0.546687
Test (Post):SelfTap (VeryGood)	0.09802	0.21917	0.447	0.654708
Test (Post):SelfTap (VeryLittle)	-1.11099	0.40116	-2.769	0.005616**

Appendix K – Experiment 2: Glaswegian Group re-levelled tables

- Re-levelling SelfTap to "Moderate"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.99582	0.29257	6.822	8.99e-12 ***
Test (Pre)	-0.27571	0.09567	-2.882	0.00395 **
SelfTap (Good)	0.18823	0.22584	0.833	0.40458
SelfTap (VeryGood)	0.0337	0.22523	0.15	0.88105
SelfTap (VeryLittle)	-0.77564	0.35258	-2.2	0.02782 *

- Re-levelling SelfTap to "Very Good"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	2.02953	0.24882	8.157	3.44e-16 ***
Test (Pre)	-0.27571	0.09567	-2.882	0.00395 **
SelfTap (Moderate)	-0.0337	0.22521	-0.15	0.88105
SelfTap (Good)	0.15454	0.16452	0.939	0.34757
SelfTap (VeryLittle)	-0.80932	0.31687	-2.554	0.01065 *

- Re-levelling SelfTap to "Very Little"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	1.22018	0.3661	3.333	0.000859
Test (Pre)	-0.2757	0.09567	-2.882	0.003954
SelfTap (VeryGood)	0.80935	0.31678	2.555	0.010622
SelfTap (Moderate)	0.77563	0.35246	2.201	0.027765
SelfTap (Good)	0.96388	0.31738	3.037	0.002389

Appendix L – Experiment 2: Individual scores and change from pre- to post-test

Individual scores and change from pre- to post-test

Chinese:

Participant	Pre-test	Post-test	Change	Mean across pre- and post-test
ct01ab	31.57894737	40.35087719	8.771929825	35.96491228
ct02ab	43.85964912	50.87719298	7.01754386	47.36842105
ct03ab	42.10526316	45.61403509	3.50877193	43.85964913
ct04ab	29.8245614	31.57894737	1.754385965	30.70175439
ct05ab	36.84210526	42.10526316	5.263157895	39.47368421
ct06ab	35.0877193	45.61403509	10.52631579	40.3508772
ct07ab	45.61403509	49.12280702	3.50877193	47.36842106
ct08ab	33.33333333	43.85964912	10.52631579	38.59649123
ct09ab	29.8245614	50.87719298	21.05263158	40.35087719
ct10ab	31.57894737	40.35087719	8.771929825	35.96491228
ct11ba	49.12280702	43.85964912	-5.263157895	46.49122807
ct12ba	24.56140351	28.07017544	3.50877193	26.31578948
ct13ba	47.36842105	47.36842105	0	47.36842105
ct14ba	43.85964912	49.12280702	5.263157895	46.49122807
ct15ba	45.61403509	52.63157895	7.01754386	49.12280702
ct16ba	24.56140351	36.84210526	12.28070175	30.70175439
ct17ba	50.87719298	50.87719298	0	50.87719298
ct18ba	31.57894737	45.61403509	14.03508772	38.59649123
ct19ba	52.63157895	52.63157895	0	52.63157895
ct20ba	40.35087719	68.42105263	28.07017544	54.38596491
cn01ab	45.61403509	50.87719298	5.263157895	48.24561404
cn02ab	40.35087719	42.10526316	1.754385965	41.22807018
cn03ab	43.85964912	45.61403509	1.754385965	44.73684211
cn04ab	40.35087719	49.12280702	8.771929825	44.73684211
cn05ab	38.59649123	38.59649123	0	38.59649123
cn06ab	50.87719298	43.85964912	-7.01754386	47.36842105
cn07ab	43.85964912	54.38596491	10.52631579	49.12280702
cn08ab	49.12280702	63.15789474	14.03508772	56.14035088
cn09ab	38.59649123	42.10526316	3.50877193	40.3508772
cn10ab	29.8245614	36.84210526	7.01754386	33.33333333
cn11ba	40.35087719	36.84210526	-3.50877193	38.59649123
cn12ba	26.31578947	31.57894737	5.263157895	28.94736842
cn13ba	28.07017544	38.59649123	10.52631579	33.33333334
cn14ba	33.33333333	22.80701754	-10.52631579	28.07017544
cn15ba	47.36842105	43.85964912	-3.50877193	45.61403509
cn16ba	54.38596491	49.12280702	-5.263157895	51.75438597
cn17ba	54.38596491	45.61403509	-8.771929825	50
cn18ba	56.14035088	52.63157895	-3.50877193	54.38596492
cn19ba	33.33333333	40.35087719	7.01754386	36.84210526

cn20ba	35.0877193	54.38596491	19.29824561	44.73684211

Canadian/US:

Participant	Pre-test	Post-test	Change	Mean across pre- and post-test
t01ab	73.68421053	70.1754386	-3.50877193	71.92982457
t02ab	57.89473684	63.15789474	5.263157895	60.52631579
t03ab	66.66666667	71.92982456	5.263157895	69.29824562
t04ab	64.9122807	61.40350877	-3.50877193	63.15789474
t05ab	77.19298246	71.92982456	-5.263157895	74.56140351
t06ab	66.66666667	66.66666667	0	66.66666667
t07ab	75.43859649	70.1754386	-5.263157895	72.80701755
t08ab	68.42105263	64.9122807	-3.50877193	66.66666667
t09ab	57.89473684	66.66666667	8.771929825	62.28070176
t18ab	71.92982456	71.92982456	0	71.92982456
t11ba	63.15789474	71.92982456	8.771929825	67.54385965
t12ba	64.9122807	68.42105263	3.50877193	66.66666667
t13ba	57.89473684	61.40350877	3.50877193	59.64912281
t14ba	75.43859649	75.43859649	0	75.43859649
t15ba	73.68421053	85.96491228	12.28070175	79.82456141
t16ba	75.43859649	73.68421053	-1.754385965	74.56140351
t17ba	56.14035088	70.1754386	14.03508772	63.15789474
t18ba	68.42105263	56.14035088	-12.28070175	62.28070176
t19ba	71.92982456	78.94736842	7.01754386	75.43859649
t21ba	66.66666667	73.68421053	7.01754386	70.1754386
nt02ab	75.43859649	80.70175439	5.263157895	78.07017544
nt03ab	66.66666667	70.1754386	3.50877193	68.42105264
nt04ab	64.9122807	66.66666667	1.754385965	65.78947369
nt05ab	71.92982456	63.15789474	-8.771929825	67.54385965
nt06ab	64.9122807	63.15789474	-1.754385965	64.03508772
nt07ab	52.63157895	56.14035088	3.50877193	54.38596492
nt09ab	71.92982456	70.1754386	-1.754385965	71.05263158
nt10ab	77.19298246	77.19298246	0	77.19298246
nt21ab	68.42105263	61.40350877	-7.01754386	64.9122807
nt11ba	75.43859649	73.68421053	-1.754385965	74.56140351
nt12ba	73.68421053	75.43859649	1.754385965	74.56140351
nt13ba	54.38596491	63.15789474	8.771929825	58.77192983
nt14ba	66.66666667	75.43859649	8.771929825	71.05263158
nt15ba	61.40350877	70.1754386	8.771929825	65.78947369
nt16ba	75.43859649	73.68421053	-1.754385965	74.56140351
nt17ba	66.66666667	77.19298246	10.52631579	71.92982457
nt18ba	66.66666667	80.70175439	14.03508772	73.68421053
nt19ba	71.92982456	68.42105263	-3.50877193	70.1754386
nt20ba	66.66666667	75.43859649	8.771929825	71.05263158
nt08ab	64.9122807	68.42105263	3.50877193	66.66666667

Glaswegian:

Participant	Pre-test	Post-test	Change	Mean across pre- and post-test
gt01ab	87.71929825	87.71929825	0	87.71929825
gt02ab	85.96491228	70.1754386	-15.78947368	78.07017544
gt03ab	75.43859649	71.92982456	-3.50877193	73.68421053
gt04ab	70.1754386	66.66666667	-3.50877193	68.42105263
gt05ab	80.70175439	78.94736842	-1.754385965	79.8245614
gt06ab	85.96491228	85.96491228	0	85.96491228
gt07ab	80.70175439	78.94736842	-1.754385965	79.8245614
gt08ab	77.19298246	84.21052632	7.01754386	80.70175439
gt09ab	85.96491228	78.94736842	-7.01754386	82.45614035
gt10ab	80.70175439	77.19298246	-3.50877193	78.94736842
gt11ba	63.15789474	75.43859649	12.28070175	69.29824561
gt12ba	73.68421053	80.70175439	7.01754386	77.19298246
gt13ba	71.92982456	87.71929825	15.78947368	79.8245614
gt14ba	75.43859649	80.70175439	5.263157895	78.07017544
gt15ba	64.9122807	66.66666667	1.754385965	65.78947368
gt16ba	66.66666667	75.43859649	8.771929825	71.05263158
gt17ba	82.45614035	82.45614035	0	82.45614035
gt18ba	64.9122807	84.21052632	19.29824561	74.56140351
gt19ba	70.1754386	78.94736842	8.771929825	74.56140351
gt20ba	64.9122807	73.68421053	8.771929825	69.29824561
gn01ab	80.70175439	71.92982456	-8.771929825	76.31578947
gn02ab	73.68421053	77.19298246	3.50877193	75.43859649
gn03ab	85.96491228	78.94736842	-7.01754386	82.45614035
gn04ab	77.19298246	78.94736842	1.754385965	78.07017544
gn05ab	75.43859649	63.15789474	-12.28070175	69.29824561
gn06ab	84.21052632	68.42105263	-15.78947368	76.31578947
gn07ab	70.1754386	68.42105263	-1.754385965	69.29824561
gn08ab	80.70175439	80.70175439	0	80.70175439
gn09ab	68.42105263	71.92982456	3.50877193	70.1754386
gn10ab	75.43859649	68.42105263	-7.01754386	71.92982456
gn11ba	66.66666667	84.21052632	17.54385965	75.43859649
gn12ba	66.66666667	77.19298246	10.52631579	71.92982456
gn13ba	70.1754386	84.21052632	14.03508772	77.19298246
gn14ba	82.45614035	84.21052632	1.754385965	83.33333333
gn15ba	71.92982456	82.45614035	10.52631579	77.19298246
gn16ba	66.66666667	82.45614035	15.78947368	74.56140351
gn17ba	70.1754386	78.94736842	8.771929825	74.56140351
gn18ba	73.68421053	85.96491228	12.28070175	79.8245614
gn19ba	70.1754386	75.43859649	5.263157895	72.80701754
gn20ba	54.38596491	73.68421053	19.29824561	64.03508772

Appendix M – Experiment 3: Materials for each training session

Session 1:

1. Hot and cold.
2. She is nice.
3. It's a lot of fun.
4. I'm hungry.
5. I'm a student.
6. It's two minutes to two.
7. My phone's very old.
8. He's just as nice.
9. Do you speak English?
10. We've seen this movie.
12. The computer is broken.
15. Anna and Molly went out for a pizza.
13. This was quite confusing.
14. I left my umbrella at the pub.
11. Can you tell me what to do?
16. They exercise every day apart from Sundays.

Session 2:

1. What's new?
2. From two till four.
3. Can you tell me what to do?
4. The computer is broken.
5. This was quite confusing.
6. They sat and talked.
7. I left my umbrella at the pub.
8. Anna and Molly went out for a pizza.
9. They exercise every day apart from Sundays.
10. She ate three eggs for breakfast.
11. Tell him we're going out now.
12. We should have seen that coming.
13. Have a good day!

14. We’re a bit tired.

15. The first three people will be rewarded.

16. The best student will get a prize.

17. This sounded very natural.

18. You should try to be quiet at the library.

19. Fructose is a naturally occurring sugar.

Session 3

Task 1.

Which sentence rhythm is the same, A, B or C?

1. Tell him.

A. He stopped.

B. Help her.

C. Talk to them.

2. I don’t want to eat.

A. She can’t have a look.

B. This woman stays at a hotel.

C. The cat went outside.

3. Just a moment.

A. We can’t go outside.

B. Wait a second.

C. Let’s have a party!

4. Let’s go and see this house.

A. We came to feed the ducks.

B. We baked a cake for her party.

C. The cat caught the mouse.

5. Would you like a cup of tea?

A. Are we going back?

B. Can the ducks swim in the sea?

C. Could we go to see this movie?

Task 2.

