Do colourless green voices speak furiously? Linkages between phonetic and visual perception in synaesthesia.

Anja C. Moos

Magistra Artium

Submitted in fulfilment of the requirements for the

Degree of Doctor of Philosophy

School of Critical Studies

College of Arts

University of Glasgow

January 2013

"Colorless green ideas sleep furiously."

This sentence was composed by Noam Chomsky in 1957 to show that sentences can be grammatically correct but semantically nonsensical.

Are they really nonsensical, or maybe rather synaesthetical?

Abstract

Synaesthesia is an unusual phenomenon, in which additional sensory perceptions are triggered by apparently unrelated sensory or conceptual stimuli. The main foci of this thesis lie in speech sound - colour and voice-induced synaesthesia. While grapheme-colour synaesthesia has been intensively researched, few studies have approached types of synaesthesia based on vocal inducers with detailed acoustic-phonetic and colorimetric analyses. This approach is taken here.

First, a thorough examination of speech-sound - colour synaesthesia was conducted. An experiment is reported that tested to what extent vowel acoustics influence colour associations for synaesthetes and non-synaesthetes. Systematic association patterns between vowel formants and colour measures could be found in general, but most strongly in synaesthetes. Synaesthetes also showed a more consistent pattern of vowel-colour associations. The issue of whether or not speech-sound - colour synaesthesia is a discrete type of synaesthesia independent of grapheme-colour synaesthesia is discussed, and how these might influence each other. Then, two experiments are introduced to explore voice-induced synaesthesia. First, a comprehensive voice description task was conducted with voice synaesthetes, phoneticians and controls to investigate their verbal voice quality descriptions and the colour and texture associations that they have with voices. Qualitative analyses provided data about the nature of associations by the participant groups, while quantitative analyses revealed that for all groups, acoustic parameters such as pitch, pitch range, vowel formants and other spectral properties influenced colour and texture associations in a systematic way. Above all, a strong connection was found between these measures and luminance. Finally, voice-induced synaesthetes, other synaesthetes and controls participated in a voice line-up, of the kind used in forensic phonetic case work. This experiment, motivated by previous findings of memory advantages in synaesthetes in certain areas, tested whether synaesthetes' voice memory is influenced by their condition. While no difference in performance was found between groups when using normal speech, voice-induced synaesthetes outperformed others in identifying a whispering speaker.

These are the first group studies on the otherwise under-researched type of voice-induced synaesthesia, with a focus on acoustic rather than semantic analysis. This adds knowledge to the growing field of synaesthesia research from a largely neglected phonetic angle. The debate around (re)defining synaesthesia is picked up. The voice description experiment, in particular, leads to a discussion of a synaesthesia spectrum in the population, as many common mechanisms and associations were found. It was also revealed that less common types of synaesthesia are often difficult to define in a rigid way using traditional criteria. Finally, the interplay of different types of synaesthesia is discussed and findings are evaluated against the background of the existing theories of synaesthesia.

Contents

List of Tables

List of Figures

Acknowledgement

A thesis of this magnitude cannot be accomplished without the help and support from and the occasional distraction of others. Therefore, I shall take this opportunity to thank the people who accompanied me over the past four years.

First and foremost, I would like to thank my supervisors, Rachel Smith and David Simmons, for their patient and encouraging support throughout the course of my PhD, through all its ups and downs. Rachel with her most understanding and balanced nature was particularly helpful with her professional advice in and beyond the field of phonetics. David gave me constant support and was especially invaluable in his guidance in the area of psychophysics and the science of colours. It was a pleasure to work with colleagues in both the Department of English Language and the School of Psychology, thanks for enlightening or relaxing chats, help and exchange of ideas. My gratitude extends to colleagues from other universities: I am grateful for the development of interesting ideas in the field of forensic phonetics with Dominic Watt, and also for the help in this field from him, Louisa Stevens and Kirsty McDougall. I also owe my sincere thanks to Sam Miller for his help with statistical analysis and John McClure and Marc Becirspahic for their support in setting up the online experiments. Furthermore, I greatly appreciate the honest work of numerous proofreaders who ploughed through various bits of my writing in all its imperfections: Cordula Klein, Brian José, Ellen Bramwell, Claire Macmillan, Sophie Holmes-Elliot, Philip Webb, Andrea Stout, Duncan Robertson and Hamish Haridas.

Thanks to everybody who took part in my experiments! A thank you also to Flor Kusnir for setting up a synaesthesia database at the University of Glasgow, and to Julia Simner for sharing the database of synaesthetic participants with me, which she set up with Jamie Ward, that allowed me to find more voice-induced synaesthetes.

I am most thankful for the Lord Kelvin Adam Smith funding by the University of Glasgow which gave me the opportunity to do my research at this reputable university and to take part in enriching conferences of various professional backgrounds.

Last but certainly not least, I would like to thank my friends and family who gave me a good time when I needed a rest from work and helped me get through difficult situations. I particularly enjoyed my time with the Ultimate Frisbee team that gave me a healthy work-life balance and all the friendships that I gained from it. The deepest gratitude is addressed to my family: Gernot, my partner, never lost his patience with me, but instead always encouraged me and kept me going when I needed a push and helped me relax when I needed a rest. My parents, my brother and Gernot always showed their proudness, respect and moral support, and never stopped believing in me. Thank you!

1 Introduction

1.1 An introduction to synaesthesia

Synaesthesia is an unusual phenomenon, in which additional sensory perceptions are triggered by apparently unrelated sensory or conceptual stimuli. It is often described as a perceptual and/or neurological condition, or a "cross-wiring" of the senses. For example, people with grapheme-colour synaesthesia see colours while reading letters or numbers, or associate colours with those graphemes. The most common types of synaesthesia are visual sensations triggered by graphemes, phonemes or words, including sequences like weekdays, numbers or months. Cross-modal sensations can exist in various combinations of modalities including auditory-visual (e.g. sound-colour), visual-visual (e.g. grapheme-colour), auditory-gustatory (sound-taste) and auditory-tactile (sound-touch). For instance, in sequence-spatial synaesthesia, months or numbers are arranged in an idiosyncratic way in space for the synaesthete, so months for example may be arranged in a circle around their body, and numbers in a spiral leading to infinity. Synaesthesia can be triggered not only by sensory stimulation such as seeing a stimulus, but also by concepts. Thinking of the weekdays, for example, may trigger seeing or imagining a spatial arrangement.

Synaesthesia can arise in one of three ways, resulting in the following categories: (1) *developmental* synaesthesia: people automatically and involuntarily experience synaesthetic perceptions from an early age; (2) *acquired* synaesthesia caused by brain injury or sensory deafferentation (e.g. losing one's eyesight); (3) *drug-induced* synaesthesia caused by hallucinogenic drugs such as LSD (Grossenbacher & Lovelace, 2001).

Traditionally, a set of key characteristics of synaesthesia has been agreed upon, although researchers' opinions differ slightly with regard to the details. The key characteristics that have usually been used to define synaesthesia, are for example detailed in Ward & Mattingley (2006):

1. It is an unusually elicited experience evoked by perceptual or conceptual stimuli.
2. Synaesthetic experiences are automatic, and difficult or impossible to suppress.
3. They are akin to a conscious perceptual event.
4. They are usually consistent over time.

These characteristics will be critically discussed in section 1.7.

As a naming convention, types of synaesthesia are usually referred to by a compound adjective. The first element of the compound is the stimulus that triggers the synaesthetic perception, and the second part is the additional synaesthetic perception. So grapheme-colour synaesthesia means that graphemes trigger additional perceptions of colour. Note that, up until the mid/late 20th century, labelling was the reverse, with the trigger in second position (cf. Simpson & McKellar, 1955). The stimulus class which triggers the synaesthetic sensation is also called the *inducer,* and the modality in which synaesthetic experiences take place is called the *concurrent* (Ward & Simner, 2003). Therefore, in synaesthesia involving taste sensations while hearing words, the word would be described as an inducer and the taste as a concurrent.

1.2 Focus of the thesis

This thesis solely investigates developmental synaesthesia. More precisely, my work examines vowel sound - colour associations and synaesthesia induced by the sound of voices. Concurrents of a complex inducer such as a voice can be astoundingly multimodal. One participant described a voice stimulus as being “Wood grain, drippy, with olive green mostly to the upper right. Outlined in thin black, except for the word ‘people’ which seems to be outlined in maroon and is taller. It’s [sic] smells like cigar smoke; this tone speaking this phrase.” Clearly, this description cannot be reduced to something as simple as colour as a concurrent. It highlights the interplay of different kinds of synaesthesia and the consequent difficulty in analysing ‘weaker’ types. My research aims to be a vital component for tackling this complex phenomenon further.

I set out to show the multifaceted nature of synaesthesia and the limitations of research methods for working with all its varieties, especially the less common ones such as voice-induced synaesthesia. In addition, I outline the difficulty of defining (less common types of) synaesthesia and where to draw the line between synaesthetes and non-synaesthetes – or, indeed, whether a line exists or the synaesthetic perceptions lie along a continuum. For this, three experiments were conducted introducing research methods new to the field of synaesthesia. In all experiments, both synaesthetes and non-synaesthetes took part to allow for comparisons in their performance. Striking similarities as well as distinct differences between the groups are revealed.

The first experiment investigates colour associations with vowel sounds among synaesthetes and non-synaesthetes. It focuses on how far vowel acoustics influence colour and grey shade associations, and whether synaesthetes and non-synaesthetes show similar association patterns. The aim of this experiment is to test for a systematic influence of vowel acoustics on colour and grey shade associations and explore how far graphemic and dialectal factors may be additional influences on vowel sounds as inducers. The introduction of precise colorimetric measures should aid the discovery of more detailed relationships between vowel acoustics and colours.

The second experiment aims to set up a comprehensive and systematic description of voice-induced synaesthesia. The voice is a complex inducer and it is unclear which aspects of the voice trigger synaesthetic perceptions. I designed an experiment using different voice qualities, which participants were asked to verbally describe, to associate with a colour and a texture, and to rate descriptively using semantic differentials. To assess which aspects of the voice function as an inducer, various acoustic analyses were carried out that unravelled distinct voice criteria, and were compared with the different concurrents. Motivational points stem from open questions triggered by mere curiosity or related research. Above all, it is necessary to find out which concurrents the sound of a voice induces. Common types of synaesthesia have a relatively clear one-to-one inducer-concurrent relation, e.g. words to tastes, graphemes to colours, etc. However, descriptions like the one cited at the beginning of this section lead us to believe that a multitude of concurrents are evoked. Several acoustic measurements were made to detect similar occurrences of synaesthetic associations with voices as found with music (e.g. Ward et al., 2006; de Thornley Head, 2006).

The third experiment searches for a potential application for synaesthesia: can voice-induced synaesthesia or other kinds of synaesthesia aid in voice recognition? This test was motivated by previous research that suggests a memory advantage in synaesthetes in areas involving their inducers and concurrents (Ward et al., 2009; Rothen & Meier, 2010; Bor et al., 2008; Luria, 1968). Here, a new type of experiment is introduced to synaesthesia research. To investigate whether synaesthetic perceptions and associations help or hinder memorising and recalling people's voices, a short voice comparison task and two voice line-ups were conducted using modal (i.e. normally speaking) and whispering voices. The voice line-ups follow the guidelines of criminal police standards as they are used for speaker identification by witnesses or victims as closely as possible.

To condense my general motivation into one sentence: I would like to introduce research on a rare type such as voice-induced synaesthesia and an under-researched inducer such as vowel sounds to the growing field of synaesthesia research, and enhance its methodologies by using acoustic and colorimetric analysis. Firstly though, a more general introduction is given in this chapter. Prevalence and types of synaesthesia are presented in Section 1.3. Section 1.4 introduces tests and definitions of synaesthesia, followed by a short discussion of why synaesthesia research is important in Section 1.5. Theories of synaesthesia are discussed in Section 1.6, followed by state-of-the-art ideas about redefining synaesthesia in Section 1.7. Section 1.8 features a closer examination of the motivation for introducing new research methods into synaesthesia research (1.8.1), and for researching vowel sound - colour (1.8.2) and voice-induced synaesthesia (1.8.3) in particular. Finally, a brief outline of the thesis is given in Section 1.9.

1.3 Prevalence and types of synaesthesia

A comprehensive study of the prevalence of synaesthesia in the general population was conducted by Simner et al. (2006). The authors circulated questionnaires to 500 students.

After having been introduced to the phenomenon, participants could indicate which types of synaesthesia they had by connecting a list of inducers and concurrents. Those who indicated that they experienced some form of synaesthesia were later tested in more detail. In addition, the authors tested 1190 visitors at a museum on their grapheme-colour associations as part of an exhibition. Simner et al. (2006) found that 4.4% of their participants showed synaesthetic perceptions of some sort. Whereas earlier research proposed that many more women experience synaesthesia than men (Baron-Cohen et al., 1996; Rich et al., 2005), Simner et al. (2006) could not confirm this sex bias in their survey. They found a female to male ratio of 1.1:1 and claim that the previously found sex bias is due to differences in self-disclosure.

Interestingly, the authors only found nine different types of synaesthesia: letter-colour, number-colour, month-colour, day-colour, word-colour, people-colour, person-smell, taste-shape and music-colour. However, strict inclusion criteria were applied, as synaesthesiae with emotions or pain as an inducer, for instance, were excluded, because these types are difficult to research and prove. Some researchers have debated whether days of the week are acceptable as an inducer; Baron-Cohen et al. (1996) for example excluded from their study two cases with days as the only inducer. Because there are only seven days, it would be very easy to remember and automatise colour associations with them, which would then be difficult to distinguish from synaesthesia proper. If this type – which is indeed the type most often reported – were to be excluded from statistics in the survey of Simner et al. (2006), the prevalence would drop to 3%. However, acceptance of days as an inducer is in line with recent changes in the definition of synaesthesia, which seem to have made the characterisation of this phenomenon more inclusive (see Section 1.7).

In fact, the understanding of types of synaesthesia has steadily grown over the past decades, or even centuries. The first scientific or medical studies that encountered synaesthesia only reported superficially on the condition and usually solely commented on synaesthetic experiences pertaining to coloured sequences, letters, vowels, sounds or music. Jewanksi et al. (2011) present an overview of the first case reports of synaesthesia in the 19th century, starting with Sachs in 1812 reporting on his own synaesthesia of coloured sequences and music. After a gap of references for many decades in the early 20th century, the interest in synaesthesia entered mainstream research in the middle of the 20th century. Simpson & McKellar (1955) reported on six developmental types of synaesthesia, alongside drug-induced ones. Fifty years later, Day (2005) listed 35 different types of synaesthesia. Sean Day is the host and moderator of an international synaesthesia emailing list, that gives synaesthetes, their families and researchers the opportunity to discuss various topics related to synaesthesia. Day keeps statistics of publications and synaesthetes' reports, and a product of his continuous work over the past decade or so is a list of types of synaesthesia on his synaesthesia webpage, which listed 65 different types of synaesthesia in summer 2012: http://www.daysyn.com/Types-of-Syn.html. Although most numbers are based on self-report, many of them have been researched and verified scientifically as well. A modified version of the list is shown in Table 1. The numbers in this table sum up to more than 100% because some synaesthetes report multiple types of synaesthesia.

Table 1: Types of synaesthesia that have been reported to date. List from the homepage of Sean Day reprinted with his permission. http://www.daysyn.com/Types-of-Syn.html, visited 01/07/2012, slightly altered (put in descending order of occurrence and changed to British spelling).

Type of synaesthesia	Frequency of occurrence among synaesthetes
graphemes -> colours	64.80%
time units -> colours	23.00%
musical sounds -> colours	20.00%
general sounds -> colours	15.20%
phonemes -> colours	9.50%
musical notes -> colours	8.80%
smells -> colours	7.00%
flavours -> colours	6.20%
sound -> flavours	6.10%
pain -> colours	5.60%
personalities -> colours ("auras")	5.60%
grapheme personification (a.k.a. ordinal linguistic personification)	4.60%
sound -> touch	4.40%
touch -> colours	4.00%
vision -> flavours	2.80%
temperatures -> colours	2.40%
orgasm -> colours	2.20%
vision -> sounds	2.00%
object personification	1.70%
emotions -> colours	1.60%
sounds -> smells	1.60%
vision -> touch	1.40%
touch -> flavours	1.00%
vision -> smells	1.00%
lexeme -> flavour	0.70%
sounds -> kinetics	0.70%

Type of synaesthesia	Frequency of occurrence among synaesthetes
smells -> touch	0.60%
kinetics -> sounds	0.50%
lexeme -> touch	0.50%
smells -> sounds	0.50%
sound -> temperatures	0.50%
touch -> smell	0.50%
emotion -> pain	0.30%
personalities -> smells	0.30%
touch -> sounds	0.30%
flavours -> sounds	0.20%
grapheme -> flavour	0.20%
musical notes -> flavours	0.20%
pain -> sound	0.20%
touch -> emotions	0.20%
vision -> temperatures	0.20%
emotion -> flavour	0.10%
emotion -> smell	0.10%
emotion -> temperature	0.10%
emotion -> touch	0.10%
flavours -> temperatures	0.10%
flavours -> touch	0.10%
grapheme -> touch	0.10%
kinetics -> colours	0.10%
pain -> flavour	0.10%
pain -> smell	0.10%
personalities -> touch	0.10%
phonemes -> flavour	0.10%
phonemes -> touch	0.10%
smells -> flavour	0.10%
smells -> temperatures	0.10%
temperature -> flavours	0.10%
temperature -> sounds	0.10%
time units -> sound	0.10%
touch -> temperatures	0.10%
vision -> kinetics	0.10%
mirror touch	not enough data

Type of synaesthesia	Frequency of occurrence among synaesthetes
non-graphemic ordinal personification	not enough data
ticker-tape	not enough data
spatial sequence (number form)	not enough data

Recently, Novich et al. (2011) conducted a study to analyse which types of synaesthesia cluster more often than others, and at the same time reported on the prevalence of the researched types of synaesthesia (cf. Novich et al., 2011, p. 356). As a database, they used Eagleman's *Synaesthesia Battery* (http://www.synaesthete.org/, described in Eagleman et al., 2007), which had 19,000 participants at the time of analysis, including non-synaesthetes. The online *Synaesthesia Battery* asks participants to self-report 22 types of synaesthesia with the option to name more; its tests focus on grapheme-colour and music-colour synaesthesia. The data of Novich and colleagues, as well as that of Day presented in Table 1, do not show the prevalence of synaesthesia in the general population, but the prevalence of the different types of synaesthesia in the participants who claimed to have at least one kind of synaesthesia. The seven most frequent types in the survey of Novich et al., each reported by above 30% of participants, are: number-colour, weekdays-colour, months-colour, letter-colour, emotion-colour, personalities-colour, and spatial arrangements.

Synaesthetes are not only classified according to their inducer and concurrent pairs, but also according to the nature of their perceptions. Some synaesthetes see their concurrents placed somewhere outside their body: for example the colour of a letter might be projected onto the letter itself, or it might be presented as a "screen" in front of their eyes. This group of synaesthetes is called *projectors*. Other synaesthetes associate their concurrents with their inducers by simply thinking of them or seeing them in front of their mind's eye. This group is called *associators* (cf. Dixon et al., 2004, for the introduction of these subtypes of synaesthesia, and Skelton et al., 2009, for a more accurate test to classify synaesthetes into projectors and associators).

It is common that synaesthetes have more than one type of synaesthesia (cf. Ramachandran & Hubbard, 2001b, p. 5). Novich et al. (2011) found five subgroups of synaesthesia that occur together more frequently than with other subgroups:

1. *coloured sequences* (letters, numbers, weekdays and months to colour)
2. *coloured sensations* (touch, orgasm, pain, temperature, personality, emotion, taste and smell to colour)
3. *spatial sequences* (on its own)
4. *non-visual sequelae* (sound and vision to smell and taste, vision to sound, sound to touch)

5. *coloured music* (pitch, chords and musical instruments to colour)

This categorisation is based on the tests of the *Synaesthesia Battery* which mainly use visual and a few musical stimuli to test synaesthesia. Therefore, more than the five subgroups may exist, but they were not tested for.

Looking back at the seven most frequent types of synaesthesia in the survey of Novich et al. (2011), we find that the first four types, namely number-colour, weekdays-colour, months-colour and letter-colour, are grouped into *coloured sequences*, suggesting that over-learned sequences are a common inducer and colours are a common concurrent. Why colours are most commonly concurrents and words or sequences inducers remains unanswered. Although there are rare cases of bidirectional synaesthesia (Cohen Kadosh et al., 2007), commonly inducer and concurrent cannot be reversed. To paraphrase:

> It remains a puzzle why synaesthetic inducers and concurrents tend to separate themselves by clustering in language and natural object features respectively. That is, orthographic and phonemic features only induce synaesthetic experience and do not appear as concurrent phenomena. Conversely, colour is a common synaesthetic concurrent but not typically reported as an inducer. (Grossenbacher, 1997, p. 166)

In summary, there are approximately 60 different types of synaesthesia. The most common inducers are graphemes, words, sequences and music; the most common concurrent is colour. The overall prevalence of synaesthesia in the general population is about 4%, and slightly more women might be affected. As mentioned in Section 1.2, this thesis also examines synaesthesia induced by the sound of voices. The attentive reader may have noticed that voice as an inducer is not even listed in any published research on prevalences and types of synaesthesia. However, based on the synaesthesia questionnaire created by Julia Simner and Jamie Ward (see https://www.survey.bris.ac.uk/sussex/syn), Simner estimates that up to 10% of the tested synaesthetes have voice-induced synaesthesia (personal communication). This discrepancy shows the magnitude and complexity of synaesthesia and how difficult it is to determine the fine-grained differences of which it is composed. It further highlights that more research is necessary to define and evaluate this curious condition further, because tests have not been sensitive or comprehensive enough to evaluate all types of synaesthesia. So far, research suggests that the common types have a relatively clear one-to-one inducer-to-concurrent relation, e.g. words to tastes, graphemes to colours (even though textural influences sometimes occur), etc. Yet, anecdotal descriptions of voice-induced synaesthesia lead us to believe that a multitude of concurrents are evoked: colours, shapes, textures, movements, to name but a few. Hence, a more detailed assessment of voices and other rare inducers is necessary.

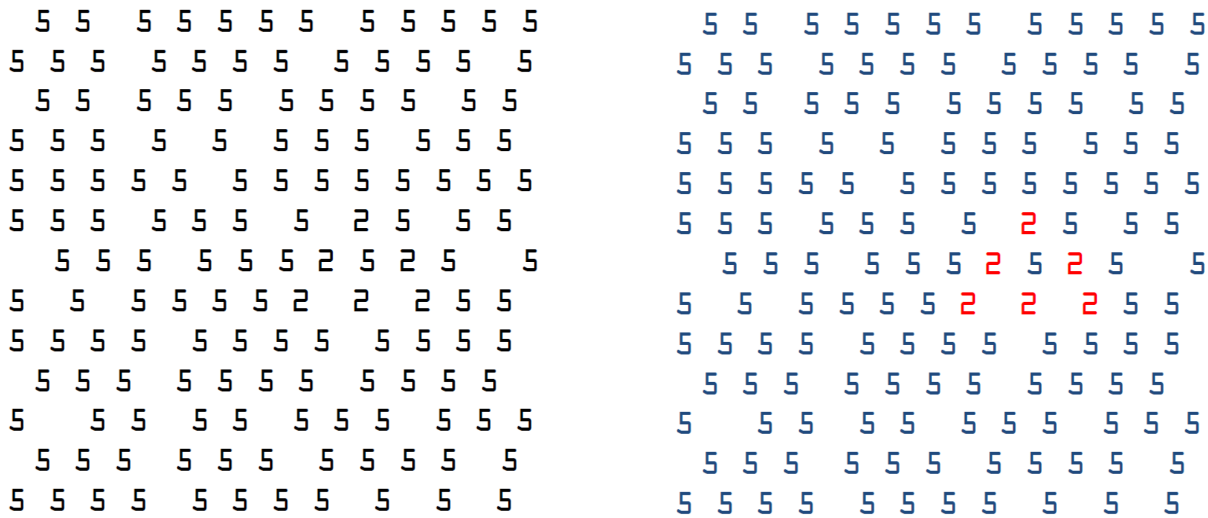


Figure 1: Visual pop-out effect. A geometrical shape of 2's is 'hidden' in a background of 5's. The display is presented in black and white (left), but some synaesthetes apparently perceive the numbers in their synaesthetic colours similar to the display on the right.

1.4 Tests of genuineness

To define synaesthesia and to test its genuineness, researchers have worked on various psychophysical approaches. Asher et al. (2006) extended an earlier "test of genuineness" introduced by Baron-Cohen et al. (1987). Both versions of the test are based on the key characteristic that synaesthesia is supposed to be consistent over time. In the TOG-R, the **revised test of genuineness**, by Asher and colleagues, colour associations with 99 acoustic stimuli are tested. The 99 stimuli consist of days of the week, months of the year, Christian names, nouns, verbs, articles, numbers, letters, animal vocalisations, environmental sounds, musical notes and chords of different instruments. Participants in the study of Asher et al. (2006) could choose one of 238 colour swatches to express their synaesthetic colour. A retest was carried out one month after initial testing for synaesthetes and one week after initial testing for controls. The authors found that synaesthetes were more consistent in their colour choices over time than controls (71% vs. 33% consistency in retests, respectively). They also found that it was mostly words which triggered synaesthetic perceptions, and only for a few participants did music trigger synaesthesia as well.

Other tests devised to prove the existence of synaesthesia include the usage of the pop-out effect in visual search tasks, and the Stroop test. Ramachandran & Hubbard (2001a) and Palmeri et al. (2002) introduced the **pop-out effect** as a measurement tool for synaesthesia. When participants were presented with a screen full of randomly placed 5's, in which a geometrical shape of 2's was 'hidden', synaesthetes found the geometrical shape more often and faster than controls in a given time limit, because the overlaid synaesthetically perceived colour made the 2's pop out from the background colour of 5's, similar to the display shown in Figure 1. However, other studies later showed that not every synaesthete is able to use their grapheme-colour synaesthesia to aid visual search (e.g. Edquist et al., 2006; Gheri et al., 2008), most likely because only a few synaesthetes actually project a colour onto the printed grapheme. Ward et al. (2009) evaluated these ambiguous results by repeating the experiment and performing additional assessment of self-reports on the colour perceptions of participants. Interestingly, they found that synaesthetes were indeed faster in detecting the geometrical shapes, but often did not experience a visual pop-out effect. Only the graphemes in the visual focus of attention seem to be “seen” in colours.

In the original **Stroop task** (Stroop, 1992, reprint of the original in 1935), a colour term was printed in an ink colour different from the one that the word represented, e.g. the word ‘green’ was printed in red. The task was to name the colour of the ink rather than to read out the colour term. As reading is an automated task, participants struggled to name the ink colour without hesitation or mistakes. Synaesthetes and controls participating in the study of Mattingley et al. (2001) were asked to name the ink colour of printed numbers (which triggered colour associations in the synaesthetic participants), similar to the original Stroop task. Here, synaesthetes were significantly slower in naming the colour when the ink colour was incongruent with their synaesthetically perceived colour than in the congruent situation. Nikolić et al. (2007) conducted a very similar experiment using coloured letters displayed on a screen, rather than numbers printed on paper. Again, the colour of the letter could either be congruent or incongruent with the synaesthetic perceptions of the synaesthetes participating. The authors also found slower reaction times in synaesthetes when the display colours of letters were incongruent with their synaesthetic colours for the letters. Moreover, the authors found that reaction time was slowest when colours *complementary* to participants’ synaesthetic colours were used (for example green, when the synaesthetic colour was red), rather than any other incongruent colours (for example yellow, when the synaesthetic colour was red).

Similarly, Ward et al. (2006) conducted a Stroop interference test for musical tone-colour synaesthesia to show that synaesthetic colours are automatically elicited. For this they tested cross-modal orienting of spatial attention: two coloured squares appeared on the screen, one of them marked with an asterisk (*). At the same time, a tone was played which participants were told to ignore. Participants had to press a key when they located the target (*). Synaesthetes were slower when the colour of the target square and the synaesthetic colour associated with the tone were incongruent compared to the congruent condition.

To sum up, (1) retests exploring the consistency of synaesthetic concurrents over time, (2) visual search tasks testing the pop-out effect, and (3) Stroop tasks testing stimuli that interfere with synaesthetic perceptions are often used to test the genuineness of synaesthesia.

However, although all three methods highlight a difference in performance between synaesthetes and non-synaesthetes, I hesitate to consider them the only methods for proving genuineness. While these tests clearly do help advance scientific synaesthesia research, synaesthesia expresses itself in so many different ways that there are synaesthetic individuals who show less consistency in their perceptions than others and/or who may be better able to shift their attention towards real or synaesthetic colours. It has also been shown that there are synaesthetes who do not project synaesthetic colours onto the graphemes but only associate colours with letters (Dixon et al., 2004; Skelton et al., 2009; Gheri et al., 2008). Tests like the pop-out effect are less suitable for these synaesthetes. In Section 1.7, definitions of the genuineness of synaesthesia are discussed in greater detail.

1.5 Why do synaesthesia research?

This question has been asked and addressed multiple times in previous literature (e.g. Ward & Mattingley, 2006). The authors state that synaesthesia research is able to inform us about intra-modal and cross-modal perceptions – be it synaesthetic or "normal" perception. Through synaesthesia research we can learn about multimodal integration at a behavioural and neurological level. In addition, this research relates to the subject of perceptual awareness, and helps differentiate between perceptual processes that we are and are not consciously aware of. Brain development and plasticity and genetic influences upon them are discussed in synaesthesia research, which might aid the understanding of other typical or atypical brain functions as well. It is interesting and useful to directly (i.e. neurologically) or indirectly (behaviourally) study brain functions in synaesthesia as a non-pathological condition; it highlights that not all healthy brains are anatomically and/or functionally equal. Barnett et al. (2008) investigated synaesthesia traits in 17 families and concluded from their study "that various types of synaesthesia are fundamentally related at the genetic level, but that the explicit associations and the individual differences between synaesthetes are influenced by other factors. Synaesthesia thus provides a good model to explore the interplay of all these factors in the development of cognitive traits in general." (Barnett et al., 2008, p. 871)

Last but not least, synaesthesia research also provokes thoughts about the overlap between perception, imagery and memory, and the influence and use of language in these respects (for example the use of metaphors). Ramachandran & Hubbard (2001b) especially dealt with these aspects in their overview, as the title of their article already suggests: "Synaesthesia – A Window Into Perception, Thought and Language". The authors think for example that metaphors are analogous to *perceptual* cross-activation in synaesthesia, because metaphors involve *conceptual* cross-activation. They also claim that metaphors share directionality with synaesthesia, so one can talk about a loud shirt (but not about a fabricky noise). This parallel in directionality might be because certain cross-activations in the brain are possible (through excess connections), but others are not. Ramachandran and Hubbard also found a larger number of synaesthetes amongst artists than amongst other professions, suggesting that synaesthetes are on average more creative than others. Links like these – between metaphors and synaesthesia, or between creativity and synaesthesia – should aid the understanding of cognitive processes and associations, whether they are of synaesthetic or semantic nature.

1.6 Theories of synaesthesia

There are multiple theories that attempt to define the origins and characteristics of developmental synaesthesia. They range from practical theories claiming learned associations to highly elaborate neurological theories defining specific brain functions. Interpretations of hyperactive imagination or metaphorical expressions are excluded from "true" theories of synaesthesia, as they lack evidencing research. A theory implicating learned associations suggests that grapheme-colour synaesthesia originates in playing with coloured refrigerator magnets (Witthoft & Winawer, 2006; Hancock, 2006): through overlearning the colour-letter pairings, synaesthetic associations are created. However, this theory is based on a few case studies only and has not received wide scientific approval. Theories which try to neurologically define the origin of the phenomenon are currently given the most credence.

Advances in neuroimaging techniques have allowed the investigation of neurological theories of synaesthesia. An accessible diagram about the umbrella theories is given in Bargary & Mitchell (2008) and reprinted in Figure 2.[1] The four models of synaesthesia are explained along two categorical axes: some are based on structural differences in the brain, whereas others are based on functional differences. Some assume a direct link between the brain areas involved, whereas others assume indirect pathways through a higher order nexus. In theories based around structural differences, additional excitatory connections are believed to be present in synaesthetes, either directly between brain areas that involve inducer and concurrent (part (a) in the diagram), or indirectly via a higher order nexus (b). In theories based around functional differences, it is believed that existing inhibitory pathways are partly inactive in synaesthetes, resulting in disinhibition, again either directly between brain areas that involve inducer and concurrent (c), or indirectly via a higher order nexus (d). In the latter case, the disinhibition would take place between the higher order nexus and the concurrent brain area.

Theories discriminating between indirect and direct pathways also distinguish between 'higher' and 'lower' synaesthetes (Ramachandran & Hubbard, 2001b). According to this distinction, lower synaesthetes perceive concurrents induced by stimuli that are processed

[1] Graphics in the Figures 2, 3, 11, 12, 13 and 14 are unaltered reproductions of the cited papers with slightly modified captures.

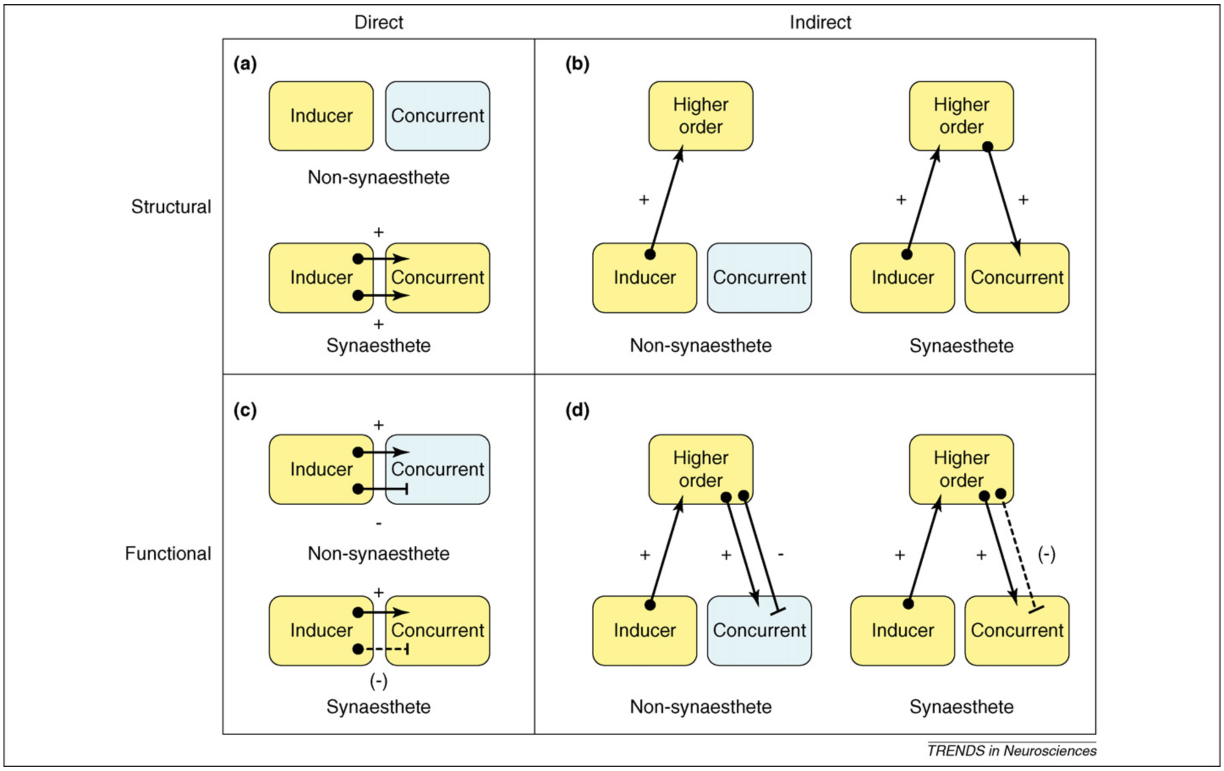


Figure 2: Neurological models of synaesthesia. Models differ in the proposed route of cross-activation (direct or indirect) between the inducer area and the concurrent area and the proposed underlying difference in synaesthetes (structural or functional). Yellow areas are active (starting with the inducer area) and blue areas are inactive. Excitatory connections are shown as arrows and inhibitory connections as blunt ended. Dashed lines represent structurally present but functionally ineffective connections. (A variation on the disinhibition model would posit a structural decrease of inhibitory connections as the reason for excess cross-activation.) Connections from the concurrent area to the higher-order area in (b) and (d) are not shown for simplicity, but note that such connections pose a problem for indirect models as they would lead to a recurrent excitatory loop. Note also that indirect connectivity via the thalamus as opposed to a higher-order cortical area is also possible. Reproduced from Bargary & Mitchell (2008, p. 337)

on an early low sensory level, whereas the concurrents of higher synaesthetes are induced by later multiple sensory or conceptual stimuli. An example of how to discriminate between higher and lower synaesthetes could be the following: synaesthetes are presented with the number '6' and its Roman equivalent 'VI'. If a synaesthete gets the same concurrent for both representations, he or she would be a higher synaesthete, as he or she reacted to the concept of the number rather than the physical appearance of the printed graphemic presentation. This is a reaction to late processing suggesting indirect pathways. If he or she had different synaesthetic reactions or reacted to only one of the representations, he or she would more likely be a lower synaesthete, who responded to early sensory processing, presumably activating direct pathways. Ramachandran & Hubbard (2001b) found that most studies point towards higher-level processing. This does not necessarily exclude the possibility that both ways of processing co-exist, but more synaesthetes seem to process synaesthetic perceptions on a higher level. Evidence favouring structural vs. functional theories, which may or may not discriminate between direct vs. indirect processing differences, are outlined in the following paragraphs.

1.6.1 Theories assuming structural differences

In an article reviewing neurological research on synaesthesia in the past ten years, Hubbard et al. (2011) seem to rule out the possibility that a higher nexus might be involved in producing concurrents, and therefore argue for **direct structural connections**. The authors used studies with EEG (electroencephalography) and MEG (magnetoencephalography) measurements to make their claim. EEG records the brain's electrical activity by measuring the ionic current flows within the neurons of the brain, whereas MEG records magnetic fields produced by the electrical currents. Both methods benefit from a relatively high temporal resolution of brain activity (while having relatively poor spatial resolution). Some EEG and MEG studies described in Hubbard et al. (2011) showed that "colour processing areas" like V4 were activated only 5ms after the visual word form areas in grapheme-colour synaesthetes. Because the time delay of activation between the two brain areas is so short, the authors assume a direct path of neuronal connections in synaesthetes rather than the involvement of a higher nexus. However, a generalisation applying these findings to all synaesthetes might be somewhat hasty, especially taking into account other studies discussed below.

Rouw & Scholte (2007) validated a particular version of the structural differences approach, the **hyperconnectivity hypothesis**, according to which synaesthetes' brain areas involved in processing inducers and concurrents have more neural interconnections than those of non-synaesthetes. To show this, they conducted a study using DTI (diffusion tensor imaging) on 18 grapheme-colour synaesthetes and controls. This method uses an anatomical brain scan to search for connectivity in white matter by tracking water diffusion in the tissue. Synaesthetes showed greater anisotropic diffusion than controls, indicating more coherent white matter. Greater connectivity was found in the inferior temporal cortex and the superior frontal and parietal cortex. An additional fMRI scan showed increased activation in "colour" and "grapheme" areas for synaesthetes. Functional magnetic resonance imaging (fMRI) is a procedure measuring brain activity by detecting changes in oxygen levels in the different brain areas. These oxygen levels vary with brain activity, as the blood flow changes accordingly. Rouw and Scholte claim that the cause of the synaesthetic perceptions in their grapheme-colour synaesthetes might be cross-connections between adjacently located grapheme and colour areas in the fusiform gyrus, as shown by the DTI. Although Rouw and Scholte focus on hyperconnectivity between "grapheme" and "colour" areas, i.e. direct structural connections, they also found hyperconnectivity reaching the frontal cortex. The authors interpret that this might give insight as to why synaesthetes perceive their condition consciously.

Maurer & Mondloch (2005) also favour the structural theory. The authors stipulate that everybody is born as a synaesthete, because neonates cannot differentiate the stimulation of different senses. One reason for this could be that brain regions in neonates' cortex are connected equivalently throughout the brain and only later differentiate through pruning, i.e. "disconnecting", or inhibition. Another reason could be that the multimodal limbic system is more mature than the cortex, which changes as the baby learns to filter the stimulation from the outside world. The authors seem to favour the **hypothesis of pruning**, which falls under the category of the structural approach – the interconnections of the brain change during development, but for synaesthetes, some remnants remain. Besides developmental changes during childhood, Maurer and Mondloch mention cortical plasticity, especially of people lacking a sense such as the blind. Although most aspects of their discussion point towards direct structural neurological manifestations of synaesthesia, they do not explicitly state whether they expect these remnant pathways to directly or indirectly link brain areas involved in processing inducer and concurrent stimuli. They also lack empirical tests to support their theory.

1.6.2 Theories assuming functional differences

Most theories that assume functional differences seem also to assume indirect neurological pathways. Nunn et al. (2002) used spoken words (and pure tones) as stimuli for ten word-colour synaesthetes who reported no synaesthesia for any other auditory stimuli. Their fMRI study showed activation in the visual brain area V4/V8 of the left occipital lobe – which is often (controversially) referred to as the "colour centre" of visual perception in both hemispheres – for synaesthetes when listening to spoken words. A control group, which had to overlearn colour associations with the words used in the study, did not show activation in that area when presented with the spoken stimuli. Usually the primary visual cortex V1/V2 is activated in visual (low level) perception, but it was not activated in the synaesthetes who saw their coloured concurrents. This indicates that the activity of the primary visual cortex is not necessary for synaesthetic colour perception. Hence, Nunn and colleagues hypothesised that synaesthesia is not a low-level perception but processed later, i.e. it is an **indirect higher-level phenomenon**.

This higher-level hypothesis is also supported by Bargary et al. (2009) using the McGurk effect. Here, a video of a face mouthing an incongruent stimulus to the audio track is played, e.g. hearing the syllable "ba" and simultaneously seeing the pronunciation of "ga", which usually results in perceiving "da" (McGurk & MacDonald, 1976). Bargary et al. (2009) used the same paradigm with monosyllabic words like "bait" and "gate", and presented auditory, visual, and incongruent audiovisual (McGurk) stimuli. When presented with the audiovisual stimulus, most synaesthetes' concurrent was induced by the multisensory McGurk stimulus ("date" in this example), rather than by sound or vision only. Hence, the authors concluded that synaesthesia is induced at a later perceptual stage in processing rather than at an early sensory one. In contrast to this, Sinke et al. (2012) found that their synaesthetic participants mostly perceived the concurrents based on the audio track when presented with audiovisual McGurk stimuli. A cause for these conflicting findings could be slight differences in stimuli: Bargary and colleagues used real words, whereas Sinke used meaningless syllables. Hence, a top-down process might have been involved in Bargary's participants but not in Sinke's.

Esterman et al. (2006) support the theory of functional differences in synaesthetes' brains. In their study, transcranial magnetic stimulation (TMS) was used on two synaesthetes. TMS uses electromagnetic induction to induce weak electric currents using a rapidly changing magnetic field, which has an effect on the firing of the neurons in the brain; this can cause temporary activation changes in the parts of the brain which TMS is applied to, allowing the functioning and interconnections of these areas to be studied. In their study, rTMS (repetitive TMS) was targeted at the angular gyrus at the junction of the posterior intraparietal sulcus and transverse occipital sulcus. This region is associated with colour-form binding in normal perception. Stimulation of these areas, but not others, resulted in an interference with synaesthetic perceptions. According to the **disinhibition hypothesis**, connections in this area should usually be disinhibited in synaesthetes, allowing them to perceive their concurrents. The rTMS interfered with this disinhibition, i.e. inhibited this connection seemingly present in the two synaesthetes. Thus, the results indicate that feedback from a multimodal association region like the right posterior parietal lobe contributes to the synaesthetic perception. Although this study is well designed, it was conducted with only two participants and the results should be treated accordingly.

Grossenbacher & Lovelace (2001) propose another theory based on disinhibition that is an example of the functional view. The authors hypothesised that the feedback from

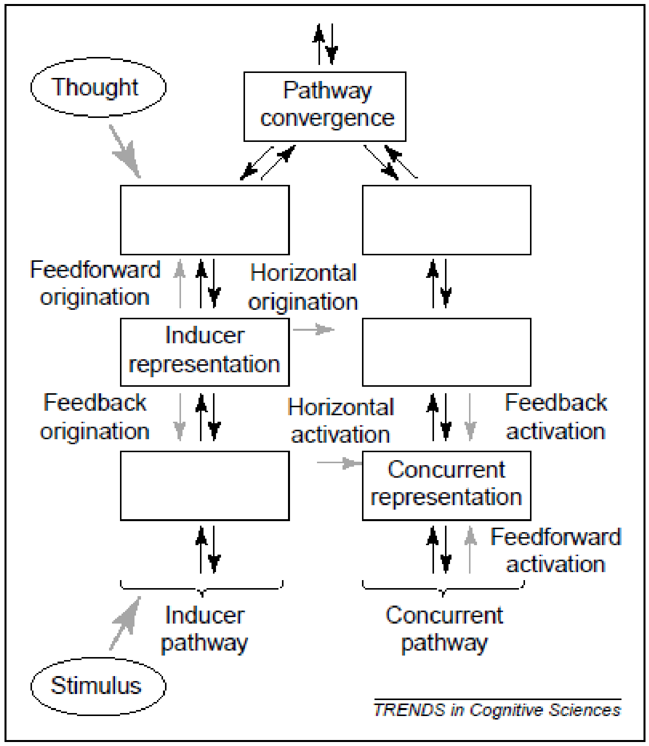


Figure 3: Schematic depiction of neural mechanisms in synaesthesia. Synaesthesia could be mediated via neural signals between an inducer pathway (left) and a concurrent pathway (right). Each box depicts a representation within a pathway (a single representation may be anatomically distributed over multiple brain areas). Afferent flow of information is conveyed by bottom-up signals via feedforward neuronal projections (upward black arrows), and top-down signals are carried by feedback connections (downward black arrows). Synaesthesia stems from activity in the inducer pathway during either synaesthetic perception of a stimulus or synaesthetic conception of a thought, and the concurrent representation could become activated either via horizontal connections between the pathways or as a result of pathway convergence. Reproduced from Grossenbacher & Lovelace (2001, p. 39).

a concurrent, which would usually be blocked, is disinhibited when an inducing stimulus is perceived. Their model (cf. Grossenbacher & Lovelace, 2001, diagram on p. 39, reproduced here as Figure 3) allows for both higher- and lower-level processing: for the higher level, they assume a feedforward origination passing through a higher nexus, which they call 'pathway convergence', resulting in feedback activation through disinhibition. For the lower level, the authors expect direct 'horizontal activation' between the area representing the inducer and the pathway representing the concurrent. Hence, their theory covers both **direct and indirect functional differences** between synaesthetes and non-synaesthetes depending on the type of synaesthesia.