1. Which sentence does this rhythm sound like?

A. I said I wanted to go there!

B. I said I wanted to go there!

C. I said I wanted to go there!

D. I said I wanted to go there!

E. I said I wanted to go there!

2. Which sentence does this rhythm sound like?

A. I'm not sure what they're planning.

B. I'm not sure what they're planning.

C. I'm not sure what they're planning.

D. I'm not sure what they're planning.

E. I'm not sure what they're planning.

3. Which sentence does this rhythm sound like?

A. She could play this song.

B. She could play this song.

C. She could play this song.

D. She could play this song.

E. She could play this song.

Session 4

Task 1: Which sentence matches this rhythm?

Rhythm 1:

A: It's for you.

B: This took a long time.

C: This essay is great!

Rhythm 2:
A: We're going away.
B: They are on holiday.
C: What time will they get here?

Rhythm 3:
A: He needs more financial support.
B: I was going to tell you about this.
C: The restaurant was closed yesterday.

Task 2: Which rhythm matches this sentence?
1. This cat is too big for this box!

Rhythm A
Rhythm B
Rhythm C

2. I want to go on a holiday to a warm and sunny place.

Rhythm A
Rhythm B
Rhythm C

3. He hates flying but he needs to do it for his job.

Rhythm A
Rhythm B
Rhythm C

Task 3: In pairs, one person is A, the other person is B. Listen and then drum these dialogues to each other.

<u>Dialogue 1</u>

A: Do you want to go?

B: Yeah, I'd love to.

A: When are you free?

B: Probably next week.

A: How about Thursday?

Dialogue 2

A: Do you want to go to the cinema?

B: Yes, I'd love to go this week!

A: When are you free this week?

B: I think probably only on Sunday.

A: Sunday would be fine. What time do you want to go?

Dialogue 3

A: I enjoy watching my cat squeeze into boxes very much!

B: That sounds really funny! Do you have any pictures of him?

A: Yes, it is a female cat though, so it's a "she". Here, look at her!

B: This is great! Why do all cats love boxes so much?

Appendix N – Experiment 3: Links to training materials online

Training materials for Experimental Group:

- Session 1: https://ewanat1.wixsite.com/experiment3
- Session 2: https://ewanat1.wixsite.com/experiment3/session-2
- Session 3: https://ewanat1.wixsite.com/experiment3/session-3
- Session 4: https://ewanat1.wixsite.com/experiment3/session-4

Training materials for Control Group:

- Class 1: https://ewanat1.wixsite.com/experiment3/clare-s-class
- Class 2: https://ewanat1.wixsite.com/experiment3/chiara-s-class

Appendix O – General background questionnaire, including language and musical background

1. What gender are you?
2. How old are you?
3. Do you have any speech, hearing or reading problems? If yes, please tell us what they are.
4. What language(s) and dialects did you learn first when you were a child? Please list ALL languages AND dialects you speak. E.g. if you speak Chinese, tell us whether it is Mandarin, Cantonese, Hakka, Foshan, Shangainese, etc etc.
5. What language(s) did your mother and/or father or other care-giver speak to you when you were a child?
6. At your home, are there/were there any other languages spoken by the members of your family? If so, which languages and which members of your family?
7. What is your main language now? This means the language you speak most now.
8. What age were you when you started learning English?
9. Do you know any other foreign languages (apart from English)? If yes, what are these languages? If not, go to question 13.
10. What age were you when you started learning the other languages?
11. Have you ever learned a foreign language in school (apart from English? If yes, what language and how many years of schooling (primary school through university) did you have in this language (or languages)?
12. Do you consider yourself proficient in your second language(s)? If so, what age did you start considering yourself proficient at it?
13. Where were you born and raised (area of the city)?
14. Have you lived in a different country (apart from the place you were born and raised in and apart from Scotland)? If so, where and how long for? Please specify which city & country.
15. How long have you lived in Scotland?
16. Have you ever lived in another English-speaking country, or anywhere else in the UK? If yes, how long for and where?
17. Do you regularly speak to or interact with any Scottish friends at the moment? If yes, when did you meet them (i.e. how long ago). If not, go to question 18.