Finally, the data of Ward et al. (2006) seem to support the account of **indirect functional differences**. Participants were asked to assign colours to musical tones of different pitch played with different instruments. As synaesthetes and non-synaesthetes showed similar patterns in their associations, the authors assumed the usage of the same heuristic for matching sounds to colours. Therefore, they find it more likely that synaesthesia stems from feedback from multi-modal regions that is simply used more excessively by synaesthetes.

1.6.3 Critical evaluation of theories

Contrary to all the studies introduced that show structural or functional distinctions of synaesthetes' brains, Hupé et al. (2011) did not find any difference in activation or structural connections of "colour areas" in grapheme-colour synaesthetes in their very carefully designed and controlled study using fMRI and voxel-based morphometry. They did however find increased white matter in the retrosplenial cortex (RSC) bilaterally, indicating enhanced structural connectivity in this area of the cingulate cortex. This area is sometimes associated with recall of episodic information, emotion processing, memory, relating objects to their contexts, switching from the external world to internal mentation, and is generally heavily interconnected with other parts of the posterior cingulate region. The authors conclude that – if there are functional differences in processing synaesthetic colours – they might not be strong or localised enough to show up in fMRI scans. The increase of white matter in the RSC might not directly explain synaesthetic colour perceptions, but may be an indicator of the diversity of synaesthesia, because this region is strongly interconnected locally as well as to distant regions. Paulesu et al. (1995) found brain areas in the frontal cortex activated in a PET study (positron emission tomography), which detects regional differences in blood flow, but no significant V4 activation in synaesthetic colour perception. When their participants heard words, the synaesthetes showed significantly greater activity than the non-synaesthetes in the prefrontal cortex, insula, and superior temporal gyrus. This finding supports the discovery of Hupé et al. (2011) stating that brain areas usually active for the perception of the concurrent do not seem to be activated in synaesthetic perceptions, but other brain areas are.

From the outcome of these two studies and the divergent results of those mentioned earlier, it can be concluded that the variety of models and proofs promote cautious interpretation of theoretical suggestions of the origins and manifestations of synaesthesia. Perhaps too simplistic a view is taken by researchers trying to isolate brain areas solely used for synaesthetic perceptions. Neurological manifestations might be as diverse and holistic as the phenomenon itself.

To sum up, multiple neurological theories about synaesthesia exist: some assume structural differences in the brain of synaesthetes, whereas others assume functional differences compared to non-synaesthetes' brains. These differences may involve the two brain areas dealing with inducer and concurrent only (direct), or an additional higher order nexus (indirect). Most theories on structural differences seem to favour direct and those on functional differences indirect pathways. Due to the wealth of contradictory results from multiple studies, it seems natural to believe that synaesthesia might neurologically express itself in different ways. This suggests that critical experiments have yet to be designed and performed to gain a clearer view of the neurological basis – or indeed bases – of synaesthesia.

1.7 (Re-)Defining synaesthesia

Behavioural and neurological tests like those described in Sections 1.4 and 1.6 help to prove the genuineness and existence of the phenomenon called synaesthesia. On many occasions, synaesthetes have not been taken seriously and have been accused of having a vivid imagination, of hallucinating, of being on drugs or simply of being crazy (Ramachandran & Hubbard, 2001b, 2003). Studies that employ scientific methods to dispute these misconceptions and that work towards a better understanding of the condition have been conducted, and debates are ongoing as to how these studies can be used to define synaesthesia. As research progresses and debates continue in many aspects of synaesthesia research (e.g. whether there is one correct neurological model of synaesthesia or whether several could jointly explain the phenomenon), the definition of synaesthesia has evolved and changed notably, retreating from classical definitions of consistency, spatial perception, homogeneity, and generalisability across types of synaesthesia. An interesting discussion took place about this very recently in the British Journal of Psychology, initiated by Simner (2012a), followed by commentaries from Eagleman (2012) and Cohen Kadosh & Terhune (2012), to which Simner responded (Simner, 2012b). Simner's points for discussion and rethinking are that synaesthesia is a merging of the senses; that synaesthetic concurrents are consistent over time; that there is a spatial mapping of synaesthetic concurrents, i.e. that they occur in the same position in space; and that synaesthesia is neurologically defined. These and related topics are discussed in this section.

The original phrasing of synaesthesia being a **merging of the senses** probably stems from the ancient Greek "syn" meaning "together" and "aisthesis" meaning "sensation". Although this literal translation is clearly a fitting description, the word might have been chosen to name the condition 'synaesthesia' prematurely, at a point when only certain aspects of the phenomenon had been described. In fact, the term 'ideaesthesia' was introduced by Nikolić (2009) with the intention to differentiate between sensorily induced (i.e. synaesthesia) and semantically induced (i.e. ideaesthesia) concurrents. However, this discrimination is not widely accepted yet and is not used in this thesis. Thus, the term 'synaesthesia' covers both meanings to date: sensorily and conceptually induced synaesthesia.

The test of **consistency** over time has been claimed to be the gold standard for verifying the condition in participants (Baron-Cohen et al., 1987, for the initial introduction of the test; Asher et al., 2006; Ward & Mattingley, 2006, for an overview). It is still in use and most certainly will be for several more years. However, Simner (2012a) disapproves of the consistency test and emphasises its risk of circularity: because synaesthetic perceptions are usually consistent over time, non-consistent participants are excluded from studies to avoid malingerers "faking" synaesthesia. Thus, true but inconsistent synaesthetes cannot be included in studies with the consequence that their manifestation of the condition cannot be included in the definition. Although Eagleman (2012) and Cohen Kadosh & Terhune (2012) agree, Cohen-Kadosh and Terhune warn against over-inclusive definitions nonetheless and suggest that researchers should behaviourally and neurologically test this manifestation of synaesthesia that has been excluded until now. It is a risk to generalise features of a subset of synaesthetes to all synaesthetes and a challenge to define where to draw the line between including and excluding criteria.

In fact, there is ongoing discussion of whether the bimodal categorisation of being synaesthetic or not should be replaced by a spectrum of being synaesthetic (Simner, 2011; Eagleman, 2012; Simner, 2012b). Eagleman suggested a **'synaesthesia spectrum'** in parallel to the autism spectrum which was introduced only after the recognition that "autism is not all-or-none" (Eagleman, 2012, p. 18). The idea of a synaesthesia spectrum finds support in several publications which highlight similarities between synaesthetes' and non-synaesthetes' colour associations to certain stimuli such as letters and digits (Simner et al., 2005; Rich et al., 2005), musical pitch (Ward et al., 2006) or vowel sounds (Chapter 3). Simner (2012b) seems particularly assertive in her publications that synaesthetes' and non-synaesthetes' responses lie on a behavioural and possibly also a neurological continuum.

Another traditionally bimodal categorisation of synaesthetes is to divide them into **projectors and associators** (Dixon et al., 2004). However, grouping participants into projectors and associators proves to be difficult because synaesthetes who see their concurrents projected into space, are still aware of the fact that they are not objects in the real world that others can see as well. Hence they technically are in front of their minds' eyes as well, which is used as an attribute of associators. Eagleman (2012) even noticed that it is easy to bias the response of the synaesthete with the phrasing of the question. For this reason, Skelton et al. (2009) introduced a more elaborate questionnaire with visual illustrations to help classify synaesthetes into projectors and associators, which reduced the number of unclassifiable synaesthetes drastically. Nonetheless, Eagleman (2012) also discussed whether a spectrum might be more appropriate to use than a bimodal categorisation. He refers to Rouw & Scholte (2007) who claimed to have found differences in structural connectivity between projectors and associators in their DTI study; but viewing their supplementary material, it is evident that their participants score along a spectrum rather than in a bimodal distribution.

Many theories have been created to try to neurologically explain the origins of synaesthesia (e.g. excess of white matter fibres (Rouw & Scholte, 2007); failure to prune abundant connections in infants (Maurer & Mondloch, 2005); disinhibition of existing pathways (Cohen Kadosh et al., 2009); see Section 1.6 and Bargary & Mitchell (2008) for discussion on structural and functional mechanisms). Instead of "choosing the right one", Simner (2012a) introduces the general term of 'hyper-association', suggesting that a biological inherited condition leads to this neurological hyper-association. These hyper-associations could take place between any brain regions; she does not specify whether these associations are of functional or structural nature. Simner (2012a) takes the risk of becoming over-inclusive in her **neurological re-definition** by giving examples such as being extraordinarily skilled at speaking languages as a "synaesthetic trait". Later, she mitigates against this general suggestion by pointing out that it is a question of debate whether these traits are "considered as types of synaesthesia in their own right, or whether they should simply be considered co-lateral features caused by a similar neurological root" (Simner, 2012b, p. 11). One of these co-lateral features might be that there are more artists amongst synaesthetes than amongst non-synaesthetes (see e.g. Rich et al., 2005; Ward et al., 2008). Eagleman (2012) states that there may be various different underlying neurological mechanisms causing the same outcome, i.e. synaesthetic perceptions (he compared it with deafness which has different causes but all in all the same outcome), whereas Cohen Kadosh & Terhune (2012) specified that different types of synaesthesia may stem from these different mechanisms. Yet, only once we have a better understanding of these mechanisms, can a precise neurological definition of synaesthesia be agreed upon. For now, "we may only be able to unite them [neurological classifications] by reference to their functional consequence, in which case, we may be flirting, once again, with behavioural classifications" (Simner, 2012b, p. 25). On these grounds, I too shall be flirting with behavioural experiments and classifications in this thesis.

There may nevertheless be one simple criterion which allows us to distinguish between synaesthetes and non-synaesthetes whatever the "strength" of their synaesthetic reactions on a spectrum. Simner noticed that "while non-synaesthetes are able to *generate* letter-colour combinations on demand [...] they have not previously been aware of these associations and do not entertain them in daily life at a conscious level" (Simner, 2012b, p. 26). Here she targets the criterion of **conscious awareness**. While non-synaesthetes do not consciously experience for example colour perceptions to sounds, synaesthetes are consciously aware of them, irrespective of the fact that both groups may lie on a continuum regarding the characteristics of their synaesthetic(-like) associations. To assess conscious awareness, a simple question such as "Have you ever (i.e., before this testing date) believed that letters had colours (or that words had tastes etc.)" (Simner, 2012b, p. 26) should make it possible to classify people into synaesthetes and others, as only synaesthetes would (truthfully) reply yes. This classification system of course relies heavily on the honesty of participants and therefore, further tests are advisable nonetheless.

In summary, this section has shown that the definition of synaesthesia is in a state of flux. Common characteristics of synaesthesia cannot be generalised to match all types of synaesthesia and maybe not even all synaesthetes within a type. Whereas synaesthetic perceptions are consistent over time for most synaesthetes, they are not consistent for all. Some synaesthetes might have different underlying neurological patterns than other synaesthetes. Many aspects of synaesthesia may be a merging of the senses, but not all inducers are sensory: they can be conceptual or semantic as well.

1.8 Motivation for studying vowel sound- and voice-induced synaesthesia

The previous sections have described the diversity of the phenomenon of synaesthesia. One aspect of this diversity is the multitude of types of synaesthesia. Most research to date focuses on common types such as grapheme-colour or spatial representations of sequences. Less common types are often ignored in research for reasons such as difficulties in finding participants, defining the type, finding the right methodologies for analysis, and so forth. Nevertheless, the diversity is tangible through synaesthetes' reports, and to phrase it in Eagleman's words, "we must treat the heterogeneity of the condition as an interesting clue rather than an inconvenience to be swept under the rug" (Eagleman, 2012, p. 18). This is why my thesis focuses on less common types of synaesthesia, using less common methods of analysis in synaesthesia research. With this, I hope to add another piece of the puzzle, contributing to the success of the interdisciplinary field of synaesthesia research.

I will be focusing on two different types of synaesthesia:

1. Vowel sound - colour synaesthesia: Inspecting colour associations with different vowel sounds by synaesthetes and non-synaesthetes.

2. Voice-induced synaesthesia: Exploring the relation of concurrents with different voice qualities, and whether voice-induced synaesthesia might enhance memory performance for voices.

For the first type, the aims are to put results in context with the types commonly named phoneme-colour and grapheme-colour synaesthesia, and to discover whether the use of fine grained acoustic analyses casts light on the phenomenon. For the second type, a more exploratory approach was taken. Two main experiments originated around this type: the first experiment explored synaesthetic associations by giving participants the opportunity to describe these associations verbally, and associate a given set of colours, textures, and descriptive words with the voices. The second experiment tested whether synaesthetic perceptions (with voices or otherwise) influenced participants' ability to memorise voices and identify them later. The aims of these two experiments were to reach a better understanding of voice-induced synaesthesia as it has not been systematically researched yet. In addition, they also seek to contribute to research on memory performance in synaesthetes, as well as testing an application of this type of synaesthesia.

The motivation for studying vowel sound - colour synaesthesia, especially in context with previous research, is described in Section 1.8.2; motivation and research questions relating to voice-induced synaesthesia are depicted in Section 1.8.3. Before going into detail about the specific kinds of synaesthesia, a general question is briefly addressed: why introduce new research methods and take a view from a different angle?

1.8.1 Why introduce new research methods?

Synaesthesia research is a field past its infancy, but still growing and developing. Hence, it is necessary to introduce new research methods and view the condition from different angles to widen its horizon. Synaesthesia research is by definition interdisciplinary: it includes not only neurological and genetic questions but also disciplines associated with the modalities of inducer and concurrent. So, in the case of music-colour synaesthesia, for example, the analysis of musical pitch, rhythm and timbre should be considered in order to define the inducer, and analysis of colours and visual perception should be considered in order to understand the concurrent. In the case of my research, it is important to analyse the acoustic and phonetic influence of the voice and speech sounds as an inducer on the concurrents. So instead of simply taking the voice as one 'unit', which influences synaesthetic perceptions somehow, it can be analysed in much more detail in order to reveal which aspects of the voice trigger and change synaesthetic perceptions. This knowledge is achieved by using research methods known to this specific area of research; acoustic phonetics in this case. Combining these with research methods from the field dealing with the concurrent, for example colour science, allows for even further detailed analysis. After this specialist analysis, findings can be related to established synaesthesia research. Only by using the specific knowledge of these fields and combining it with previous findings can the field of synaesthesia research grow.

1.8.2 Vowel sound - colour synaesthesia

The aim of this experiment was to investigate how synaesthetes and non-synaesthetes match vowel sounds to colours and grey shades. The experiment and its setup were motivated by three main issues: firstly, it has previously been found that colour associations to vowels, letters or words are most often induced by graphemes, i.e. orthography (Simner, 2007), and only for a few synaesthetes by phonemes, i.e. speech sounds (Day, 2005) (which in English often have a rather opaque correspondence to graphemes). By covering a large range of vowel sounds, including those which are not present in the participants' native language, it should be possible to analyse whether there are indeed two subgroups of grapheme-induced and speech sound-induced synaesthetes, or whether there is a different underlying mechanism, or some sort of overlap or combination (cf. Smith et al., 2011). Secondly, literature suggests that the colour which is associated to vowel sounds might be systematically influenced by vowel acoustics (Marks, 1975) or articulatory settings (Jakobson, 1962). Although both references go into great detail describing which acoustic characteristics or settings of articulators induce which colours, the authors do not support their ideas with empirical measurements. Thus, I introduce empirical data to qualify their claims. Finally, I am interested in finding similarities and differences between responses of synaesthetes and non-synaesthetes. Research on music-colour synaesthesia by Ward et al. (2006) and grapheme-colour synaesthesia by Simner et al. (2005) suggests that both groups may use similar mechanisms in their associations, but synaesthetes may be more consistent.

In summary, I examine vowel sound - colour associations to test whether vowel acoustics influence colour and grey shade associations in a systematic way for both synaesthetes and non-synaesthetes, and to make comparisons with previous research on grapheme- and phoneme-induced synaesthesia.

1.8.3 Voice-induced synaesthesia

Voice-induced synaesthesia is anecdotally reported, and listed in the synaesthesia questionnaire of Jamie Ward and Julia Simner (https://www.survey.bris.ac.uk/sussex/syn), which they use to assess people's synaesthesia in their data pool of synaesthetes. But this variant has received very little attention from researchers to date. The case of one synaesthete whose perception of familiar voices induces gustatory experiences is described by Richer et al. (2011), but the assessment stays at a qualitative descriptive level. Another study of voice-induced synaesthesia includes quantitative analysis and the comparison with control participants, but remains a single case study (Fernay et al., 2012). My aim is to introduce systematic research on this type of inducer. The first experiment on this topic explores the relationship of concurrents to voice qualities; the second experiment analyses the influence of voice-induced synaesthesia on memorising voices. By introducing both qualitative and quantitative research on voice-induced synaesthesia, I hope to add another piece of knowledge to the riddle of synaesthesia.

Synaesthetic associations with the voice

The voice is a complex inducer and it is unclear which aspects of the voice trigger synaesthetic perceptions. An exploratory experiment analysing colour and texture associations with, and verbal descriptions of, different voice qualities was designed to address this topic. To assess which aspects of the voice function as an inducer, various acoustic analyses were carried out that unravelled distinct voice criteria, and were compared with the different concurrents.

Because no previous research has looked at this type of synaesthesia, motives for pursuing this stem from related research or open questions triggered by pure curiosity. The first goal is to explore which concurrents this type of synaesthesia induces. Considering the complexity of the stimulus *voice*, a variety of concurrents are expected to occur. Second, certain parallels can be drawn between the sound of a voice and music. The pitch of the voice, for example, can be compared to musical pitch; the sound of a voice can be compared to the timbre of an instrument. For this reason, I consulted publications in musical synaesthesia, e.g. Ward et al. (2006) and de Thornley Head (2006), to compare their results with mine, seeking similar regularities in synaesthetic associations to different sounding voices which may be detected on an acoustic level.

Memorising and identifying voices in a voice parade

In this study, a voice parade, or voice line-up, was conducted. Similar to the procedure of a line-up as a means of identification of a perpetrator by an eyewitness, a voice line-up can be conducted. Instead of having a line of people standing in front of the eyewitness, a "line" of voices is played to an earwitness. Why is this curious aspect of forensic casework introduced to synaesthesia research? Previous research on the memory of synaesthetes has revealed memory advantages in synaesthetes compared to non-synaesthetes, especially in areas involving their kind of synaesthesia (e.g. Yaro & Ward, 2007; Rothen & Meier, 2009; Luria, 1968; Smilek et al., 2002). Number-colour synaesthetes, for example, are on average better at memorising and reproducing number matrices than controls. Hence, I would like to test whether the same enhanced performance is present in voice-induced synaesthesia, i.e. whether people experiencing synaesthetic perceptions induced by voices are better at remembering voices. Some studies debate whether there is a general memory advantage in synaesthetes or not (Edquist et al., 2006; van Campen, 2009; Rothen & Meier, 2010), or whether synaesthesia can indeed hinder performance (Rehme & Zedler, 2010). To examine this dilemma further, three participant groups were invited to take part in the voice parade experiment: voice-induced synaesthetes, synaesthetes with inducers other than voices, and non-synaesthetes. Results of this study could be beneficial not only for gaining knowledge about performance in the rare case of synaesthetes being earwitnesses, but also for judging performance in everyday tasks like recognising a voice on the telephone or radio.

1.9 Outline of the thesis

Chapter 2 introduces the reader to **conceptual preliminaries** in the fields of phonetics, colour science and textures that are important in order to understand the studies described in the following chapters.

In Chapter 3, a thorough examination of **vowel sound - colour synaesthesia** is presented. For this, an experiment on colour and grey shade associations with vowel sounds by synaesthetes and non-synaesthetes is described. The experiment analyses to what degree vowel acoustics influence colour and grey shade associations, which are carefully measured by colorimetric standards. Consistency of colour associations over time and group differences and similarities are revealed. Finally, results are compared with other studies on vowel sound - colour associations, commonalities between synaesthetes and non-synaesthetes and possible reasons for this are discussed, as are influences of accents, music and neurological aspects.

A brief excursion is taken on **texture perception** in Chapter 4. A small additional experiment was conducted to facilitate the analysis of texture choices of participants of the experiment described in Chapter 5. A set of textures was used as a visual response display. To allow quantitative analysis of these textures, a different group of participants rated them on semantic differentials, i.e. bidirectional descriptions. These ratings are presented in this chapter.

Chapter 5 is about research on **voice-induced colour and texture associations** in synaesthetes, phoneticians and controls, and voice descriptions by these participant groups. Qualitative data analyses of the verbal descriptions and individual differences are presented here. Further, detailed acoustic analyses of the voices are also obtained. These are related to colour and texture associations with the aim of finding which aspects of a voice trigger which kinds of synaesthetic perceptions. Results of a retest inform us about the consistency in colour and texture associations and the use of descriptive semantic differentials. Finally, results are compared with related research and their implications are discussed.

In Chapter 6, a practical implication of synaesthesia is presented, namely how participants perform in **voice identification** tasks. In the first part, a voice comparison task is analysed, testing synaesthetes, phoneticians and controls in their voice identification performance when comparing different voice qualities. In the second part, two voice line-ups are introduced. Voice-induced synaesthetes, other synaesthetes and non-synaesthetes are tested on how they perform when identifying a previously heard voice, in whisper or modal phonation, from a set of voices. Results are interpreted with reference to memory performance in synaesthetes and voice line-up regularities and identification performance in forensic case work.

Chapter 7 **"General discussion"** summarises the findings and puts them in the broader context of synaesthesia research and cross-modal perceptions. The co-existence of types of synaesthesia and its possible redefinition is discussed. Challenges and limitations of both my studies and synaesthesia research in general are debated.

There is a **glossary** at the end, briefly explaining terms and abbreviations that are not expected to be common knowledge.

2 Conceptual preliminaries

The aim of this chapter is to introduce the key terminology from the fields of phonetics, colour science and texture analysis that is needed to understand the content, analyses and results of the studies described in the following chapters.

2.1 Linguistic background

A large part of this section deals with the terminology and concepts of phonetics. In order to fully understand the role of potential inducers, terms such as speech sounds, phonemes, phones, graphemes, and voice qualities are defined. The concepts of the vowel quadrilateral, formants and how they work in combination or in comparison, are additionally explained to facilitate comprehensibility of the analyses that were conducted to explore the relations between inducers and concurrents.

2.1.1 Graphemes

The Concise Oxford Dictionary of Linguistics defines a grapheme as a "character in writing, considered as an abstract or invariant unit which has varying realizations. For example, the grapheme <A> subsumes the variants or 'allographs' 'A' (Roman capital), 'a' (Roman minuscule), and so on" (Matthews, 1997). In the case of English, the units or characters are letters of the alphabet, numerical digits and punctuation marks. The essay of Henderson (1985) on the usage of the term extrapolates three main groups of definitions: "(i) a letter or cluster of letters that can usefully be regarded as corresponding with a phoneme, (ii) the minimal distinctive unit of a writing system, and (iii) an abstract letter identity in all its physical realizations" (Henderson, 1985, p. 146). Definition (i) means that clusters like <sh> as in *shin* are also regarded as graphemes, even though they consist of two letters (called a digraph), because they refer to one speech sound. Definition (ii) is in agreement with the Oxford Dictionaries (2010) definition which describes a grapheme as "the smallest meaningful contrastive unit in a writing system" that can change meaning. Definition (iii) appears to refer to the visual configuration of a grapheme, that is, its abstract visual appearance, where <a>, <A> and <ɑ> all represent the same grapheme. Although the author emphasises that (ii) is perhaps the most accurate definition, publications on grapheme-colour synaesthesia often do not refer to this definition directly. In fact, it is often not clear which definition they use or whether they use a single definition exclusively. In most cases in synaesthesia research, the term *grapheme* appears to be interchangeable with the term *letter*. In this thesis, the term *grapheme* is used to refer to the letters of the alphabet in lower case standard font, unless otherwise specified. When citing literature on grapheme-colour synaesthesia, it may be unclear how the term is defined.

Following convention, graphemes will be presented between arrow brackets: <>.

2.1.2 Speech sounds

In research into synaesthesia, it is often unclear whether reference is being made to *phonemes*, i.e. abstract linguistic units, or *phones*, i.e. the acoustic realisations of these units. For this reason, the current research uses the term *speech sounds*, or more precisely *vowel sounds* (as the experiment presented in Chapter 3 is on vowels only), as an umbrella term. Nonetheless, it is important to know how to discriminate between phonemes and phones; this distinction, as it relates to synaesthetic inducers, is addressed in the general discussion in Chapter 7.

Phonemes "The phoneme is the smallest unit of sound which can differentiate one word from another: in other words, phonemes make lexical distinctions" (Ogden, 2009, p. 4). For example, *tree* and *free* differ in terms of /t/ and /f/, two distinct phonemes in English. Phonemes are not necessarily a single sound; affricates as in the last phoneme of *edge* /ɛd͡ʒ/ are also called phonemes, but here consist of a plosive and a fricated part. Phonemes are phonological, i.e. abstract units, because they relate to linguistic structure and the organisation of a given language (Ogden, 2009). A phonemic transcription records "all and only the variations between sounds that cause a difference in meaning" (Ladefoged, 1993, p. 26) in a given language and is placed between oblique lines: //. Although what is transcribed as /a/ in French may sound slightly different from what is transcribed as /a/ in English, the phonemic transcription is the same, as they refer to the speech sound as it is used in this language. Dictionaries, as well as the studies I am referring to in this thesis, use phonemic transcriptions as opposed to phonetic transcriptions (see Subsection "Phones" below). This can sometimes cause difficulties or inaccuracies when comparing speech sounds across languages.

Phonemes can be realised as allophones. Two allophones of /p/ for example are given in *pin* [pʰɪn] and *spin* [spɪn], as they are pronounced slightly differently, either aspirated (with a puff of air after the burst) or not. Allophones do not change the meaning of the word when interchanged, but only make it sound as if pronounced by a non-native speaker of English.

Phones A phone is a speech segment representing "the smallest discrete segment of sound in a stream of speech" (Oxford Dictionaries, 2010), or a "speech sound which is identified as the realization of a single phoneme" (Matthews, 1997). It has distinct physical and perceptual properties. Phones can be used to transcribe speech sounds in a language unknown to the transcriber, i.e. when the transcriber is unaware of the phonemic inventory of the language. They are also used to transcribe allophonic differences, i.e. the different realisation of phonemes by different speakers or in the context of co-articulation, where different surrounding speech sounds may influence the pronunciation. Phonetic transcriptions are

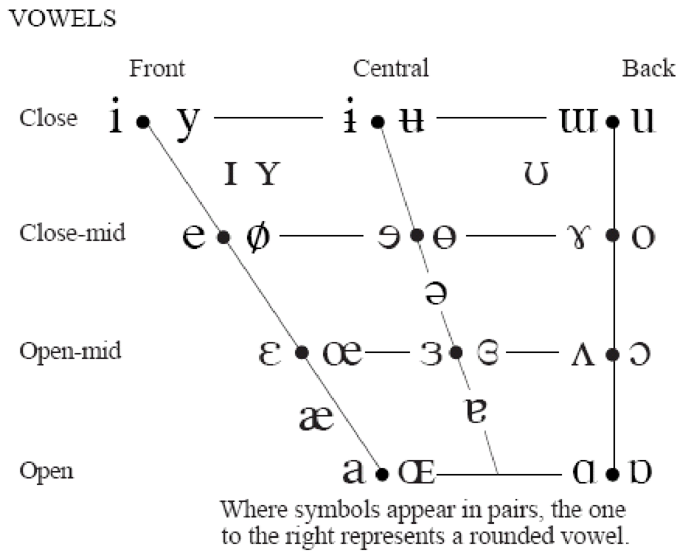


Figure 4: The vowel chart of the International Phonetic Alphabet (2005). Source: http://www.langsci.ucl.ac.uk/ipa/vowels.html, visited 18/12/2012.

given between square brackets: [].

2.1.3 The vowel quadrilateral

The vowel quadrilateral that is part of the International Phonetic Alphabet (International Phonetic Association, 1999) is displayed in Figure 4. It shows phonetic symbols as they are used to transcribe vowel sounds in an articulatory setting. The articulatory settings are indicated by the words above and to the left of the quadrilateral. The dimension close to open refers to the setting of the mouth, tongue and jaw. The dimension front to back refers to the setting of the tongue body. The vowel quadrilateral is a schematic cross-section of the oral vowel space, with the speaker facing left. Here, the [i] is close to the lips and the [u] close to the pharynx. (Imagine superimposing the vowel quadrilateral onto the mouth in Figure 7 on page 48.) The third dimension represented is that of lip rounding. Each point in the quadrilateral shows two symbols, for example [i] and [y] in the top left. The left of each of these pairs represents the vowel sound with the lips spread or unrounded. The right of each of these pairs represents the vowel sound with rounded lips.

Examples of the pronunciation of key vowels from the quadrilateral are as follows. To produce the vowel [i], an exaggerated version of the vowel in *bee*, the tongue is tense and bunched up in the front of the mouth as high and as far forward as possible without causing friction (Catford, 1994) and the lips are spread. To produce the vowel [u], an exaggerated version of the vowel in *moon*, the tongue is as high and as far back as possible without causing friction and the lips are rounded. To produce the vowel [ɑ], an exaggerated version of the vowel in *palm*, the tongue is as low and as far back as possible without causing friction and the lips are in a neutral position.

The points in the quadrilateral on which the transcriptions are placed are reference points only, so the quadrilateral needs to be regarded as an abstraction. The reference vowels at the periphery of the quadrilateral are called *cardinal vowels* (CVs). The eight primary cardinal vowels are [i e ɛ a ɑ ɔ o u]; [y ø œ ɶ ɒ ʌ ɤ ɯ], with opposite lip-rounding status from the primaries, are the secondary cardinal vowels. CVs are not phonemes but are "a descriptive device, not something that occurs in a language" (Abercrombie, 1967, p. 154) and "represent possibilities of the human vocal tract" (Ogden, 2009, p. 57).

2.1.4 Formants

Formants are created by the resonances of the vocal tract (Clark et al., 2007). This means that they change as the configurations of the vocal tract changes. As the lips, jaw and tongue move during the production of different speech sounds, this modifies the acoustic energy produced by the source of sound (usually the larynx). Formants are defined as spectral peaks of intensity at different frequencies (usually measured in Hz) in the frequency spectrum of the sound. A vowel sound contains several formants. They are numbered upwards from the lowest in frequency. So the first formant, or F1, is the lowest spectral peak of intensity, F2 the second lowest and so forth. F1 and F2 are the main formants defining the vowel quality, although higher formants influence the quality of the vowels too. Borden et al. (2003) describe the relation between vowel formants and articulatory settings:

> The first formant [...] is most responsive to changes in mouth opening. Speech sounds requiring small mouth openings have low-frequency first formants. Conversely, open-mouth sounds are characterized by relatively high-frequency first formants. (The first formant at its highest frequency value is still the lowest resonance of the tract.) The second formant [...] is most responsive to changes in the size of the oral cavity. Tongue backing or lip rounding may lower the frequency of this formant, as these constrictions occur in areas of high velocity, but any tongue or jaw activity that narrows the region in the oral cavity where the pressure is relatively high causes an increase in the frequency of the second formant. (Borden et al., 2003, p. 87)

Representative formant values for the primary cardinal vowels are given in Catford (1994). A diagram using these formant values is displayed in Figure 5.

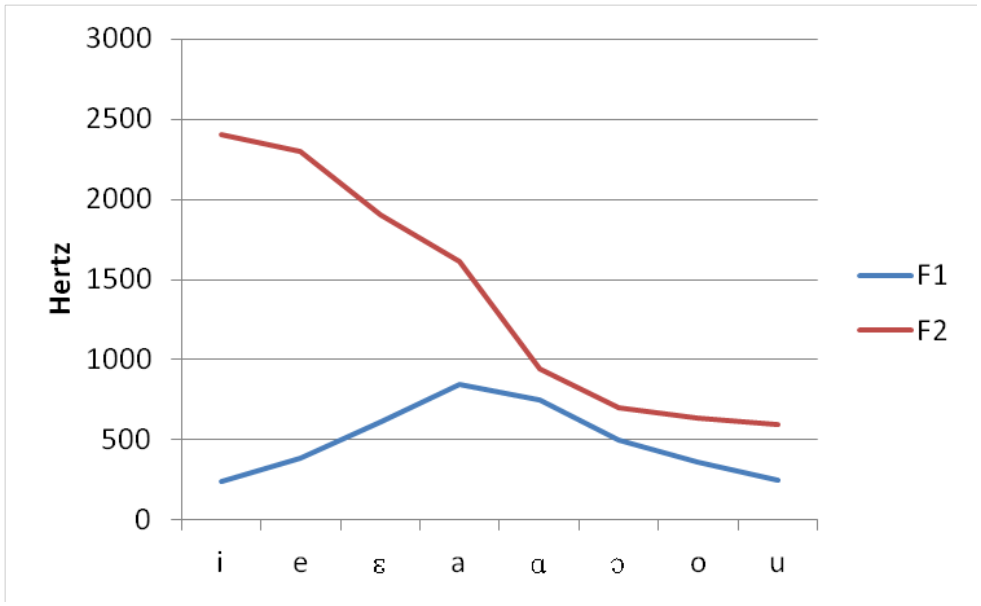


Figure 5: Representative formant values of F1 and F2 for the primary cardinal vowels for a male speaker.

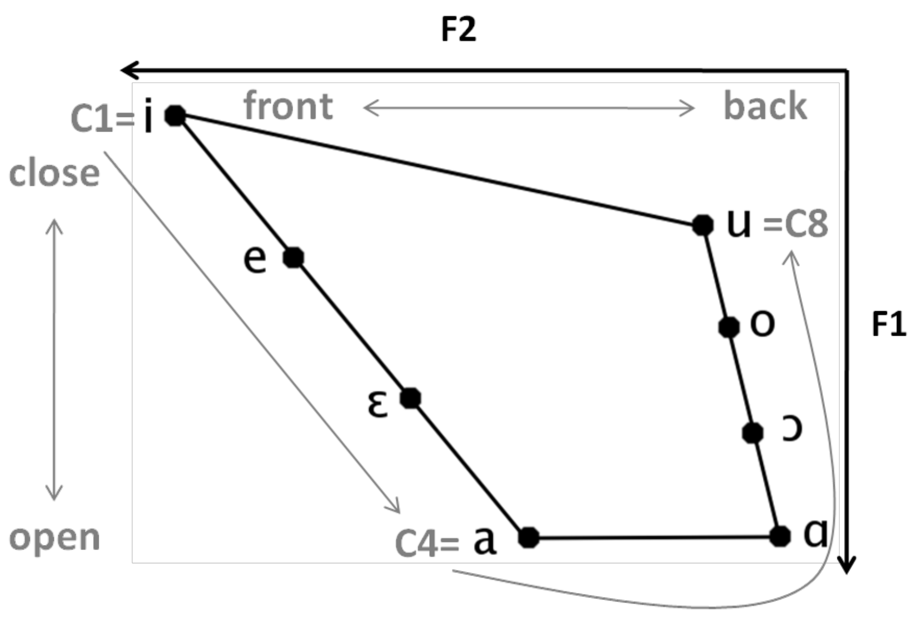


Figure 6: Schematic vowel quadrilateral of the eight primary cardinal vowels C1 to C8, positioned in the frequency space of the first two formants F1 and F2. Articulatory settings are marked in grey, open and close referring to the opening of the mouth, front and back referring to tongue position.

Phonetic transcription	Cardinal vowel	F1	F2	Articulatory setting
[i]	C1	240	2400	close front
[e]	C2	390	2300	close-mid front
[ɛ]	C3	610	1900	open-mid front
[a]	C4	850	1610	open front
[ɑ]	C5	750	940	open back
[ɔ]	C6	500	700	open-mid back
[o]	C7	360	640	close-mid back
[u]	C8	250	595	close back

Table 2: List of vowels: their transcriptions, names of cardinal vowels (CVs) as their reference points, representative formant values for F1 and F2 in Hz, and articulatory settings.

2.1.5 Putting it all together: The cardinal vowels from articulatory and acoustic perspectives

Figure 6 was created to show the relationship between transcription symbols, articulatory settings, formant characteristics and the numbering of the cardinal vowels. In this thesis, vowels are referred to by all of these means, but transcription symbols will be used most often. The transcriptions of the primary cardinal vowels are shown as symbols in Figure 6. The primary cardinal vowel [i] is referred to as C1 (displayed in grey), [e] is referred to as C2, following around the quadrilateral anti-clockwise, up to [u] as C8. Neighbouring CVs are auditorily equidistant. Articulatory settings are indicated by the grey arrows ranging from close to open and front to back. Finally, the acoustic definitions of the vowel formants are indicated by the coordinate system with its origin in the top right corner. F1 increases with a more open jaw setting; F2 increases with a more fronted setting of the tongue body. As depicted in Figure 6, for the vowel [i] as in an exaggerated version of *bee*, the phonetic transcription is [i], it is called cardinal vowel C1, it is a close front vowel, and it has a low F1 and a high F2. Both articulatory as well as acoustic correlates are kept abstract in this schema as small variations occur between individuals. C1-C4 are termed front vowels, C4-C5 open vowels, and C5-C8 back vowels, according to the articulatory settings of the mouth and tongue. The correspondence between vowel sound transcriptions, CVs, formant values and articulatory settings is additionally given in Table 2. This list displays the primary cardinal vowels, that is C1-C5 with spread lips and C6-C8 with rounded lips. Auditory examples of C1-C8 can be found on the supplementary CD under "Ch3_Vowel_sound-colour_associations".

2.1.6 Monophthongs vs. diphthongs

Monophthongs are vowel sounds which exhibit a steady state throughout articulation, for example in the words *mask* (/ɑ/) or *feet* (/i/). They are characterised by little to no

Letter	SSBE	ScE	German	Polish
<a>	/eɪ/	/e/	/a/	/a/
<e>	/i/	/i/	/e/	/ɛ/
<i>	/aɪ/	/ae/	/i/	/i/
<o>	/əʊ/	/o/	/o/	/ɔ/
<u>	/ju/	/jʉ/	/u/	/u/

Table 3: Pronunciation of vowel-graphemes as expected for Standard Southern British English (SSBE), Scottish English (ScE), German and Polish.

articulatory movement, and the formants are relatively stable throughout. "Diphthongs are monosyllabic vowels which have two discernibly different points, one at the start and one at the end." (Ogden, 2009, p. 64) They are characterised by articulatory movement and, in consequence, formant movement and a change of auditory quality during the course of producing one vowel, for example in the words *play* (/e͡ɪ/) or *mouse* (/a͡ʊ/). There is articulatory movement during the vowel production of a diphthong.

2.1.7 Relation between graphemes and vowel sounds

The organisation of speech into units of sound is distinct from the organisation of writing into letters. The English language exhibits notoriously opaque correspondences between phonemes and graphemes. A single vowel phoneme in Standard Southern British English, /i/ for example, can be represented by the graphemes <e> (*we*), <ea> (*bead*), <ee> (*seed*), <e.e> (*effete*), <ei> (*ceiling*), <ey> (*key*), <ie> (*chief*), or <y> (*happy*). Conversely, a single grapheme, e.g. <o>, can be associated with the vowel phonemes /ɒ/ (*long*), /u/ (*do*), /ʌ/ (*won*), /ə/ (*observe*), /əʊ/(*so*, or <o> produced in isolation), /ɪ/ (*women*), or /ʊ/ (*woman*). An extensive but not necessarily comprehensive list of correspondences of the phonemes of Standard Southern British English (SSBE) and graphemic representations of them in words is given in Appendix A (Francis Nolan, personal communication). However, correspondences between graphemes and vowel sounds may change depending on the accent of English. While the vowel in *way* is pronounced [eɪ] in SSBE, it is pronounced [e] in Scottish English; the vowel in *nose* is [əʊ] for SSBE and [o] for Scottish English. An extensive but again, not necessarily comprehensive list of different pronunciations of vowels in Scottish English, RP, Northern Irish, Northern English and General American is given in Appendix B, a list created by Mike MacMahon (2011).

While the correspondence of vowel sounds and graphemes is somewhat clearer in other languages such as German or Spanish, there is no language with a unique one-to-one relationship between graphemes and phonemes in continuous speech. The pronunciation of vowel graphemes in isolation in the case of SSBE, Scottish English, German and Polish[2]

[2]Note that the transcription for <e> in Polish is usually given as /e/ and for <o> as /o/, but are actually much closer to the cardinal vowels [ɛ] and [ɔ], as indicated in figures by Warth (2007) and

is given in Table 3, as these languages are discussed in Chapter 3.

2.1.8 Voice quality

Abercrombie (1967) describes voice quality as "those characteristics which are present more or less all the time a person is talking: it is a quasi-permanent quality running through all the sound that issues from his mouth" (Abercrombie, 1967, p. 91). The term 'quality' is not used as a value judgement, i.e. whether a voice is good or bad, but rather as a characterisation of the voice. Laver (1980) also described voice quality as the "characteristic auditory colouring of an individual speaker's voice"[3] (Laver, 1980, p.1) and "as a major vehicle of information about physical, psychological and social characteristics of the speaker" (Laver, 1980, p.2). A speaker's voice quality informs us not only about their anatomy and physiology, but also about their organic and acquired settings and how they make use of their speech organs.

Figure 7 depicts a vocal tract. Whereas the larynx, glottis and vocal folds play a role in phonatory or laryngeal settings, everything above the larynx is termed supralaryngeal. Voice quality refers to "the overall auditory quality which characterizes an individual's speech, including supralaryngeal and phonatory features" (Stuart-Smith, 1999, p. 211). The first features mentioned here, **supralaryngeal** ones, refer to settings of the articulators in the vocal tract. These include the settings of the lips, for example whether a speaker tends to protrude them habitually; the settings of the tongue, for example whether a speaker tends to speak with the tongue tip reaching between their front teeth; the settings of the jaw, for example whether a speaker tends to not open their mouth very much during speech; the settings of the larynx, for example whether it is raised or lowered; and the settings of the velum, for example whether a speaker releases a lot of air through the nose while speaking. **Phonatory** features refer to **laryngeal** settings, i.e. the activity of the larynx, for example whether and how the vocal folds vibrate to produce voicing. These include phonation types such as falsetto, whisper, creak and modal voice. Ten examples of different voice qualities, as they are used in the experiment described in Chapter 5, are briefly outlined below. The first six voice qualities are mainly affected by phonatory settings, the last four mainly by supralaryngeal settings (cf. Laver, 1980).

Modal Modal describes a voice quality for which all settings are neutral and regular vocal fold vibration takes place, so that no friction or other irregularities are audible. No specific feature is explicitly changed or added and the chest register is used.

Falsetto Falsetto is characterised by a high fundamental frequency and is often referred to as 'head voice'. "The 'thin' auditory quality which is characteristic of falsetto derives

Wrembel (2007).

[3]Note already here the cross-modal use of 'colouring', which is further investigated in my work.

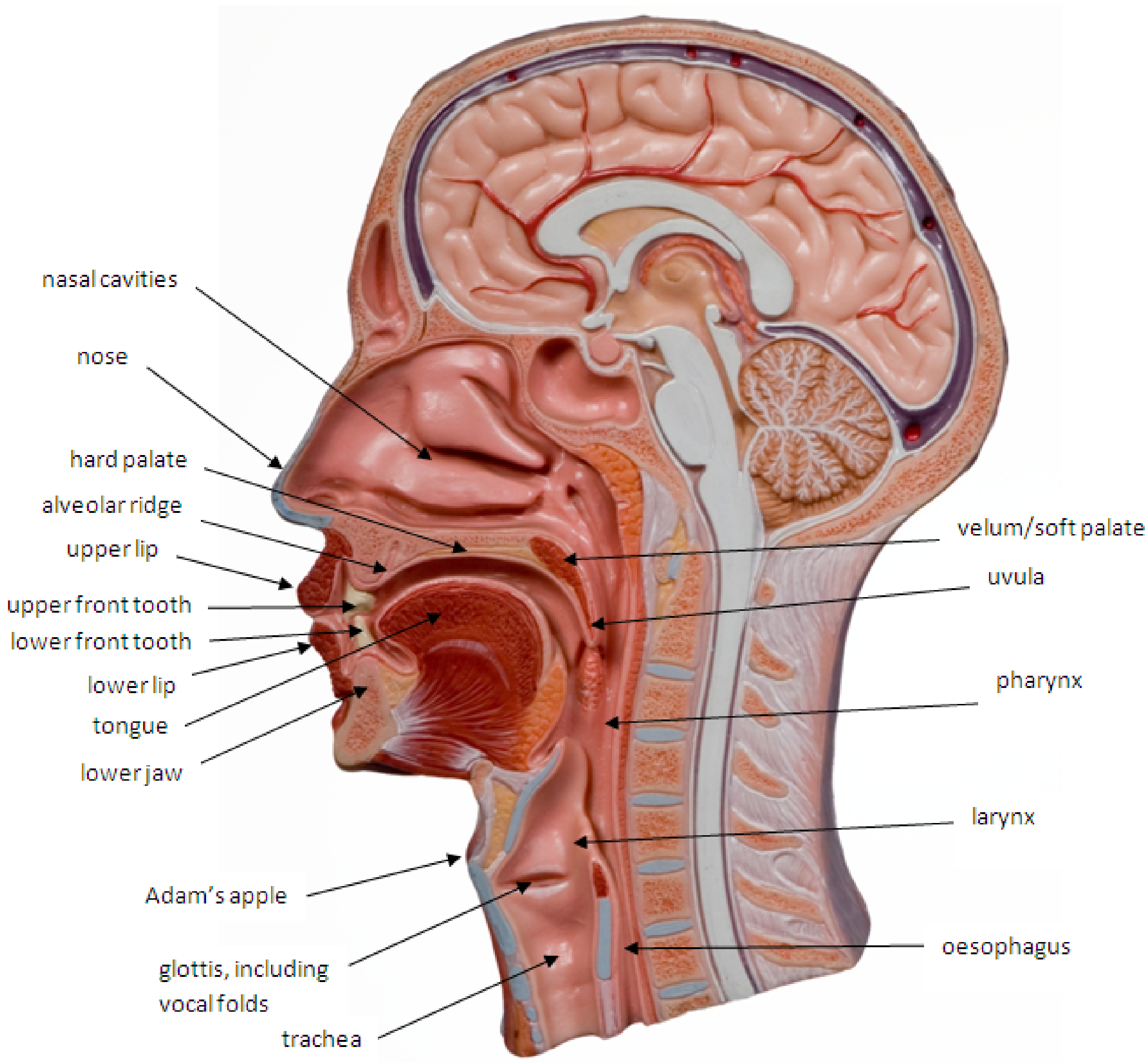


Figure 7: The vocal tract in sagittal view.

from the interaction of high fundamental frequency and the mode of vibration of the vocal folds" (Laver, 1991, p. 200f), as they are stretched and therefore thinner.

Whisper Whisper describes a voice quality in which speech is produced without any voicing. There is a sense of constriction and friction in the voice, as turbulence occurs when the airflow passes through the small opening of the glottis. It is of low volume.

Creak Creak emerges from slow and somewhat irregular vocal fold vibration (small periodic bursts), resulting in a very low fundamental frequency.

Harsh "Harshness results from a high degree of aperiodicity or frequency 'jitter' in the glottal wave" (Nolan, 1983, p. 141). It is accompanied by laryngeal hyper-tension and a perception of 'roughness' and increased volume and intensity.

Breathy A breathy voice is characterised by an inefficiently large escape of air while speaking, which is due to a lack of tension in the larynx and an incomplete closure of the vocal folds when vibrating. The perception of breathy voice is soft and of lower intensity and volume than modal.

Nasal A nasal voice quality is given when the velum is lowered throughout speech to allow parts of the air stream to be released through the nose. (The velum is always lowered when breathing through the nose.)

Denasal A denasal voice quality is given when the velum is raised throughout speech to stop the nasal airflow. Denasal speech can for example be found in people with a blocked nose.

Raised larynx Muscles around the larynx elevate it, which results in a higher fundamental frequency if no compensatory adjustment is made and a somewhat strained auditory impression.

Lowered larynx Muscles around the larynx lower it, which results in a lower fundamental frequency if no compensatory adjustment is made and a rather relaxed auditory impression.

2.2 Colour

There are a variety of different colour systems. This section outlines the ones most commonly used in research on synaesthesia. The analyses in my studies are based on the CIELUV colour space (Section 2.2.3).

2.2.1 Berlin and Kay typology

In the book "Basic Color Terms: Their Universality and Evolution", first published in 1969, Berlin & Kay (1999) established a typology of colour terms which they suggested was universal. Although the authors revisited their system after further research in more cultures (Kay et al., 2009), most studies still refer to the traditional typology. Since all studies cited in this work refer to the traditional typology, this is the one outlined here. This system suggests that the kinds of colour terms used in a culture can be predicted by the number of colour terms present in that culture. According to this rule, there are between two and 11 basic colour terms in every language, depending on which stage a culture falls into. As a language evolves, it acquires additional basic colour terms in the order as follows:

Stage I: Dark-cool and light-warm (this covers a larger set of colours than English "black" and "white")

Stage II: Red

Stage III: Either green or yellow

Stage IV: Both green and yellow

Stage V: Blue

Stage VI: Brown

Stage VII: Purple, pink, orange, and/or grey

It is assumed that languages at a higher stage include the colour terms of the lower stages as well; for example a language at stage IV should have black and white (or similar), red, green and yellow as basic colour terms. For English and many other languages of industrialised countries, the basic colour terms are white, black, red, green, yellow, blue, brown, orange, pink, grey and purple.

2.2.2 RGB colour space

RGB stands for red, green and blue. The RGB colour space is an additive colour model of light defined by the three chromaticities of the red, green and blue primaries. This is the basis for displaying colours on a screen (e.g. television or computer). In the models usually used for computer screens, the red, green and blue units range from 0-255. White is achieved by combining the full light of all three primaries, whereas black is defined by the absence of light. A model of the RGB colour space is presented in Figure 8, with the primary axes of red, green and blue indicated in bold. Every point within this cube is

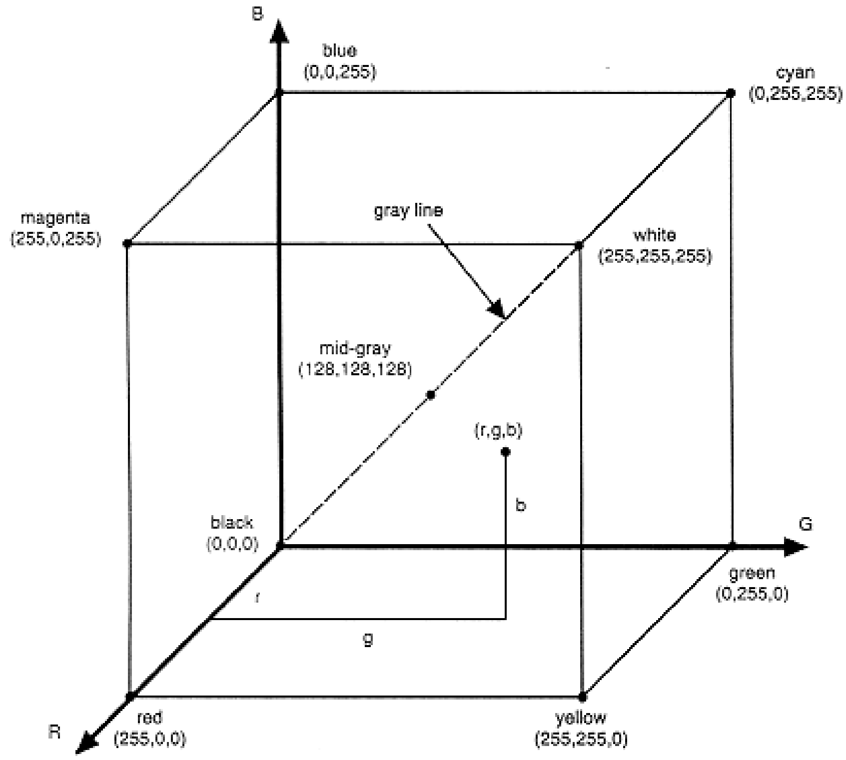


Figure 8: RGB colour space. Source: http://www.viz.tamu.edu/faculty/parke/ends375f03/notes/sec1_4.html, visited 01/09/2012.

defined by a combination of three numbers (0-255 for red, green and blue), so the number (255,0,0) indicates red and (128,128,128) mid-grey. Grey scale values are achieved by keeping all three numbers identical. The RGB space is not perceptually uniform, i.e. the change from (20, 30, 40) to (25, 35, 45) may be perceived as more or less strong than the change from (240, 240, 200) to (245, 245, 205) (Ford & Roberts, 1998). The RGB system is a device-dependent colour model because spectral characteristics of the phosphors differ between products such as different computer screens. Also the gamut, i.e. the size of the colour space, varies depending on the purity of the light source in the screen (Malacara, 2011). Hence, a colour with a specific RGB value may look different on different screens.

2.2.3 CIELUV colour space

CIE stands for *Commission Internationale de l´Eclairage*, or International Commission on Illumination. The commission publishes standards relating to the science and art of light and lighting, colour and vision, photobiology and image technology (http://www.cie.co.at). CIE coordinates use a colour-opponent colour space. CIE 1976 (L* u* v*), abbreviated CIELUV, coordinates are used for self-luminous colours such as those displayed on a computer screen. The CIELUV colour space attempts perceptual uniformity. Perceptual uniformity is given when a change of a colour value produces a change of about the same visual importance, regardless of where in colour space the change occurs. There are three

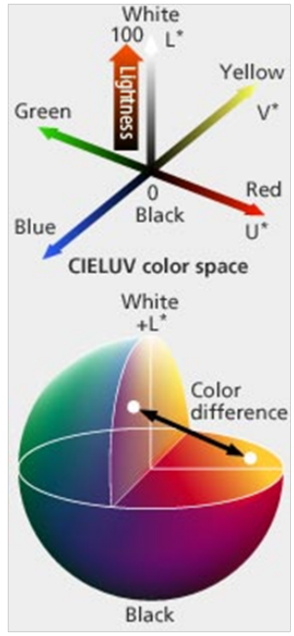


Figure 9: CIELUV colour space. Source: http://www.nikon.com/about/feelnikon/light/chap03/sec01.htm, visited 11/09/2012.

values: L* for lightness or luminance, u* which approximates the redness-greenness axis, and v* which approximates the yellowness-blueness axis (Fairchild, 1998). An example of the CIELUV colour space is displayed in Figure 9. L* values usually lie between 0-100, when the reference point is white (Anya Hulbert, personal communication). Measures, however, may lie outside this area, when another colour is chosen as reference point. The visual displays in my experiments used a grey background, so that L* values fell beyond the standard values. u* and v* values in standard images range between +/-100, but can exceed these values in both directions.

The CIELUV colour space is intended to cover the whole colour space perceivable to the human eye, whereas the RGB space only covers a subsection of it.

2.2.4 Munsell

Munsell codes refer to the colour space introduced by Albert Munsell (1858-1918), who was the first to separate *hue*, *value* (lightness), and *chroma* (saturation) into perceptually uniform and independent dimensions in the early 20th century. He introduced a three-dimensional space of colours for illustrative purposes. Munsell's colour system is based on visual perception, as his model was created by extensive empirical testing. It is therefore a perceptual colour space rather than a physically defined space.

The Munsell colour tree, or Munsell colour system, as shown in Figure 10, is based on five principal colours, the primaries yellow, red, green, blue and purple, with complementary colours on the opposite side of the tree. The units for the *hues* are as follows (Anderson Feisner, 2001): every hue receives the number 5 plus the initial letters. Thus,

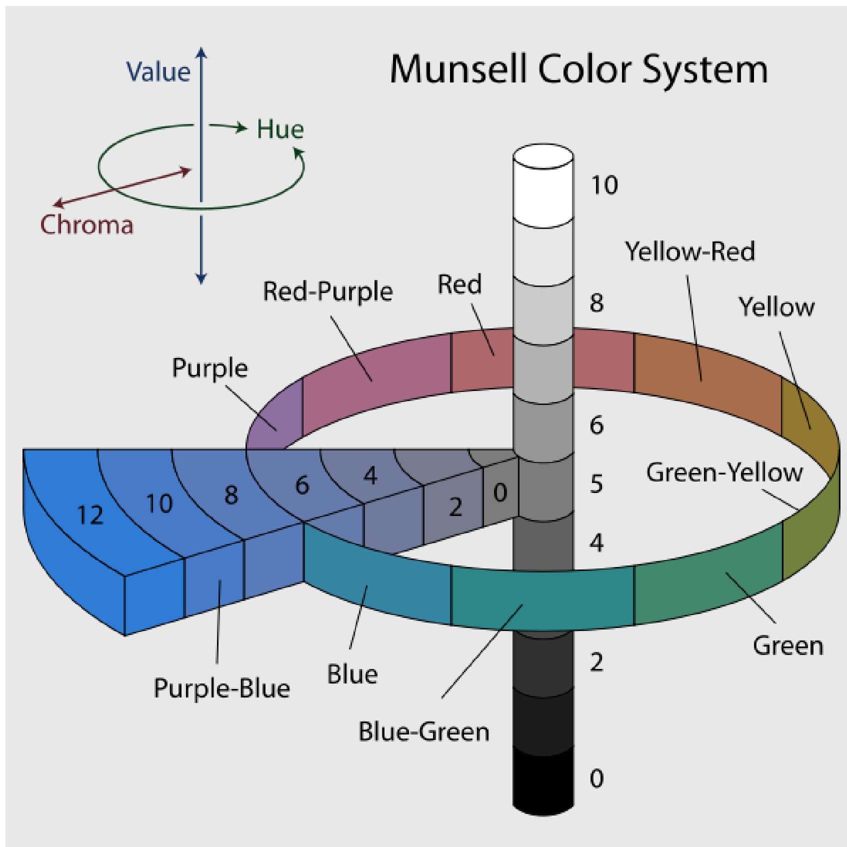


Figure 10: Munsell colour system. Source: http://en.wikipedia.org/wiki/File:Munsell-system.svg, visited 01/09/2012.

5R stands for red and 5YR for yellow-red, what is generally referred to as orange. Lightness, also referred to as the *value*, is arranged along the tree trunk from dark (0) at the bottom to light (10) at the top. *Hue* changes as you go around the tree. Branches growing from the trunk outwards measure the *saturation/chroma,* or purity of a hue, with achromatic hues in the centre (0) and most saturated or purer hues at the outside edges (12). The *saturation,* or *chroma,* is sometimes referred to as the "brightness" or vividness of a colour, not to be mistaken with the lightness which describes the value dimension from black to white.

2.2.5 Luminance

Luminance is defined as "the intensity of light emitted from a surface per unit area in a given direction" (Oxford Dictionaries, 2010). A light grey therefore has a higher luminance and dark grey has a lower luminance. The unit usually used is candela per square metre (cd/m^2). It is the objective correlate of brightness. It relates to L* in the CIELUV colour space and to *value* or *lightness* in the Munsell system.

2.3 Texture

The Oxford English Dictionary provides several different definitions for the term texture, depending on the context in which it is used. The general definition reads as "the feel, appearance, or consistency of a surface or a substance" with the examples of "skin texture and tone" and "the cheese is firm in texture" (Oxford Dictionaries, 2010). In the arts, the term is also known to describe "the tactile quality of the surface of a work of art" (Oxford Dictionaries, 2010). Even in the concise definitions that a general dictionary can offer, it goes as far as describing "the quality created by the combination of the different elements in a work of music or literature: [e.g.] a closely knit symphonic texture" (Oxford Dictionaries, 2010) – clearly a metaphorical, or cross-modal, use of the term.

Today, research on texture often refers to visual texture, and less often includes tactile or other information (Djonov & Leeuwen, 2011; Clarke et al., 2011; Rao et al., 1996). As the focus of research and life in general more and more shifts towards the use of digital media, textures are often represented visually, often with the aim to create the illusion of tactility. Djonov and Leeuwen emphasise the development and interdisciplinary use of the term:

> The term 'texture' originally referred to the art of weaving and the qualities of woven materials, but gradually expanded to encompass the tactile, material quality of objects generally and the synaesthetic interaction of tactile, visual and aural features. When used in describing images, 'texture' suggests an illusion of tangibility, brought about visually, by shifts in focus and colour and

> by patterns of lines and shapes. Applied to sound, texture refers to gradable timbral qualities such as tension, roughness, vibrato and pitch register. In short, texture, as the term is used now, applies across different media and can have tactile as well as visual and aural manifestations, and plays an increasingly important role in the semiotic landscape. (Djonov & Leeuwen, 2011, p. 541)

The use of textures in my experiment is visual only. The visual identity of textures is defined by their surface-, depth- and illumination discontinuities (Gurnsey & Fleet, 2001). From visual cues alone, we can detect both the surface characteristic (e.g. that something looks like wood) and its tactile sensations (e.g. that something looks slippery) (Landy & Graham, 2004).

2.4 Summary

This chapter has explored the core definitions that are relevant to this study of graphemes, speech sounds, voice quality, colour and texture. In short, the term grapheme usually refers to letters in this thesis; the term speech sound comprises both phonemes (i.e. abstract speech sound units of a specific language) and phones (i.e. concrete realisations of those); the relation between graphemes and speech sounds is rather opaque in the English language. The term voice quality describes the characteristics of a speaker's voice, including both anatomical and habitual aspects. Colours can be described using various different systems; the typology of Berlin & Kay (1999) aims at systematising the use of basic colour terms in different languages and cultures; the RGB colour space is an additive colour model of light which is mainly used for displaying colours on a screen, defining colours by adding values of red, green and blue; CIELUV colour space describes self-luminous colours with the aim of perceptual uniformity, defining colours by adding values of luminance, red-green and blue-yellow; Munsell codes use reflecting surfaces and are based on the visual perceptual space, defining colours by adding hue, value (lightness), and chroma (saturation). The term texture is versatile but will in this thesis mainly refer to visual features such as patterns.

3 Vowel sound - colour associations

3.1 Introduction

What colour do you see when you hear the vowel sound [a] – the red of roses, or maybe the blue of a clear sky? Vowel sound - colour synaesthetes have automatic associations like this. Also non-synaesthetes can often easily name a preferred colour for a particular vowel sound, even though this association does not automatically and consciously occur to them. How are vowel sounds and colours related, and to what extent do synaesthetes and non-synaesthetes use similar mechanisms for these associations? This chapter presents an experiment designed to uncover and compare the vowel sound - colour associations of both synaesthetes and non-synaesthetes.

3.1.1 Naming the inducer: graphemes vs. speech sounds?

As previously described in the introduction in Section 1.3, most synaesthetic inducers are linguistic units such as words, graphemes, speech sounds or sequences such as days of the week, months of the year or numbers (cf. e.g. Day, 2005; Simner et al., 2006; Rich et al., 2005). Concurrents are most often colours (c.f. e.g. Novich et al., 2011; Grossenbacher, 1997). Terms for inducers vary widely across studies, which can make it difficult to determine exactly what researchers believe the inducers of synaesthetic perceptions to be. Rich et al. (2005), for example, defined words, letters and numbers as lexical inducers in their assessment of 'lexical-colour' synaesthetes. In other studies, letters of the alphabet and numbers as inducers for colour perceptions carry the cover term 'sequence-colour' synaesthesia (e.g. Novich et al., 2011). Other studies examine 'grapheme-colour' synaesthetes; some of these studies presented the graphemes by pronouncing them out loud to the participants (e.g. Beeli et al., 2007), whereas others showed the grapheme on paper or on a screen (e.g. Simner et al., 2005; Smilek et al., 2007; Rich et al., 2005). It is a challenge to determine which aspects of the inducer actually influence the concurrents. Is it the semantic meaning of the word, the physical appearance of the letter, what it sounds like when pronounced, or something entirely different? Few studies address this issue by analysing and comparing semantic, visual and auditory influences.

Studies trying to tease apart graphemic from phonemic influences usually use words that contain different graphemes which are pronounced the same (i.e. homophones), or the same graphemes which are pronounced differently (i.e. homographs). Baron-Cohen et al. (1993) used words with the same initial phoneme spelled differently, such as *writer/rice* or *fish/photograph,* for this purpose. Their nine participants all seemed to be influenced by graphemes rather than phonemes, as they associated similar colours to words sharing the same initial letter rather than the same initial phoneme. Barnett et al. (2009) tested these influences by comparing multilingual synaesthetes' associations between colours and words

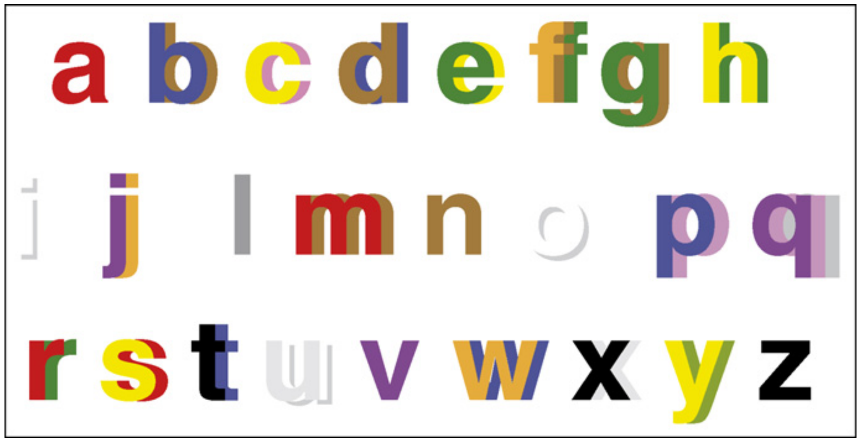


Figure 11: Representation of the 'prototypical' coloured alphabet of 70 grapheme–colour synaesthetes. Letters in two or more colours represent those with more than one significant shared colour preference; reported colour terms are represented as focal colours. Reproduced from Simner (2007, p. 25).

(such as weekdays and numbers) according to the language that the words were presented in. They reasoned that synaesthetic colours are induced on a semantic level when words with the same meaning trigger the same colour across languages, on an orthographic level when the visual properties of the words are similar across languages, and on a phonological level when the auditory properties of the words are similar. Unfortunately, the authors do not mention whether stimuli were presented in written or spoken form, or both. Barnett et al. came to a similar conclusion as Baron-Cohen and colleagues. Stimuli beginning with the same grapheme across languages were more likely to have consistent colours than those beginning with the same phonemes, which were nevertheless more likely to have consistent colours than those sharing neither initial graphemes nor phonemes.

In sum, stimuli that share the same initial grapheme have been found to be more consistent in colour associations than stimuli that share the same initial phoneme. This finding suggests that the orthographic input might play a more important role than the phonological form. The authors concluded that "at least for sequence words, it seems that colours induced by newly acquired languages are largely related to the low-level (mainly visual) properties of the new stimuli and do not necessarily reflect the semantic associations between the colours and stimuli from the primary language." (Barnett et al., 2009, p. 1352).

Graphemes Although synaesthetic associations are usually idiosyncratic, commonly shared associations can be found in both isolated graphemes and phonemes. Simner et al. (2005) tested colour associations to the visually displayed letters of the alphabet of 70 synaesthetes with English as their native language. Although no synaesthete had the same coloured alphabet as another, a significant amount of agreement on letter-colour

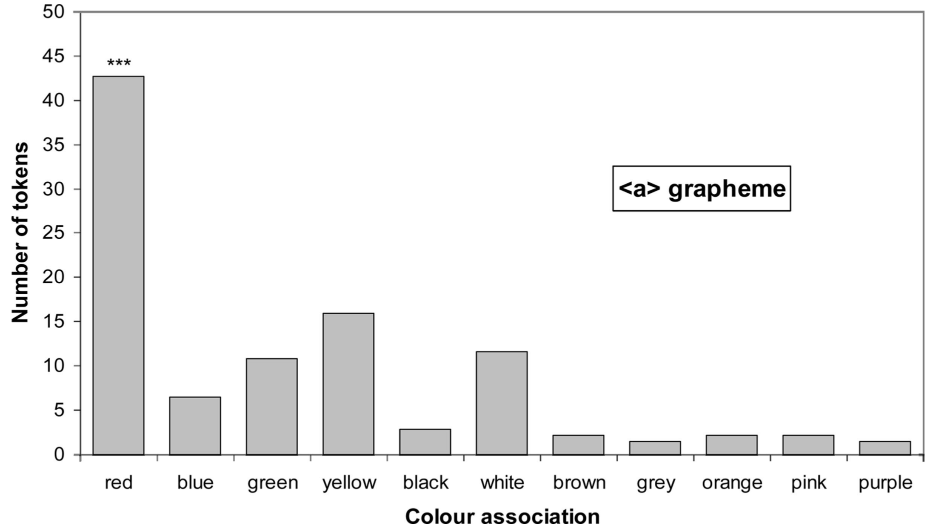


Figure 12: Frequency distribution of colour associations for the grapheme <a> in synaesthetes. Reproduced from Simner et al. (2005, p. 1074).

associations could be found in 25 letters (all but <k>), as displayed in Figure 11. A particularly strong agreement amongst synaesthetes occurred for the letter <a>, as shown in Figure 12. Red was chosen nearly three times as often as the second most associated colour yellow. Note that, despite the strong agreement on red for <a>, there is still a wide dispersion of idiosyncratic colour associations, as each of the 11 basic colour terms referring to the 11 focal colours (Berlin & Kay, 1999) was chosen by at least one person.

Rich et al. (2005) conducted a large-scale study on lexical-colour associations, amongst other things, and tested the grapheme-colour associations of 150 'lexical-colour' synaesthetes and 45 controls, also using the 11 basic colour terms for analysis. Vowel-colour associations which were shared either within or between the groups of synaesthetes and controls were as below. (There were no significant results for <e>.)

<i>	white for synaesthetes and controls
<a>	red for synaesthetes and controls
<o>	white for synaesthetes and orange for controls
<u>	grey for controls

Speech sounds Thoughts about the colouring of vowel sounds reach back a long time. Stumpf (1926) pondered about the lightness and colouring of vowel sounds, motivated by earlier work by von Helmholtz (1863) and Köhler (1924). Similar to the vowel quadrilateral as shown in Figure 6, a vowel triangle was assumed (which is indeed more appropriate for the German vowel sounds that they researched). Stumpf stated that, according to his perception experiments, front vowels are light and back vowels are dark, and put the main vowels in descending order of luminance going around the triangle: [ieaou]. In his research, Stumpf (1926) used the term "Vokalfärbung", i.e. vowel colouring. Similar to Jakobson (1962), he found that vowels become more colourful with a wider jaw opening. While Stumpf experimented with synthetic vowel sounds by playing musical notes with lingual organ pipes, which are supposed to resemble formants, he did not explain how these notes or formants exactly relate to colours. Marks (1975), on the other hand, described this extensively in his meta-analysis of numerous studies. Although his descriptions of vowel sound - colour associations and influences of formants on colour and luminance are detailed, his study faces other methodological weaknesses. It is not clear how many participants were included in his meta-analysis, whether they had synaesthesia and indeed whether the vowels were presented orally or in written form. His analysis was based on the following grapheme - vowel sound correspondences, as defined in Marks (1975, p. 311): /a/, /e/, /i/ and /o/ to <a>, <e>, <i> and <o>, respectively, in all languages included; /u/ to the French <ou> and the German <u>; /y/ to the French <u>. As data from different languages was used, his grapheme-vowel sound translation may not always be consistent, but this inconsistency is probably counter-balanced by the large amount of data and studies included.

Marks (1975) agreed with Jakobson and Stumpf that front vowels yield bright colours and back vowels dark colours, and explained this in terms of their 'intrinsic vowel pitch' which is thought to be aligned with the second formant, F2. Jakobson (1962) assumed an additional influence of the first formant, F1: the first two formants of light vowels are far apart from each other, whereas for dark vowels, they are close to each other.

Marks did not restrict his predictions and analyses to lightness; he also made assumptions about formant influences on colour associations. To prepare the data of the meta-study for colour analysis, Marks chose nine colour categories to which the data were reduced: yellow, red, green, blue, violet, white, grey, brown and black. Orange, for example, would be assigned to yellow and red in equal amounts. He found the following associations:

/a/	red or blue
/e/, /i/	yellow or white
/o/	red or black
/u/	blue, brown or black
/e/, /i/	brightest
/o/, /u/	darkest

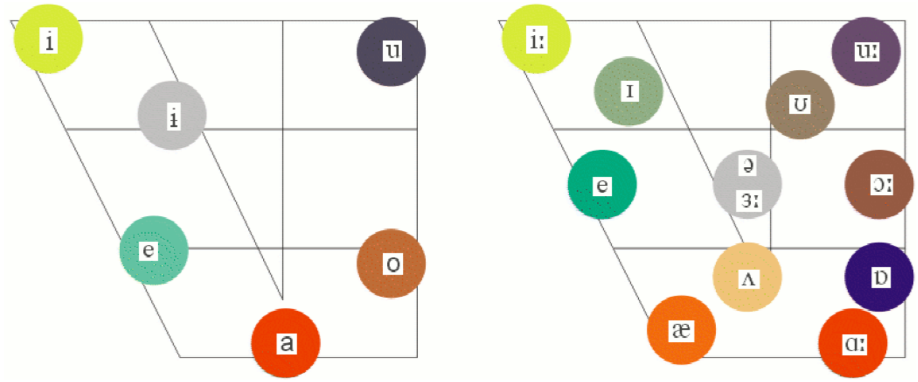


Figure 13: Coloured vowel charts for Polish (left) and English (right). The data represents the most frequent colour associations of Polish non-synaesthetes. Reproduced from Wrembel (2007).

For statistical analysis including colour and formant values, colour groups were transformed by assigning them to three dimensions: red-green, yellow-blue and white-black. When using vowel formants for the analysis, I assume he used standard values similar to those introduced in Table 2 on page 45 for the cardinal vowels, since he would not have been able to measure formants from stimuli of the original experiments. About the relation between formants and colours, he said:

> The greenness and redness of induced colors depend systematically on the ratio of the frequencies of the second to the first formants and, thus, on the distinctive feature compactness. Compact vowels yield red colors; diffuse vowels yield green colors. (Marks, 1975, p. 312)

'Compact' describes vowels with the first two formants close to each other, whereas for 'diffuse' vowels, F1 and F2 are far apart from each other (Jakobson & Waugh, 1987). From his literature review, Marks (1974) additionally found that yellowness is more associated with high frequency and blueness with low frequency. From experiments he further concluded that correlations exist between whiteness-blackness and 'vowel pitch' across participants (Marks, 1975).

These ideas were largely confirmed by Wrembel (2007), who asked 29 Polish students and eight phonetics teachers (although not tested for synaesthesia, they were presumably non-synaesthetes) to choose a colour from a palette of 11 colours to match the Polish vowel sounds. Results were compared with the results of a study with Polish participants who learned English and took part in a similar experiment with English vowel sounds (Wrembel & Rataj, 2008; Wrembel, 2009). Figure 13 shows the results of the two studies. For the Polish vowels, red was chosen significantly more often than any other colour for /a/; green and yellow for /i/; green and blue for /e/; blue, orange and brown for /o/; blue and brown for /u/; and grey for the central vowels /ɪ ɨ ə/.

Although a comparison across languages is challenging, overall, it seems there is large agreement between Marks' and Wrembel's data, particularly in defining front vowels as light, [a] as red and back vowels as dark.

Graphemes and speech sounds in comparison In this section, an attempt is made to compare the results of grapheme-colour studies with those of vowel sound - colour studies. To be able to compare letters with vowel sounds, the grapheme-phoneme correspondences have been defined in Table 2 on page 45 in the previous chapter for English, German and Polish for the pronunciation of isolated vowel graphemes. Recall that for German this is a straightforward process because letter-phoneme correspondences use the same symbols, i.e. <e> is pronounced /e/ etc. The same order of lightness that Stumpf (1926) found for vowel sounds, namely /i e a o u/ in descending order, was found by Beeli et al. (2007) for graphemes in Swiss German.

English vowel graphemes do not match the phonemes so neatly. For most vowel letters produced in isolation, the pronunciation in Standard Southern British English (SSBE) involves a diphthong rather than a monophthong. The letters are also pronounced differently depending on the syllabic context or even the context of the whole word. The clearest grapheme-phoneme relation is found in the grapheme <e>, which corresponds to /i/. In the case of <e> and /i/, the associated colours in the three studies by Simner, Marks and Wrembel are largely in agreement with each other: for Simner's and Wrembel's participants, the letter <e> (or the vowel sound /i/) was associated with yellow and green; for Marks' participants, it was associated with yellow and white. For other vowels, the relation between graphemes and phonemes is less clear, especially because most of them are pronounced as diphthongs, whereas Wrembel and Marks presumably used monophthongs in their studies. Still, it can be concluded that there is little agreement between Wrembel's and Marks' phonemes and Simner's graphemes <i>, <o> and <u>: Simner's data suggest white associations, whereas white is never present in Wrembel's graphs – and only for /i/ and /e/ in Marks' data, which are phonemes that should not be present in any accent of English when pronouncing <i>, <o> or <u>. Marks listed colour associations not only with transcribed phonemes, but also with graphemes, presumably pronounced as if reading out the alphabet. As it is not always clearly defined whether he referred to <a> or /a/, for example, I hesitate to compare his data with Simner's.

The role of music Additional acoustic influences on synaesthetic colour perceptions can be found in music. De Thornley Head (2006) tested six pitch-colour synaesthetes and 32 controls on their colour associations with synthesised clarinet tones. The stimuli consisted of one octave of semitones plus five quartertones (i.e. intermediate tones between the normal semitones). The author found that there was a large difference in the variance of RGB values chosen for the same pitch across subjects. Within-subject consistency of colour associations for musical tones was much greater for synaesthetes than for controls. Hue and shade were matched more strongly for pitch class than for neighbouring tones, meaning that Cs of different pitches had more similar hues and shades than C and D. This conclusion of the author appears to be based only on a test of one octave, i.e. with only two Cs and with every other note presented only once. The result should therefore be interpreted with appropriate caution. For quartertones, synaesthetes chose the intermediate colour which lies between the hues of neighbouring semitones, suggesting a localised pitch-colour correspondence. However, de Thornley Head did not find a continuous colour scale along the complete musical scale. In contrast to synaesthetes, controls had a random attribution to hue, but a significant linear whiteness scale resulting in a whiteness-frequency effect.

Ward et al. (2006) tested ten sound-colour synaesthetes, who additionally had grapheme-colour synaesthesia, and controls on their colour associations with ten notes ranging from C1 to Eb6 (1 being the lowest octave used and 6 the highest). Stimuli were presented as pure tones, piano and string notes; colour choices were made by using the standard Windows API Choose Colour dialog box. The synaesthetes' associations were significantly different from that of controls in the Munsell value, hue and chroma for the various single tones played with the piano, strings or as a pure tone. In other words, synaesthetes chose different colours from controls and were more consistent in their colour choices. For both controls and synaesthetes, there was a positive correlation between pitch and lightness. Chroma peaked at mid pitch range, suggesting that participants perceived mid pitch notes as most colourful. Timbre also had a significant effect on chroma: piano and string notes were literally more colourful than pure tones. In general, synaesthetes customised their colours more than controls by adjusting the colours shown on the screen to their synaesthetic colours using sliders.

To sum up the results of the studies by de Thornley Head (2006) and Ward et al. (2006), synaesthetes were more consistent and more specific in their colour choice than controls, but seemed to employ an identical strategy to non-synaesthetes for sound-colour mappings like pitch-lightness.

Marks (1974) also investigated the relation between loudness, brightness, volume (as in the size of an object) and pitch in non-synaesthetes. His participants saw a grey shade and had to adjust the loudness and/or pitch of a pure tone to match it. While ten participants set increasing brightness to increasing pitch, two set it to decreasing pitch. That is, even though there was no agreed direction of associating pitch with brightness, all followed the systematicity of a decreasing or increasing scale. If there was a constant brightness but an increasing pitch, most participants lowered the loudness. All in all, this suggests that most people match auditory to visual brightness. Figure 14 shows the schematic associations of pitch, loudness, brightness and volume: with increasing brightness, participants increased the pitch and loudness; with increasing volume, the pitch

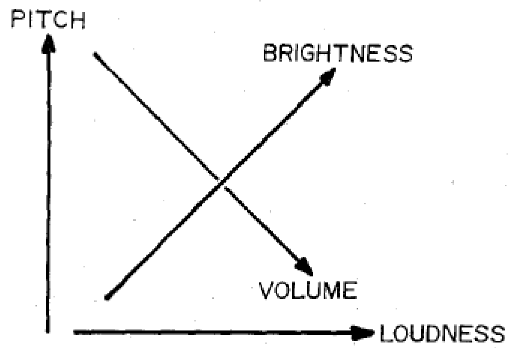


Figure 14: Schematic representation of the psychological space for simple sounds. Reproduced from Marks (1975, p. 319)

decreases but the loudness increases. According to this graph, combined with findings by de Thornley Head (2006) and Ward et al. (2006), it can be expected that front vowels with a higher intrinsic vowel pitch should be associated with lighter colours and grey shades than back vowels, which resemble a low intrinsic vowel pitch. Colour associations by participants of de Thornley Head (2006) and Ward et al. (2006) seemed to be more systematic and consistent in synaesthetes, leading to a similar expectation for vowel sounds.

3.1.2 Motivation and hypotheses

As described in Chapter 1.8, the experiment was motivated by three different issues: (1) to find out if there is any systematic relation between vowel formants and colour and lightness associations, (2) to explore group differences in the association patterns and the consistency of associations between synaesthetes and non-synaesthetes, and (3) to explore whether synaesthetic participants can be divided into vowel sound- and grapheme-induced synaesthetes, or whether these types co-exist or overlap in some way.

(1) The relation of vowel formants and colours Jointly, the studies by Stumpf (1926); Jakobson (1962); Marks (1975) and Wrembel (2007), described in the previous section, suggest systematic relations between vowel formants and colours. However, each of them has some weaknesses. The work of both Stumpf and Jakobson was mainly theoretical in nature and their experimental parts lacked perceptual tests on participants other than the authors. Marks' data is based on participants' vowel-colour associations, but it is often unclear how stimuli were presented, whether vowel sounds were heard or vowel graphemes seen, and how the colour choice was made, for example whether participants replied verbally or picked from a colour display. It also remains unclear whether his formant calculations were based on hypothetical or measured formant frequencies. Wrembel (Wrembel, 2007, 2009; Wrembel & Rataj, 2008) addressed some of these issues, conducting perception experiments using audio recordings of vowel sounds and displaying the 11 focal colours. However, she did not undertake acoustic measures to analyse vowel formants, nor did she define the colours in a quantitative fashion. I intend to fill these gaps in the literature with this study.

To address the questions of lightness and colouring of vowel sounds separately, both colour and grey shade displays were used in the present study. Based on previous findings, I expect luminance associations to relate to the acoustic feature of F2, the so-called intrinsic vowel pitch, and maybe also with F1 which relates to the openness of vowels, the so-called vowel height. Recall that vowels with a high F1 are produced with a very open mouth setting, and those with a low F1 with a closed mouth setting. Vowel sounds with a high F1 are expected to be perceived as colourful and, in particular, to be perceived as red.

(2) Differences in association and consistency between synaesthetes and non-synaesthetes Although synaesthetic perceptions are usually idiosyncratic, general tendencies have been found in grapheme-colour associations such that many synaesthetes perceive <a> as red and <o> as white (Simner, 2007). Similar to the coloured alphabet in Figure 11 on page 57, Wrembel (2007) introduced a coloured vowel chart for English and Polish based on the associations of non-synaesthetes (Figure 13 on page 60). These two studies indicate a common basis of colour associations across people, here for graphemes among grapheme-colour synaesthetes and for vowel sounds among non-synaesthetes. By testing colour associations of both synaesthetes and non-synaesthetes using the same setup, differences or commonalities within and across groups can be detected.

It is also claimed that synaesthetes' associations are consistent over time. Should general agreement on colour associations across groups be found during the course of the present research, one differentiating criterion could be that synaesthetes' associations will be more consistent than those of non-synaesthetes. In other words, based on findings from previous publications (e.g. Asher et al., 2006; Rich et al., 2005; Simner et al., 2005; Ward & Mattingley, 2006), I expect that synaesthetes will assign the same or similar colours to the same vowel sound repeatedly. For non-synaesthetes, I expect the results to be less consistent, though not necessarily unsystematic. It will also be of interest to examine how consistency varies between individuals within groups.

(3) Speech sounds and graphemes as (co-?)inducers Neurological theories about cross-activation of adjacent brain areas (Hubbard et al., 2011) and about the lack of developmental pruning (Maurer & Mondloch, 2005) are not in agreement. While the former favours a dominance of graphemic influences on colour associations, the latter suggests a stronger influence of speech sounds because a child first encounters words and speech sounds phonetically, before he or she becomes literate and learns to correspond those sounds to letters. There are studies supporting both theories. Thus, more detailed studies on these topics are necessary to test to what extent those theories are supported. Perhaps one theory describes the situation of some synaesthetes, whereas the other describes the situation of other synaesthetes. Or it could be that neither is correct and some combination of them or some alternative is more suitable.

According to previous research, it is more common that a graphemic representation of a speech sound functions as inducer rather than its phonemic representation, even when spoken stimuli are used (Simner, 2007; Ward et al., 2005; Day, 2005; Baron-Cohen et al., 1993; Barnett et al., 2009). However, cases of 'real' speech sound - colour synaesthesia are also reported (Day, 2005). With this study, I aim to explore whether there are indeed two subgroups of grapheme-induced and speech sound-induced synaesthetes, or whether there is a different underlying mechanism, or some sort of overlap or combination (cf. Smith et al., 2011). By offering a large range of vowel sounds as inducers, some of which relate to English phonemes while others do not, it should be possible to find regularities in the vowel sound - colour associations that suggest stronger influences of graphemes or speech sounds. As it is common for synaesthetes to have more than one type of synaesthesia, it could be the case that a person with grapheme-colour *and* sound-colour synaesthesia reacts both to English vowel sounds according to their orthographic representation, and to vowel sounds which do not exist in their native language according to their sounds.

Hypotheses

Based on the combined findings of the studies reviewed in this section, the hypotheses are as follows:

1. Open vowels will be associated with reddish colours, front vowels with greenish colours.

2. Front vowels will be perceived as light, back vowels as dark. In other words, the lightness of associations will increase with increasing F2.

3. Synaesthetes and non-synaesthetes will have similar patterns in associating colours with vowels but synaesthetes will behave more consistently.

3.2 Methods

3.2.1 Experimental design

Previous studies have used various different methods in their behavioural experiments for looking at synaesthetic perceptions in general and at sound-to-colour associations in particular (e.g. Ward et al., 2006, for musical pitch- and timbre-colour synaesthesia; de Thornley Head, 2006, for musical pitch-colour synaesthesia; Wrembel, 2007, and Wrembel & Rataj, 2008, for coloured vowel charts for second language acquisition; Marks, 1975, for vowel-colour synaesthesia). To decide which method was best for this experiment, several considerations had to be taken into account. A balance between the need of synaesthetes to have a broad enough colour choice and the risk of non-synaesthetes being overwhelmed had to be found. In this study, a display of 16 colours was used, which was intended to give people enough choice to find a colour close to their synaesthetically perceived one. This way, psychophysical constraints were taken into account as well, as discussed in Section 3.2.3. Another important issue, also discussed in Section 3.2.3 on page 69, is the arrangement of colours in a random way, so that colours are displayed in context of varying neighbouring colours in a varying spatial setup.

To test consistency of vowel-colour associations, two designs are possible: either to present a limited set of vowel sounds to participants on multiple occasions, or to present a large set of (repeating) vowel sounds to participants on one occasion. The latter design was chosen to avoid drop-out. To test within-subject consistency, each vowel sound was presented 16 times. The so called "Type 1 index 1 sequence" was chosen (Aguirre, 2007; Nonyane & Theobald, 2007). In this sequence, every stimulus is followed by itself as often as by any other stimulus. This rules out any carry-over effects that might be there when stimuli are presented in a less balanced context. Because every vowel appears in the context of every vowel sound, context does not bias the results.

3.2.2 Participants

There were 34 participants (14 synaesthetes and 20 non-synaesthetes). They all had normal hearing, normal or corrected-to-normal eyesight and normal colour vision, apart from one synaesthete who had a colour deficiency.

Synaesthetes

The 14 synaesthetes were recruited either through personal contacts, or through the psychology department subject pool, whose form asks subjects whether they are synaesthetes or not with a two page synaesthesia questionnaire (Appendix C). If their responses indicated grapheme-colour or phoneme-colour synaesthesia, I invited the participants to take part in the experiment. Since recruitment was via self-referral in this way, participants were asked to fill out an extended questionnaire created by Jamie Ward and Julia Simner (Appendix D, the predecessor of https://www.survey.bris.ac.uk/sussex/syn) after having taken part in the experiment, to further specify their synaesthesia.

2 participants were male, 12 female. The mean age was 26, ranging from 18 to 69. 1 synaesthete was colour deficient. 6 were native speakers of Scottish English, 5 were native speakers of other varieties of English (including 2 bilingual French), 1 was German, 1 Polish and 1 Slovenian.

12 synaesthetes described themselves as musical and 13 "musically active", i.e. they played at least one instrument each and/or sang. 10 synaesthetes were involved in the

visual arts.

Non-synaesthetes

20 people were recruited from the general population of the University of Glasgow. Most participants were naive to the specific purpose of the experiment; only 3 of them were likely to have been aware that the research concerned synaesthesia. From the replies to the questionnaire (Appendix C) and personal communications it was found that no participant in this group had grapheme- or phoneme-colour synaesthesia.

9 participants were male, 11 female. The mean age was 24, ranging from 19 to 37. 18 participants had a university degree or were university students; 2 were college students. All but 2 participants were native speakers of Scottish English, 1 of English English and 1 of Northern Irish English. 1 participant was raised bilingually with English and Punjabi. All participants had lived in Scotland for most of their lives.

10 participants described themselves as musical, playing at least one instrument each. 5 participants were involved in the visual arts.

3.2.3 Stimuli

Auditory stimuli

16 auditory stimuli were used in the experiment. 8 of them represent the primary cardinal vowels of the International Phonetic Alphabet (International Phonetic Association, 1999), see Figure 6 on page 44. The other 8 were created using a morphing technique to produce intermediate vowels between neighbouring cardinal vowels, as described below. Audio samples of the stimuli are included on the supplementary CD under "Ch3_Vowel_sound-colour_associations".

Recording A trained male phonetician provided the recordings. He was 65 years old and grew up in East Yorkshire. He was asked to produce the cardinal vowels three times with a steady pitch. Each vowel was sustained for approximately 1 second. The recordings were made in a high-specification double-walled booth in mono with a sampling rate of 44 kHz using a high-quality Sennheiser cardioid condenser microphone, a Symetrix pre-amplifier, an Edirol AD/DA converter and a PC (located outside the studio to exclude noise).

Selection and Morphing The eight primary cardinal vowels [i e ɛ a ɑ ɔ o u], i.e. C1 to C8, were used. Of all recorded repetitions, those that sounded most clear and correct with respect to the cardinal vowels and the sustained pitch were chosen, one for each primary cardinal vowel. Each neighbouring pair of the eight cardinal vowels produced by the speaker was morphed, creating eight morphed vowels in addition, resulting in 16 vowel sounds in total. Morphing C2 and C3, for example, results in a vowel which I shall refer to as C2.5. Two MatLab scripts were written to automatise the morphing. The scripts were based on a modification and synthesis procedure called STRAIGHT (Speech Transformation and Representation based on Adaptive Interpolation of weiGHTed spectrogram) (Kawahara, 2003), which is mainly used for voice quality analysis and perception experiments. The technique was designed for manipulating sound while retaining a natural quality. The morphing program treats both files as equally important and interpolates between them. It uses five dimensions to calculate the morphing process:

1. F0 (fundamental frequency): logarithmic interpolation of the fundamental frequency by measuring the harmonic structure of both sound files.

2. Spectral amplitude: logarithmic interpolation of frequency- and time-dependent amplitude of the spectrogram of both sound files.

3. Aperiodicity: linear interpolation of aperiodicity is achieved by measuring the energy of inharmonic frequencies, which is normalised by setting it in relation to the total energy.

4. Time: linear interpolation of the length of both sound files.

5. Frequency: logarithmic interpolations for frequency frames by measuring formants of both sound files as detailed below.

The first four measurements were done automatically by the program. Because the formants were difficult to measure in the recordings obtained, the automatic detection was prone to measurement errors. Therefore, manual formant measurements in addition to three different automatic methods were taken using Praat (Boersma & Weenink, 2012). The manual measurement was done by eye by myself, a trained phonetician, using the spectrogram. In cases of doubt, another trained phonetician was asked for advice. For the automatic methods, an LPC spectrum (using linear predictive coding) for representing the spectral envelope, a spectral slice viewing the frequency spectrum at a point of time with stable formants, and formant tracking using Burg's algorithm were used, in the standard settings as provided in Praat. The value that found most agreement between those four measurements (prioritising the manual measure by eye if still unclear) was taken for the morphing procedure and manually fed into the script. To morph the frequency structure and formants adequately across time, the formant measurements were taken at two time points: at roughly 200ms after the beginning and 200ms before the end of each sound file, varying slightly depending on the quality of the spectrogram. As stimuli consisted of vowels spoken with a steady pitch, no large f0 variation was expected throughout a sound file.

The morphed vowels sounded as natural as the recorded vowels. No participant noted any differences in quality of the vowels.