If yes, how often do you talk to them? Type in one of the following: Rarely, Sometimes, Often, Very often

18. How often do you listen to music? Type in one of the following: Never, Rarely, Sometimes, Often, Very often

If yes, please tell us what genres (e.g. Blues, Jazz, Electronic, Rock, Pop ,Indie etc) :

19. About how many HOURS PER WEEK do you spend listening to music?

20. How often do you dance? Type in one of the following: Never, Rarely, Sometimes, Often, Very often

21. How would you rate your overall sense of rhythm compared to other people? Type in one of the following: Poor, Below average, Average, Good, Excellent

22. Do you participate in sports or other physical activities? Type in one of the following: Never, Rarely, Sometimes, Often, Very often

22a. If yes, what sports or activities:

23. In general, how would you rate your physical coordination? Type in one of the following: Clumsy, Below average, Average, Good, Excellent

24. Do you have any musical training? If not go to question 29.

25. Please list what instruments, including singing, you have studied, and for how long.

26. Are you still playing an instrument?

Yes -> Which instrument; how many hours per week do you practice?

No -> How long ago did you stop?

If you no longer practice, how many hours per week did you practice?

27. Do you currently take or have you ever taken music classes? If yes, how long for, how many hours per week and what instruments (or if singing, please specify).

28. Can you read music, i.e. musical notes?

29. Do you think you can tap along on time with music? Please rate your ability tap (i.e. tap your finger) on the beat. Type in one of the following: no ability at all / very little / moderate / good / very good

30. Please rate your ability to dance in time with music. Type in one of the following: no ability at all / very little / moderate / good / very good

Appendix P - Motivation Questionnaire

In this section, you will be asked for YOUR OPINION on some things to do with English and learning English. Your answers can only be a number from 1 to 5. You will read some sentences, e.g. like this one:

Eating vegetables is very good for me.
Your answer can be one of these:
1 - this means you "Strongly disagree"
2 - this means you "Slightly disagree"
3 - this means you "Neither agree nor disagree", i.e. you have a neutral opinion on this
4 - this means you "Slightly agree"
5 - this means you "Strongly agree"

Many people know and believe that eating vegetables is good for them, so their answer to this question would be 5. However, if you for some reason disagree with this and think that vegetables are bad for you, you can pick a different answer, e.g. 1 or 2, and that is OK if that is what you really think. It is important for us to know what your own opinion is.

Try a few sentences first:

Smoking is bad for my health.
Your answer:

Doing sports makes me stronger.
Your answer:

2. Knowing English will be helpful for my career.
3. I need to know English to study at university.
4. English will broaden my future options.
5. Knowing English will make it possible for me to talk to others about my own ideas and thoughts.
5. I want to have native English friends.
6. Knowing English will help me achieve.
7. English will can lead to my success in life.
8. Knowing English will make me an educated person.
9. English helps me/will help me learn about other cultures, e.g. books, music, movies.
10. Being proficient in English will make other people respect me.
11. Knowing English will make me a knowledgable and skilful person.
12. Knowing English helps me to understand the ways of life of native English speakers.
13. I want to study English so I can behave more like English native speakers.
14. Knowing English makes me a more open-minded and social person.
15. I am determined to study English as best I can to achieve maximum proficiency.
16. My parents want me to learn English.
17. I actively seek help to do my English homework or get things explained to me outside the class.
18. Outside the class, I watch films, listen to songs or read in English.
19. I like studying English.
20. Even if I was not going to Glasgow University, I would still like to learn English.

Appendix Q - Experiment 3: re-levelled tables

Re-levelling Reduction Type to "Preposition"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-0.9256	0.3509	-2.638	0.008346**
conditione	-0.7144	0.1861	-3.839	0.000124***
UKLive	0.3588	0.1429	2.51	0.012075*
Reduction Type DET	-0.8283	0.4066	-2.037	0.041643*
Reduction Type VERB	-1.707	0.6778	-2.519	0.011782*

Re-levelling Reduction Type to "Verb"

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	-2.6326	0.6144	-4.285	0.0000183***
conditione	-0.7144	0.1861	-3.839	0.000124***
UKLive	0.3588	0.1429	2.51	0.012075*
Reduction Type PREP	1.707	0.6776	2.519	0.01176*
Reduction Type DET	0.8787	0.6431	1.366	0.171816