Equalising sound files Certain adjustments were made on the sound files in order to keep the sound as natural as possible while controlling for as many variables as possible. The intensity and length of the sound files were adjusted because loudness and length might have a confounding influence on the choice of colour or grey shade. Marks (1974), as previously mentioned, found that his participants matched a loudness scale to a brightness scale. The original average intensity of the 16 vowel sounds was 83.3 dB_{SPL} (ranging from 78.56 to 85.92) with a SD of 2.5. They were equalized to 80 dB_{SPL} using Praat. The length of all sound files was adjusted to the mean length, 1049ms, using PSOLA (pitch synchronous overlap add) in Praat. In this method, f0 stays constant and the durations are changed linearly by averaging adjacent f0 periods which overlap in the time domain (Charpentier & Moulines, 1989). It was not considered necessary to correct the fundamental frequency, which only ranged from 120 Hz to 124 Hz.

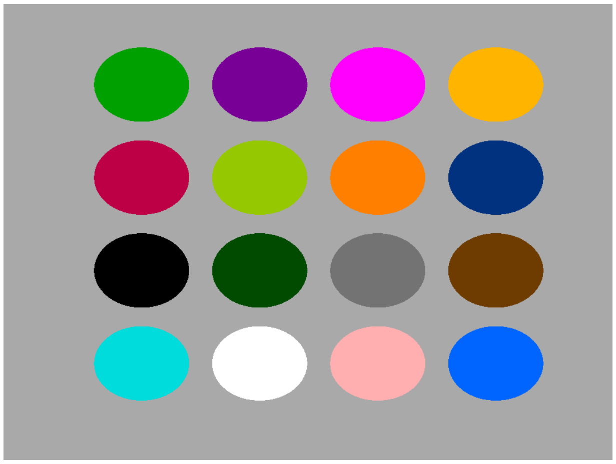
Visual response display

It was difficult to find a balance between the needs of the synaesthetes for a colour choice that is as detailed as possible and the need to obtain analysable data and meet certain psycho-physical constraints. There were three main alternatives for presenting the colours:

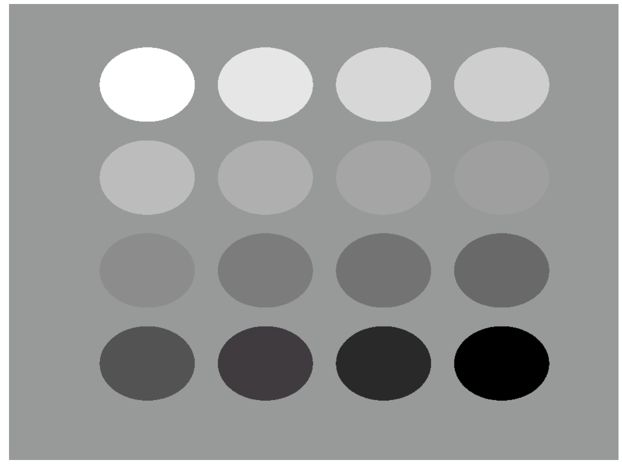
(1) A colour wheel. Although a colour wheel is easy to use and allows for accurate colour choices, it has some disadvantages for an experiment such as this one. As colours are always arranged the same way for every stimulus in a colour wheel, choices could be based on a spatial mapping rather than on colours per se. Participants might, for example, subconsciously choose colours from the top of the wheel for close vowels, and colours from the bottom for open vowels. Even if the wheel was randomly turned per stimulus, colours would still always be presented in the same context.

(2) A colour slider. A colour slider allows for very accurate colour choices. It presents the whole colour spectrum and can start at any point in the spectrum. However, as with a colour wheel, colours would be presented in the same colour context for every stimulus. It might also take subjects a long time to find the exact representation of the colour that they have in mind, which could have detrimental effects on the results of the experiment. Pisoni (1973, 1975) found that vowel sound perception shifts with delay: the memory of a vowel sound is more abstract than the immediate perception of the vowel sound.

(3) A limited set of colours. A limited set of colours might not always satisfy the synaesthetes, but it has advantages over the first two options. Firstly, it facilitates fast responses and therefore responses based on the immediate auditory perception of the vowel sound rather than its abstract mental representation, which was important in this



(a) One randomised representation of the colours.



(b) Grey shades.

Figure 15: Visual response displays. (The quality presented here may differ from the quality presented on the experimental screen).

experiment because of the fine grained differences of the vowel sounds. Secondly, the colours can be randomised on the screen so that they will appear in different contexts and at different places in space for every auditory stimulus. Hence, colours will not be confounded with colour context or spatial mapping.

16 colours were deemed to provide a broad enough choice for the synaesthetes and a narrow enough choice to allow fast responses. Colours were set up in a 4x4 matrix in front of a mid grey background. They consisted of the 11 focal colours (white, black, blue, green, yellow, red, grey, brown, orange, pink, purple; cf. Berlin & Kay, 1999 for the 11 basic colour terms of English), plus dark green, light green, rosa (= pale pink), cyan and dark blue to fill the gaps in colour space as evenly as possible. The colours were randomly assigned to the different positions on the screen for every stimulus. The displays and randomisations were created using MatLab scripts. A sample image is shown in Figure 15a.

For the second part of the experiment, another 4x4 matrix was created with 16 grey shades gradually changing from white to black. These grey circles were not randomised but stayed in the same order for every stimulus (white in the top left to black in the bottom right), see Figure 15b, representing a luminance continuum. More details about the colours and grey shades are given in Section 3.2.5.

3.2.4 Procedure

Participants were verbally introduced to the task, using an introduction sheet as a guide. Synaesthetes were additionally asked about their kinds of synaesthesia and asked to describe their perceptions. After that, a colour deficiency test using Ishihara's "Test for Colour-Blindness" (Ishihara, 1978) was carried out. A subset of 13 cards was used for this experiment, two each of the six subsets recommended by the author for the general test of identifying colour deficiency (by presenting coloured numbers on a differently coloured background), and one control plate testing whether participants were willing and able to do the test (here, the colours of the number and the background were always contrastive, no matter what colour deficiency the viewer may have).

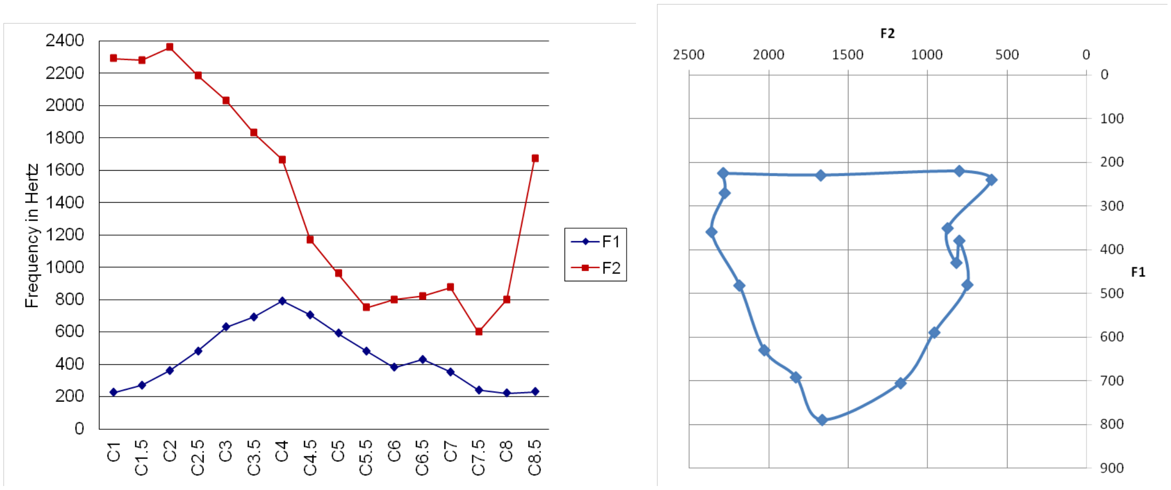
The experiment took place in a high-specification double-walled booth and was run from a PC located outside the booth to eliminate noise. Participants wore high-quality Sennheiser headphones and looked at a monitor from a distance of approximately 65 cm on which the colours had been calibrated prior to the experiment (see "Colour tuning and measurements" on page 72 for the calibration procedure). Before the experiment started, the light was switched off. It was explained that the darkness was necessary for an undistorted perception of the colours, as the colours are shown as a source of light themselves, and the ceiling light would interfere with their correct display.

The main experiment was conducted using DMDX (Forster & Forster, 2003), a program which is used to present auditory and visual stimuli concurrently and collect responses via mouse clicks. The colour and the grey shade experiment each consisted of eight blocks. In each block, 32 stimuli were presented (in the first block 33); thus 257 stimuli in total for each task (colour and grey shade). Participants were offered short breaks between each block. With the eighth block, they finished the colour experiment and had a longer break with optional refreshments. After the break, the same procedure was repeated with grey shades instead of colours.

The instruction for the task was "Please match each sound with the colour that goes best with it" for colour circles, and "Please match each sound with the shade that goes best with it" for grey circles respectively. The response display was presented at exactly the same time as the audio stimuli started to play. Participants chose a colour by clicking on the circle. After a circle had been selected, the blank grey background was shown for 2 seconds to give the after-image time to disappear before the next stimulus. Then, the next stimulus pair appeared and the procedure was repeated. If no circle had been chosen after 15 seconds, a null response was recorded and the next stimulus pair was displayed.

After the grey shade experiment, all participants filled out a short form asking about their language, musical and artistic background and other demographical data. Additionally, synaesthetes were given the extensive questionnaire designed by Ward and Simner (see Appendix D) and asked to fill it out in their own time. Information from this questionnaire was helpful for disentangling grapheme from phoneme synaesthetes and to more precisely define their synaesthesiae.

The experiment lasted for 1 $\frac{1}{2}$ to 2 hours, and participants were compensated for their time either with money or – in case of psychology students – with course credits. The 11 synaesthetes who filled in Ward and Simner's questionnaire were additionally paid for the



(a) (b)

Figure 16: Values of the vowel formants F1 and F2 for the 16 vowels used as stimuli.

time it took them to do so.

3.2.5 Analyses

To be able to analyse vowel sound - colour relationships in a quantitative fashion, vowel stimuli were analysed acoustically by measuring formants, and colour responses were analysed in CIELUV colour space, before these measures finally underwent statistical tests.

Vowel formants

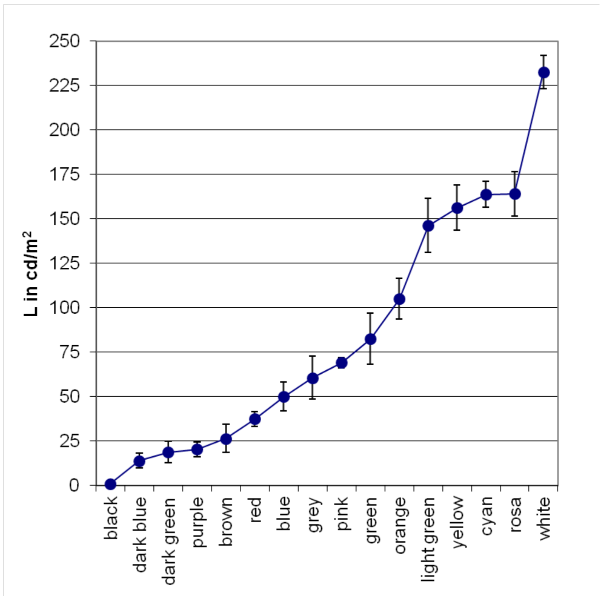
The procedure for measuring formants is described in Section 3.2.3 on page 67. The vowel formants F1 and F2 are displayed in Figure 16. In Figure 16a, the vowels are lined up along the x-axis, and the frequency is given in Hz along the y-axis, so the values can be compared to those in the schematic representation in Figure 5 on page 44. In Figure 16b, the formant values of F1 and F2 are plotted against each other, resulting in a shape similar to that of the vowel quadrilateral of the IPA (1999) shown in Figure 4 on page 42. C1 or /i/ is located in the top left corner, going around anti-clockwise with C4 or /a/ at the bottom (refer back to Figure 6). Additionally, the values of F1, F2 and F3 are listed in the table in Appendix E.

Colour tuning and measurements

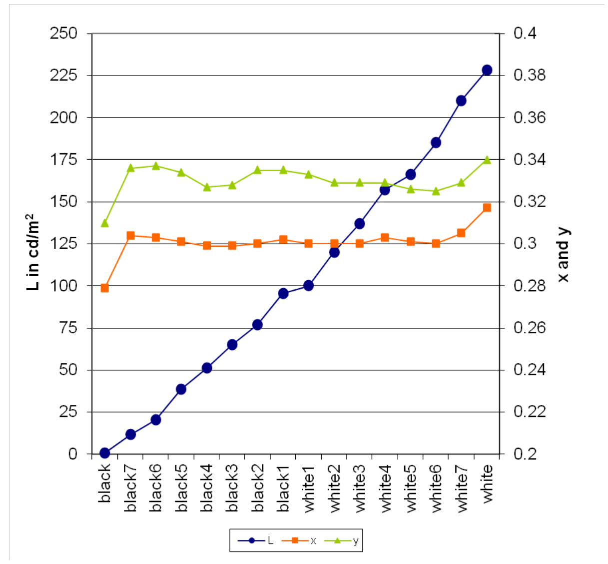
The colours were chosen very carefully. The experiment used 11 focal colours relating to the 11 basic colour terms (Berlin & Kay, 1999) and another five distinct colours. To find the focal colours, the Munsell codes defined by Rosch Heider (1972) were used and converted into CIE coordinates (cf. Westland & Ripamonti, 2004, for conversion). As every screen displays colours slightly differently, the colours had to be measured on the monitor which was used for the experiment. For this measurement, a chromameter was used, the settings of which were: calibration preset, measuring mode abs., response slow, cd/m^2 (candela per square metre). The chromameter physically measures L for luminance, and x and y (CIE 1931 coordinates) of light sources like monitors. The luminance L ranged from dark (=0 cd/m^2) to bright (=232 cd/m^2). The higher the x value, the redder the colour; the higher the y value, the greener the colour. Where the colours did not match the focal colours from the reference, adjustments were made by changing the RGB values in the user interface of the computer. Every colour was measured in every position on the screen to check for irregularities of the LCD display. The grey of the background was used as the 'reference white' in the transformation formula, resulting in some L* values above 100. Some variation of luminance was found across positions on the screen, with a standard deviation between 3 and 15 cd/m^2 depending on the colour. Figure 17a displays the luminance value of every colour; the error bars show the standard deviation caused by the different positions on the screen. A test of correlation between the positions of the colour displays on the screen and the luminance values of these positions revealed that luminance did not significantly correlate with the position on the screen (p=0.733 with a Spearman's rho of 0.021), i.e. the screen position did not significantly influence the luminance of a colour. Thus, the monitor was suitable for use.

The dispersion of the L, x and y values of each colour depending on the position on the screen is shown in Figure 18. Most importantly, there are no overlaps between different colours, which means that, independent of the position on the screen, each colour has its unique colour values. The average values of L, x and y, that were used for the stimuli, are displayed in Table 4. Westland & Ripamonti (2004) list formulas to transform the CIE coordinates L, x and y, taken by the chromameter, into CIELUV coordinates resulting in L*, u* and v*, the measure used for self-luminating displays, also presented in Table 4.

The grey shades were also measured on the experimental monitor with a chromameter. Values of L* for luminance and u* and v* for hue were obtained from the adjusted shades, for subsequent analyses. The L, x, y and L*, u*, v* values of the grey shades are displayed in Table 5. Some grey shades looked a bit bluish, reddish or greenish when displayed on the experimental monitor using DMDX. This was corrected by adjusting x and y to 0.33 as closely as possible, because 0.33 is supposed to be a neutral grey setting without coloured tints. While the value for y was close to 0.33, the value of x was around 0.30 for all grey shades. The L, x and y values of the final grey shades are displayed in Figure 17b. The endpoints of the x and y curves displayed in Figure 17b fall out of alignment. This most likely reflects measurement errors because white and black are very difficult to measure with a chromameter due to their extreme L values.



(a) Colours. Luminance only for the sake of clarity.



(b) Grey shades. Scaling for L on the left, for x and y on the right.

Figure 17: L, x and y values for visual response displays. L = luminance, x roughly representing redness, y roughly representing greenness. Error bars represent 1 SD.

	L	x	y	L*	u*	v*
white	232.4	0.317	0.340	137.7	-2.871	9.452
rosa	164.1	0.339	0.324	120.8	32.151	0.135
cyan	163.8	0.220	0.325	120.7	-96.514	-26.387
yellow	156.2	0.449	0.474	118.6	50.181	120.246
light green	146.2	0.368	0.529	115.7	-40.487	125.382
orange	105.0	0.504	0.433	101.9	110.028	96.030
green	82.5	0.289	0.576	92.8	-89.262	103.520
pink	69.0	0.326	0.170	86.5	111.587	-135.557
grey	60.6	0.301	0.327	82.1	-7.914	-4.350
blue	50.0	0.170	0.160	76.1	-48.698	-153.412
red	37.3	0.532	0.292	67.5	169.598	11.574
brown	26.4	0.488	0.424	58.4	57.897	50.894
purple	20.3	0.260	0.126	52.2	42.523	-126.287
dark green	18.8	0.286	0.550	50.4	-46.805	51.632
dark blue	13.9	0.166	0.132	44.1	-23.830	-109.153
black	1.0	0.295	0.304	8.8	-0.361	-2.072

Table 4: CIE coordinates of the 16 colours used for the visual response display. Values are averaged across the measures in the 16 different positions on the screen.

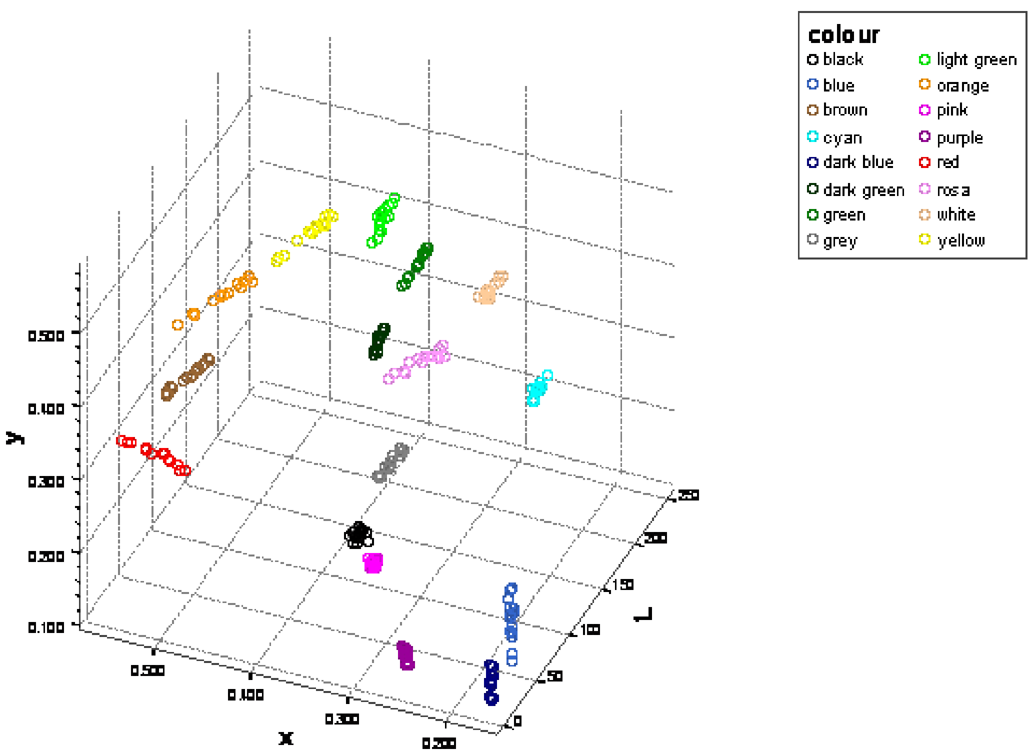


Figure 18: Colour values of L, x and y of the 16 colours measured in the 16 different positions on the experimental monitor in three-dimensional space.

	L	x	y	L*	u*	v*
white	228.0	0.317	0.340	136.7	-2.112	9.085
white7	210.0	0.305	0.329	132.5	-9.322	-4.166
white6	185.0	0.300	0.325	126.4	-12.199	-8.733
white5	166.0	0.301	0.326	121.3	-11.188	-7.294
white4	157.0	0.303	0.329	118.8	-10.493	-4.190
white3	137.0	0.300	0.329	112.8	-13.002	-4.626
white2	120.0	0.300	0.329	107.3	-12.361	-4.398
white1	100.0	0.300	0.333	100.0	-13.368	-1.334
black1	95.4	0.302	0.335	98.2	-12.281	0.407
black2	77.0	0.300	0.335	90.3	-12.897	0.031
black3	65.0	0.299	0.328	84.5	-10.101	-4.215
black4	51.0	0.299	0.327	76.7	-8.811	-4.363
black5	38.7	0.301	0.334	68.5	-8.866	-0.314
black6	20.3	0.303	0.337	52.2	-6.540	1.024
black7	11.5	0.304	0.336	40.4	-4.523	0.596
black	0.8	0.279	0.310	6.9	-1.507	-1.511

Table 5: CIE coordinates of the 16 grey shades used for the visual response display.

Statistical analyses

The effects of the vowel formants F1 and F2 on the visual responses were tested by conducting a repeated-measures analysis of covariance (ANCOVA) using SAS software version 9.2. An ANCOVA was preferred over the analysis of variance (ANOVA), because a relationship between formants and colour choices was to be expected a priori according to the findings of previous research (Jakobson, 1962; Marks, 1975, 1974; Wrembel, 2009). L*, u* and v* values from the colour experiment and L* from the grey experiment were the dependent variables. Fixed effects were participant group (synaesthete vs. non-synaesthete), F1, F2, interactions of F1*group and F2*group, with between-subject random effects for intercept, F1 and F2. The between- and within-subject variances were estimated separately for each group, as differences between synaesthetes and non-synaesthetes were expected. Equality of variances between groups was assessed using F-tests. The ANCOVA automatically corrects for multiple comparisons.

To be able to visualise vowel sound - colour associations for all the vowel sounds and participant groups, graphs were plotted based on a LOESS analysis (local polynomial regression fitting) using SAS software version 9.2, returning a smoothed regression curve. For this test, L*, u* and v* values of the vowel sound - colour associations were averaged across the groups of Scottish synaesthetes, other synaesthetes, synaesthetes in total and non-synaesthetes for each vowel.

To compare response consistency between synaesthetes and non-synaesthetes, Mann-Whitney U tests were conducted using SPSS version 15, testing the number of times a participant selected his or her most frequently chosen colour (or grey shade) for each vowel.

For example, if a participant chose green 9 times, cyan 3 times, yellow 2 times, and orange 2 times for vowel C1, he or she would score 9 for number of times the most frequent colour was selected for this vowel. A non-parametric test was chosen for this analysis because the data of synaesthetes and non-synaesthetes was not expected to be normally distributed.

Consistency was further addressed by exploring the random effects structure of the repeated-measures ANCOVA results, using F ratios to compare the between- and within-subject variances across participant groups. It also allowed for the comparison of both individual within-subject consistencies, and between-subject agreement within groups. The individual within-subject consistency was expected to be greater for synaesthetes than for non-synaesthetes and should be in agreement with the result of the Mann-Whitney U test. The between-subject agreement within groups describes the variation of consistency within the group of synaesthetes or non-synaesthetes. Here, I do not necessarily expect a greater consistency in the group of synaesthetes because of the idiosyncrasy of synaesthetic perceptions and because of the possibility that synaesthesia is a continuous condition.

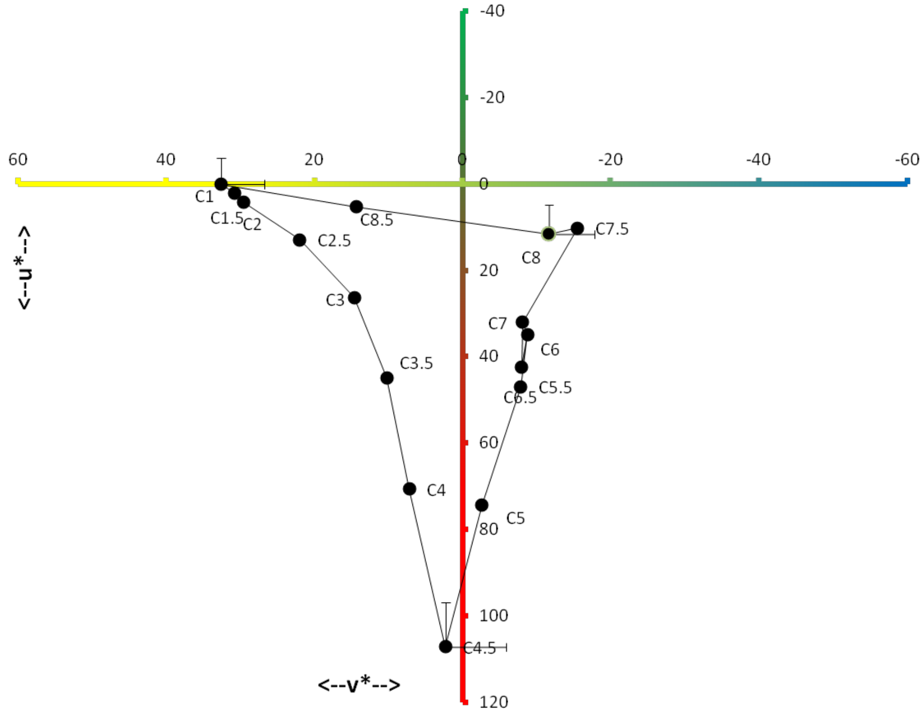
For some analyses, the synaesthetes were split into two subgroups based on their linguistic background: the Scottish synaesthetes and other synaesthetes. This division was made to analyse the influence of the native language or, more precisely, the accent of the participants. The group of the other synaesthetes was very heterogeneous, but the large variety of linguistic backgrounds in that group did not allow for a further subdivision.

3.3 Results

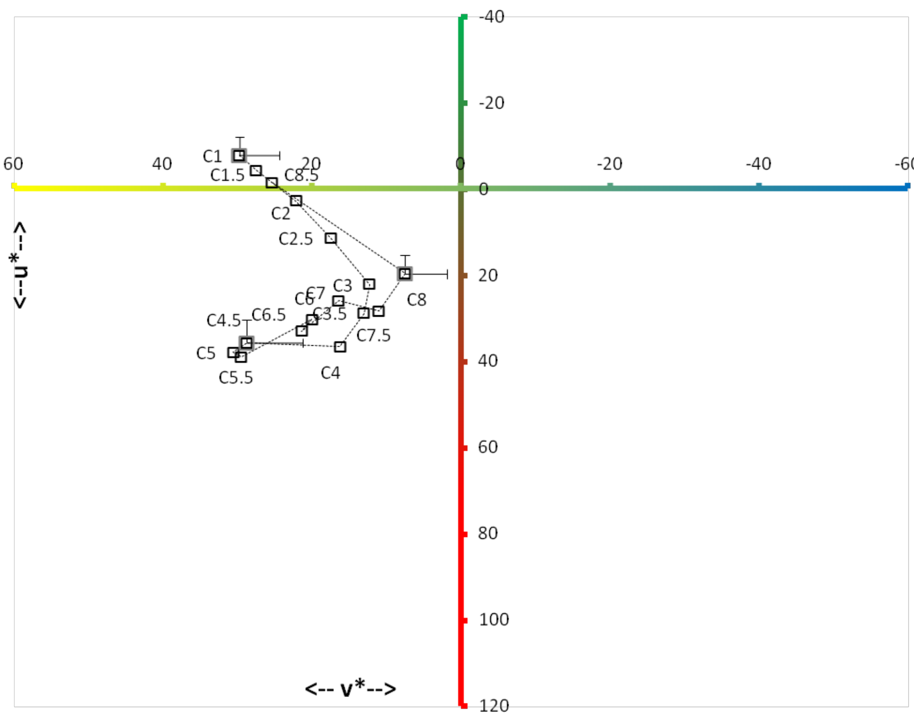
3.3.1 Colour associations

It was hypothesised that open vowels would be associated with reddish colours and front vowels with greenish colours (hypothesis 1). Figure 19 shows the averaged data of synaesthetes (a) and non-synaesthetes (b) for their colour choices per vowel projected onto the u^* and v^* axes of the CIELUV colour space. The colouring of the axes is given to guide the reader in the perception of the colour space: a high u^* represents reddish colours, a low u^* greenish colours. A high v^* represents yellowish colours, a low v^* bluish colours. Note that usually, the CIELUV colour space has a different spatial orientation: here, the space was rotated by 90 degrees and mirrored along the u^* axis. This was done to highlight the resemblance of the responses of the synaesthetes (Figure 19a) and the vowel quadrilateral shown in Figure 6 on page 44. Whiskers of 1 standard error (SE) were inserted for the vowels C1, C4.5 and C8 to show the variation of responses for these corner vowels.

While the average response of synaesthetes shows a strong similarity to the vowel quadrilateral – indicating a systematic correspondence between the acoustic or articulatory characteristics of vowel sounds and colour choices – the average responses of non-synaesthetes suggest less correspondence between the characteristics of vowel sounds and



(a) Synaesthetes.



(b) Non-synaesthetes.

Figure 19: Colour choices per vowel sound projected onto the CIELUV colour space, averaged across synaesthetes (a) and non-synaesthetes (b). Note that the colour space is rotated by 90 degrees and mirrored along the u* axis to resemble how the synaesthetes' responses are similar to the vowel space shown in Figure 6.

colour choices. In other words, non-synaesthetes seem to be less systematic in their colour associations with the vowel space than synaesthetes. There seems to be an influence of the frontness of vowels, or F2, on the v* scale which appears relatively strong for synaesthetes and weak for non-synaesthetes: a higher F2, as found in front vowels, seems to be associated with more yellowish colours, whereas a lower F2 seems to be associated with more bluish colours. F1, or the openness of vowels, seems to be more associated with the u* scale: open vowels with a high F1 are perceived as redder than close vowels, which are perceived as greener. The dispersion of the colour choices in the CIELUV colour space appears much larger for synaesthetes than non-synaesthetes.

Interestingly, not all of these group differences are reflected in the statistical analysis of the repeated-measures ANCOVA. As shown in Table 6, neither L* nor the hue variables u* and v* show a significant effect of participant group, and L* and v* fail to show any significant influence from vowel formants. However, there is an interaction of group and F1 for u*. This significant interaction reflects the much larger dispersion along the u*-axis of synaesthetes' responses to the different vowel sounds. As shown in Figure 19, synaesthetes' lowest average u* score is 0.1 and their highest is 107.2, whereas for non-synaesthetes the range only spans from -7.3 to 39. Additionally, both F1 and F2 have significant main effects on u*; the effect of F1 is much larger than that of F2 (F = 32.87 for F1 and F = 4.90 for F2). On average, redder colour associations were made across groups for vowel sounds with a higher F1, for example [a], and greener associations for vowel sounds with a lower F1, for example [i] or [o]. This confirms hypothesis 1. For F2, higher values induced slightly greener colour choices than lower values. However, the effect is not as strong as for F1 and is difficult to detect in the graphs.

3.3.2 Grey shade associations

It was hypothesised that, in general, front vowels would be perceived as light and back vowels as dark (hypothesis 2). Figure 20 shows the group averages of grey shade associations using luminance values (L*) with the different vowel sounds, for Scottish synaesthetes (red squares), other synaesthetes (orange circles), synaesthetes in total (pale pink triangles) and non-synaesthetes (blue diamonds). Vowel sounds are listed along the x-axis, starting from C1 and going around the vowel quadrilateral anti-clockwise. Luminance is displayed on the y-axis with lighter grey shades giving higher L* scores. The graph is based on the LOESS analysis of the L* values of the average vowel sound - grey shade associations per participant group. Although group differences occur, it can be seen that, overall, front vowels were associated with lighter grey shades than back vowels. This pattern is most strongly represented for Scottish synaesthetes (red) and weakest for other synaesthetes (orange). The weaker manifestation in other synaesthetes could be due to the heterogeneity of the group regarding participants' cultural and linguistic background. It is interesting

Experimental run	Measure	Effect	F	df	p
Colours	L*	group	0.72	1, 25.6	0.4032
		F1	1.28	1, 18.3	0.2727
		F1*group	4.07	1, 18.3	0.0586
		F2	2.82	1, 22.3	0.1071
		F2*group	0.05	1, 22.3	0.8210
	u*	group	0.60	1, 21.6	0.4479
		F1	**32.87**	**1, 32.4**	**<0.0001**
		F1*group	**4.56**	**1, 32.4**	**0.0404**
		F2	**4.90**	**1, 21.9**	**0.0375**
		F2*group	0.01	1, 21.9	0.9147
	v*	group	2.82	1, 30.9	0.1031
		F1	0.11	1, 21.2	0.7409
		F1*group	0.03	1, 21.2	0.8621
		F2	2.32	1, 17.6	0.1456
		F2*group	1.48	1, 17.6	0.2395
Greys	L*	group	2.23	1, 29.4	0.1462
		F1	**5.73**	**1, 22.7**	**0.0254**
		F1*group	0.90	1, 22.7	0.3538
		F2	**16.54**	**1, 24.7**	**0.0004**
		F2*group	1.23	1, 24.7	0.2787

Table 6: Effects of F1, F2, participant group and interactions on colour and luminance choices, from repeated-measures ANCOVAs. Significant results are in bold.

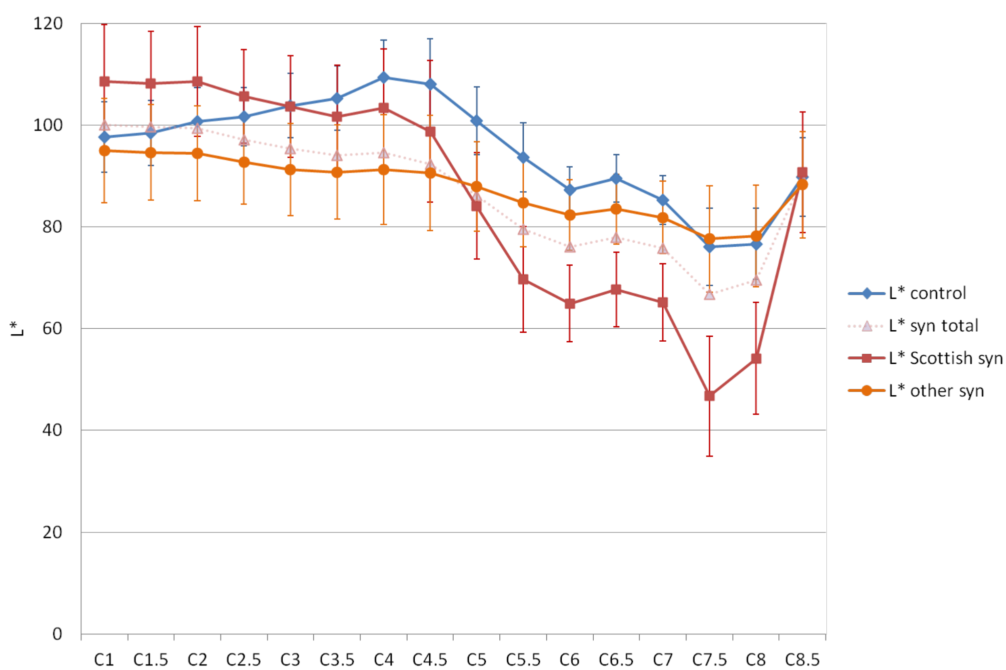


Figure 20: Luminance values of grey shade associations with vowel sounds for all synaesthetes, Scottish synaesthetes, other synaesthetes and non-synaesthetes. Error bars represent 1 SD.

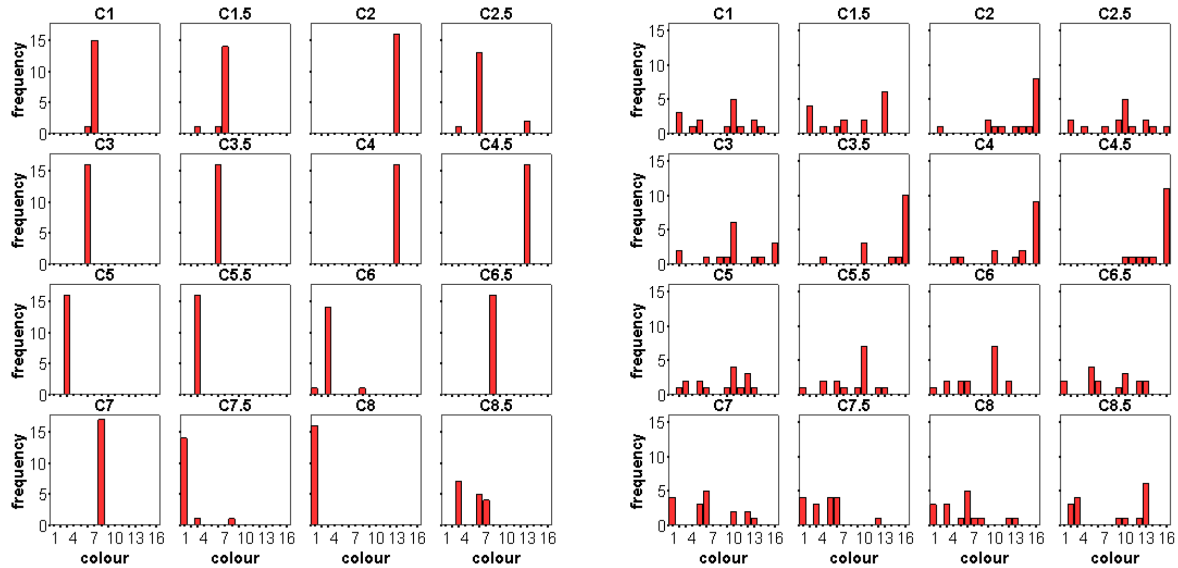
to note that, for non-synaesthetes, the lightest grey shades are actually associated with open vowels. It could therefore be suggested that the lightness was more influenced by F1 or the openness of vowels for non-synaesthetes, whereas it was more influenced by F2 or the frontness of vowels for synaesthetes. Nonetheless, all groups share the pattern of associating darker grey shades to (close) back vowels, where F1 and F2 are closest to each other and the intrinsic vowel pitch is presumably low. Considering luminance associations of individual participants explains some variations within the group of non-synaesthetes: while most participants followed the general pattern of associating lighter grey shades to front vowels and darker grey shades to back vowels, a few participants behaved differently. Four non-synaesthetes associated darker shades with front vowels and lighter shades with back vowels.

The effects of the formants F1 and F2, participant group and interactions thereof on luminance associations with vowel sounds were analysed using a repeated-measures ANCOVA. Results of this test are displayed at the bottom of Table 6. The analysis was conducted using two participant groups only, namely synaesthetes and non-synaesthetes, to preserve a large enough number of participants per group. No significant effects of participant group or interactions of participant group with formants could be detected. That is, participant groups did not differ significantly from each other in their luminance associations to vowel sounds when using a grey shade scale.

The repeated-measures ANCOVA returned significant results for both formants, indicating that both F1 and F2 had an influence on the luminance in grey shade associations with vowel sounds. Both a higher F1 and a higher F2 resulted in lighter grey shade choices, with the effect being stronger for F2 ($F = 16.54$ for F2, and $F = 5.73$ for F1). Knowing that open vowels have a high F1 and front vowels have a high F2 (cf. Figure 16b), the interpretation of the line chart in Figure 20 visualises these effects of the formants: front and open vowel sounds are associated with lighter shades than back vowels. Thus, hypothesis 2 is confirmed for the luminance of front and back vowels. It is an unexpected finding that open vowels, with their high F1 value, seem to be perceived as similar in lightness to (close) front vowels, with their low F1 and high F2 values: the relationship between lightness and the vowel sequence C1-C8 was hypothesised to be roughly linear. Figure 20 suggests that open vowels are somewhat lighter than close front vowels for non-synaesthetes, but slightly darker for synaesthetes, but as interactions between groups and formants were not significant, this pattern only suggests a tendency.

3.3.3 Consistency of colour and grey shade associations

To test whether participants associated similar colours and grey shades to each vowel sound consistently across the 16 stimulus presentations, Mann-Whitney U tests were conducted to compare the most frequently chosen colours and grey shades, and the random effects



(a) A synaesthete. (b) A non-synaesthete.

Figure 21: Consistency of colour choice in 16 repetitions per vowel C1-C8.5 for a typical synaesthete and non-synaesthete. Colours arranged in alphabetical order: 1 black, 2 blue, 3 brown, 4 cyan, 5 dark blue, 6 dark green, 7 green, 8 grey, 9 light green, 10 orange, 11 pink, 12 purple, 13 red, 14 rosa (pale pink), 15 white, 16 yellow.

of the repeated-measures ANCOVA for L*, u* and v* for colours and L* for grey shades were analysed.

Consistency of colour associations

Synaesthetes selected the same colour for repetitions of a given vowel sound between 4 and 17 times (mean = 12.4), while non-synaesthetes repeatedly selected the same colour for a given vowel sound between 3 and 16 times (mean = 7.8).[4] This is a significant difference ($U=20.5$, $z=-4.183$, $p<0.001$), thus establishing that synaesthetes are, indeed, considerably more consistent in their vowel sound - colour associations than non-synaesthetes.

Figure 21 displays the number of times each colour was chosen for each vowel sound by a typical synaesthete (panel a) and a typical non-synaesthete (panel b). In each panel, each box shows the responses for one vowel, C1 to C8.5. The x-axis lists the colours in alphabetical order. The height of the bars reflects how often a colour was picked for that vowel. The figure illustrates how, typically, synaesthetes were much more consistent in their colour choices with individual vowel sounds than non-synaesthetes. Although there is a preferred colour for many vowels for the non-synaesthete depicted in the graph, for example yellow for C4, she is still much less consistent than the synaesthete.

[4]Note that one vowel was presented 17 instead of 16 times, so that the type 1 index 1 sequence could be used correctly. Each vowel sound had to be followed by every other vowel sound, including by itself.

Another way of testing consistency in colour associations with vowel sounds is to analyse the effect of vowel formants F1 and F2 on L*, u* and v* values of participants' colour choices. For this parametric approach, the random effects structure of the ANCOVA results reported in Sections 3.3.1 and 3.3.2 was explored. Results of this test are given in Table 7. **Between-subject F ratios** describe the degree of between-subject agreement for synaesthetes and non-synaesthetes (i.e. whether the inter-individual within-group consistency of synaesthetes is different from the inter-individual within-group consistency of non-synaesthetes). "Between-subject mean" describes group differences in the intercept, i.e. the height of the L*, u*, v* values; "between-subject F1" describes influences of the first vowel formant on group differences; and "between-subject F2" describes influences of the second vowel formant on group differences. **Within-subject F ratios** describe residual group differences in within-subject consistency (i.e. whether the average intra-individual consistency of each synaesthete's responses is different from the average intra-individual consistency of each non-synaesthete's responses); these within-subject ratios relate more closely to consistency as usually defined in synaesthesia research. For both between and within-subject f ratios, $F < 1$ implies that synaesthetes are less consistent than non-synaesthetes, whereas an $F > 1$ implies that synaesthetes are more consistent than non-synaesthetes. Hence, F=1 implies equality of variance between the two groups.

The within-subject F ratio for L* is significant for both grey shades and colours. Interestingly, however, results of the colour associations are the opposite of what was found for grey shades. For grey shades, it was found that non-synaesthetes were slightly more consistent in their intra-individual luminance perception of vowel sounds than synaesthetes ($F < 1$), whereas for colours, it was vice versa: F=1.25 suggests that synaesthetes are less variable, i.e. more consistent, than non-synaesthetes in the luminance values of their colour associations with vowel sounds. This may indicate that non-synaesthetes used L* in the grey shade task as a scale, but could not use this dimension in the colour task. Also the between-subject F1 returned a significant F value for L* in the colour task; the low F value means that synaesthetes showed less between-subject agreement in the luminance values in response to F1, while non-synaesthetes were more consistent with one another of F1 on the luminance values of their colour choices.

Both within-subject F ratios for u* and v* are larger than 1, indicating that the non-synaesthetes show less intra-individual consistency than synaesthetes in their colour associations with vowel sounds. Therefore, the results confirm hypothesis 3 and are in line with the results of the Mann-Whitney U test. In contrast, inter-individual differences for the effects of F1 and F2 on between-subject consistency was smaller for synaesthetes than for non-synaesthetes, indicated by the small and significant F values for v* displayed in Table 7. This means that synaesthetes show larger idiosyncrasies in their colour perceptions than non-synaesthetes. Although this seems counter to what was expected at

Experimental run	Measure	Variance parameter	F	df	p
Colours	L*	Between-subject mean	0.62	19.3, 13.8	0.3322
		Between-subject F1	**0.21**	**17.1, 13.6**	**0.0028**
		Between-subject F2	0.42	18.7, 13.7	0.0832
		Within-subject	**1.25**	**5073, 3552**	**<0.0001**
	u*	Between-subject mean	0.38	18.7, 13.7	0.0551
		Between-subject F1	1.32	18, 12.4	0.6274
		Between-subject F2	0.40	18.9, 13.7	0.0674
		Within-subject	**1.09**	**5073, 3552**	**0.0044**
	v*	Between-subject mean	1.06	18.3, 13.4	0.9286
		Between-subject F1	**0.35**	**17, 13.6**	**0.0410**
		Between-subject F2	**0.17**	**16.5, 13.8**	**0.0010**
		Within-subject	**2.06**	**5073, 3552**	**<.0001**
Greys	L*	Between-subject mean	0.92	19.6, 13.7	0.8522
		Between-subject F1	0.45	19.1, 13.7	0.1027
		Between-subject F2	0.57	19.4, 13.7	0.2497
		Within-subject	**0.90**	**5075, 3553**	**0.0010**

Table 7: Results of consistency analyses comparing synaesthetes and non-synaesthetes using the random effects obtained in repeated-measures ANCOVAs assessing relationships of F1 and F2 with L*, u* and v*. The "F" column gives F ratios for the differences in variance between participant groups for the intercept ("between-subject mean"), for specific fixed effects ("between-subject F1, F2") and residually ("within-subject"). An F ratio of 1 implies equality of variance between participant groups; $F < 1$ implies that synaesthetes are more variable than non-synaesthetes; $F > 1$ implies that non-synaesthetes are more variable than synaesthetes. Significant differences between participant groups are in bold.

first sight, it is in fact a logical result: synaesthetes show a high within-subject consistency and highly personalised colour perceptions. Thus, there should be more variation in colour choices within the group of synaesthetes than non-synaesthetes. Because non-synaesthetes chose colours less consistently in general, the between-subject variation is less strong in their group.

Consistency of grey shade associations

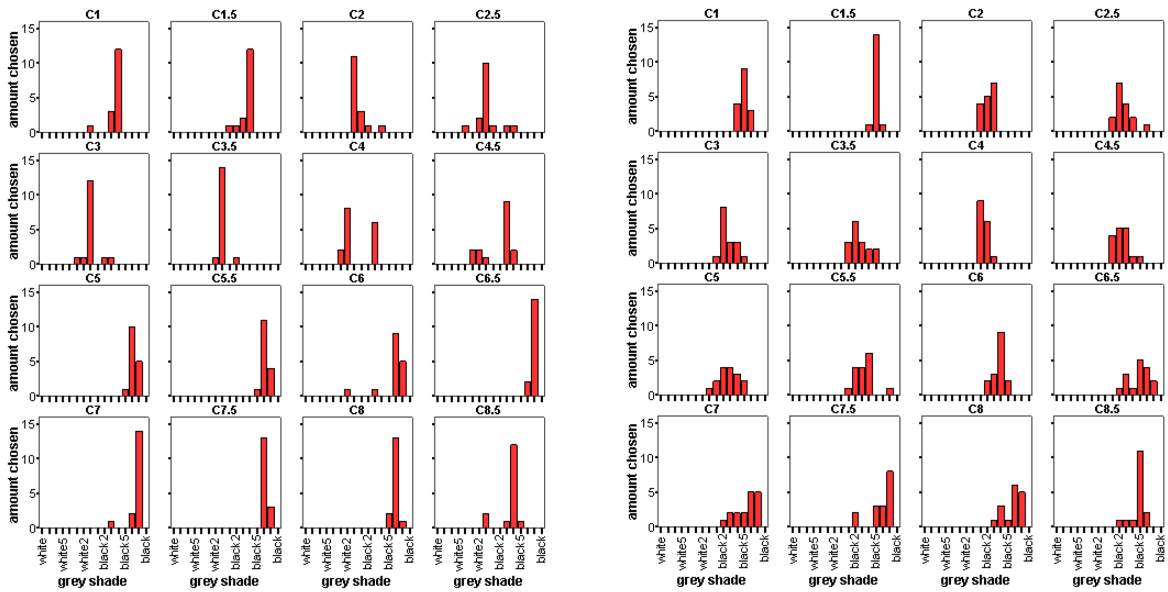
Synaesthetes selected the same grey shade for repetitions of a given vowel sound between 2 and 17 times (mean = 9.2), while non-synaesthetes repeatedly selected the same grey shade for a given vowel sound between 2 and 16 times (mean = 7.5). This is not a significant difference (U=93.5, z=-1.628, p=0.1), thus establishing that synaesthetes and non-synaesthetes chose the most frequent grey shade for a particular vowel sound with a similar degree of consistency. The similar behaviour across groups may be explained by the fact that the display represented a grey scale, so synaesthetes and non-synaesthetes could access similar processing mechanisms for the associations.

Figure 22 displays the number of times each grey shade was chosen for each vowel sound by a typical synaesthete and a typical non-synaesthete. In contrast to the consistency comparison of colour associations in Figure 21, the grey shade responses of the two participants are relatively similar. Although synaesthetes and non-synaesthetes did not differ significantly in their most frequently chosen grey shades, the data of the two participants shown in Figure 22 are representative of the trend for non-synaesthetes to use a slightly broader range of neighbouring grey shades.

The effects of a vowel's F1 and F2 values on the luminance (L*) values of participants' responses, according to the ANCOVA, are summarised in Table 7. Only the within-subjects F ratio is significant for grey shades (shown in the bottom part of the table): intra-individual consistency of non-synaesthetes' associations of a given vowel sound with a given grey shade is slightly larger than the intra-individual consistency within the group of synaesthetes ($F = 0.9$), when L* is taken as measurement for the grey shades and formants for the vowel sounds. This slightly more differentiated method than that of the Mann-Whitney-U test came to a different conclusion, although the group difference is rather weak.

3.3.4 Within-group differences

Consistency within the group of non-synaesthetes One supposition was that performance consistency of non-synaesthetes is better in the grey shade associations than in the colour associations, because the grey shades represent a scale, whereas colour associations are essentially arbitrary. Indeed, most non-synaesthetes felt more comfortable with the grey shade task than with the colour task. They found it easier to match vowels to a



(a) A synaesthete.

(b) A non-synaesthete.

Figure 22: Consistency comparison of a prototypical non-synaesthete's and synaesthete's grey shade associations with vowel sounds.

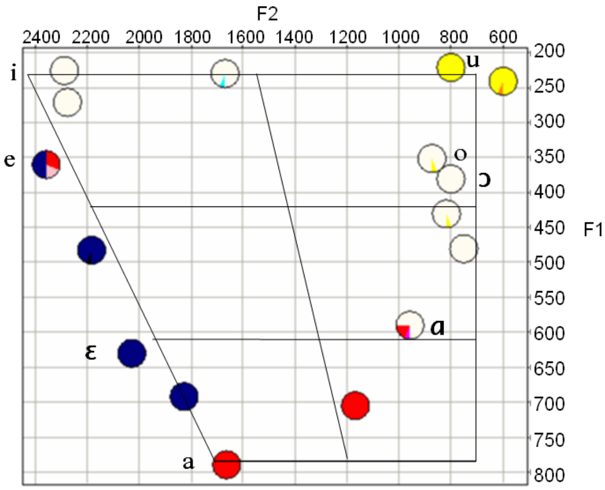
white to black continuum than to random colours. However, the consistency scores did not reflect their preference: participants were no more consistent in their responses in the grey scale experiment than they were in their responses in the colour experiment. On average the non-synaesthete group chose the same grey shade 7.5 out of 16 times for each vowel sound compared to 7.8 for colours, which is a non-significant difference ($U=148$, $z=-1.407$, $p=0.159$). Thus, the impressionistic judgements of the participants are not in line with the consistency of their responses. This suggests that colour associations might be accessed subconsciously, whereas the grey shade choices represented a consciously accessible luminance scale.

Consistency within the group of synaesthetes One might expect that synaesthetes would have less consistent grey shade associations than colour associations with vowel sounds because crucial aspects of colour perception, namely hue and chroma, were taken away for the grey shades display. Indeed, most synaesthetes reported that they found it more difficult and less enjoyable to do the task with grey shades than with colours. Although the available colours often did not exactly match their synaesthetic colour associations, the colour choice still offered an approximate match to their synaesthetic perceptions. The grey shade display, on the other hand, was incongruent with their synaesthetic perceptions because it lacked essential criteria of hue. These feelings and influences are reflected in their consistency performance: synaesthetes were much less consistent in their replies to the grey scale compared to the colours. On average this group chose the same grey shade 9.2 out of 16 times for each vowel sound compared to 12.4 for colours ($U=46$, $z=-2.39$, $p=0.017$).

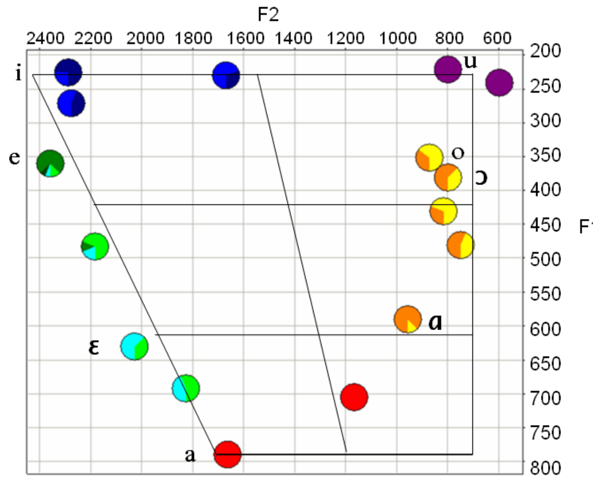
Individual responses of synaesthetes As expected, the colour associations of individual synaesthetes were highly consistent, confirming hypothesis 3. Nevertheless, associations and consistency varied across participants within the group. Figure 23 displays the colour choices for each vowel sound for three synaesthetes who represent the range of consistency observed in the group of synaesthetes. The vowel quadrilateral is superimposed on the F1-F2 formant space to make it easier to relate the data points to the vowel sounds. Each pie chart is positioned according to the formants of the 16 vowel sounds starting with C1 (= [i]) in the top left corner, going around the vowel quadrilateral anti-clockwise up to C8.5 in the top centre. A pie chart indicates the proportion of colours chosen by a participant for all 16 repetitions of this particular vowel sound.

Some participants only used a total of three or four different colours for all vowel sounds, as shown in Fig. 23a. Participant 6 chose the same colour for a number of neighbouring vowels resulting in a block of one colour for a series of adjacent vowels. Other participants sometimes chose two colours for one vowel equally often, resulting in "bi-coloured" vowels (see Fig. 23b). This could indicate that their synaesthetically perceived colour lay in-between two offered colours, so they picked them alternately. An alternative interpretation is that they did not perceive the vowel as carrying the same colour in every trial, for example depending on the previous vowel sound. [o] might have triggered yellow in the context of a previous [a] but orange in the context of a previous [u]. Most often, two to three neighbouring vowels were associated with the same colour, resulting in the use of approximately six different colours, as shown in Figure 23c. This sub-figure represents a common distribution of the colour associations across the vowel quadrilateral regarding the number of colours used and the groupings of colours for adjacent vowel sounds.

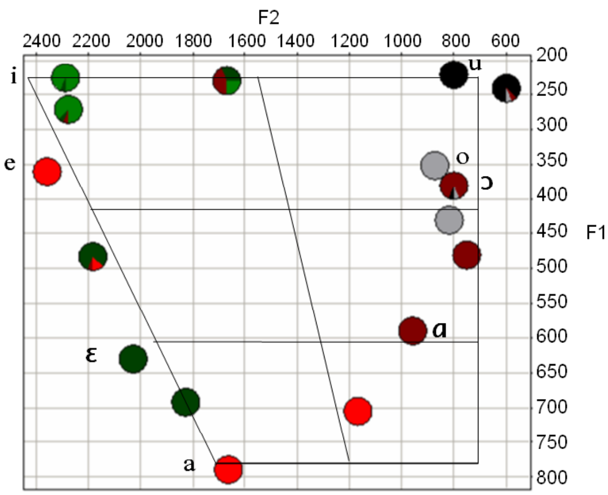
Scottish synaesthetes As previously mentioned, to tentatively explore the influence of participants' cultural and linguistic background, the synaesthetes were split into 'Scottish' and 'other' synaesthetes. The group of Scottish synaesthetes consisted of six participants and is undoubtedly the more homogeneous of the two groups. In a similar fashion to the previous Figure, Figure 24 shows the responses of the six Scottish synaesthetes in one coloured vowel quadrilateral. Each pie now corresponds to $6 * 16 = 96$ responses. Although colours varied more than for the individuals presented in the previous section, a general trend is clearly visible: front vowels were associated with greenness, open vowels with redness and back vowels with blueness. An interesting occurrence is the colouring of [e]. Although it was generally the trend that adjacent vowel sounds were associated with similar colours, [e] built an "island" of red between its otherwise mainly green neighbours.



(a) Synaesthete 6: blocks of colours for adjacent vowels.



(b) Synaesthete 12: "Bi-coloured" vowels.



(c) Synaesthete 2: typical number and distribution of colours.

Figure 23: Coloured vowel quadrilaterals of three different synaesthetes, projected onto the reversed F1-F2 space with superimposed schematic vowel quadrilateral.

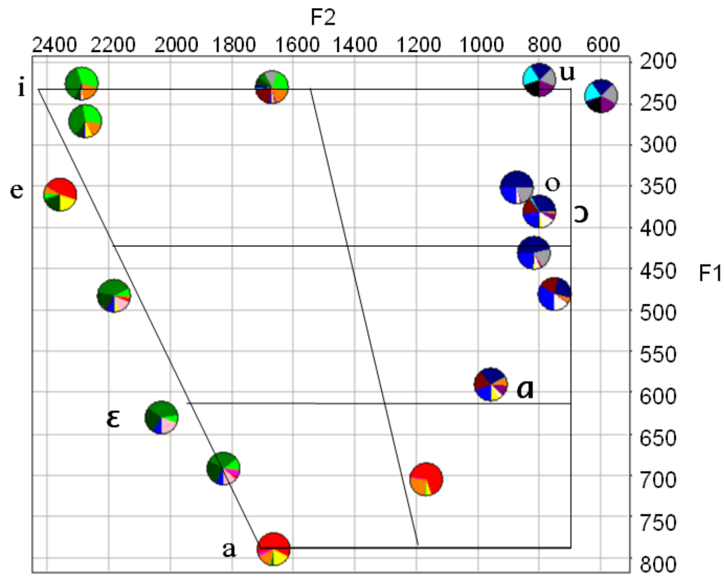


Figure 24: Coloured vowel quadrilateral for six Scottish synaesthetes, projected onto the reversed F1-F2 space with superimposed schematic vowel quadrilateral.

This phenomenon may have a linguistic explanation: in Scottish English, the standard pronunciation of the letter <a> is not a diphthong as in Standard Southern British English, [ɛɪ], but rather a monophthong close to [e]. Thus, it is reasonable to believe that hearing the vowel sound [e] makes the Scottish synaesthetes think of the letter <a>, which is often perceived as red (Simner, 2007).

To pursue this possibility, a short test was conducted with ten Scottish participants, who did not take part in any of my other experiments. They listened to all 16 vowel sounds used in this experiment and were asked to assign a vowel grapheme to the sound (forced choice of <aeiou>). Associations of the graphemes to the vowel sounds are depicted in the vowel quadrilateral in Figure 25.[5] Indeed, all ten participants associated the vowel sound [e] with the letter <a>. C4 and C4.5 were also associated with <a> by all participants. This confirms the assumption that [e] might share the same associations as [a] for Scottish participants. This pattern could not be found in other synaesthetes.

Table 8 shows colour associations by the six Scottish synaesthetes with the graphemes <a, e, o> derived from the questionnaire by Simner and Ward (Appendix D) and those associated most often with the vowel sounds closest to the pronunciation of these letters, i.e. [e, i, o]. Most agreement between and within participants was found for <e> and [i]. For <a> and [e], four participants showed within-subject agreement between grapheme and corresponding speech sound, whereas one showed completely different colour associations; for one, the colour association for <a> (and <o>) is missing. For <o> and [o], there was no within-subject agreement. This finding indicates a co-existence of graph-

[5]If one grapheme is displayed, at least 70% of the participants chose that vowel. If two graphemes are shown, each one was chosen at least by 40% of the participants.

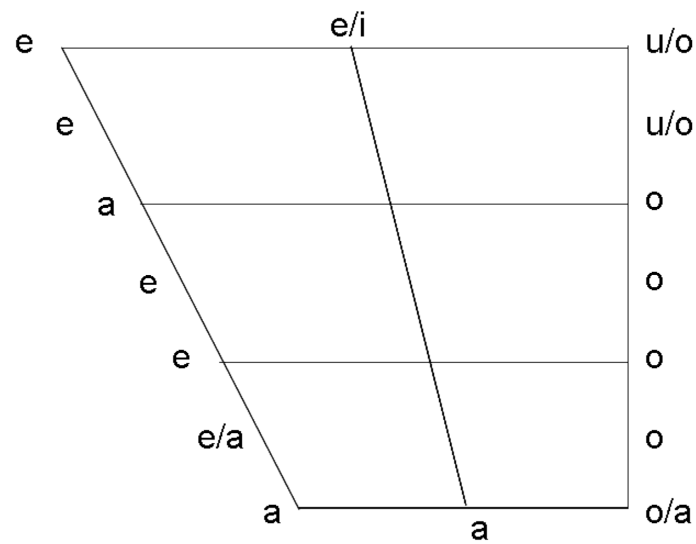


Figure 25: Vowel quadrilateral with the graphemes that Scottish English speakers associate with the vowel sounds.

ID	<a>	[e]	<e>	[i]	<o>	[o]
1	yellow	yellow	bright orange	orange	bright pink	dark blue
2	red	red	green	green	black	grey
3	scarlet	dark green	jade green	light green	white	blue
5	n/a	red	blue	light green	n/a	purple
9	yellow	yellow	dark blue	cyan	white	dark green
14	orange-yellow	orange, red	light green	green	grey	dark blue

Table 8: Colour associations with graphemes and with corresponding Scottish pronunciations by Scottish synaesthetes.

emic and phonemic influences for some vowel sounds, whereas for other vowel sounds, acoustic-phonetic influences seem to dominate.

3.4 Discussion

3.4.1 Summary

This experiment explored vowel sound - colour associations in synaesthetes and non-synaesthetes. Participants listened to 16 different vowel sounds and chose out of a range of 16 different colours and grey shades those that they thought matched the vowel sound best. It was hypothesised that (1) open vowels would be associated with red colours and front vowels with green colours, (2) front vowels would be perceived as light and back vowels as dark, (3) synaesthetes and non-synaesthetes might have similar patterns in associating colours with vowels but synaesthetes would be more consistent. The main results of this experiment can be summed up as follows:

1. There was a significant influence of the vowel formants F1 and F2 on the colour associations, so that u* – a measure of greenness to redness – reached higher, i.e. redder, values with decreasing F2 and particularly strongly with increasing F1. As open vowels have the highest F1, with C4 having the highest F1 in my stimuli, the findings confirmed the first part of hypothesis 1: open vowels are associated with red colours. F2 is higher in front vowels than back vowels. As u* decreased with rising F2, it confirmed the second part of hypothesis 1: front vowels are associated with greenish colours. No significant results were found for v* which represents the blue-yellow axis. The influences of F1 and F2 on u* hold true across participant groups, but the influence of F1 was stronger for synaesthetes than for non-synaesthetes. Although individual differences in colour associations with vowel sounds were found, the systematic acoustic influences of the vowel sounds nonetheless proved to be significant, on average.

2. There was a significant influence of the vowel formants F1 and F2 on luminance associations. L* increased with both increasing F1 and increasing F2 in the grey shade condition and showed a trend to do so in the colour condition. As F2 is higher for front vowels than back vowels, this confirmed hypothesis 2: front vowels are perceived as light and back vowels as dark. The weaker influence of F1 requires a more precise statement of these results: the openness of vowels also influences luminance perception, indicating that open front vowels, in particular, are perceived as lighter than back vowels, which have both a lower F1 and a lower F2. The influence of F1 and F2 on L* in grey shades holds true across participant groups.

3. There was a strong significant difference in the consistency of associating colours

Synaesthetes				Non-synaesthetes		
Day, 2004	Marks 1975	my data		my data	Simner et al. 2005	Wrembel 2007
black	*yellow*	green	i	*green*	*blue*	*yellow*
yellow	*white*	blue		pink	*yellow*	*green*
white				*blue*		
white	**yellow**	red	e	*blue*	*pink*	*green*
yellow	*white*	green		*pink*		*blue*
grey		**(yellow)**		*green*		
red	**red**	**red**	a	blue	*red*	*red*
(brown)	blue			green		
				pink		
white	*red*	*blue*	o	*blue*	yellow	*orange*
black	*black*	*red*		green	*orange*	brown
red, blue				(pink)		*blue*
blue	**blue**	purple	u	green	**blue**	**blue**
green	brown	**blue**		**blue**	*brown*	*brown*
	black	grey		pink		

Table 9: Comparison of vowel sound - colour associations across studies. Agreement of two studies within a participant group printed in italic, agreement of three studies in bold. The 11 basic colour terms are used.

with vowel sounds: synaesthetes were much more consistent in their colour associations than non-synaesthetes, meaning that they repeatedly chose the same or similar colours for individual vowel sounds. However, no robust differences in the consistency of associating grey shades with vowel sounds between participant groups were found.

3.4.2 Comparing vowel sound - colour associations across studies

Results confirmed a systematicity of colour associations with vowel sounds, largely as predicted. To assess whether these results can be generalised, they need to be compared with findings of other studies. Table 9 gives an overview of results of studies looking at vowel sound - colour associations of synaesthetes and non-synaesthetes. All of these studies covered at least the five phonemes /ieaou/. The cells of the table contain up to three colours that were associated most often with the corresponding phonemes per study. Vowel sound - colour associations that could be found in all three studies per participant group are printed in bold; those associations that could be found in two of the three studies are printed in italics.

It has to be noted that, while I used cardinal vowels, the other studies used vowels occurring in the languages examined that come close to those cardinal vowels but do not necessarily resemble them. Wrembel (2007) clarifies this issue by superimposing the vowel sounds onto the vowel quadrilateral used in the IPA (cf. Figure 13 on page 60), visualising that /e/ is actually [ɛ] for Polish, and /o/ is actually [ɔ], as also shown in Warth (2007). The data of Simner et al. (2005) represents grapheme-colour associations of native speakers of German. Here, a close match with cardinal vowels can be expected. Day (2004) and Marks (1975) do not define the vowel sounds further, using only transcriptions. Since different language backgrounds are used in the different studies, a certain variation in vowel quality has to be assumed, but probably does not dramatically affect the results. Another discrepancy among studies is in the means of obtaining the participants' colour choices. In some cases, the options were displayed visually (the present study, Wrembel, 2007) and in other cases, the participants were asked to name their colour choices (Simner et al., 2005); in still other cases, the methods of obtaining the participants' responses are unclear (Day, 2004, and Marks, 1975). Finally, the studies differ in the number of analysed colours: Marks (1975) is based on nine colours, whereas Day (2004); Simner et al. (2005) and Wrembel (2007) are based on the 11 focal colours (Berlin & Kay, 1999). For the comparisons in Table 9, the colour categories of the present study were also reduced to the 11 focal colours. Thus, the category *blue* now combines blue, dark blue and cyan; the category *green* combines green, dark green and light green; *pink* includes both pink and rosa (or pale pink). This means there is a risk that these categories might be more frequent because they were represented by two or three colour patches in the response display instead of solely by one.

Leaving aside these methodological differences, it can be seen that only two vowel sounds share the same most frequent colour associations across participant groups and across studies (cf. Table 9): [u] is often perceived as blue and [a] as red (or in the case of non-synaesthetes in this study, pink). Although there is no agreement between synaesthetes and non-synaesthetes on the front vowel [i], similarities can be found as both groups mainly chose light colours such as white and yellow. Within the group of synaesthetes, there is additionally strong agreement that [e] is yellow, and more generally that front vowels are associated with light colours such as white and yellow, while back vowels are associated with dark colours such as blue, brown and black. Within the group of non-synaesthetes, less regularity is found in the colour patterns, although [o] is mostly perceived as blue or orange.

For analysing both the inducer and the concurrent, previous studies have used categorical data, i.e. colour categories and vowel categories. None of the studies used for comparison here (Day, 2005; Marks, 1975; Simner et al., 2005; Wrembel, 2007) defined the focal colours according to their Munsell values, nor in fact any other way. The lack of acoustic definition of the vowel sounds used in the studies further hamper detailed analysis and comparisons. While Wrembel (2007) at least defines the speech sounds in the articulatory space of the vowel quadrilateral used in the IPA, Day (2004) and Simner et al. (2005) do not give any specification. To circumvent disadvantages associated with a purely qualitative analysis, colorimetry was used in this study for the concurrent, allowing for more accurate colour measures and giving the option of a more detailed analysis, and acoustic formant measures for the inducer, introducing the option of relating vowel sounds to each other in the vowel space. Thus, the quantitative and acoustic analyses of my study show the systematic influences of acoustic characteristics of vowel sounds on colour associations in a physically quantifiable way.

3.4.3 Influences of both speech sounds and graphemes

Most of the synaesthetes taking part in this study considered themselves to have letter-colour associations rather than vowel sound - colour associations. Yet, they showed systematic, fine-grained influences from the acoustic-phonetic structure, as described above. Is grapheme-colour synaesthesia therefore fundamentally a phonetic phenomenon? In fact, there seem to be influences of both graphemic representations and acoustic-phonetic sound structure. Suggestive supporting evidence comes from participants' responses to different sounds that can correspond to the grapheme <a>. In Scottish English, the commonest pronunciation of <a> in words like *hat* resembles C4.5; but the name of the letter <a>, and its pronunciation in words like *hate*, resembles C2, or [e]. For several Scottish synaesthetes in my sample, C2 had the same colour as C4.5, and formed a "colour-island" distinct from the more finely-graded colours of phonetically-neighbouring vowels, suggesting its colour was influenced by the pronunciation of <a>. Thus, I propose that acoustic-phonetic influences co-exist with established effects of graphemes, such as graphemes' frequency of occurrence (Simner et al., 2005; Beeli et al., 2007; Smilek et al., 2007). This graphemic influence seems at times to override the general tendency of synaesthetes to map the vowel space onto the u* v* colour space, at least in cases where an easy vowel sound - grapheme match was possible.

If phonetic and graphemic influences do co-exist, the presentation modality in specific experiments may be key to understanding the balance among them. Studies presenting the stimuli as graphemes will inevitably trigger stronger graphemic influences, whereas studies presenting the stimuli auditorily will trigger stronger acoustic-phonetic influences. Online or paper surveys eliciting responses to visually presented letters usually use few repetitions of the stimuli. They also cannot cover the various different pronunciations of letters, unless embedded in words, in which case contextual influences may add to the synaesthetic perceptions. The multiple presentations of subtly-varying sound qualities in this study were a good method to detect acoustic-phonetic influences on colour associations.

3.4.4 Commonalities between synaesthetes and non-synaesthetes, and sound symbolism

Although my results show differences in the vowel sound - colour associations between synaesthetes and non-synaesthetes, similarities could nonetheless be found. Agreement was found in the grey shade experiment in that front vowels were associated with lighter grey shades than back vowels by all participant groups – a pattern that previously has been found, for example by Marks (1974); Spence (2011); Wrembel (2010) and Beeli et al. (2007). These findings also accord with findings that lightness associates with higher musical pitch and "clearer" timbre (e.g. Ward et al., 2006; Marks, 1975). Here, a parallel can be drawn between the intrinsic vowel pitch represented by F2 and musical pitch. Although pitch-lightness correspondences have been frequently documented, it is unclear whether they result (1) from structural similarities between neural processing of different stimulus dimensions, such as stimulus intensity or the sensory modality in which they were presented; or (2) from environmental statistical regularities, such as the way the resonance of objects depends on their size; or (3) from semantically-mediated associations, for example the way that words such as *low* and *high* refer to both pitch and elevation/situation in space (Spence, 2011).

Congruities between participant groups were also found in the colour domain. Even non-synaesthetes showed group patterns of consistent vowel sound - colour associations, suggesting similar underlying processes or strategies that drive these associations. Other studies have come to similar conclusions. Ward et al. (2006), for example, tested synaesthetes and controls on their musical note-colour associations, and additionally, synaesthetes on their orientation of spatial attention: they had to detect a coloured square, accompanied by a musical note played alongside which was congruent or incongruent to their synaesthetic perception. The authors found that both groups shared a strategy for sound - colour mapping like pitch-lightness correlation, but that synaesthetes were more specific and more consistent in their colour choices, and slower in the incongruent Stroop task condition. They concluded:

> This suggests that synaesthesia uses the same (or an analogous) mechanism of exogenous cross-modal orienting as normal perception. Overall, the results support the conclusion that this form of synaesthesia recruits some of the same mechanisms used in normal cross-modal perception rather than using direct, privileged pathways between unimodal auditory and unimodal visual areas that are absent in most other adults. (Ward et al., 2006, p. 264)

Similar patterns of systematic colour associations were found by de Thornley Head (2006), who found that controls projected tones onto a lightness scale with random hue attribution. Synaesthetes were significantly more consistent in their colour associations to tones than controls. Interestingly, the synaesthetes associated quartertones with colours which neighboured the RGB values of the adjacent semitones. The quartertones can be considered analogous to the morphed vowels used in this study. Similar to the colour data distribution de Thornley Head found for the associations with quartertones, synaesthetes partaking in my study also chose adjacent colours in colour space for morphed and other adjacent vowels as displayed in Figure 19a on page 78.

The relation between F1 (or openness) and the redness-greenness axis allows for interpretation at various levels. The most universal result found was the association of open vowels with red colours. Many species make themselves appear "larger" both visually and auditorily to threaten adversaries (Harris et al., 2006). Mouth and jaw opening, which contribute to perceived threat, increase both frequency and intensity of F1, as well as overall vowel intensity (Fairbanks et al., 1950). Participants in Wrembel (2010) also rated the near-open vowels /æ/ and /ɜ/ as aggressive in opposition to quiet (/a/ was not presented in her study). The association of open vowels with red might thus reflect a shared semantic association with threat, dominance or warning (Humphrey, 1976). A related phenomenon occurs as sound symbolism, or phonaesthesia, where vowels such as [a] seem to have an augmentative force as in *vast* or *large*, whereas [i] has a diminutive force as in *wee* or suffixes such as *-ling* and *-ie* (Reay, 2006). As something large is naturally more threatening than something small, this could explain the association with red. It may also be possible to support the theory of dominance and threat at a psychophysical level. There is a tendency for red stimuli to appear closer in space to the perceiver than non-red stimuli (Simmons, 2011). At the same time, [a] is perceived as larger (Wrembel, 2010), while [i] is perceived as smaller (Reay, 2006). Thus, the interplay of large and threat seems to support the [a]-red association.

Alternatively, /a/ is one of the commonest vowels cross-linguistically (Liljencrants & Lindblom, 1972) and red is one of the most prototypical colours (Kay et al., 2009). This statistical correspondence might underlie their association, as has been found for grapheme frequency in English and German (Simner et al., 2005; Beeli et al., 2007).

Perhaps all these alternatives have some influence on these cross-modal associations, depending on the context or on the individual experiencing them; or the influences may be intertwined. The environmental regularities of size and resonance of objects and associated emotions may have an influence on neural processing, resulting in cross-modal activations. Those, in turn, may have formed semantics and meanings of words, so that we now use words such as low and high for both pitch and situation in space. In fact, the language that somebody speaks has an impact on these associations – no completely universal sound symbolism has been found yet (Reay, 2006). Rather, related languages seem to share sound symbolism. It would therefore be interesting to research less common languages of less known cultures to find out whether their vowel sound - colour associations are similar to ours.

In conclusion of this chapter, systematic influences of acoustic-phonetic influences on colour associations with vowel sounds have been found. These findings were related, amongst other things, to musical pitch - colour associations, indicating that the influence of acoustics is not restricted to vowel sounds but extends to other topics involving sound. This leads us to another area involving sound and speech, namely that of voices. Synaesthetes have reported synaesthetic perceptions induced by the sounds of voices. Often, the inducer 'voice' seems to trigger more concurrents than merely colour. Thus, the experiment presented in Chapter 5 researched this more complex type of synaesthesia, in which acoustic parameters of voices are taken into account as inducers, and colours and textures as concurrents. As the use of textures in the form of visual displays is a new method in synaesthesia research, a small additional study was carried out which was designed to facilitate analyses of voice-texture associations and which is described in the following chapter.

4 Interlude: Texture rating experiment

4.1 Introduction

The sounds of voices often evoke texture perceptions in addition to colours, as has been reported by synaesthetes. Examples of synaesthetic descriptions of voices including texture are: "smooth but granulated voice", "dark grey, ribbed, touches of white hiss around the sides", "grey voice with slightly rich but soft centre and very slight fuzziness around the outside", "Well, any kind of voice, even dogs barking, I get a similar kind of 'picture' that is like a 'voiceprint"'. Despite our knowledge of perceptions like these, no systematic investigation of texture perceptions in synaesthesia has yet been conducted – which is perhaps not surprising considering that it is not easy to quantify texture and to relate this quantification to perceptual categories (Clarke et al., 2011; Petrou et al., 2007). Taking into account how variable synaesthetic texture concurrents can be makes it even more difficult to analyse the associations in a systematic and quantitative fashion. This quantification, however, is essential in order to find systematic relations between acoustic features of voices and their texture concurrents.

Clarke et al. (2011) attempted to define the relation of quantitative texture features and perceptual categories by correlating perceptually grouped texture sets with those classified by four different texture classification algorithms. In their perception experiment, 30 participants were asked to group a set of 334 textures into an unspecified number of subgroups comprised of similar textures. Results of human groupings compared with classifications made by machine-driven measures showed that inter-class distances did not correlate well between the classifications made by humans and machines. Petrou et al. (2007) compared the human groupings of 56 textures rated by 11 participants with machine-driven measures and also found that different mechanisms/criteria were used to rank textures in each case. Since studies on synaesthesia mainly deal with perceptual data, it was therefore decided to quantify the textures used here by solely perceptual data, gathered from participants who did not take part in the synaesthesia experiment proper. The results of experiments conducted by Rao & Lohse (1993, 1996); Rao et al. (1996) and Tamura et al. (1978) served as the basis for the rating task as described in the Methods section. Rao & Lohse (1993) asked their participants to group the photographs of Brodatz (1966) into as many classes as desired. Based on various statistical methods, the authors labelled the classes that emerged from hierarchical cluster analysis as *repetitive*, *oriented/directional*, *disordered* (i.e. neither repetitive nor directional), *grainy* and *unclassified*. Later, Rao & Lohse (1996) identified three dimensions with the following criteria: *repetitive vs. non-repetitive*; *high-contrast and non-directional vs. low-contrast and directional*; *granular, coarse and low-complexity vs. non-granular, fine and high-complexity*. Tamura et al. (1978) qualitatively identified the dimensions *coarse vs. fine*, *high contrast vs. low con-*

trast, *directional vs. non-directional*, *linelike vs. bloblike*, *regular vs. irregular* and *rough vs. smooth*. These dimensions were used to create the semantic differentials in a way that seemed most appropriate for the selection of textures offered in this experiment, as described in Section 4.2.3.

The results of the texture rating experiment outlined in this chapter form part of the analyses of the voice description experiment described in the following chapter, where participants were asked, among other tasks, to pick textures from a choice of 16 to associate with the voices. Therefore, this piece of research contributes indirectly to the overall aim of analysing synaesthetes' and non-synaesthetes' associations of voice quality with texture.

4.2 Methods

4.2.1 Materials

The image selection for the experiments was based on synaesthetes' verbal descriptions of their textural perceptions which were provided in emails from Sean Day's international *synesthesia list* (http://www.daysyn.com/Synesthesia-List.html) and via personal communication with me. The textures most often named were: rough, velvet, liquid/fluid, smooth, shiny, hard, dry, soft, bumpy, sharp, bubbly, milky, transparent, metallic, and textiles like linen, flannel, corduroy, plaid, and felt. Sometimes it is debatable whether a descriptive term can be regarded as a texture or not. Shiny, for example, describes a surface that is so smooth that it reflects the light. So, indirectly, it contains information about textural consistency. That is why descriptions like these were also taken into account. Based on these descriptions, a set of 16 textures was created. Each texture was designed to be close to the descriptive words used by the synaesthetes, but distinct from the other textures. The selection is depicted in Figure 26.

Specific restrictive regulations, or uniformity criteria, were made to control for comparability between texture images: textures were left as grey-scale images to ensure that participants' choices were based on texture rather than colour. Also, textures were uniform with respect to their simulated viewing angle. Some pictures were used from Fraser Halley and colleagues (Clarke et al., 2011; Halley, 2011), namely *rough-ish*, *jeans* and *stripes*, as each of these was photographed under the same conditions and standardised in terms of dimension, quality etc. *Rough* and *bumpy* were taken from Brodatz's well known photographic database of textures (Brodatz, 1966). Unfortunately, there was not enough variability in these two sets to cover the synaesthetes' various descriptions. Thus, images of other textures collected from the internet[6] were added. These were selected to be as close to the uniformity criteria as possible. An example of a violation of a uniformity criterion would be the folding pattern in the *velvet* picture, as here a three dimensional

[6]These were taken from homepages without copyright limitations, and manipulated to accommodate the uniformity criteria.

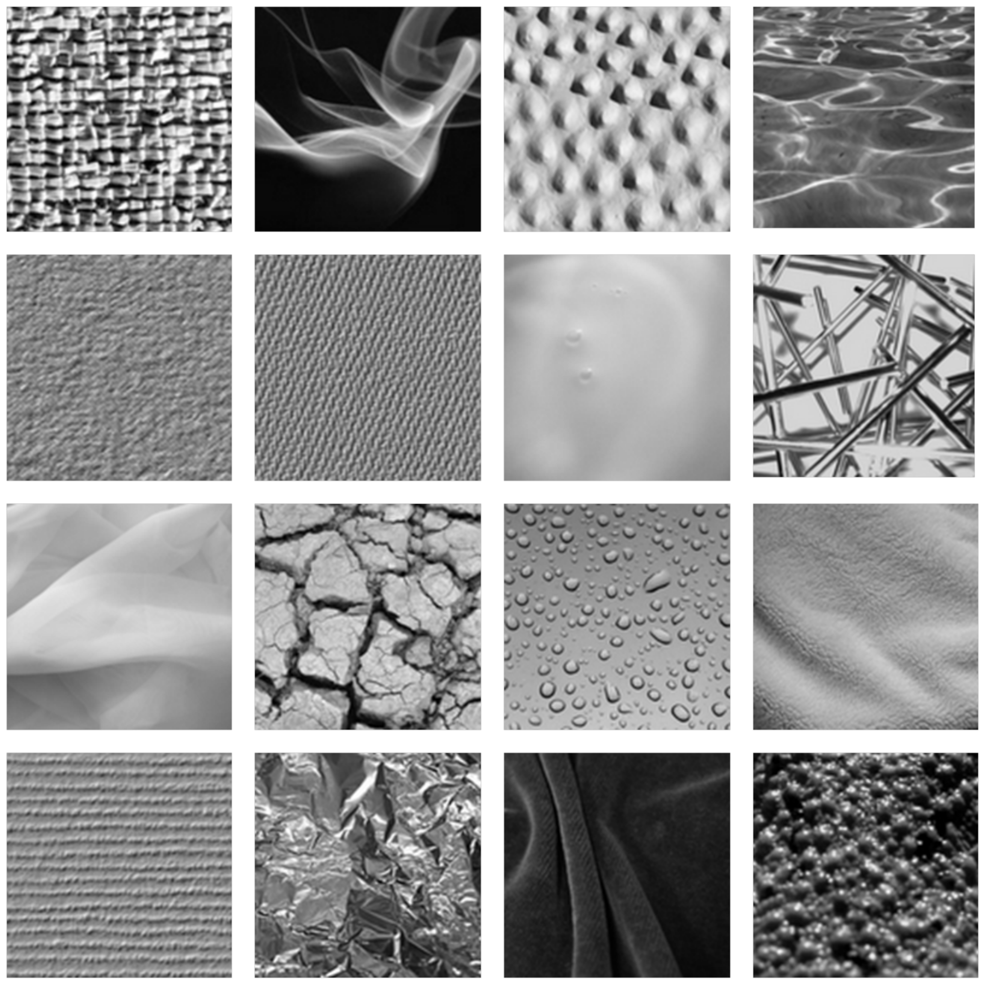


Figure 26: The 16 textures used in the texture rating and voice description experiment. In writing, these are referred to (from left to right) as: 1. *rough*, 2. *smoke*, 3. *bumpy*, 4. *water*, 5. *rough-ish*, 6. *jeans*, 7. *milk*, 8. *sharp*, 9. *net*, 10. *dry*, 11. *drops*, 12. *fleece*, 13. *stripes*, 14. *foil*, 15. *velvet*, 16. *bubbly*.

effect is added that does not describe the three dimensional pattern of velvet itself, but of the way the piece of fabric used in this photograph is folded. However, because velvet seemed to be a key word in synaesthetes' descriptions and no other suitable image could be found, this picture was used as a compromise.

For multiple reasons, the selection is unlikely to be exhaustive or comprehensive enough to accommodate synaesthetes' needs: their perceptions do not resemble exactly the pictures shown, but rather might consist of a blend of multiple textures; it is unlikely that the textures are static; they might move or fade into each other, for example. Nonetheless, a limited set was offered for two main reasons: (1) to display the selection in an assessable fashion that would not overwhelm non-synaesthetic participants, (2) to make the data of the response display manageable for analysis.

4.2.2 Participants

Ratings were gathered from 32 native English speakers, none of whom took part in the voice description experiment described in Chapter 5. 13 male and 19 female participants took part, with an average age of 26 years and a standard deviation of 8 years. 27 participants were British, 2 Irish, 2 US American and 1 Canadian. One British participant grew up in the context of two cultures, possessing both English and Cantonese as native languages. Native speakers were chosen for this task to provide cultural similarity with the people who participated in the next experiment. Some textures might be categorised differently in different cultures, as occurs with colours: for example, some colours are classified as a shade of the basic colour green in one culture and blue in another (Kay et al., 2009). The use of native speakers also ensured that participants fully understood the semantic differentials involved. It was considered that names and translations of textures might differ slightly across languages; hence interpretations might differ as well.

4.2.3 Procedure

The method of human rating was chosen over technical measures or algorithms for this study because human ratings match the perceptual space of textures better than computer algorithms (Clarke et al., 2011), making this choice more suitable for synaesthesia research.

To be able to analyse the set of 16 textures quantitatively, in addition to having the 16 nominal categories of textures themselves, these were converted into numerical data. For this, participants were asked to rate each texture on eight semantic differential scales, which were created after previous studies of textural classification and categorisation by human beings (Rao & Lohse, 1993, 1996; Tamura et al., 1978). The dimensions and clusters of studies by Rao and Lohse, and Tamura and colleagues, were used to create the semantic differentials in a way that seemed most appropriate for the selection of textures offered in my experiment. These were presented on a horizontal slider with descriptive words at opposing ends, as follows:

- rough – smooth
- fine – coarse
- low contrast – high contrast
- high complexity – low complexity
- repetitive – non-repetitive
- non-directional – directional
- line-like – blob-like
- regular – irregular

The experiment was conducted online with the software LimeSurvey (LimeSurvey project team, 2009). The ratings, i.e. the positions on the sliders, were transposed into a scale from 0-100, with the left-most position being 0 and the right-most position being 100. These numerical data points were then used for the analysis. Textures were presented in random order, with semantic differentials also appearing in random order below each texture image. Participants were tasked with moving all sliders before they could proceed to the next image. Each texture was rated once by each participant on all eight semantic differentials, resulting in 128 new data points per participant (16 textures times 8 semantic differentials) for use in analysis of the following experiment. To reduce this great quantity, results were fed into a factor analysis. This analysis tested whether some semantic differentials were redundant due to their ratings being very similar to one another, or correlating strongly (either positively or negatively) with a latent variable that could become the factor. It can be expected, for example, that 'rough-smooth' and 'fine-coarse' will have a strong negative correlation, and therefore a high regression coefficient with the latent variable, so should be combined into a single factor.

4.3 Results and discussion

Average scores on the semantic differentials, including standard deviations for each texture image, are displayed in Figures 27 to 30. Average ratings above 75 or below 25 on a scale will be commented upon in the text.

Unsurprisingly, *bubbly* was judged to be 'blob-like', as was *bumpy*, which was also rated as fairly 'regular'. *Drops* were perceived as 'blob-like' and 'smooth'. The attributes 'rough' and 'coarse' were assigned to the *dry* texture. None of the average ratings for *fleece* approached the extreme ends of the semantic differentials. *Foil* was perceived as 'irregular' and 'non-repetitive'. The texture of *jeans* was rated as 'repetitive', 'directional', 'line-like' and 'regular'. For *milk*, participants had clear preferences on the given semantic differentials being rated as 'smooth', 'fine', 'low contrast', 'low complexity', 'non-repetitive', 'non-directional', 'blob-like' and 'irregular'. *Net* was perceived as both 'smooth' and 'fine'. The image of the *rough* texture was naturally described as 'rough' and 'coarse', but also as 'line-like'. *Rough-ish* resulted in less extreme choices; no ratings over 75 or below 25 were found. Participants tended towards perceiving it as 'low contrast'. 'High contrast' and 'line-like' attributes were assigned to the *sharp* texture image. Clear associations were found for *smoke*: 'smooth', 'fine', 'high contrast', 'non-repetitive' and 'irregular'. The *stripes* pattern was rated as 'repetitive', 'directional', 'line-like' and 'regular'. The attribute 'line-like' was assigned to the *velvet* texture. *Water*, finally, was judged to be 'smooth'.

Most of these ratings seem intuitively sensible whereas some are more difficult to understand. The lack of scores at the extreme ends of the semantic differentials for *fleece,* for example, suggests that either the chosen semantic differentials did not represent the texture description for fleece very well, or participants did not agree with each other in their ratings. The choice of rating 'high contrast' for *smoke* might stem from the way the photograph was taken: white smoke in front of a black background. In a room full of smoke, as a counter example, I suspect people would give ratings of 'low contrast'. Rating *velvet* as 'line-like' might be because of the way the fabric is folded in the photograph and not because of an actually visible line-like pattern in the structure of velvet. I expected *velvet* to be perceived as 'smooth' but the folding pattern of the fabric might have interfered with this quality. The 'irregular' and 'non-repetitive' ratings for *foil* are likewise probably due to the creasing pattern rather than the surface property.

All participants' ratings were fed into a factor analysis to reduce the amount of data that needed to be dealt with in the analysis of the following experiment (see Chapter 5). To do this, a factor analysis with promax rotation was carried out using R (R Foundation for Statistical Computing, 2010). Promax is an oblique rather than an orthogonal rotation, so correlations between factors are allowed. Data reduction through factor analysis was achieved by finding semantic differentials whose scores were strongly correlated with a common latent variable or underlying dimension, so that they could be reduced into one group, i.e. a factor. This correlation with an underlying dimension is expressed by using a regression coefficient, which indicates the strength of relationship between the predictor, here the semantic differentials, and the outcome, here the factor. The factor can then be given a semantically relevant label in accordance with the qualities of the semantic differentials which comprise it. This label clarifies which qualities the variables in each

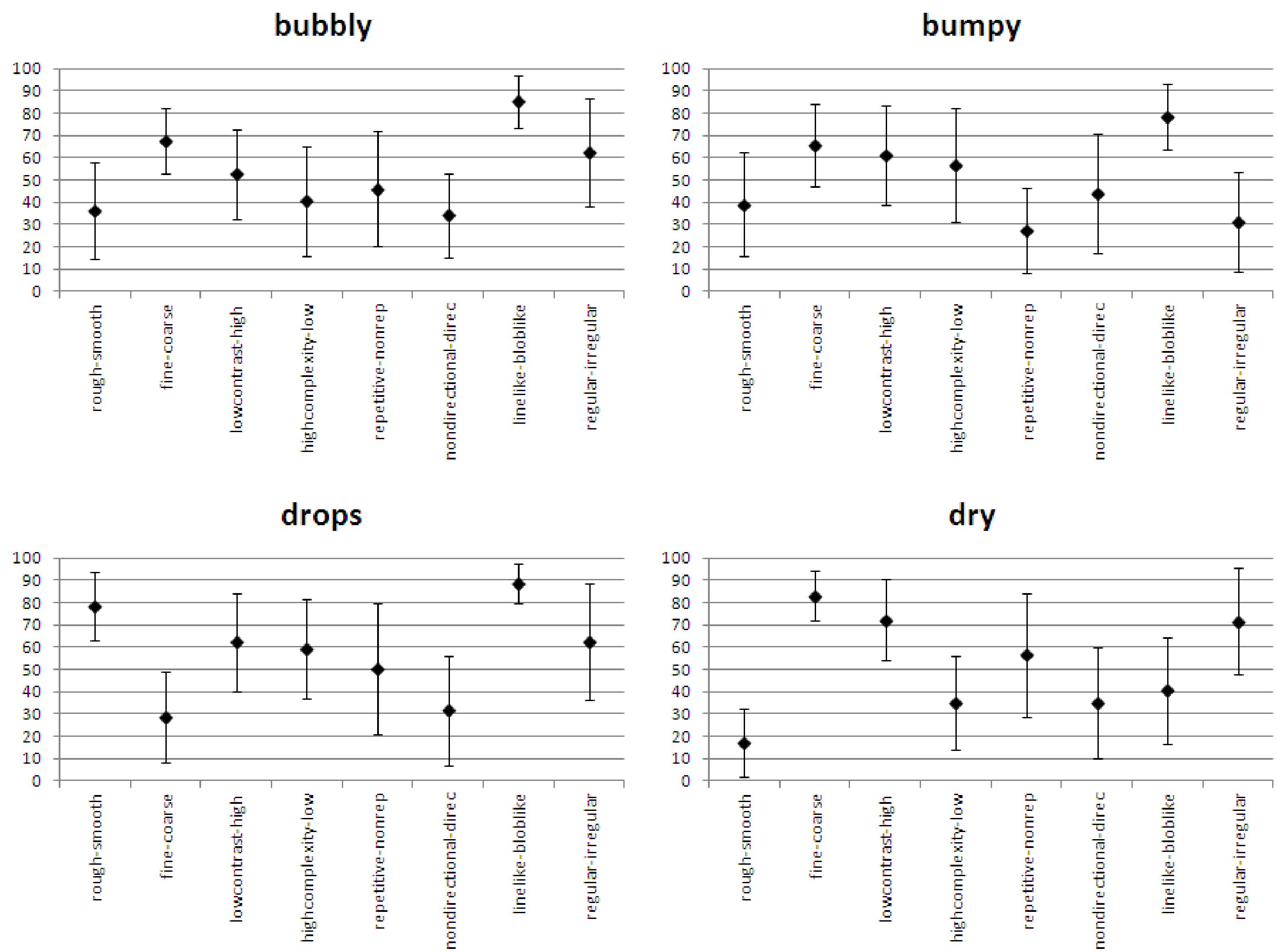


Figure 27: Average values of semantic differentials for the textures *bubbly*, *bumpy*, *drops* and *dry*. Error bars indicate 1 SD.

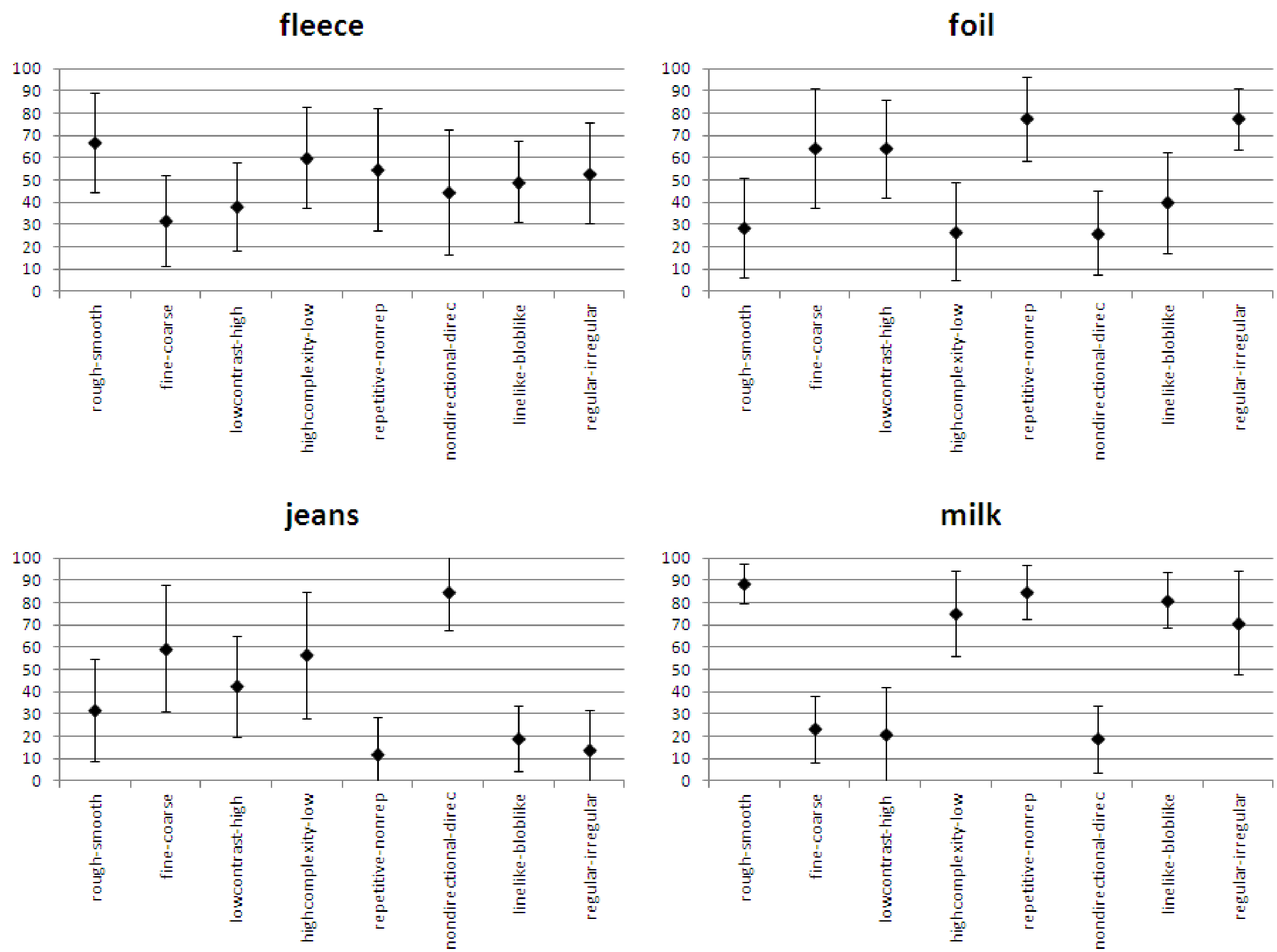


Figure 28: Average values of semantic differentials for the textures *fleece, foil, jeans* and *milk*. Error bars indicate 1 SD.

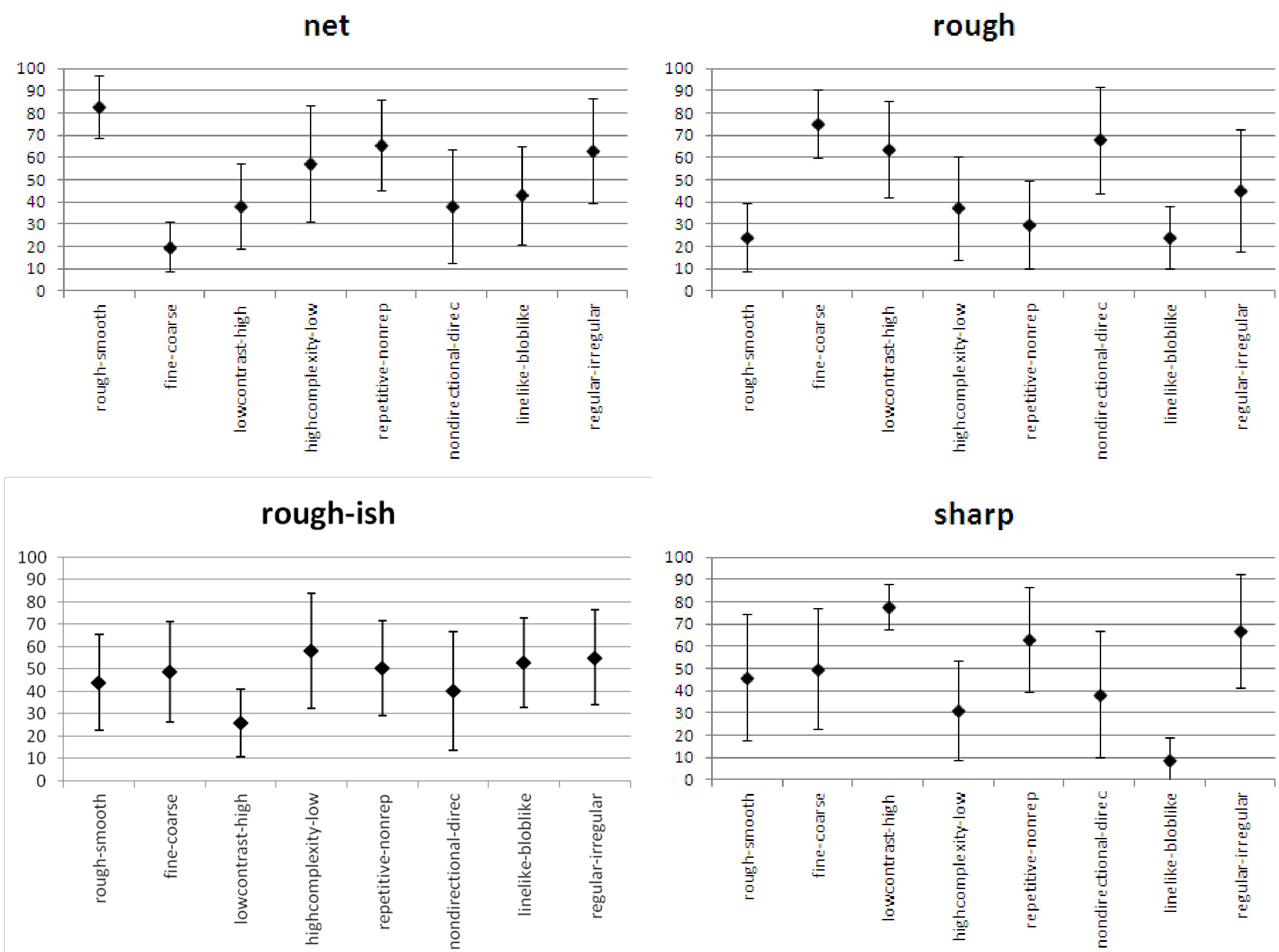


Figure 29: Average values of semantic differentials for the textures *net*, *rough*, *rough-ish* and *sharp*. Error bars indicate 1 SD.

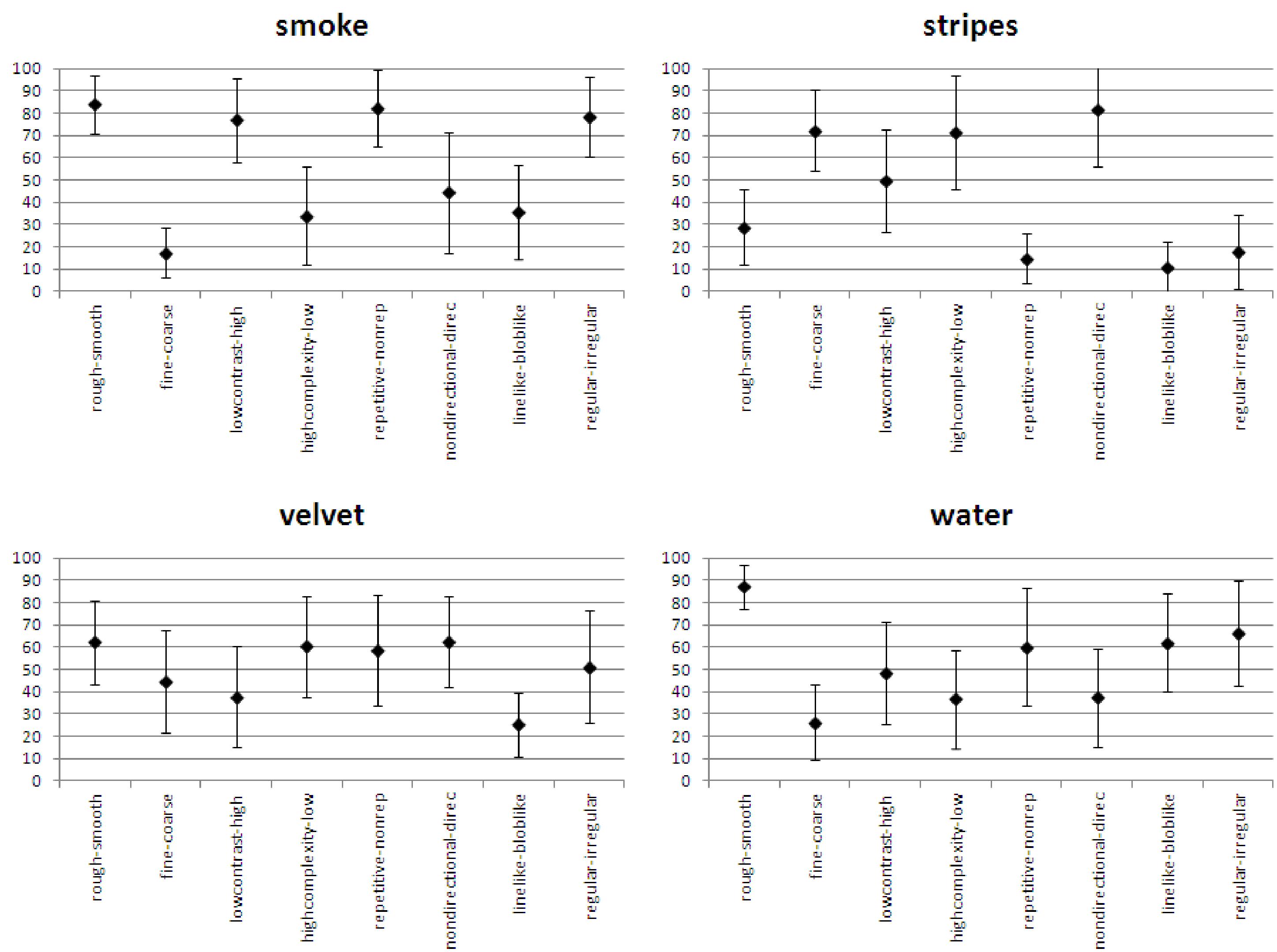


Figure 30: Average values of semantic differentials for the textures *smoke*, *stripes*, *velvet* and *water*. Error bars indicate 1 SD.

	Factor 1	Factor 2	Factor 3	Factor 4
	Repetitiveness	Roughness	Complexity	Line- vs. bloblike
rough–smooth		***0.962***		
fine–coarse		**-0.952**		
low–high contrast	-0.154		**0.862**	
high–low complexity	-0.183		***-0.877***	
repetitive–non-repetitive	***1.015***	0.123		-0.214
non-directional–directional	**-0.791**	0.130		-0.390
line-like–blob-like				***0.998***
regular–irregular	**0.906**		0.204	

	Factor 1	Factor 2	Factor3	Factor 4
SS loadings	2.542	1.874	1.562	1.208
proportion variance	0.318	0.234	0.195	0.151
cumulative variance	0.318	0.552	0.747	0.898

Table 10: Factor analysis with promax rotation of texture rating data. Top: rotated factor loadings of the semantic differentials for four factors. Bottom: sums of squares (SS) loadings of the above factor loadings and their proportional and cumulative variance of the data. Bold typeface indicates a strong regression coefficient of the semantic differential with a factor. The semantic differential with the strongest regression coefficient, chosen as representative of a factor, is indicated in italics.

factor have in common. Each factor was named according to the semantic differential that had the strongest regression coefficient with it.

Table 10 (top) shows the rotated factor loadings of the semantic differentials on four factors. Because the promax rotation allows the data to significantly correlate with more than one factor, some semantic differentials show more than one regression coefficient. The bold typeface indicates which semantic differentials had strong regression coefficients with which factor. So 'rough-smooth' and 'fine-coarse', for example, are strong predictors for factor 2. Italic typeface marks the strongest regression coefficient within one factor. The name of the semantic differential with the italicised regression coefficient was used to name this factor. It was found that the following four factors fitted the data best:

Repetitiveness: repetitive – non-repetitive, non-directional – directional, regular – irregular

Roughness: rough – smooth, fine – coarse

Complexity: low – high contrast, high – low complexity

Line- vs. blob-like: line-like – blob-like

The sums of squares (SS) loadings in the bottom part of Table 10 explain the eigenvalue for each factor. A higher eigenvalue means that the spread of data points is small and can be explained by a strong correlation of variables (here the semantic differentials) in one factor. A smaller eigenvalue means that the variables within a factor are correlated less strongly, i.e. data points of the semantic differentials in this factor are more spread. Factors are sorted descending by "SS loadings". This means that the first factor explains most of the variance within the data, here 31.8%, because this factor has the least-spread data points. The weakest factor still explains 15.1% of the data. Summed up, the four factors explain 89.9% of the variance in the data, highlighting the good fit of choosing four factors for this data reduction analysis.

Rao & Lohse (1996) also reduced their 12 rating scales to four factors using a principal component analysis with an orthogonal design. Eight of the rating scales used in Rao & Lohse (1996) are very similar to the eight semantic differentials used here. Their first factor also comprises 'repetitive', 'directional' and 'regular dimensions', and additionally the scales 'random', 'locally oriented' and 'uniform'. Moreover, the dimensions correlating most strongly with their second factor are in accordance with mine: 'smooth', 'fine', and additionally 'granular'. Only the third and fourth factors differ slightly between the two studies. Rao and Lohse's data on 'density' and 'complexity' have a significant correlation with factor 3, and their 'contrast' and 'local orientation' with factor 4. My factor analysis, however, placed contrast in factor 3 along with 'complexity', whereas factor 4 mainly comprises 'linelike vs. bloblike' – a semantic differential which is not used by Rao & Lohse (1996), which explains the slight restructuring of significant correlations with the factors 3 and 4. On the whole, the results of my study are almost identical with those of Rao and Lohse, suggesting that the factors represent an established model.

Data from the four semantic differentials which have the strongest correlations with their factors will be used for analysis in the following experiment (Chapter 5); that is, 'rough – smooth', 'high – low complexity', 'repetitive – non-repetitive', and 'line-like – blob-like'. These texture ratings will function as a new research tool to observe the relationship between perceptually descriptive texture criteria and acoustic voice quality measures.

5 Voice-induced colour and texture associations

5.1 Introduction

Voice-induced synaesthesia is a rare type of synaesthesia. According to the database compiled by Simner and Ward using an extensive questionnaire (https://www.survey.bris.ac.uk/sussex/syn), up to 10% of the synaesthetes filling out the form have voice-induced synaesthesia (Julia Simner, personal communication). In this type, people experience synaesthetic perceptions induced by the sound of people's voices. This form of synaesthesia has never been researched systematically (but see Fernay et al., 2012 for a case study). Informal reports from voice synaesthetes illustrate its multiple facets and complexity: some synaesthetes claim to see the voice better and more strongly when the person is singing rather than speaking, and that the colours of the notes merge with that of the voice. For some synaesthetes, who associate colours with voices, the colours do not vary much between different voices whereas some find the colours very idiosyncratic. While some individuals only experience synaesthetic reactions to familiar or famous voices, others find that every sound triggers visual perceptions. Still others find that voices trigger textural and shape impressions instead of colours. Some synaesthetes can apprehend that the pitch of the voice influences the tint of colour or texture, whereas others cannot find any criteria that define any regularities in the assignments of synaesthetic perceptions to voices. There have also been reports of synaesthetic perceptions changing when the voice is transmitted through the radio or telephone (similar to the auditory impression changing as well).

Can the influences of voice quality (VQ) features be determined in a systematic way? Are there regular patterns in non-synaesthetes' associations between colours, textures and voice quality, and are they similar to synaesthetes' perceptions? These questions, amongst others, are addressed for the first time in this exploratory study on voice-induced synaesthesia. While informal reports (like those mentioned above) can be found in popular science articles, synaesthesia forums, blogs and email lists, no scientific publication has dealt with this kind of synaesthesia to date. However, studies on related types of synaesthesia can be used as a guideline for research questions and methods, such as work on speech sounds, timbre of instruments, or other sounds. Some of these studies shall be described in the following paragraphs, and their implications shall be referred to when introducing my own experiment.

The study of Ward et al. (2006) dealt with **synaesthetic perceptions of the timbre of musical instruments, musical pitch and composition** (single notes or dyads) by synaesthetes and control participants. The authors found that higher notes were associated with lighter colours and lower tones with darker colours. Additionally, and in closer relation to voice quality, they found that timbre affected lightness choice. When analysing chroma values, it was found that piano and string notes were judged more colourful than pure tones. In a further experiment by this research group, participants had to associate colours with the sound of ten different instruments. Although results for chroma and lightness did not show significant differences in colour associations between the different instruments, it could still be observed that instruments like the harp and flute were given light shades, whereas the tuba and didgeridoo, for example, were assigned dark shades. Comparing single notes with dyads, the authors found that dyads often triggered more than one colour perception. While no significant differences were found between synaesthetes and non-synaesthetes for all these associations, the specificity of colour mapping was more precise for synaesthetes and more noisy for non-synaesthetes. It seems that synaesthetes' associations are automatic whereas those of non-synaesthetes are strategic. This goes hand in hand with the nature of synaesthetic perceptual experiences: synaesthetes have conscious, explicit associations whereas non-synaesthetes have no or implicit colour experiences.

De Thornley Head (2006) conducted a similar study in which colour associations to quartertones (i.e. tones that lie between the commonly used semitones) within an octave were measured for pitch-colour synaesthetes and controls. He found that synaesthetes were significantly more consistent in their colour associations to tones than controls. Interestingly, the synaesthetes chose colours for quartertones whose RGB values lay between the RGB values of the adjacent semitones. Controls projected the tones onto a lightness scale with random hue attribution.

Hubbard (1996) tested the **mapping of lightness onto musical pitch** by a group of undergraduates, presumably non-synaesthetes. He found that lighter shades were associated with ascending musical intervals, whereas darker shades were associated with descending intervals. Similarly, Ben-Tal & Sagiv (2010) tested synaesthetes' and non-synaesthetes' colour associations to rising and falling tones, amongst other things. They found that rising tones were associated with lighter colours by both groups, regardless of average pitch, and descending tones with darker colours. They also looked at the effect of the spectral centroid, which indicates where the centre of mass in the frequency spectrum lies; they called this spectral brightness because this spectral measure is similar to the perceptual feature of acoustic "brightness". Spectral brightness of the tones did not affect colour choice. Their study is the only one to have researched the effects of dynamics of sound and spectral features of sound on synaesthetic perceptions. Other studies on music-induced synaesthesia usually analyse pitch only or chords, or acoustically less-defined features like 'timbre' of an instrument.

Finally, **colour associations with speech sounds** can be investigated in order to find a basis for the research methods on voice quality associations. As described in Chapter 3, Marks (1975) hypothesised that formant settings influence the colour associations with vowels; Jakobson (1962) had the same expectations but based his hypotheses on articulatory settings. The results of my study (cf. Chapter 3) empirically confirm these ideas. The vowel formants, and therefore articulatory settings, influence both synaesthetes and non-synaesthetes in their colour associations in a systematic way (refer back to Figure 19 on page 78). The study showed that acoustic measures (here, formant measures) can be used to analyse and explain inducer-concurrent relations.

While the above-mentioned studies have researched sound-induced colour associations and illustrated relevant methods, none of them has analysed the sounds of voices. The experiment presented in this chapter is a comprehensive approach to investigate both colour and texture associations with different VQs in synaesthetes and non-synaesthetes, and to examine how they verbally describe voices. As this study embarks upon a new area of research, I avoid stating hypotheses but rather formulate research questions (see below). Nevertheless, the studies mentioned above lead to certain expectations. From the work of Ben-Tal & Sagiv (2010), I assume that intonation could confound the perception of voice quality, as the dynamic tones can be compared to the rising and falling pitch of intonation. On the other hand, the effect might not be very strong when stimuli are similar in intonation. The lack of evidence showing influence of spectral brightness on colour choice suggests using other acoustic measures than that of the spectral centroid or centre of gravity. Therefore, I used a wide range of acoustic measures which avoid averaging across the whole spectrum. As Ward et al. (2006) found that dyads often trigger multiple colour sensations, it could be speculated that voice stimuli might induce more than one colour perception as well: one colour accounting for the VQ and another for the intonation pattern for example. Another inference that can be drawn from the study of Ward and colleagues is that the timbre of instruments can be compared with voice qualities. This could lead to the expectation that some VQs might be judged to be more colourful than others.

From the work on vowel sound to colour associations (Marks, 1975, Moos et al., in preparation, Chapter 3 of this thesis) it is expected that acoustic features like formant measures will prove to be a good tool for measuring inducer-concurrent relations in this context. All the studies mentioned lead me to expect that synaesthetes and non-synaesthetes will have similar colour associations with the VQs to some degree, and even that they might use similar underlying mechanisms in their cross-modal perceptions. The literature further suggests that a higher pitch, be it musical or the fundamental frequency of a voice, triggers lighter hues and a lower pitch darker hues.

To my knowledge there is no research investigating sound-texture synaesthesia. Eagleman & Goodale (2009) note that people who took the tests of their *Synesthesia Battery* not only describe plain colour terms for their synaesthetic associations with graphemes, but also mention material properties like matte, soft, bubbly, metallic, velvet etc. While the authors try to explain this phenomenon neurologically – there is an overlap of brain areas active in colour processing that are also active in texture processing – they point out the lack of behavioural studies which research textural perceptions with graphemes. Day (2005) lists the types of synaesthesia he surveyed in a sample of 572 synaesthetes in his chapter on demographic and socio-cultural aspects of synaesthesia; in these 25 types, texture is not mentioned once. This leads to two questions: (1) Are texture concurrents so uncommon that no synaesthete in Day's study reported them? Or (2) are the texture perceptions hidden within another type? Touch is mentioned as a concurrent which seems to be a good candidate for including texture sensations. Djonov & Leeuwen (2011), for example, emphasise the growing importance of texture describing visual and aural features, and research the influence of textural backgrounds in vision. However, my study is the first to look at synaesthetic texture concurrents under controlled conditions.

Taking into account these studies on related types of synaesthesia, more general studies of single aspects like textures, and the informal reports by voice synaesthetes, the following **research questions** were formulated:

1. How do participants describe different voice qualities and what are their associations with them?

2. What is the precise relationship between acoustic features of voices and participants' perceptions? Or to elaborate: are there relationships between acoustics on the one hand, and colours, textures, and descriptive attributes on the other?

3. How do synaesthetes' associations and descriptions compare to those of non-synaesthetes (both laypeople and phoneticians)? Do they have a common ground?

4. Are participants' perceptions and associations with the voices consistent over time?

To help answer these questions, especially as no strong claims about outcomes were made, qualitative as well as quantitative analyses were carried out. The types of analyses are described in Section 5.3, after the methods of the experiment are introduced in Section 5.2. Research question 1 is addressed in the Results Section 5.4.1 concerning the free verbal descriptions of the voices. Research question 2 is discussed in Sections 5.4.2, 5.4.3 and 5.4.4 which look at colour and texture associations with voices and ratings of descriptive attributes in the form of semantic differentials. Research question 3 is dealt with in all these sections, and returned to in Section 5.5. Research question 4 is addressed in Section 5.4.5 which presents the results of a retest. A brief discussion concludes this chapter in Section 5.5.

5.2 Methods

5.2.1 Experimental design

As voice-induced synaesthesia is a rare condition, the experiment was conducted online to reach as many voice-induced synaesthetes as possible. Besides voice-induced synaesthetes and controls, phoneticians were also invited to take part to investigate whether professional expertise in the analysis of voice sounds would manifest itself as a systematic influence in their responses. Since this was an exploratory experiment, two aspects were taken into account to allow for successful control over the measurements of the stimuli: first, the sound samples were controlled recordings of phoneticians producing different voice qualities. The aim of using these professional recordings was to have a range of VQs which are both perceptually and acoustically clearly distinct from each other, similar to different vowel sounds, musical notes or instruments as used in previous research. Secondly, the visual response displays were a limited set of colours, textures, semantic differentials, and a text field for verbal descriptions. The text box for verbal descriptions was chosen as a tool for assessing how synaesthetes and others describe the voices, and to give participants the option to comprehensively describe what they associate with them in case the colour and texture displays did not offer a wide enough range. The other displays were offered to allow for quantitative analyses. A retest was offered with a subset of the stimuli to check for consistency in participants' associations.

5.2.2 Sound stimuli

The stimuli used were recordings of two male phoneticians reading "The rainbow passage", a short story commonly used in phonetic recordings, in various different voice qualities. One set of recordings is of John Laver (recorded in 1976) and the other one of Francis Nolan (recorded in 1979). These recordings served the purpose of illustrating to students of phonetics the range of VQs in moderate to extreme manifestations. These recordings were chosen because they comprise a large number of different VQs. The production of different VQs by two professionals was chosen over normal voices of different speakers to control for acoustic variation, such as the availability of a large range of phonatory and supralaryngeal settings. Most sentences of the rainbow passage contained colour terms. To avoid the presence of colour words in the stimuli, I used the two sentences that did not contain any colour information:

1. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon.

2. People look, but no-one ever finds it.

Ten VQs were chosen based on the criterion that they are perceptually maximally different. These were:

1. *Modal*
2. *Nasal*
3. *Denasal*
4. *Whisper*
5. *Falsetto*
6. *Harsh*
7. *Breathy*
8. *Raised larynx*
9. *Lowered larynx*
10. *Creak*

Recall that these are VQs characterised by articulatory changes in both laryngeal and supralaryngeal settings, as described in Section 2.1.8. The mode of vibration of the vocal folds in the larynx is changed for *whisper* (no voicing), *falsetto, harsh, breathy* and *creak*. Supralaryngeal settings are alterations of the vocal (i.e. oral and nasal) tract, which are changed for *nasal, denasal, raised* and *lowered larynx*. *Modal* voice is defined by having a neutral setting in both the larynx and the vocal tract.

It was decided to use mp3 sound files because these were least likely to cause difficulties in the online experiment, with respect to downloading speed and technical issues with online players. Van Son (2002) suggests that quality loss is negligible when the correct settings, especially for bit rate, are used for conversion. Originally, the recordings were made on a reel-to-reel tape. Laver's recordings have been digitized onto a CD at 44 kHz and Nolan's recordings at 11 kHz. (They were digitized by different people at different times.) For comparability, I downsampled Laver's recordings to 11 kHz using Praat (Boersma & Weenink, 2012). The intensity was also equalised for all sound files to 70 dB_{SPL} using Praat, to avoid changes in volume. To get a bit rate high enough to prevent quality loss, sound files were upsampled using Audacity (Audacity Team, 2010) to 32 kHz, allowing for a bit rate of 192 kbps, which according to van Son (2002) has no significant quality loss compared to wave files. Originally, I considered equalising the speaking rate as well, so that all stimuli had the same length. This plan was discarded because of the risk that the PSOLA algorithm would heavily distort the sound, which might have had a bigger impact on synaesthetic associations than the speaking rate itself. The average length of

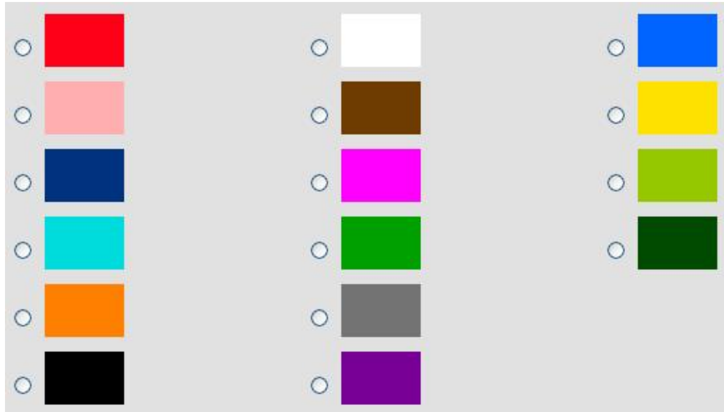


Figure 31: Colour choice response display.

the recordings of the long sentence (sentence 1 above) was 8.06 seconds and of the short sentence (sentence 2 above) 2.66 seconds.

With two speakers, two sentences and ten voice qualities, there were 40 stimuli in total, which were presented in random order. Audio samples of the stimuli are included on the supplementary CD under "Ch5_Voice-induced_associations".

5.2.3 Colours

A response display was offered with 16 different colours, comprising the 11 focal colours of English (Kay et al., 2009) plus five more to offer a larger range. The colour choice is based on that of Chapter 3, but adjustments were made for those two that did not seem to be satisfactory probes of the focal colours: red was adjusted to lean more towards orange than dark pink, and yellow was adjusted to be lighter and less brownish. In the online survey, the colour display looked similar to Figure 31. The colour order changed for every display. Specifications of the colours are given in Section 5.3.3.

5.2.4 Textures

The 16 textures used in this experiment are described in Chapter 4, and displayed in Figure 26 on page 100. As mentioned in Chapter 4, the texture selection was based on synaesthetes' verbal descriptions of their textural concurrents that I found in forum posts and through personal communication. Eagleman & Goodale (2009) state: "Quantitatively testing these prevalences [of texture concurrents] will be a challenge: it is straightforward to develop a user-friendly color chooser [...], but not so with the multidimensional varieties of texture." (Eagleman & Goodale, 2009, p. 291) This difficulty was addressed by offering 16 textures only. Participants were asked to find the texture that matched best. This is the first time that texture perceptions in synaesthesia have been systematically analysed in a closed design.

5.2.5 Participants

Participants were clustered into three groups: synaesthetes, phoneticians and controls. Phoneticians and controls were non-synaesthetes, and had been identified as such through a synaesthesia questionnaire (Appendix F). Potential participants were advised not to take part if they were under-age or had severe sight or hearing difficulties. One participant was excluded because it was unclear which group she belonged in, even after several email conversations. One further participant was excluded because his verbal descriptions suggested he failed to understand the task. The data of the participants described below is reported on in this chapter. All participants were native speakers of English. The number of participants in each group varied because of their availability.

Synaesthetes Synaesthetes were recruited through an announcement on Day's synaesthesia email list and by individually contacting voice-induced synaesthetes on Simner and Ward's database.[7] 14 voice-synaesthetes took part in this study: 11 females and 3 males. Their mean age was 34 (SD 19). 9 were British, 3 US American, 1 Indian and 1 Swedish. The Indian and Swedish participants were trilingual with Bengali/Hindi and Swedish/Arabic as their other native languages. 2 other participants reported speaking at least one other language, and 10 did not speak any second languages. 5 of the participants were students, the rest came from a variety of professional backgrounds. 12 replied they were musical and 11 that they were artistic (they could reply yes to one, both or none).

Participants were asked to fill out a questionnaire listing any other types of synaesthesia that induce colour and/or texture perceptions. They reported 0-18 other kinds of synaesthesia out of a list of 19 with the option to name non-listed types (see Table 11). On average, they had 10 different kinds of synaesthesia. There appears to be no data published on how many types of synaesthesia a synaesthete has on average. Novich et al. (2011) analysed how likely clusters of certain types are without actually stating how many synaesthesiae their participants reported on average. Ten kinds of synaesthesia seems to be a rather large but not unusual number. The last column of the table lists the frequency of the different types in my participants (apart from voice-induced synaesthesia). Music is the most common inducer besides voices, which is not surprising as both share having acoustic stimuli as inducer. Letters, months and numbers, in short sequences, are common additional inducers too. Unfortunately, I can only partly compare my data to the clusters found in the work of Novich et al. (2011), because their data only offer non-visual concurrents for the inducer 'sound' and only musical inducers in the auditory domain for the concurrent 'colour'. However, their cluster of instrument/timbre, musical pitch and chord to colour seems to be the cluster that a voice-colour synaesthete would fall into as

[7] The collaboration protocols of the *Sussex-Edinburgh Database of Synaesthete Participants* were followed carefully.

ID:	5	8	9	10	11	25	47	101	102	103	104	105	106	107	SUM
letters		y		y	y	y		y	y	y	y	y	y	y	11
music		y	y	y	y	y	y	y	y	y	y	y	y	y	13
words		y		y	y	y		y	y		y	y		y	9
days		y		y	y			y	y		y	y	y	y	9
months		y	y	y	y	y	y	y	y		y	y	y	y	12
pains		y		y			y			y		y	y	y	7
touch		y		y						y		y	y		5
body posture				y											1
noises		y		y	y	y		y	y	y		y		y	9
smells				y		y	y	y	y	y		y	y		8
tastes				y		y		y	y	y			y		6
colours				y		y		y		y	y				5
shapes				y	y	y		y		y	y		y		7
emotions			y	y				y		y	y		y		6
fingers															0
faces			y	y				y		y					4
places				y			y	y			y	y	y	y	7
names		y		y	y			y	y		y	y	y	y	9
numbers		y		y	y			y	y	y	y	y	y	y	10
SUM	0	10	4	18	9	9	5	15	10	12	11	12	13	10	

Table 11: Voice synaesthetes' self-reports on what other kinds of synaesthesia trigger colour and/or texture perceptions.

well, seeing the similarity in the auditory nature of the inducers.

Phoneticians 10 phoneticians took part: 7 females and 3 males. Their average age was 40 (SD 15.9). 5 of them were British, 3 US American, 1 Irish, 1 Australian. 9 of them claimed to speak at least one second language. 3 of them were PhD students, the rest were professionals. 6 individuals said they were musical and 4 said they were artistic.

Controls There were 28 control participants: 17 females, 11 males. Their average age was 23 (SD 3.5). 18 of them were British, 6 US American, 2 Singaporean, 1 Swedish and 1 New Zealandish. 1 was bilingual with Cantonese as the other native language, 1 was trilingual with Arabic and Swedish. 10 other participants claimed to speak at least one second language. 24 participants were students and the rest were professionals. 15 said they were musical and 17 said they were artistic.

5.2.6 Procedure

In the introduction and instruction part of the experiment, the tasks were described and participants were asked to use the best possible audio equipment at their disposal, usually headphones or external speakers. After that, a consent form was filled in. Then, the participants listened to the first audio file which contained the stimulus to be rated. This and all subsequent audio files could be played as many times as needed. For every stimulus the following response order was given:

1. Text field: “What are your first impressions of and associations with this voice? Please describe the voice in your own words.”

2. “Please rate the voice on the following sliders.” The 8 sliders were (presented in random order per stimulus):

    (a) low – high

    (b) colourful – grey

    (c) light – dark

    (d) sharp – fuzzy

    (e) tense – relaxed

    (f) hard – soft

    (g) dry – fluid

    (h) smooth – rough

3. “Which colour matches the voice best?” A choice of 16 colours was given in random order per stimulus (see Figure 31 in Section 5.2.3).

4. “Which texture matches the voice best?” A choice of 16 textures was given in random order per stimulus (see Figure 26 in Chapter 4).

5. Optional comments: “On a scale from 0 to 9 (where 0 is nothing and 9 very intense), how intense are your colour and texture experiences? Is there anything more you want to add?”

After 20 stimuli had been rated, participants had to give demographical data about themselves, e.g. their age, occupation, where they grew up etc. At this midpoint of the experiment, participants could save the data and take a break, before proceeding to the next 20 stimuli. After the final stimulus, there was an ABX task which is described in Chapter 6. The experiment concluded with a synaesthesia questionnaire. The total procedure lasted for about 1.5-2 hours.

For most participants a retest was conducted after 5-8 months, in some cases as early as 2 months after initial testing because those were recruited later than most other participants. In the retest, 10 of the 40 stimuli were presented again. Each voice quality was presented once, each speaker five times and each sentence five times. The questions and response displays were set up in the same way as in the initial test described above. There were no questions regarding demographical data or synaesthesia and no ABX task. Participants needed 15-25 minutes to complete the retest.

5.3 Analyses

Due to the exploratory nature of this experiment, various different measures were taken allowing for both qualitative and quantitative analyses. The verbal responses of participants were analysed qualitatively by coding them into different categories (see Section 5.3.1). Many acoustic measures were taken to find the most useful ones in the context of this new research (see Section 5.3.2). Colours were quantified by putting them into CIELUV colour space (see Section 5.3.3). The way the textures have been analysed has been described in Chapter 4. The ratings on the semantic differentials were simply translated into scales from 0-100, where the left-most slider position is 0 and the right-most is 100. These numerical data could then be used for data analysis.

5.3.1 Qualitative data analysis

In order to explore people's reactions to different VQs, regardless of whether these reactions were synaesthetic or not, a qualitative data analysis was conducted in addition to the quantitative data elicitation. It was assumed that the verbal descriptions of the participants would give insights into why certain associations were made. The qualitative analysis was also used to find out which classes of words and concepts the different participant groups used for describing someone's voice, particularly the extent to which the synaesthetes actively used their synaesthetic perceptions.

To analyse participants' verbal descriptions of the voices and their associations and impressions, a categorisation system was needed. Because this is a first exploratory experiment, no clear hypotheses could be made on how synaesthetes would describe the voices; therefore, the Grounded Theory (Strauss & Corbin, 1998) was considered. This is a systematic methodology involving the discovery of a theory through the analysis of data. Rather than starting with a hypothesis and analysing the data accordingly, data is coded during analysis and grouped into concepts and categories; once a pattern emerges, a theory can be formulated. Following the Grounded Theory, codes were created while reading through the participants' replies. During that process it was necessary to go through every reply at least twice, because the creation of codes might not be finished until every reply had been attended to; new codes or alterations might take place during the coding process. Later these codes were grouped into different categories. Finally, depending on the distribution of the codes and categories in the three different participant groups, a theory was developed explaining the data, especially the differences in the data between groups. I created 25 codes which were grouped into six different categories as listed below:

1. Associations

person: associations with real or fictitious people

anno: time period of recording

pre: comparing the speaker to one previously heard

2. Description of person

age: age

sex: sex of speaker

occupation: occupation

look: physical appearance, clothing

health: state of health

personality: character, habits, attitudes

emotions: emotional state of the speaker, feelings of the speaker

3. Feelings in listener

feelings: emotions or feelings evoked in the listener

4. Phonetics

voice quality: professional terminology or layperson's description of voice quality

phonetics: any terminology related to phonetics (other than relating to VQs), e.g. pitch or speaking rate

accent: regional area, accent

fake: disguised or pretend voice

evaluation: evaluation of the voice, e.g. "good voice", "could be better if..."

style: speaking style, e.g. "story telling", "newsreader", "telling off"...

5. Synaesthetic perceptions

colour: any colour terms

texture: any texture terms

shape: any terms describing a shape

space, movement: any terms describing where in space something is positioned and/or whether it moves

taste: any gustatory terms

smell: any olfactory terms

temperature: any terms related to temperature

6. Unclassified

misc: terms that did not fall under any of the categories mentioned above

After these codes and categories were created, I became aware of a student's essay investigating VQ labelling by linguistically untrained listeners (Burns, 2010, unpublished). The author and I had independently grouped the participants' replies into very similar categories, although we gave the categories different names. The categories in her study were as follows (Burns, 2010, pp. 6-7):

- Comments on Pitch (high, low, deep)
- Health (drunk, ill, head cold, sore throat, lisp)
- Sincerity (put on, pretending to be a woman, deliberate)
- Character (pervert, educated, scary, timid, threatening, posh)
- Clarity (clear, unclear, slurred, informative)
- Celebrities & Characters (Muppets, Sean Connery, BBC newsreader, German soldier)

This similarity between the two independent studies on VQ descriptions indicates the universality of the verbal descriptions made. Only an equivalent to the synaesthetic category is missing in Burns' categories. That, however, is not surprising as she probably did not have synaesthetes amongst her participants.

5.3.2 Acoustic analyses

The voice recordings were chosen based on the criteria that different VQs offer a wide range of 'different sounding voices' but at the same time offer a controlled setting. Because VQs are defined by certain articulatory settings as well as acoustic characteristics (Laver, 1980; Nolan, 1983), acoustic measurements are a good tool to analyse differences in voices. In this section, all acoustic measurements taken on the stimuli are listed. They fell into several categories:

- Pitch measures: minPitch, maxPitch, pitchRange, f0_mean
- Formant measures: LTF1, LTF2, LTF3, LTF1SD, LTF2SD, LTF3SD
- Harmonicity/noise measures: H/N
- Measures relating to the mode of vocal fold vibration: H1*-H2*, H1*-A1, H1*-A3*

The large number of acoustic measures in these four categories was taken to extract information about as many different acoustic aspects of voices as possible for this exploratory research.

Measures related to pitch and harmonic-to-noise ratio were taken across the whole sentences; formant measures for long-term formant distributions (LTF) were taken across all vowels; and all other measures on formants and harmonics were taken on two specific vowels – the vowel in the second syllable of 'app<u>a</u>rently' for the first sentence, and the initial vowel in '<u>e</u>ver' for the second. These two vowels were chosen because they appear in an accented syllable and are reasonably similar in quality. All measurements were performed using Praat (Boersma & Weenink, 2012) except for the LTF measurements, which were performed using WaveSurfer (Sjölander & Beskow, 2005).

minPitch 'minPitch' informs us about the lowest fundamental frequency that the speaker used during the production of the sentences. The minimum pitch, or lowest f0 per sound sample, was measured using Praat's pitch tracking feature in Hz with the standard setting, i.e. showing the f0 between 50 and 400 Hz. If the pitch tracker made measurement errors, the upper or lower limit of the frequency range was adjusted or the minimum pitch was identified manually.

maxPitch 'maxPitch' informs us about the highest fundamental frequency that the speaker used during the production of the sentences. The highest f0 per sound sample was measured in the same way as 'minPitch'.

pitchRange 'pitchRange' describes the variability of f0 in a speaker. It was calculated by subtracting minPitch from maxPitch in Hz. The data were then transposed into semitones using the following formula: $12 * LN(maxF0/minF0)/LN(2)$. Semitones are often used to compare ranges across speakers with different fundamental frequencies because, for example, a 50 Hz range around 100 Hz is perceived as larger than around 300 Hz due to nonlinear relationship between pitch and fundamental frequency (Clark et al., 2007) – an artefact which is taken care of when transposed into semitones.

f0_mean 'f0_mean' stands for the mean fundamental frequency that the speaker used during the production of the sentences. For the mean f0, all pitch measures that were tracked in Praat (and corrected as mentioned under 'minPitch') were used to calculate the mean.

LTF1, LTF2, LTF3 LTFs, or long-term formant distributions, display the mean of all vowel formant measures per recording – here per sentence – per formant 1, 2 and 3. LTF was chosen in preference to single formant measures because the stimuli were sentences and not isolated vowels. LTF was introduced by Nolan & Grigoras (2005) and further explored by Moos (2010). The term LTF is defined in the following way:

> Long-Term Formant Distribution (LTF) is a method used to determine average formant values of a speaker. For each formant, all formant measurements of all vowels produced by a speaker are averaged (across the entire recording or appropriate sub-portions of a recording). This average is the LTF value for this formant. That means that every speaker has one LTF value and a standard deviation (SD) per formant which shall be called LTF1, LTF2 and so on. It is a frame-by-frame measurement, meaning that long vowels carry more weight than short vowels. (Moos, 2010, p. 7)

When LTF measures are used to compare speakers using spontaneous speech, they only become reliable when a certain amount of data is given (at least 6 seconds of pure vocalic stream) to even out the variation in vowel qualities. Here, although failing to reach this amount of data, the measures can still be used because all my stimuli consist of the same two sentences, i.e. the same vowels. The selection criteria for vocalic portions used were:

- Clear and visible formant structure of the first three formants (intensity settings were sometimes increased to find F3, especially for back vowels which tend to have a higher spectral tilt)
- Laterals and approximants were included in the analysis
- Creaky voice was included in the analysis if vocalic
- No nasals or strong nasality (because of anti-formants at 2-3 kHz)
- No vowels spoken with a very high pitch as harmonics rather than formants were visible in high pitch

The segmentation of vowels as well as the hand correction of inaccurate formant tracking was done using WaveSurfer (Sjölander & Beskow, 2005). The automatic formant tracking was set to four formants, an LPC order of 12, a frame interval of 0.01 seconds and a nominal

F1 of 500 Hz. Data concerning F4 were not used. Manual correction of the automatic formant tracking was sometimes necessary because the segmentation procedure could yield unnatural formant jumps without any transition, which are not predicted accurately by the prediction algorithm.

LTF1SD, LTF2SD, LTF3SD These are standard deviations of the long-term formant distributions. They show the variations of formant distributions within each voice recording.

H/N H2N, or harmonic-to-noise ratio, shows how much aperiodic noise, such as frication, can be found in the signal. The more aperiodic noise is found, the lower the ratio; whisper, for example, has a low ratio because there is no voicing, or periodicity, in the signal. 'H/N' was measured in Praat. The formula $10 * log10(\frac{energy\, of\, periodicity}{energy\, of\, noise})$ was used.

H1*-H2* The difference in amplitude between the first harmonic and the second harmonic is one way to define the slope of the spectral tilt in the low frequency range. (In a periodic speech signal, such as a vowel sound, a harmonic is a component frequency of the signal that is a whole-number multiple of the fundamental frequency. In speech, f0 is usually regarded as the first harmonic. Thus, with an f0 of 100 Hz, the first harmonic (H1) is 100 Hz, H2 is 200 Hz, etc.) Energy at higher frequencies is less damped in a shallow spectral tilt and more damped in a steep spectral tilt. In other words, if the energy of H2 is much lower than that of H1, a steep spectral tilt can be expected, as found e.g. in breathy voice quality. If H1 and H2 are of equal energy, a shallow tilt can be expected, as found e.g. in harsh voice quality.

Measurements were taken from one stressed vowel per sentence, as described above. Three measures per vowel were taken, 20ms apart, at the most stable part of the vowels. Corrections were made to the measures of the harmonics so that measures were comparable across speakers, across VQs, and across vowels. Corrections were based on Hanson's suggestion: "The quantity of $20log_{10}[F1^2/(F1^2 - f^2)]$ is subtracted from H1 and H2, where f is the frequency at which the harmonic is located" (Hanson, 1997, p. 480). The formant measures that were needed for this formula, were taken manually using FFT and LPC spectra in conjunction with a wideband spectrogram.

H1*-A1 Following Hanson (1997), the corrected amplitude of the first harmonic minus the amplitude of F1 can be used to approximate the bandwidth of F1. A low value suggests a properly closed glottis at the end of a vibration cycle of voicing (e.g. in modal voice), whereas a higher value suggests incomplete closure of the glottis throughout the cycle (e.g. in breathy voice). The amplitude of F1 was measured in Hz using an FFT spectrum.

H1*-A3* The corrected amplitude of the first harmonic minus the corrected amplitude of F3 is another measure that approximates the spectral tilt. A low value suggests an abrupt closure of the glottis during a cycle of voicing (e.g. in harsh voice), while a high value suggests incomplete closure (e.g. in breathy voice). The same vowels were used for the acoustic measurements as above, using the same measurement criteria. As Hanson points out, "[...] the amplitude of the first harmonic relative to that of the third-formant peak (H1-A3) may be a suitable acoustic correlate of spectral tilt, if certain corrections are made" (Hanson, 1997, p. 474). Corrections of H1 are as explained above. Corrections for A3 are done using the following formula as suggested by Hanson (1997, p. 480), (500 is used as F1 of the neutral vowel schwa, and 1500 as F2 of schwa):

$$A3 + 20log_{10}(\frac{[1-(\frac{F3}{F1})^2][1-(\frac{F3}{F2})^2]}{[1-(\frac{F3}{500})^2][1-(\frac{F3}{1500})^2]})$$

A Praat script measured 'f0_mean', 'minPitch', 'maxPitch' and 'H/N' (called harmonicity in Praat) using Praat's standard settings, except that for *creak* the upper limit for f0 measurements was set to 150 Hz to avoid measurement errors. For some VQs certain measures could not be taken. 'minPitch', 'maxPitch', 'f0_mean' and 'pitchRange' were not measured for *whisper* because whisper is by default unvoiced. To avoid missing data in the analyses, these missing values were replaced with the mean measures of all other VQs. Formant related measures, i.e. 'LTF1', 'LTF2', 'LTF3', 'LTF1SD', 'LTF2SD', 'LTF3SD', were not measured for *falsetto*. Because of the high pitch of *falsetto* it was not possible to detect formants accurately; the spectrogram showed harmonics mainly. To avoid missing data in the analyses, these missing values were replaced with the mean LTF measures of all other VQs.

It was expected that some acoustic measures would correlate with each other, so it was deemed redundant – and too time consuming – to use all 14 acoustic measures for analyses. To reduce the number of independent variables for later analyses, a factor analysis with promax rotation (an oblique rotation that allows correlation between factors) was conducted. Four factors best captured the dimensions of variation among the stimuli; they explained 89% of the variability within the data. Table 12 shows the significant correlation coefficients of all acoustic features with each of the factors. If an acoustic feature correlated with one or more factors, the strongest correlation is printed in bold face. The representative of each factor, which is later used for the analyses, is additionally printed in italic typeface. The features that correlated most strongly with the factors were:

1. 'f0_mean' (mean f0 was preferred as representative of Factor 1 over 'minPitch',

	Factor 1	Factor 2	Factor 3	Factor 4
H1*-H2*	**0.665**	0.118		
H1*-A1			**0.782**	
H1*-A3*	0.190		***0.805***	
minPitch	**0.988**			-0.357
maxPitch	**0.935**			0.204
pitch range			-0.145	***0.984***
f0 mean	***0.947***			0.119
H/N	0.454	-0.197	**0.581**	
LTF1	0.295	0.428	**-0.769**	0.130
LTF2		***0.882***	-0.206	
LTF3	-0.136	**0.747**	0.340	0.187
LTF1 SD		**0.552**	-0.150	-0.211
LTF2 SD		**0.868**		
LTF3 SD		0.157	0.227	0.171

Table 12: Correlations of acoustic measures with the four factors from the factor analysis with promax rotation. Strongest correlation per acoustic feature in bold, representative of each factor in italic.

because it is a more common measure whose correlation coefficient is nearly as high as that of 'minPitch')

2. 'LTF2' (mean F2)

3. 'H1*-A3*' (spectral tilt)

4. 'pitchRange'

These features were later used as predictors in the linear mixed effect modelling to predict colour and texture choices and ratings of the semantic differentials.

Table 13 shows the values of the four representative acoustic measures of the ten different VQs, averaged across the two speakers and across the two sentences. Values of the fundamental frequency are in the expected range that characterise the VQs accordingly: *creak* has the lowest mean f0 whereas *falsetto* has the highest. Although f0 does not necessarily have to rise when using a raised larynx setting, speakers often lack compensation mechanisms to keep f0 low while shortening the vocal tract by raising the larynx (Laver, 1980). Hence, the f0 of the *raised larynx* stimuli were higher than in any setting apart from *falsetto*. I am not aware of any research studying the pitch range of different voice quality settings. However, it can be assumed that a high degree of relaxation or tension of the pharyngeal muscles restricts pitch variation. This is the case for *breathy*, where a relaxed pharynx is expected, and *harsh*, where pharyngeal tension is expected. *Whisper* shows the highest LTF2, caused by slightly different vocal tract settings (Tartter, 1991; Higashikawa et al., 1996); perhaps a slight pharyngeal tension. The lower values of LTF2

VQs	f0_mean (Hz)	LTF2 (Hz)	H1*-A3* (dB)	pitchRange (st)
modal	119	1339	20.32	12.22
raised lx	156	1245	13.19	13.68
lowered lx	124	1247	19.83	8.89
nasal	110	1352	19.49	9.30
denasal	114	1341	17.63	9.14
falsetto	232	n/a	n/a	10.24
breathy	109	1319	23.33	7.46
whisper	n/a	1463	0.30	n/a
harsh	106	1408	-0.52	6.42
creak	92	1311	14.96	9.47

Table 13: Average acoustic values for the different voice qualities. Only the measures representing a factor, and thus used for later analysis, are displayed: f0, LTF2, H1*-A3* as a measure of spectral tilt, and pitch range. lx = larynx, st = semitones.

for *raised larynx* compared to other settings are in agreement with the measures of the average F2 values published in Laver (1980), which are based on parts of the recordings I used, namely Laver reading the rainbow passage. Although a rise in F2 would be expected with a shorter vocal tract length, the lowering of F2 can be explained in terms of the pharyngeal tension which takes place by raising the larynx. Also, the LTF2 value for *lowered larynx* is similar to the average F2 given by Laver (1980).

As stated earlier, H1*-A3* represents the spectral tilt. A shallow spectral tilt – that is, equally distributed energy across the frequency range – can be observed for *whisper* and *harsh*. As expected, *breathy* shows the steepest spectral tilt: breathiness is produced by incomplete closure of the glottis with a slackness of the vocal folds, which in turn leads to a decrease in energy in high frequencies (Hanson, 1997). *Nasal* settings dampen spectral peaks especially in the frequency range of F1 and above F3, and widen the formant bandwidths (Laver, 1980). Since the flattening takes place across the frequency range, a change in spectral tilt is not necessarily expected, but rather a general attenuation of energy.

5.3.3 Colour analyses

The 11 focal colours of the English language (Kay et al., 2009) plus an additional five as outlined above, were presented as a forced choice display (see Figure 31 on page 116). The colour rectangles for the response display were created by entering the RGB values. To quantify them for analysis, they were measured with the Minolta chromameter CS-100 and converted into CIELUV colour space (Westland & Ripamonti, 2004). The chromameter was used to measure the L, x and y values on ten different computer screens in different lighting conditions to get an estimate of the variation of settings that participants would use. The average of these ten measures was then used to convert the numbers into L*, u*,

	R	G	B	L*	u*	v*
black	0	0	0	15.3	2.8	2.5
red	255	0	25	73.9	206.0	52.4
grey	115	115	115	74.4	-5.4	-6.3
white	255	255	255	129.8	-1.6	6.0
dark green	0	75	0	45.1	-30.3	42.5
yellow	255	225	0	121.9	30.8	125.5
pink	255	0	255	81.3	92.0	-124.2
olive	150	200	0	105.4	-29.2	111.5
blue	0	100	255	69.6	-39.8	-154.9
brown	110	60	0	48.1	51.7	38.4
pale pink	255	175	175	109.4	44.3	4.4
cyan	0	220	220	110.0	-80.6	-23.0
green	0	160	0	82.1	-67.5	88.9
orange	255	128	0	92.6	124.6	83.2
purple	120	0	150	48.7	32.5	-107.4
dark blue	0	50	128	41.5	-19.9	-95.1

Table 14: RGB values of the 16 colours used for creating the colour patches in the online survey. L*, u*, v* values calculated from the average L, x, y measures of ten randomly selected computer screens.

v* – the CIELUV colour space – using the formula published in Westland & Ripamonti (2004, p. 50f). Values of the background grey were used as the "reference white", so that colour values were put in the correct colour context. This occasionally resulted in L* values being above 100, which is usually the upper limit when white is used as reference. The means of these transposed measures were then used for statistical analysis. Both RGB and L*, u*, v* values are listed in Table 14.

5.4 Results and discussion

In this section, the results of the verbal descriptions, colour and texture associations, semantic differential ratings and the retest are presented in the respective sections. Since these five sections are mostly independent from each other, results of these subanalyses are discussed within the sections; only results and conclusions spanning across the five subfields are discussed in 5.5.

5.4.1 Verbal descriptions

How did participants characterise the different voices when asked to verbally describe them and name their associations to and impressions of them? Most participants used descriptions that are accessible to the general population like "This voice sounds fatigued and exasperated almost", "nasal, male, tired, uninterested". Some synaesthetes employed their synaesthetic perceptions to describe the voice, like "Orange, yellow and red, creases and jagged splashes", "Rather blue and the shape of an O that has been compressed from left to right". Phoneticians tended to use more technical terms in their descriptions, like "tense creaky", "denasalized male voice". Results of the qualitative data analysis are summarised per group for all 25 codes and the six categories which they are grouped into in Table 16. The percentage per group, i.e. per column, amounts to 100% (rounding errors permitted) in the columns titled "in % per group". That means that for example 22.2% of all codes for synaesthetes are on colour terms; or in other words, out of every 100 verbal descriptions by synaesthetes, 22 were (on) colour terms. Absolute numbers of participant responses are given in the right-most columns. Here, columns add up to different amounts for two reasons: first and foremost, there is a different number of participants per group; second, depending on the complexity of a description, a different number of codes is assigned to one voice description.

Figures 32 and 33 visualise group differences in the verbal descriptions. Figure 32 shows the codes assigned to the verbal descriptions per participant group, sorted according to the frequency of usage within each group. The frequency of terms used to describe the voices varies according to group: synaesthetes use colour terms most often, phoneticians VQ terms and controls most frequently describe the personality of the speaker. The figure also illustrates how all groups frequently use words from the 'phonetics' category: *evaluation*, *phonetics* and *voice quality* are always amongst the six most frequently assigned codes. Also the *personality*[8] is very frequently used across groups. Both could be expected because the task probably biased participants towards using words with these codes, as it asked participants to describe the voice and name impressions and associations.

To increase readability and simplify group comparison, Figure 33 groups the 25 codes into the six categories as introduced in Section 5.3.1. Figure 33 shows very clearly that the category 'synaesthetic perceptions', which includes codes like *colour*, *texture*, *temperature*, etc., is most often assigned to the verbal descriptions given by synaesthetes and very rarely to those of phoneticians and controls. The category 'description of person' is frequently used by all three groups but most often by controls. Evidently, a large proportion of my participants deduced aspects such as a speaker's character, physical appearance or age simply from hearing his voice, and thereby created a mental image of the speaker. Emotional reactions in the listener, 'feelings in listener', were not mentioned frequently. This category was mostly reserved to synaesthetes' descriptions. Perhaps their emotional reactions are stronger because they experience multiple perceptions at once; or they have

[8] *Personality* is called "description of personality" in Figure 32 to aid discriminability between this code and *person* which falls under the category 'association'.

Categories	Codes	Synaesthetes		Phoneticians		Controls		Synaesthetes		Phoneticians		Controls	
		in % per group						in absolute numbers					
	person	1.2		6.5		3.4		13		42		75	
Associations	pre	3.2	4.6	2.7	9.9	2.3	7	31	47	13	58	43	144
	anno	0.2		0.7		1.3		3		3		26	
	age	0.4		1.5		2.1		11		19		129	
	sex	1.3		3.7		3.8		49		54		293	
	occupation	0.3		0.5		1.4		7		6		60	
Description of person	look	3.0	19.8	2.0	20.0	1.2	33.3	24	286	10	199	46	1181
	health	1.6		3.4		2.6		30		30		107	
	personality	10.9		8.4		16.5		128		74		414	
	emotions	2.3		0.5		5.7		37		6		132	
Feelings in listener	feelings	5.1	5.1	1.6	1.6	2.4	2.4	61	61	11	11	70	70
	VQ description	6.7		26.6		8.9		135		288		394	
	phonetics	5.2		11.4		9.9		77		54		280	
Phonetics	accent	1.2	26.3	3.9	61.2	3.0	51.5	12	324	18	435	64	1141
	fake	1.7		4.5		3.5		14		23		62	
	evaluation	8.1		8.5		12.9		52		23		113	
	style	3.4		6.3		13.3		34		29		228	
	colour	22.2		0.5		1.2		273		5		64	
	texture	9.0		2.3		1.9		160		29		101	
	shape	5.0		0		0.1		72		0		5	
Synaesthetic perceptions	space, movement	4.9	42.7	0	2.8	0	3.3	49	594	0	34	0	177
	taste	0.1		0		0		1		0		0	
	smell	0.5		0		0		7		0		0	
	temperature	1.0		0		0.1		32		0		7	
Unclassified	misc	1.8	1.8	4.5	4.5	2.3	2.3	19	19	16	16	37	37
Total		100	100	100	100	100	100	1331	1331	753	753	2750	2750

Table 16: Codes assigned to verbal voice descriptions by synaesthetes, phoneticians and controls in percent per participant group in the middle and in absolute numbers on the right.

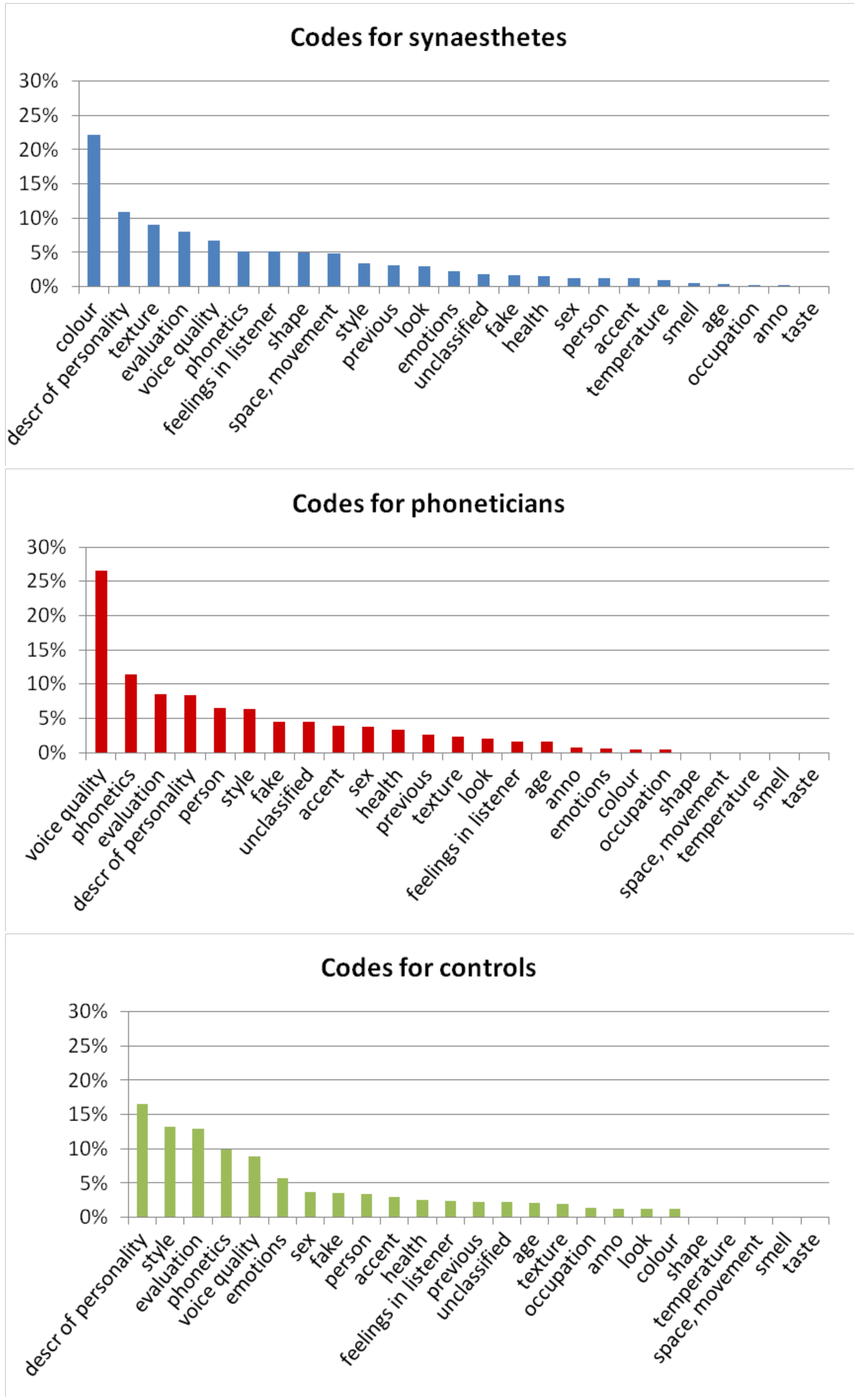


Figure 32: Codes assigned to verbal voice descriptions of synaesthetes, phoneticians and controls, sorted from most to least frequently used per group. Group percentages displayed.

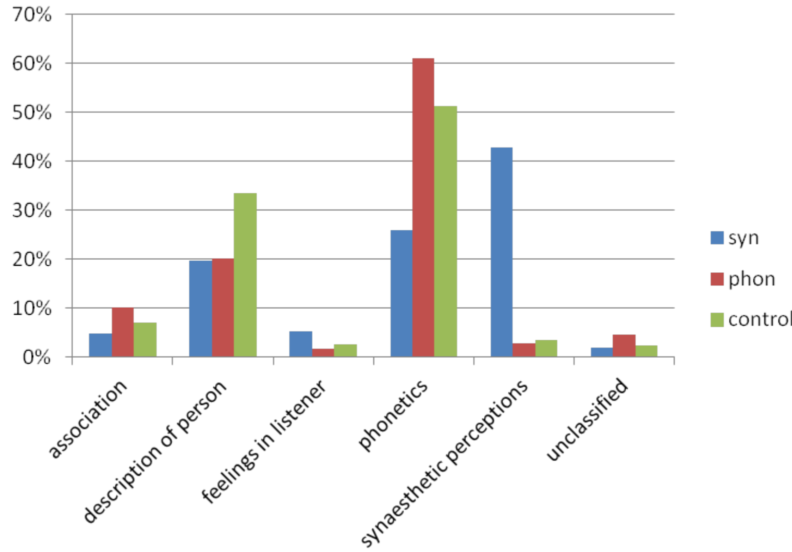


Figure 33: Categories of the verbal descriptions of the speakers' voices by synaesthetes, phoneticians and controls in percent.

emotional reactions to their synaesthetic perceptions, for example to a pleasant colour or unpleasant texture. The category 'phonetics' is used most often. It includes the codes *accent*, *evaluation*, *fake*, *phonetics*, *style* and *voice quality*. It was expected that most codes in this category would be assigned to phoneticians and much less to the other groups, but control participants also used this category surprisingly often. This is mainly due to two codes, *evaluation* and *style*. *Evaluation* describes how listeners find there is something unusual about a voice, e.g. "This voice is unpleasant to listen to; it could be used in a much better way". Because the VQs used in this experiment are of extreme manifestation, it seems a logical conclusion that participants evaluate these "near-pathologies". *Style* stands for the speaking style, such as newsreader, bedtime story etc. Most, if not all, people encounter different speaking styles in their daily lives and thus find it easier to access this information to use in their verbal descriptions. Both these codes are not phonetic in a strict sense and therefore skew the result of the frequent usage of the category 'phonetics' slightly towards a broader interpretation of the term phonetics here. Synaesthetes were found to have used terms of the category 'phonetics' least.

To test how far the use of verbal descriptions differed between the three groups, a MANOVA was carried out. Testing the difference between the six categories showed four significant differences (Gabriel's test was used for post hoc tests because group sizes were slightly different; see Field (2009, pp. 374-375)):

1. Synaesthetic perceptions: $F=22.8$, $p<.001$; mean difference between synaesthetes and controls 38.18, $p<.001$, between synaesthetes and phoneticians 38.74, $p<.001$

2. Phonetics: $F=14.904$, $p<.001$; mean difference between synaesthetes and controls

-22.87, $p<.001$, between synaesthetes and phoneticians -35.18, $p<.001$

3. Description of person: $F=4.084$, $p=.023$; mean difference between synaesthetes and controls -13.53, $p=.039$

4. Feelings in listener: $F=4.689$, $p=.014$; mean difference between synaesthetes and phoneticians 4.12, $p=.017$

5. Association and Unclassified did not yield any significant results.

A difference value of 38.18 means that synaesthetes used synaesthetic descriptions on average 38.18% more than controls. This confirms that synaesthetes use more synaesthetic expressions to describe the voices than non-synaesthetes. Words of the category 'phonetics' are used significantly more often by phoneticians and controls than by synaesthetes. Controls used words describing the speaker as a person significantly more often than synaesthetes. Synaesthetes were found to express slightly more feelings than phoneticians in their verbal descriptions.

Inter-individual differences in synaesthetes' verbal descriptions

The use of synaesthetic descriptions varied strongly within the group of synaesthetes (see Figure 34). The results for each synaesthete are plotted individually along the x-axis, sorted by the amount of usage of the category 'synaesthetic perceptions'. On one side, there are synaesthetes like ID 10 and 104, who use 'synaesthetic perceptions' almost exclusively to describe the voices. On the other side, there are also synaesthetes like ID 11 and 25, who, despite reporting having voice-induced synaesthesia, hardly used 'synaesthetic perceptions' in their verbal descriptions of the voices. The question arises whether those synaesthetes with low frequency usage of synaesthetic terms do indeed possess the type of synaesthesia investigated in this study or possibly one or more other forms of synaesthesia instead. Participant 25, a low frequency user of 'synaesthetic perceptions', stated: "Regardless of what is being said, I do associate colours and to a lesser extent textures to the voice that I hear. [...] voices generally translate into different shades of purple, yellow or brown in my head but are not specifically restricted to these colours. In regard to textures, I mostly envisage stone with different degrees of roughness depending on the voice, from smooth pepples [sic] to the most asperous rocks." As the statement is very differentiated, it can indeed be assumed that the participant does have voice-induced synaesthesia. Reasons why ID 25 chose not to describe the voices in synaesthetic terms can be manifold: he could be used to hiding his synaesthetic perceptions due to common reactions of irritation and incredulity from his peers, as is frequently described in anecdotal reports of synaesthetes in forums; he recognised the synaesthetic associations as secondary ones and only named the

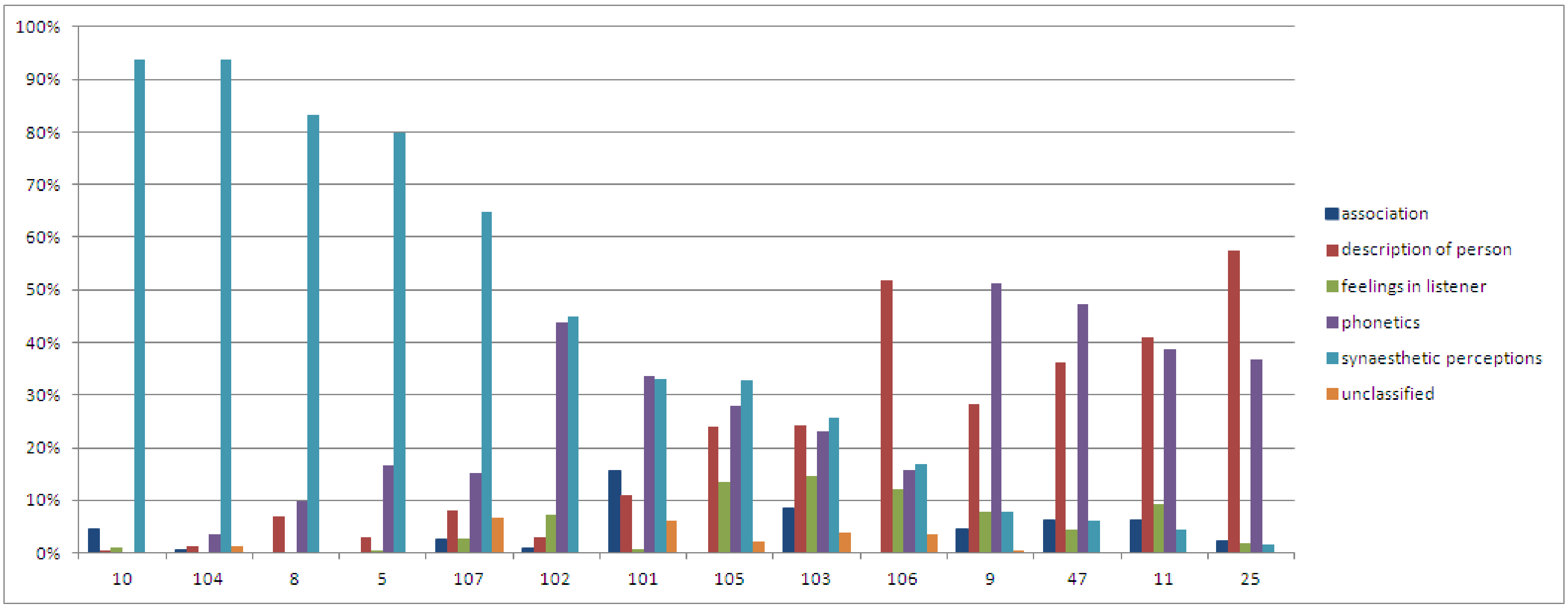


Figure 34: Usage of the six different categories in the verbal voice descriptions by individual synaesthetes in percent. Synaesthetes along the x-axis, named by their ID.

primary ones; as his colour and texture range seems limited, he might not have thought it reasonable to use these terms to describe the voices; etc. Although the nature of the experiment suggests a bias towards using the synaesthetic descriptions more than others (despite not having been asked explicitely to use these), some synaesthetes chose otherwise.

As introduced in Section 1.3, synaesthesia expresses itself in various different ways. Therefore types of synaesthesia or individual manifestations do not always fit the traditional classification scheme. It may thus be difficult to apply definitions of synaesthesia, which are based on the grapheme-colour type, to the voice-induced types of synaesthesia, which clearly have a more complex inducer. Seeing that the synaesthetic traits within my group of self-referred voice synaesthetes vary and may lie on a continuum, I have tried to interpret the individual differences by relating their usage of synaesthetic words to their demographics and their ratings of the intensity of their synaesthetic perceptions. The participants had various other kinds of synaesthesia, as shown in Table 11 on page 118, yet no relation to other types of synaesthesia seems to explain their individual differences in using synaesthetic descriptions for voices. Generally, those who use synaesthetic descriptions least (ID 9, 47, 11, 25) state that they have fewest types of synaesthesia. This would lead towards the interpretation that participants with fewer kinds of synaesthesia are more reluctant to make use of their perceptions in verbal descriptions. However, ID 5 for example claims to have no other types that induce colour or texture perceptions but uses synaesthetic terms frequently.

It was expected that there would be a relation between the low and high frequency users of synaesthetic verbal descriptions and their intensity rating depending on how strong their synaesthetic associations were, because I assumed that synaesthetes with stronger sensations would feel more compelled to express their perceptions. The average rating of the high frequency users (ID 10, 104, 8, 5, 107) was 6.7 on a scale from 0-9 (where 9 was the strongest intensity). Unfortunately, only one of the low frequency users rated the intensity, as this rating was optional, so no conclusions can be drawn from this comparison. However, her mean rating of 6.4 suggests no difference in the felt intensity of synaesthetic associations between high and low frequency users.

5.4.2 Colour choices

To detect colour choice preferences across participants, I first tested how often each colour was chosen in total. Some colours were chosen much more frequently than others (see Figure 35). Although one could hypothesise that vivid or generally preferred colours were chosen most often, this behaviour was not found. Instead, darker and more muted colours were chosen. This bias might be explained by the fact that only male voices were used, and a lower pitch is found to be associated with lower luminance (de Thornley Head, 2006;

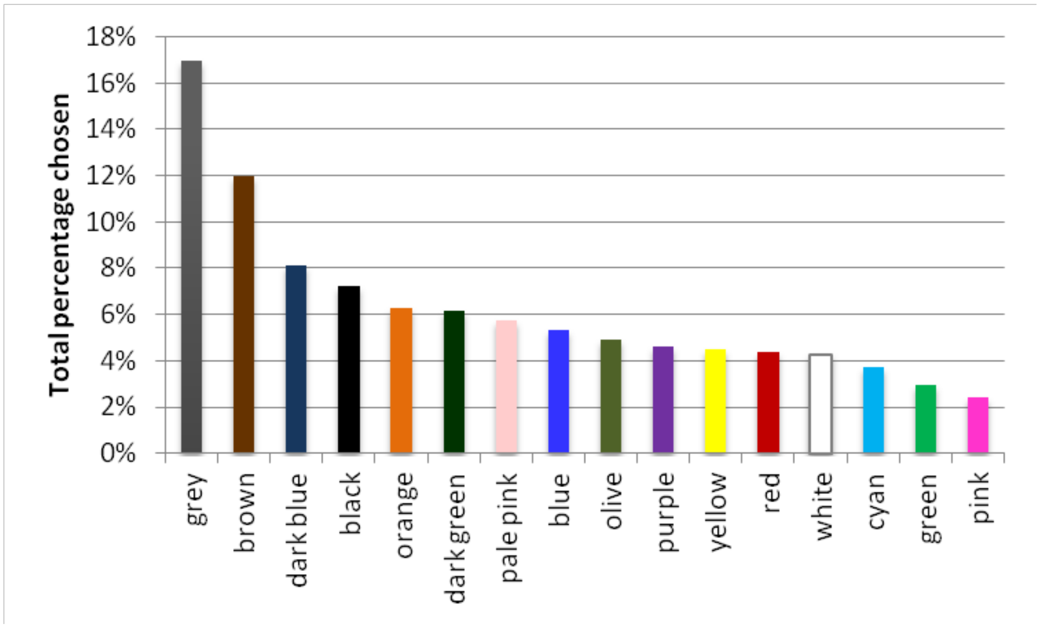


Figure 35: Frequencies of overall colour choices, averaged across VQs and participant groups, in percent.

Marks, 1974, 1987). The percentage of colour usage in synaesthetes or non-synaesthetes hardly diverged from the general pattern.

To address research questions 2 and 3 (cf. page 113), linear mixed effect modelling was carried out using R (Baayen, 2008). L*, u* and v* were used as dependent variables. The four acoustic features nominated in the factor analysis were used as predictors, as was participant group, and the interactions of group with each acoustic feature; participants were included as random effects. This way it could be examined how far mean f0, pitch range, LTF2 and the measure for spectral tilt (H1*-A3*) influenced the colour choices, split into the measures for luminance (L*), reddish-greenish (u*) and yellowish-bluish

		t	p	Qualitative explanation
L*	f0 mean	12.57	<0.001	higher f0, higher L*
	pitch range	6.58	<0.001	larger range, higher L*
	LTF2	2.46	0.014	higher LTF2, higher L*
	group (control vs. phon)	-3.52	<0.001	phon darker
	LTF2*group (control vs. phon)	3.67	<0.001	phon>control
u*	f0 mean	5.30	<0.001	higher f0, redder
v*	spectral tilt	-2.31	0.021	steeper tilt, bluer
	pitch range	2.50	0.012	smaller range, bluer

Table 17: Results of linear mixed effect modelling testing acoustic influences on participants' colour associations, measured on the luminance (L*), reddish-greenish (u*) and yellowish-blueish (v*) scales.

(v*). Recall that a positive u* value stands for red tint and a negative one for green tint; a positive v* value stands for yellow tint and a negative one for blue tint (cf. Table 14). For each dependent variable, the model-fitting procedure was as follows: first the simplest model was fitted, with participants as random effects. Then, predictors were introduced into the model one at a time and were retained only if they contributed significantly to the fit of the model, as assessed by log-likelihood tests. This procedure was continued until a final model with all variables contributing significantly to it had been obtained. Table 17 shows the results of the linear mixed effect modelling in detail. Refer back to Table 13 on page 128 to see the values of acoustic measurements for each VQ. The key results are:

- A higher f0, a higher LTF2 and a larger pitch range led to lighter colour choices across groups, whereas a lower f0, a lower LTF2 and a smaller pitch range resulted in darker colour associations.

- A higher f0 led to redder colour choices across groups, whereas a lower f0 resulted in greener colour associations.

- A steeper spectral tilt and a narrower pitch range led to bluer colour choices across groups, whereas a shallower spectral tilt and a larger pitch range resulted in yellower colour associations.

The influence of f0 on luminance associations is in line with findings in musical tone-colour synaesthesia (Ward et al., 2006), and the influence of LTF2 on luminance associations is in line with findings of the experiment on vowel sound - colour associations. There is a group difference between controls and phoneticians on the luminance scale: phoneticians associated the voices with significantly darker colours than controls. LTF2 interacts with group in the following way: phoneticians use the luminance scale more extensively when driven by LTF2 than controls; there is no significant difference between synaesthetes and others. Perhaps the phoneticians are more sensitive to the intrinsic vowel pitch, or to formant frequencies in general, than controls.

Figure 36 illustrates how colour associations are distributed across VQs. The figure shows a pie chart of the colour choices of all participants across groups, for all four *creak* stimuli on the left and all four *falsetto* stimuli on the right. It illustrates that a higher f0 (*falsetto*) both triggered lighter and redder colour associations in participants compared to the darker associations for the low pitched *creak*. Additionally, for reference, Appendix G displays pie charts of the colour choices per participant group for all voice qualities separately.

5.4.3 Textures

The overall usage of textures was more evenly distributed across the 16 choices than was

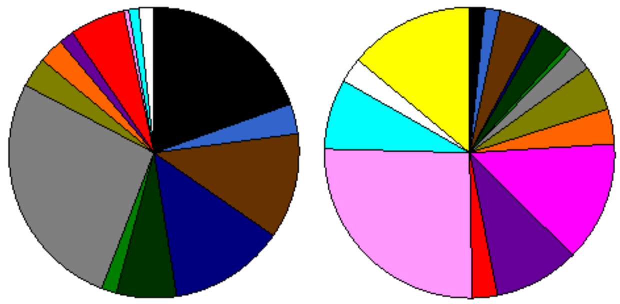


Figure 36: Colour associations with *creak* (left) and *falsetto* (right) across all participants. Stimuli of these associations include both speakers and both sentences for *creak* and *falsetto*.

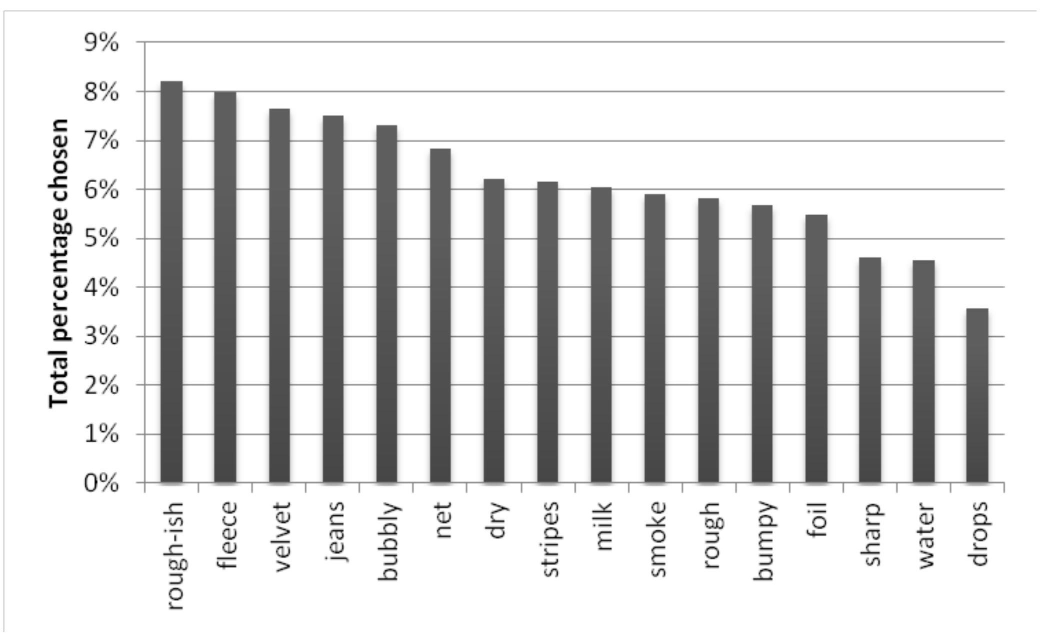


Figure 37: Frequencies of texture choices overall, across VQs and participant groups, in percent.

the case with colours. While the most frequently chosen colour was used 16.5% of the time and the least frequently chosen only 2.6%, the span of usage for textures lay between 3.8% and 8% (see Figure 37). In other words, the tendency found in colour choices, namely for participants across all groups to favour certain colours over others, could not be found for textures. One reason for this could be that textures are more difficult to classify than colours. For colours, there are clear categories – the 11 focal colours (Kay et al., 2009) – which then can be refined further if needed. For textures, we do not have such a defined system. There is no established terminology because the words people use to describe textures vary much more greatly than, for example, those used for colours. This can only be changed by introducing textures as a standard definition set of objects, as is the case with colours in present-day western culture, where children learn a set of colour terms from an early age.

To address research questions 2 and 3 (cf. page 113), I used the texture data elicited in the texture description experiment that was described in Chapter 4 for the quantitative analysis. This analysis aims to determine which acoustic features triggered what kind of texture choices, and whether synaesthetes and non-synaesthetes differed in these choices. The four semantic differentials that were chosen to be used for further analysis were 'rough–smooth', 'high–low complexity', 'repetitive–non-repetitive' and 'line-like vs. blob-like'. Those differentials were used as dependent variables in the linear mixed effect modelling, representing the texture choices of participants. The four acoustic features were again used as predictors to find out how far differences in the acoustics influence texture associations. Detailed statistical results, with brief qualitative explanations in the last column, are shown in Table 18. Analyses were carried out separately for each semantic differential. Acoustic measures that did not yield significant predictions were not included in the final models. Key results are summarised in Table 19. It was found that both higher pitch and higher LTF2 were associated with textures that were 'smoother', 'more complex' and 'less repetitive'. A steeper spectral tilt resulted in 'smoother', 'less complex' and 'blob-like' texture choices, whereas a larger pitch range triggered choices of textures with 'more repetitive' patterns. A small but significant group difference in the usage of textures was found for the complexity scale: synaesthetes chose 'more complex' textures than phoneticians and controls. This could be due to the fact that synaesthetic concurrents are on average more complex in their structure than associations of non-synaesthetes which are only consciously present when triggered by visual input.

A comparison with previous research is difficult because no study has looked at the relationship between texture attributes and acoustic features of voices, sounds or music. Gomez & Danuser (2004) for example presented an extensive study on affective and physiological responses to environmental sounds and music. Arousal and valence were analysed in great detail, but sounds were not quantified in any way, i.e. the results are

		t	p	Qualitative explanation
rough-smooth	f0 mean	2.63	0.009	higher f0, smoother
	pitch range	2.30	0.022	larger range, smoother
	LTF2	4.30	<0.001	higher LTF2, smoother
	spectral tilt	5.18	<0.001	steeper tilt, smoother
high complexity-low complexity	f0 mean	-3.02	0.003	higher f0, more complex
	pitch range	4.33	<0.001	smaller range, more complex
	LTF2	-2.13	0.034	higher LTF2, more complex
	spectral tilt	7.11	<0.001	shallower tilt, more complex
	group (syn vs. con)	-2.40	0.016	synaesthetes more complex
	group (syn vs. phon)	-2.46	0.014	synaesthetes more complex
repetitive-non-repetitive	f0 mean	4.40	<0.001	higher f0, less repetitive
	pitch range	-2.48	0.013	smaller range, less repetitive
	LTF2	5.93	<0.001	higher LTF2, less repetitive
linelike-bloblike	spectral tilt	3.61	<0.001	steeper tilt, more bloblike

Table 18: Results of linear mixed effect modelling testing acoustic influences on participants' texture associations.

Pitch		**Spectral tilt**	
Low	**High**	**Shallow**	**Steep**
rougher	smoother	rougher	smoother
less complex	more complex	more complex	less complex
more repetitive	less repetitive	line-like	blob-like

Pitch range		**LTF2**	
Small	**Large**	**Low**	**High**
rougher	smoother	rougher	smoother
more complex	less complex	less complex	more complex
less repetitive	more repetitive	more repetitive	less repetitive

Table 19: Summary of key influences of acoustic features on texture associations.

all related to nominal sound and composition categories, thereby no assumptions about acoustic values and hence no comparison can be made. Another study looked at the categorisation of emotion and affect. Russell created a circle with four bipolar scales describing affective meaning (see e.g. Russell & Pratt, 1980): pleasure, arousal, interest and relaxation. One acoustic scale of my stimuli coincides with the 'interest' scale of Russell – laypeople describe a small pitch range as boring or monotonous and a large pitch range as interesting. 'Boring' and 'interesting' are the extremes of the 'interest' scale. A neighbouring axis in the circumplex to the scale 'interest' is the scale 'pleasure', with 'unpleasant' next to 'boring' and 'pleasant' next to 'interesting'. Neighbouring affections such as 'pleasant' and 'interesting' are likely to correlate. This might suggest that voices with a larger pitch range are associated with more interesting and pleasant textures than voices with a smaller pitch range. The other three acoustic criteria observed here – mean f0, spectral tilt and LTF2 – unfortunately remain uncompared to other findings as they have not been researched in conjunction with textures before.

5.4.4 Semantic differentials

Participants' ratings on the eight semantic differentials were analysed using linear mixed effect modelling with the four acoustic features and participant group being the predictors. One semantic differential at a time was analysed. Significant results are summarised in the following eight subsections. The respective t and p values are listed in Table 20.

Table 20: Significant results of linear mixed effect modelling with acoustic features and participant groups as predictors of the semantic differential ratings. The table shows significant differences between the ratings of acoustic features, between groups, and interactions of acoustic features and groups. Non-significant results are omitted. Syn = Synaesthetes, con = controls, phon = phoneticians.

		t	p
light-dark	mean f0	-18.758	$<$0.001
	pitch range	-7.042	$<$0.001
	LTF2	-3.024	0.002
	group: syn vs. con	-2.651	0.008
	LTF2*con	2.710	0.008

		t	p
low-high	mean f0	15.835	<0.001
	pitch range	2.315	0.021
	LTF2	6.471	<0.001
	group: syn vs. con	-2.417	0.016
	f0*phon	3.328	<0.001
	pitch range*syn vs. con	-2.399	0.017
colourful-grey	mean f0	-10.69	<0.001
	pitch range	-5.48	<0.001
	group: con vs. syn	-4.02	<0.001
	pitch range*con vs. syn	2.04	0.041
smooth-rough	pitch range	-8.51	<0.001
	spectral tilt	-17.56	<0.001
	group: con vs. phon	2.30	0.022
dry-fluid	mean f0	4.889	<0.001
	pitch range	3.932	<0.001
	LTF2	-2.396	0.017
	spectral tilt	4.345	<0.001
	group: con vs. syn	3.191	0.001
	group: phon vs. syn	2.808	0.005
	f0*con vs. syn	-3.266	0.001
	f0*con vs. phon	-3.302	0.001
	LTF2*con vs. syn	-2.718	0.007
	LTF2*syn vs. phon	2.926	0.004
	spectral tilt*con vs. phon	2.685	0.007
hard-soft	mean f0	4.161	<0.001
	pitch range	2.638	0.008
	spectral tilt	10.849	<0.001
	group: con vs. phon	-3.764	0.008
tense-relaxed	mean f0	-6.269	<0.001
	pitch range	2.912	0.004
	LTF2	-3.852	<0.001
	spectral tilt	17.583	<0.001
	group: syn vs. phon	-3.008	0.003
	group: con vs. phon	-3.066	0.002
	f0*syn vs. phon	-2.177	0.030
	f0*con vs. phon	-2.118	0.034
	LTF2*syn vs. phon	3.108	0.002

		t	p
sharp-fuzzy	mean f0	-3.051	0.002
	LTF2	-5.638	$<$0.001
	spectral tilt	2.398	0.017
	group: syn vs. con	-2.203	0.028
	group: syn vs. phon	-2.282	0.023
	LTF2*syn vs. con	2.390	0.017
	LTF2*syn vs. phon	2.417	0.016

light–dark

f0 mean: The higher f0, the *lighter* the ratings.

Pitch range: The larger the pitch range, the *lighter* the ratings.

LTF2: The higher LTF2, the *lighter* the ratings.

Group: Controls rated voices on average *lighter* than synaesthetes.

Interaction LTF2*group: Controls rated voices with higher LTF2 *darker*, whereas phoneticians rated them *lighter*. Synaesthetes were not influenced in their lightness rating by LTF2.

These ratings confirm the luminance data results of the colour choices. This suggests that people do not actually need to see a colour or have other visual luminance input, it is sufficient to think of the concept of lightness. That phoneticians rate voices with higher LTF2 as *lighter* in colour, whereas controls do it vice versa, suggests that phoneticians are well trained in listening to the spectral frequency details in higher frequencies than f0, because they show the same behaviour for increasing LTF2 as for increasing pitch. As F2 of a vowel is sometimes also referred to as the vowel's 'intrinsic pitch', this seems to be a congruent behaviour of the experts.

low–high

f0 mean: The higher f0, the *higher* the ratings.

Pitch range: The larger the pitch range, the *higher* the ratings.

LTF2 The higher LTF2, the *higher* the ratings. This was a weak effect only.

Group: Controls rated the voices on average *lower* than synaesthetes.

Interaction f0*group: Although all groups rated higher f0 as *higher*, phoneticians used the low–high scale more extensively.

Interaction pitchRange*group: Controls used the scale slightly more extensively than synaesthetes.

The main effects of these ratings are the same as those for the light–dark scale. This suggests that participants judge these two scales to express similar things; they seem to be semantically intertwined. The more extensive use of the scale by phoneticians reflects their professional experience in estimating the fundamental frequency of voices.

colourful–grey

f0 mean: Higher f0 was rated as more *colourful* by phoneticians and controls; synaesthetes were less influenced by f0.

Pitch range: The larger the pitch range, the more *colourful* the ratings.

Group: Synaesthetes rated the voices as more *colourful* than controls.

Interaction pitchRange*group: Controls rated a larger pitch range as more *colourful*; this effect is weak for synaesthetes.

One could hypothesise that 'colourfulness' goes with colour warmth. Yellow and red are perceived as warmer than blue and green (Ou et al., 2004). Seeing that a larger pitch range was rated as yellower and a higher f0 as redder, this reflects participants' colour choices in the colour response display.

smooth–rough

Pitch range: The larger the pitch range, the *smoother* the ratings.

Spectral tilt: The steeper the spectral tilt, the *smoother* the ratings.

Group: Phoneticians rated the voices on average as *rougher* than controls.

These ratings can be compared to pleasantness. A *breathy* or *modal* voice is usually perceived as pleasant to listen to (see e.g. Reich & Lerman, 1978, for pleasant ratings on acoustic attributes associated with breathy and modal), and those VQs show a steep spectral tilt. Moreover, a smooth texture is usually perceived as pleasant; individuals rate the touch of fur or velvet as more pleasant than that of sandpaper (e.g. Rolls et al., 2003).

dry–fluid

f0 mean: The higher f0, the more *fluid* the ratings.

Pitch range: The larger the pitch range, the more *fluid* the ratings.

LTF2: The higher LTF2, the *drier* the ratings. This was a weak effect only.

Spectral tilt: The steeper the spectral tilt, the more *fluid* the ratings.

Group: Phoneticians and controls rated the voices on average as *drier* than synaesthetes.

Interaction f0*group: Although the associations of all groups go the same way (higher f0 means more *fluid*), this effect is found to be stronger for controls than for phoneticians and synaesthetes.

Interaction LTF2*group: Although the associations of all groups go the same way (lower LTF2 implies more *fluid*), this effect is found to be stronger for synaesthetes than for phoneticians and controls.

Interaction spectral tilt*group: Although the associations of all groups go the same way (steeper tilt means more *fluid*), this effect is found to be stronger for phoneticians than for controls.

It seems intuitive that acoustic features that are characteristic for *harsh* and *creak*, such as low f0, small pitch range and shallow spectral tilt, provoked *dry* ratings; these voice qualities make the speaker's throat seem like it is dry and irritated which leads to irregularities.

hard–soft

f0 mean: The higher f0, the more *soft* the ratings.

Pitch range: The larger the pitch range, the more *soft* the ratings. This was a weak effect only.

Spectral tilt: The steeper the spectral tilt, the more *soft* the ratings.

Group: Phoneticians rated voices on average as more *hard* than controls.

A breathy voice could commonly be referred to as soft spoken. So again, associations are intuitively appropriate as *breathy* has a steep spectral tilt. Lower f0 might have been rated as more *hard* because the speakers needed to put more vocal effort into producing VQs with low f0, or maybe because the speaker sounds more masculine – two adjectives that are often associated with one another as a consequence of human stereotypes.

tense–relaxed

f0 mean: The higher f0, the more *tense* the ratings.

Pitch range: There is a weak effect of a larger pitch range being rated as more *relaxed*.

LTF2: The higher LTF2, the more *tense* the ratings.

Spectral tilt: The steeper the spectral tilt, the more *relaxed* the ratings.

Group: Phoneticians gave on average more *tense* ratings than synaesthetes and controls.

Interaction f0*group: Although all groups rated voices with higher f0 as more *tense*, this effect is found to be stronger for phoneticians than synaesthetes or controls.

Interaction LTF2*group: The lower LTF2 is, the more *relaxed* synaesthetes rated it. This effect is non-existent for phoneticians.

As the stimuli were produced by two male speakers, it is a natural process that they have to tense their larynx and vocal tract to achieve higher fundamental frequency. Hence, the results of rating a higher f0 as more *tense* is a congruent reaction of the participants, showing their passive knowledge of vocal tension. The fact that phoneticians use this rating to a stronger degree once again demonstrates their expert knowledge of voice quality perception. It is also interesting to compare this perception with findings by Russell & Pratt (1980), who found that the attribute *tense* has a distressing quality, and *relaxed* a relaxing quality in their bipolar space of affective meaning.

sharp–fuzzy

f0 mean: The higher f0, the more *sharp* the ratings.

LTF2: The higher LTF2, the more *sharp* the ratings.

Spectral tilt: The steeper the spectral tilt, the more *fuzzy* the ratings. This finding applies mainly to synaesthetes and controls.

Group: Synaesthetes and phoneticians rated the voices on average as more *sharp* than controls.

Interaction LTF2*group: Although all groups rated voices with higher LTF2 as more *sharp*, this effect is found to be weaker for synaesthetes.

High energy in higher spectral frequencies (which are less damped in a shallow spectral tilt and more in a steep one) is often perceived as sharp, for example a shrill voice or an alarm. This is congruent with the ratings of a higher f0, higher LTF2 and a shallower spectral tilt as being more *sharp*.

Many significant results were found by asking participants to rate the voices on eight semantic differentials. This decisively suggests that associations are processed in a conceptual manner, in front of the mind's eye; no visual stimulation is needed to bring it onto a perceptual level. This was found across groups. In summary – looking at the data from the perspective of acoustics rather than semantic differentials individually – a higher frequency value in f0 or LTF2 resulted in more *light, high, tense* and *sharp* ratings. A higher f0 was additionally judged to be more *colourful*, *fluid* and *soft*, whereas a higher LTF2 was perceived as more *dry*. A larger pitch range or a steeper spectral tilt were associated with more *smooth*, *soft* and *fluid* features. A larger pitch range was additionally perceived as more *light*, *low* and *colourful*.

The phoneticians seem more inclined to assign those attributes to the voices, the majority of which seem to be associated with negative connotations: more *dark, rough, dry, hard, tense, sharp* and *colourful* (with the latter being an exception from having a negative connotation). For controls, the connotations seem to be fairly mixed: more *light, low, grey, smooth, dry, soft* and *fuzzy*. Synaesthetes relate more positive connotations to the voices: more *high, colourful, smooth, fluid, relaxed* and *sharp* (with the latter being an exception from having a positive connotation). To interpret these diverse reactions is difficult. One possibility could be that phoneticians perceive the VQs in this extreme manifestation as pathological or dysfunctional voices, and therefore the connotations with the voices are rather negative. Synaesthetes, on the other hand, might enjoy being able to express their synaesthetic associations, rather than ignoring them in order to listen to the content of what the voice is saying; hence, the usage of the sliders is skewed towards the more positive side.

5.4.5 Retest

12 synaesthetes, 10 phoneticians and 20 controls took part in the retest which consisted of a subset of ten stimuli. Table 21 shows how often participants chose exactly the same colour and texture again for exactly the same speech recording. Although the percentages displayed in Table 21 illustrate that synaesthetes chose the same colours and textures in the retest more often than the other groups (apart from phoneticians who showed a higher consistency in texture associations), a Pearson χ^2 test did not reveal significant group differences: χ^2=5.066, p=0.079 for colours; χ^2=1.882, p=0.390 for textures. If synaesthetes and controls are compared in a χ^2 test excluding the group of phoneticians,

	Colours	Textures
synaesthetes	25%	15%
phoneticians	21%	19%
controls	15%	13%

Table 21: Consistency results: percentage of the time that exactly the same colours and textures were chosen in main test and retest, per participant group.

synaesthetes did indeed choose the same colour significantly more often than controls (χ^2=4.923, p=0.027). As it cannot legitimately be argued that it would be appropriate to exclude one group from the analysis, this result can only be interpreted as a tendency of synaesthetes to be more consistent in choosing the same colour in a retest than controls, but not significantly so.

Another χ^2 test was carried out to test whether there was a relationship between the frequency of using synaesthetic terms in the verbal descriptions and the consistency of colour and texture associations. The synaesthetes were grouped into high, medium and low users of synaesthetic descriptions in their verbal descriptions. ID 10, 104, 8, 5 and 107 were in the high user group; 102, 101, 105, 103 and 106 were in the medium one; 9, 11, 47 and 25 were in the low user group. Unfortunately, 9 and 25 did not do the retest, so they could not be included in this test. It would be interesting to know whether this was a random drop-out of participants, or whether they dropped out because they were low frequency users and assumed the participation in the experiment to be meaningless for them. Results of the test revealed no significant differences in consistency of colour or texture associations between high, medium and low frequency user groups (Pearson χ^2=1.25, p=0.535).

To find out whether in the retest participants chose colours, textures and semantic differentials *similar* to their choices in the initial test, a MANOVA was conducted. L*, u* and v* were used as the measures for colour. As mentioned on page 120, the slider values of the semantic differentials and texture ratings were scores between 0-100, with the left-most slider position being 0 and right-most position 100. Participant groups were the fixed factors and the response measures dependent variables. For the post hoc test, Tukey's and Gabriel's methods were used; Tukey is more powerful while having good control over the Type I error, and Gabriel takes different group sizes into account (Field, 2009). As results were very similar, only the results of the Tukey's test are shown. First, significant results of the consistency across VQs are displayed. However, since *whisper* displayed several significant results in terms of consistency, it is discussed separately.

Phoneticians were more consistent in choosing 'rough-smooth' texture (t=6.40, p=0.04), 'fine-coarse' texture (t=4.50, p=0.053) and in using the 'smooth-rough' scale of the semantic differentials (t=4.12, p=0.035) than synaesthetes. It could be argued that these three scales are very similar to voice quality ratings (as in the vocal profile analysis of

Beck, 2005), for example for harsh or creaky, and that they are basically translated from texture into acoustic phonetic perception. Phoneticians may therefore be most consistent in these measures due to their profession demanding them to rate VQs on fine scales, so they learn to be more accurate. Looking further at semantic differentials, synaesthetes were significantly less consistent on the colourful-grey scale than phoneticians ($t=7.89$, $p<0.001$) and controls ($t=6.08$, $p=0.002$). This is opposite to what I had speculated. This may be because saturation or chromaticity of their synaesthetic perceptions changed, for example they may have listened to the stimuli in a different volume or in a different context. All group differences for consistency in colours were non-significant. However, v^* approached significance between synaesthetes and controls ($t=12.37$, $p=0.063$), indicating that synaesthetes are marginally more consistent in the yellow-blue dimension.

Looking at consistency in individual VQs, no coherent pattern could be found. For six VQs there was a significant difference between groups for one out of the 19 measures,[9] and these consistencies seemed to be arbitrarily distributed across groups. For two VQs there were no group differences at all. Yet, for *whisper*, group differences in consistency were found on six semantic differential scales, with four of them showing higher rating consistency for the phoneticians, compared to the two other groups. Phoneticians were more consistent in their colour choice than controls ($t=-45.48$, $p=0.018$ for u^*; $t=-45.03$, $p=0.010$ for v^*). Also synaesthetes are more consistent than controls on the red-green scale ($t=-38.24$, $p=0.038$). Interestingly, consistency is reversed for *colourful-grey*; here controls are more consistent than synaesthetes ($t=22.26$, $p=0.001$), in agreement with the result found across VQs. In the texture domain, phoneticians show larger consistency compared to synaesthetes on 'rough-smooth' ($t=20.28$, $p=0.021$) and 'fine-coarse' (16.75, $p=0.033$) scales for *whisper*.

While trying to disentangle and interpret these mixed results, I can only refer back to the conclusion in Section 1.7: the definition of synaesthesia has not yet reached its final state. As Simner (2011) points out, synaesthetic perceptions are usually consistent in synaesthesia, but not always. Because voice as an inducer has not been researched systematically yet, it cannot necessarily be concluded that this type of synaesthesia shows as strong consistency patterns as other types.

5.5 Further discussion

An exploratory study on voice-induced synaesthesia was presented using both qualitative and quantitative analyses. The qualitative approach – coding and analysing the verbal descriptions – gave insights into the different ways participants perceive the voices, and which words they used to describe them. This was a necessary first step towards understanding this under-researched type of synaesthesia. As conjectured, it showed that

[9] They consist of 3 measurements for colours (L^*, u^*, v^*), 8 for textures and 8 for semantic differentials.

synaesthetes regularly use their synaesthetic perceptions to describe voices, phoneticians used more technical terms, and controls focus on describing personal characteristics of the speaker.

Semantic differential ratings and colour and texture choices were used for linear mixed effect modelling as a systematic quantitative approach to test how participants' responses were influenced by acoustic criteria of the different voice qualities. Higher fundamental frequencies were associated with lighter colours across groups. This is in line with the findings of Ward et al. (2006) on musical pitch, but group differences were found by de Thornley Head (2006). For his synaesthetes, pitch changes did not affect the lightness of their colour choices. A higher f0 also led to redder colour choices in my participants. Although not statistically demonstrated in de Thornley Head's article (2006), the graphs on p. 170 suggest that his participants also associated redder colours with higher pitched tones. A larger pitch range resulted in lighter and yellower colour choices across groups, while a steeper spectral tilt triggered more blue colour associations across groups. A recent study on pitch-luminance mapping found that even chimpanzees prefer to match white to high pitched sounds and black to low pitched sounds (Ludwig et al., 2011). This finding suggests a common underlying mechanism of sensory processing in primates, which seems to be hard-wired rather than acquired through culture or language. Neurons in the auditory cortex are organized according to the frequency of sound to which they respond best in both humans and other primates (Talavage et al., 2004; Bendor & Wang, 2005; Lauter et al., 1985). This means that f0 of different frequencies should be processed at slightly different regions of the primary auditory cortex, similar to the arrangement of sound processing in the cochlea. This frequency map in the auditory cortex may possibly relate systematically to a luminance map in V4.

A high f0, high vowel formants, a larger pitch range and a steeper spectral tilt resulted in associations with textures that were 'fluid', 'smooth' and had a pleasant connotation, whereas a shallow spectral tilt resulted in 'rough' and 'line-like' texture choices and those with unpleasant connotations. Additionally, consistency in people's associations was measured for textures, colours and semantic differentials by doing a retest. Although not statistically significant, synaesthetes seem to be more consistent in some of their associations than controls, mostly in colour associations. Phoneticians also show regular patterns in some measures that underline their expert knowledge of assessing VQs, e.g. the extensive use of the *smooth-rough* scale of the semantic differentials.

A study on the affect of texture and colour was published by Lucassen et al. (2011). The authors list colour and texture features within semantic differentials as well, namely warm-cool (for colours only), feminine-masculine, hard-soft and light-heavy. They found that more complex textures are more masculine, hard and heavy. Light weight is associated with light colours and heavy with dark colours. Soft is associated with less saturated colours than hard. Feminine is rated as more pink and masculine more blue, green and dark. Warm is perceived as more red and brown, whereas cool is more blue and green. A comparison between their and my experiment is not explicit, because they looked at relations between colours and textures, whereas I investigated relations between acoustic features of voices on one side and colours and textures on the other. Nonetheless, parallels can be found for the feminine-masculine scale and ratings for f0 by my participants: high f0 – *falsetto* reaching frequencies of a female voice – were associated with redder colours than low f0, matching their findings exactly. Furthermore, one synaesthete, ID 5, had temperature concurrents with the voices which she expressed in the verbal descriptions. It is noticeable that her comments of warm temperature are often accompanied by red colour choices, and cold temperature by green or sometimes blue colour choices.

With this experiment, we obtained insight not only into the idiosyncratic differences but also similarities of participants' perceptions of different voice qualities, both within and across participant groups. The surprising lack of strong consistency in synaesthetes' choices leaves two main interpretations: (1) Voice-induced synaesthesia is not as easily defined as other types of synaesthesia; ergo, the definition of synaesthesia needs to be revised, especially regarding consistency as a main criterion, as has recently been suggested by Simner (2012a), as discussed in 1.7 and mentioned on page 136; (2) Perceptions might be influenced by other types of synaesthesia to a certain degree, and it is impossible to separate these influences in the analysis.

It is in the nature of exploratory research that multiple findings are made; some of them are easy to understand, others are more difficult to interpret and need to be discussed. The influence of other synaesthesiae is one of the largest difficulties faced in this experiment. All but one synaesthete reported having at least four other types of synaesthesia. Of the 14 synaesthetes, 13 had music as an additional inducer, 11 had letters and 9 had words. These will undoubtedly be the types of synaesthesia interfering most strongly with the sound of a voice. The influence of words on the synaesthetic perceptions could have been avoided by using nonsense syllables. However, some word synaesthetes also had perceptions with meaningless words or syllables that sound similar to a word. Moreover, participants would have reported the same associations for all stimuli containing the same sentence if the interference of words as an inducer had been overwhelmingly strong. This scenario was not found.

The influence of letters is even more difficult to eliminate. One synaesthete, who did not participate in the experiment because she found she could not reply to my "rigid" questions, offered to paint some of the stimuli. The painting of her visualisation of the sentence "People look, but no-one ever finds it" read by Laver with a raised larynx is shown in Figure 38. In this painting, an influence of letters can be observed in three places: the first letter <P>, then the <l> whose sound is prolonged due to a clash of the end of

Figure 38: Painting of a *raised larynx* production of John Laver saying "People look, but no-one ever finds it" by Tessa Verrecchia.

"People" and the beginning of "look", and finally the <k> of "look" which sounds rather pronounced due to a long phrase break after this word. Influences of letters could have been avoided by using hummed stimuli, for example. However, this would not have been a good solution because on the one hand, it is impossible to hum different voice qualities because of limited manipulation scope for both laryngeal and supralaryngeal features, and on the other hand, it did not seem a realistic representation of the inducer 'voice'. Besides, although the influence of letters is clearly visible in this painting, it does not necessarily interfere with the perception of the voice; it may simply be superimposed on it. The voice appears to be blue to her, and intensity seems to fade towards the end (towards the right), after it had taken a wavy pattern that might represent the intonation pattern and/or volume.

Music can be regarded as one of the most similar inducers to voice quality sounds. In fact, when asking potential participants whether they have synaesthetic perceptions with the sound of a voice, many people thought of singing voices. Although it was mentioned that speaking voices would be researched, it could be that some participants failed to read this information and instead meant their music-related synaesthesia when confirming they have voice-induced synaesthesia. Many synaesthetes who have additional sensations with a singing voice also have sensations with a speaking voice, though to a lesser degree, as synaesthetes mentioned (personal communication). This attenuation could be a reason why the synaesthetes' responses were on average not as consistent as expected.

It could also be argued that the choice of stimuli was not ideal. Two speakers might not offer enough variation even if they produced different VQs. This would at least be the case for familiar voices, as some phoneticians recognised the speakers (both well known phoneticians), even through their voice quality "disguises". Conversely, Kuwabara & Takagi (1991) found that shifting the first three formants by only 5% affected the perception of voice-individuality to such a degree that their participants perceived the manipulated recordings as a different voice. Pitch and bandwidth manipulations, on the other hand, did not change the perception of voice-individuality to such an extent. I calculated the range of +/–5% for LTF2 of all 40 stimuli and tested whether each stimulus lay within or outside the 5% range of every other stimulus. Of the 1600 pairwise comparisons, 812 lay outside the range, indicating that approximately half of the stimuli have an LTF2 that is more than 5% different from the others. Keeping in mind that each VQ was presented twice per speaker (and assuming that the LTF value was similar for these two productions), it can be deduced that on average every speaker had a different LTF2 to a degree that each VQ realisation sounded different from the others. The largest differences were found for *whisper*, *raised larynx*, *lowered larynx* and *harsh*.[10]

Occasionally, some synaesthetes complained about the limited set of colours. However, using a more complicated colour response display was deemed cumbersome, as it would have made the already lengthy experiment even longer. Also, the set of textures was not very comprehensive. It remains untested how to present textural displays more optimally to synaesthetes, as no other study has investigated this so far. One option would be a browsing environment similar to that presented by Halley (2011) and Clarke et al. (2011), although this would be very time consuming. There were fewer comments by synaesthetes about the limited set of textures than about that of colours. Potentially, their synaesthetic reactions were less clear or more indifferent for textures than for colours; or they found it easier to match their concurrents to the given set because their textures were more similar to those on display.

Finally, there are general influences of an online survey that cannot be controlled by the experimenter. For this experiment, sound quality of participants' audio devices is a crucial factor. Participants were asked to use the audio device with the highest standard they have, i.e. headphones or external speakers. Although the sound quality of their devices was unknown, I assume that in any case it was good enough to allow them to distinguish between the different VQs. While the sound quality should have been the same for every participant throughout the experiment, there is a very high probability that they differed between participants. This might have an unwanted effect on the comparison across participants. Some participants might have given slightly different ratings had they heard the stimuli on another participant's audio device. With the technical standard being high nowadays, it is assumed that the effect is small. Also a difference in the colour display can be expected between participants' monitors. As the variation in the colour display was restricted to 16 colours, I do not expect that the different screen settings had a large impact in this respect.

In summary, it can be concluded that most results show similarities across participant groups, feeding into the discussion that being synaesthetic lies on a continuum. This suggests common underlying mechanisms in associations, which synaesthetes access on a conscious and non-synaesthetes on a subconscious level. The results highlighting the

[10]LTF2 of *falsetto* could not be measured because the spectrogram only displayed harmonics. Otherwise this would have probably been most different from other LTF2 measures.

individual differences in the verbal descriptions and in the consistency data within synaesthetes suggest that more emphasis should be put on these differences within synaesthetes, and they should be taken into account when 'classifying' synaesthetes. In fact, a categorisation of synaesthetes and non-synaesthetes might not be achievable in the same way for the various types of synaesthesia; the entanglement of multiple types of synaesthesia within one individual has to be taken into account.

6 Voice identification performance by synaesthetes and others

6.1 Introduction

We identify familiar voices on a daily basis, for example when answering the phone or listening to the radio. The "hello" of a dear friend on the phone is often enough by which to identify her. The task of identifying a voice becomes much more difficult when it has only been heard once, i.e. if it is an unfamiliar voice. This type of voice identification is important in forensic casework. Similar to being an eyewitness to, or victim of, a crime who might have to identify a suspect in a line-up, one could be an earwitness who might have to identify a suspect in a voice line-up, or voice parade. One might be asked to identify a perpetrator by his or her voice only, because the person could not be seen for various reasons such as the perpetrator being masked, overhearing the crime scene through an open window, getting a threatening phone call, the crime being committed under conditions of darkness etc. Recognising a voice heard only once could also be useful in a less threatening context, such as remembering the voice of a character in an audio book to help follow the story line (before it becomes familiar). Identifying an unfamiliar voice may be based on paying attention to characteristic features of the voice such as the accent, the pitch and basic criteria like sex and age of the speaker. It may also be based on memorisation techniques such as associating the voice, or aspects of it, with the voice of a friend. Or it may simply be based on the talent and training of one's auditory skills and memory capacity (Hollien et al., 1995).

The two experiments presented in this chapter were conducted to test the ability of synaesthetes and non-synaesthetes to identify unfamiliar voices. For this, stimuli spoken with different voice qualities were used, including whisper, which is known to be particularly difficult for identification purposes (e.g. Yarmey et al., 2001; Evans & Foulkes, 2009). The aim was to find out whether additional synaesthetic perceptions help or hinder voice identification. No hypothesis supporting either of these claims was formulated as information exists that supports both of them. On the one hand, reports of synaesthetes in forums and via personal communication state that some synaesthetes are overwhelmed by too much sensory information, when many or strong synaesthetic perceptions take place. As a consequence, this does often disrupt their concentration and hinder their memorisation and identification skills (see Rehme & Zedler, 2010 for a case report). On the other hand, other personal reports and publications suggest that synaesthetic perceptions can be used as a mnemonic (Luria, 1968; Rothen & Meier, 2010, 2009; Smilek et al., 2002) or that synaesthetes have a better memory performance for other reasons, such as enhanced retention of colours in general or tendency to empathise, which would improve memory by giving more idiosyncratic importance to situations or things to memorise (Yaro & Ward, 2007; van Campen, 2009). More details about these studies are presented in Section 6.1.2 on page 163. First, a short introduction to voice identification is given.

6.1.1 Voice identification

The term speaker identification, a core aspect of forensic phonetics, is subdivided in professional terminology into *voice comparison*, *voice profiling*, and *speaker or voice identification by laypeople* (witnesses or victims) (Jessen, 2008; Watt, 2010). A forensic phonetician is asked to do *voice comparison* when recordings of the perpetrator during the crime and other recording(s) of an unknown person, who is suspected to be the perpetrator, are available. The task of voice comparison is then to estimate how likely it is that the voices in the recordings belong to the same person, i.e. the perpetrator, or different speakers (Jessen, 2008; Rose, 2002; Morrison, 2009). The comparison is usually done by acoustic and/or auditory analysis, and sometimes using automatic speaker identification; statistically, a Bayesian approach is recommended. *Voice profiling* describes a task in which recordings of the perpetrator during the crime are present, but no other recordings of suspects are given. In this case, the forensic phonetician is asked to create a profile of that voice, similar to a facial composite/photofit. The profile would usually include estimation of age, sex, medical conditions that affect speech, native language, regional accent and social class, and might include idiosyncrasies such as unusually fast speech or high pitch (Jessen, 2008). The focus of this section lies on *voice identification by laypeople*. Speaker or voice identification by laypeople is necessary when no recordings of the perpetrator exist, hence only a victim or witness knows what the voice of the perpetrator sounded like, and therefore no professional analysis can be carried out (Jessen, 2008).

Voice identification by a layperson Voice identification by a witness or victim usually involves a *voice line-up*, or *voice parade*. For this, audio recordings of the suspect alongside recordings of so-called foils are made and later played to the witness or victim. A foil is a person who is not involved in the crime but whose sex is the same as that of the suspect and whose voice features are similar in age, social and educational background, nationality, accent, etc. (Broeders & van Amelsvoort, 1999, 2001; Jessen, 2008). The task for the witness is to identify a voice from the line-up as that of the perpetrator, or state that the perpetrator was not present in the line-up. Voice parades are not conducted very often (Hollien, 1996; Broeders & van Amelsvoort, 1999) and may be seen as a "last resort", because the correct identification or rejection of a voice by a layperson varies in its reliability due to many different factors (see below). This might be a reason why earwitness procedures are relatively under-explored compared to eyewitness procedures (Hollien et al., 1995; Broeders et al., 2002). Guidelines on how to conduct a voice parade fairly exist nonetheless (see for example Broeders & van Amelsvoort (1999, 2001); Hollien (1996) and Nolan (2003)), which vary in detail but present a similar global picture. The voice experiment that I set up for my study is most strongly oriented towards the McFarlane guidelines for good conduct of voice parades published in Nolan (2003). The most important aspects are summarised here:

- Recordings should have a duration of about one minute.
- A line-up should ideally include eight foils.
- The witness has to be made aware that the suspect may or may not be present.
- The samples should consist of short unrelated utterances of (semi)spontaneous speech, from which no theme emerges, ideally from an interview. All samples need to be of the same style.
- Foils should be chosen from around 20 samples of people with the same ethnic and regional background, social group, sex and age. Samples should be of similar accent, inflection, pitch, tone and speaking style.
- A mock voice parade should be conducted with volunteers unrelated to the crime and witness as listeners. After having been introduced to the type of crime, they should be asked which person they think is guilty of the offence. If the suspect is identified by a few volunteers, more appropriate foils are needed. This is to allow a fair line-up and to control against any general bias that might occur towards the suspect.
- Witnesses should listen to each sample at least once before making a decision.
- The procedure should be undertaken within 6-8 weeks after the alleged crime because of memory decay.

Influences of stimulus characteristics on identification performance

Voice identification performance is influenced by many different factors. Identification performance may depend on the **speaking style**: correct identification rates are higher when spontaneous speech is used compared to read speech (Hollien et al., 1995). A reason for this could be that spontaneous speech reveals more idiosyncrasies of the speaker, as read speech is more standardized. Performance also depends on the **length of samples**: Longer speech samples result in more correct identifications (Hollien et al., 1995, see also Kerstholt et al., 2004, where this was only found when the line-up was after one week, not immediately after exposure). Studies have compared samples of various lengths, ranging between one syllable (Tartter, 1991), one or more words (Ladefoged & Ladefoged, 1980;

Brungart et al., 2001), one or more sentences (Cook & Wilding, 2001, 1997b; Yarmey, 1991), and recordings of up to 8 minutes (Orchard & Yarmey, 1995; Hammersley & Read, 1985). The general consensus is that, depending on circumstances such as familiarity of the voices or voice quality, longer speech samples usually result in more reliable listener performance.

While there have been various studies on identification performance on foreign **accents** or languages, where either the speakers used a non-native language (e.g. Goldstein et al., 1981; Goggin et al., 1991) or listeners heard recordings in a language foreign to them (e.g. Goldstein et al., 1981; Philippon et al., 2007; Goggin et al., 1991), not many studies have addressed the influence of regional accents within a language. The latter case, however, is of most interest for the experiment presented in this chapter, as all participants were native speakers of English but originated from various different regions and countries. Kerstholt et al. (2006) researched this case for Dutch and found that listeners' performance was slightly better when speakers used a standard accent rather than a strong local accent; but the difference in performance was only marginally significant. Vanags et al. (2005) tested this with Australian English and English English stimuli on Australian listeners; it was found that significantly more correct identifications took place when the listeners' own accent, i.e. Australian English, was spoken.

Relatedly, speaker identification of **familiar voices** is usually more accurate than that of **unfamiliar voices** (Hollien, 1996; Jessen, 2008; Yarmey et al., 2001). Yarmey et al. (2001) found that highly familiar voices of men or women were identified correctly 85% of the time, moderately familiar voices 79% of the time, those of low familiarity 49% of the time and unfamiliar voices 55% of the time in a line-up of four speakers. The authors also found that more speech material (i.e. longer recordings) is needed to perform well in the task for less familiar or unfamiliar voices. Hollien et al. (1995) explain the difference in correct identification performance for familiar and unfamiliar voices with a comparison of the cognitive tasks involved: whereas the identification of familiar voices resembles pattern recognition, the identification of unfamiliar voices resembles feature analysis – so instead of comparing them to a known pattern, the features of the voice sample have to be detected, "mentally described", and only then are they ready for comparison.

Speaker identification becomes even more difficult when **whisper** is used instead of **normally voiced** samples. In the longest speech samples of Yarmey et al. (2001) – 2 minutes of spontaneous speech – correct identification of normally voiced and whispering voices were: 89% vs. 77% for high familiarity, 75% vs. 35% for moderate familiarity, 66% vs. 22% for low familiarity and 61% vs. 20% for unfamiliar voices. False identifications, i.e. when a foil is wrongly identified as the target voice, were also much higher for whispered voices than for normally voiced stimuli. This is because pitch, an important criterion for reliable pattern recognition, is lost when whispering. But at the same time, performance drops drastically as familiarity decreases in the case of whisper. Comparable performance for familiar speakers was found by Evans & Foulkes (2009) with female whispered voices: correct identification was made 71% or 87% of the time, depending on the stimulus length of 4 vs. 16 syllables, from a pool of nine voices. This pool consisted of six familiar female voices, two unknown female and one unknown male voice. Correct identification of the two unknown female voices in this pool was 26% and 33%, respectively.

In the literature, it is often claimed that the accuracy of speaker identification performance does not only vary with familiarity of the speakers, but also that **distinctive voices** are easier to identify (Philippon et al., 2007; Jessen, 2008). Upon exploring this topic further, however, I found that only one publication describes the results of voice parades comparing listeners' performance on line-ups with either unfamiliar distinctive or unfamiliar non-distinctive voices (Orchard & Yarmey, 1995). The distinctiveness of a voice was defined as the extent to which it was "highly striking and not likely to be confused with other voices" (Orchard & Yarmey, 1995, p. 252). Voices for the experiment were rated according to this definition on a 10-point Likert scale by independent judges. Surprisingly, they found contradictory results to what was expected. In their experiment, participants either heard a 30 second or 8 minute recording of the target voice, before a 6-person line-up was conducted. For the short initial target voice recording, correct identification occurred 3 out of 13 times for distinctive voices and 7 out of 13 times for non-distinctive voices. For the long initial target voice recording, correct identification occurred 10 out of 13 times for distinctive voices and 13 out of 13 times for non-distinctive voices. These results clearly contradict the expected outcome. When using whispered voices, however, distinctive voices were identified slightly more often than non-distinctive voices: 7 vs. 5 out of 13 times for the short presentation period and 9 out of 13 vs. 7 out of 14 times for the long duration.

Other publications do not use adequate empirical data to support their claim about the influence of distinctive voices on identification performance: Philippon et al. (2007), for instance, cite the study of Orchard and Yarmey just described, which actually shows the opposite of what has been claimed. The study of Yarmey (1991) is also cited as supportive work on identification of distinctive voices. However, Yarmey (1991) actually investigates how people make *verbal descriptions* of voices and compares the quality of those descriptions over time when using distinctive and non-distinctive voices. Although this study's outcome points in the expected direction, namely that distinctive voices are more reliably correctly described over a retention period, it tests *description* performance, not *identification* performance. Further, Philippon et al. (2007) cite a study by Meissner et al. (2005) to support the claim of better identification performance with distinctive voices. Yet again, the support cannot be upheld because Meissner's study is on *face* perception and does not mention voices at all. As a conclusion from this small survey on the influence of distinctiveness of voices on identification performance, no clear statement can be made regarding identification performance with differently distinctive voices.

Finally, voice identification performance may be influenced by the presence of moving or static images of the speaker. Research suggests that voice identification is negatively influenced by simultaneous presentation of the speaker's face, which is consequently called the face **overshadowing effect**. This term was introduced and defined by Cook & Wilding (1997a, 2001) in analogy to the verbal overshadowing effect. The *verbal* overshadowing effect describes the negative influence on voice identification performance when a verbal description of the voice has been given before the identification process (Vanags et al., 2005; Perfect et al., 2002). The *face* overshadowing effect describes the negative influence of the visual presence of the speakers' face on identification performance. In a series of experiments by Cook & Wilding (2001), participants were exposed to different set-ups of recordings of the target voice for a voice parade: two conditions were auditory only, three were auditory plus visual. In one multisensory condition, participants were exposed to a video recording of the target speaker saying a sentence. In another, they were exposed to a video recording of the target speaker saying three sentences. And in the final condition, the video recording of the speaker's face was already presented 5 seconds before the auditory stimulus onset. The face overshadowing effect took place in the one-sentence condition with a simultaneous onset of auditory and visual display, but disappeared with both longer stimulus presentation (three-sentence condition) and presentation of the face before the onset of the auditory stimuli. The authors concluded that the face overshadowing effect may be due to involuntarily prioritising the visual input of the face over the voice; once the face has been processed, more attention can shift to the voice, whereas earlier, participants would focus on the content of the visual input, at the expense of the sound of the voice.

The impact of the face overshadowing effect had already been researched before it received its name. Melara et al. (1989) found 60% correct identifications in a 6-person target-present line-up in the auditory-only condition in their study, which dropped to 41% when a photograph was presented simultaneously. Auditory stimuli in the voice line-up were four sentences about the speakers' demographics. However, this result was not interpreted as an advantage for presenting stimuli auditorily only, because the rate of false identifications (i.e., of an innocent person) was much higher for the auditory-only condition as well (63% vs. 38%). McAllister et al. (1993) found similar identification performance: 33% correct identification in the auditory-visual and 61% in auditory-only condition, in a 6-person target-present line-up, also using photographs as the visual stimulus and 10-second recordings for the voice line-up. The rate of false identifications was the same in both conditions. Heath & Moore (2011) established that the face overshadowing effect is true for both masked and unmasked faces by using videos with normally visible faces or those wearing a balaclava. Their participants performed better in the auditory only condition but did not show significant differences in the two auditory+visual conditions.

Other influences on identification performance

Not only do the stimuli themselves influence participants' identification performance in a voice parade, but both participants' abilities and the circumstances of presentation also play an important role. Hollien et al. (1995), for example, states that identification performance is better when participants know they need to remember the voice, i.e. when **conscious attention is drawn to the voice**. However, Cook & Wilding (1997b, 2001) found that instruction to attend to the voice did not affect performance in their participants. Divergent results have been found for the influence of **speaker and listener sex** as well. Cook & Wilding (1997b) found better identification performance when listeners were of the same sex as the talkers. In a study by Thompson (1985), male speakers were identified correctly more often than female speakers, independent of participants' sex. However, most studies did not find a gender bias (e.g. Clifford, 1980; Kerstholt et al., 2006; Yarmey et al., 2001). It has also been reported that the **age of listeners** influences performance (cf. Watt, 2010). Some studies by Clifford and colleagues looked at the influence of age in adult listeners. Participants were usually divided into the three age groups 16-20, 21-40 and 41 years and older.[11] On average, the oldest group performed worse than the younger ones (Clifford, 1980; Bull & Clifford, 1984), although it is emphasised that there is overlap in the performance of individuals across age groups. There could be two explanations for the performance decline with increasing age: Both hearing and general memory abilities decline with age. Unfortunately, the authors fail to discuss or interpret the result in this respect.

The level of auditory **training** also influences listeners' abilities. Identification performance is enhanced when people have been trained to identify particular voices, through the process of either familiarisation with the voice(s) or through professional phonetics training in voice identification (e.g. Yarmey et al., 2001; Schiller & Köster, 1998; Brungart et al., 2001). In other words, the identification of an unfamiliar voice by a layperson is the least reliable. However, there can be large individual differences in the performance within the group of laypeople. Simply put, some listeners have better natural ability at identifying voices than others (Hollien et al., 1995, p. 147).

Apart from their training, age and sex, the style of analysing the voice samples may be of relevance. Vanags et al. (2005) tested whether an analytic **cognitive style** was better than a holistic one. Cognitive styles were defined by the response speed in two different geometry tasks. A holistic style was assumed when participants showed a fast reaction time when tasked with stating whether two complex geometrical shapes were identical or

[11] Unfortunately there is some inconsistency in their publications in that they sometimes set the age boundary between middle-aged and old adults at 40 years and sometimes at 50 years of age.

not. An analytic style was assumed when participants showed a fast reaction time when given the task to determine whether a simple shape is contained within a complex figure or not. The authors found that an analytic style was superior in the good sound quality condition, whereas a holistic style was superior for telephone quality speech. It remains unclear, however, whether participants' cognitive style in visual tasks generally transfers to an auditory task. As they were not asked to describe how they (think they) analysed the speech samples, care should be taken when assuming the cognitive style for visual tasks to be the same as that for auditory tasks. Jessen, for example, makes a statement opposing the findings of Vanags and colleagues: "Since lay listeners are not trained in the systematic analysis of language and speech, their voice perception is essentially holistic." (Jessen, 2008, p. 695 et seq.)

In summary, voice identification performance by laypeople may depend on many factors: the speaking style, the length of exposure to the voice, the language and accent used, familiarity with the speaker, the voice quality and other simultaneous stimulation such as seeing the face of the speaker. But it may also depend on the age of the listener, whether their attention is drawn to the voice, whether they received appropriate training and which cognitive style they use. In this chapter, the reader will find out whether synaesthesia is another factor to be added to this list.

6.1.2 Enhanced memory performance in synaesthetes?

Single case studies of people who became famous for their **outstanding memory skills**, like Shereshevsky (Luria, 1968) and Daniel Tammet (Baron-Cohen et al., 2007; Bor et al., 2008), or studies of other synaesthetes (Smilek et al., 2002; Mills et al., 2006), have publicised extraordinary memory skills of individuals with synaesthesia. In all four cases, the participants stated that their use of synaesthetic perceptions was as a mnemonic. While the memory performance of the synaesthetes described in these studies is significantly above average ability, the fact that Daniel Tammet can remember 22,514 decimal places of π is beyond the imagination of most. Although it is known that mnemonic strategies like the method of loci[12] aid the memorising of items – and synaesthetic colour associations could probably be classified as one such associative method – it still remains a miracle that this method can work for more than 22,000 items. Tammet is not only a synaesthete, but also a savant with Asperger syndrome. The combination of these conditions seems to enhance his memorising skills: as a person with Asperger syndrome he is likely to focus on local detail; as a savant he is likely to have a prodigious memory in a narrow specific field; and as a synaesthete he can further use his synaesthetic perceptions as a memory

[12]The method of loci is a mnemonic device. Items can be remembered because they are mentally associated with specific physical locations, for example body parts from top to bottom.

aid. Or as Bor and colleagues state: "The propensity to focus on local detail, in concert with a form of synaesthesia that provides structure to all digits, may account for DT's exceptional numerical memory and calculation ability." (Bor et al., 2008, p. 311)

It is also noteworthy that, although these synaesthetes have increased memory skills in certain areas, it seems usually **limited to areas involving their synaesthesia**: for Tammet, the special area is memorising numbers, for Shereshevsky it is memorising numbers and letters, for a participant identified as C in Smilek et al. (2002) it is numbers, and similarly for synaesthete MLS in Mills et al. (2006) it is words, names and numbers. The fact that their performance in memory tasks outside their specialised areas is often not above average was tested in most cases. In the case of Shereshevsky, it was reported on informally. MLS had superior memory performance in verbal tests but only average performance in visual memory tests (see Mills et al., 2006, for details). While C had enhanced memory for digits, her performance was not significantly better than that of controls when shapes were used (see Smilek et al., 2002, for details). While Tammet showed superior memory performance with numbers, he had average scores for a general short-term memory test and showed poor performance in a facial memory task (see Baron-Cohen et al., 2007, for details). His synaesthesia helped Shereshevsky remember lists of words, letters or numbers but hindered him in remembering faces, as his synaesthetic experiences altered with every change in facial expression.

After publication of these cases of astounding memory performance, the idea that synaesthesia might be used more widely as a mnemonic became popular. Although the individual cases seemed very convincing, the criticism of a selectional bias towards particularly skilled individuals could not be ignored (see for example Rothen & Meier, 2009). It is easy to draw attention to something phenomenal – but a few single case studies do not allow for generalisation of the advantage of synaesthesia. Hence, **group studies** have been conducted as well to test this phenomenon on a wider scale. Yaro & Ward (2007) found that their synaesthetic participants reported having better-than-average memory and could indeed prove this for word-lists and colours. However, synaesthetes did not perform better at a memory task using abstract figures. These findings suggest a subjective superiority in memory performance in general and an objective superiority in memory performance involving those modalities that are synaesthetic inducers or concurrents. Similar findings were made by Rothen & Meier (2009), who discovered a moderate advantage in grapheme-colour synaesthetes compared to controls in visual search task performance (to find a set of a certain grapheme in a cloud of another grapheme), but no advantage in episodic memory. The same researchers found slightly contradictory results in a follow-up study (Rothen & Meier, 2010), where synaesthetes performed better in episodic memory tests using the Wechsler Memory Scale, but had no advantage in short-term memory tests. Still, the performance of synaesthetes in episodic memory was within the normal range, i.e. significantly but only moderately better than that of controls. Hence, Rothen and Meier "downgrade" the finding of extraordinary memory skills of synaesthetes to be rather ordinary, despite their richer world of perception and the possibility of using their synaesthesia as mnemonics. In line with their conclusion, another synaesthete and researcher made a similar statement based on her own experiences: "It certainly makes life more interesting and is mildly useful for such things as remembering telephone numbers and people's names." (Kay & Mulvenna, 2006, p. 204).

On the other hand, some reports suggest a **negative influence of synaesthesia** on people's performance. It has been mentioned already that Shereshevsky had difficulties remembering faces, because his synaesthetic associations altered whenever the facial expression changed. In addition, he had voice-induced synaesthesia. Luria writes:

> 'What a crumbly, yellow voice you have', he once told L.S. Vygotsky while conversing with him. At a later date he elaborated on the subject of voices as follows:
> 'You know there are people who seem to have many voices, whose voices seem to be an entire composition, a bouquet. The late S. M. Eisenstein had just such a voice: listening to him, it was as though a flame with fibers protruding from it was advancing right toward me. I got so interested in his voice, I couldn't follow what he was saying...
> But there are people whose voices change constantly. I frequently have trouble recognizing someone's voice over the phone, and it isn't merely because of a bad connection. It's because the person happens to be someone whose voice changes twenty to thirty times in the course of the day. Other people don't notice this, but I do.' (Record of November 1951.)
> 'To this day I can't escape from seeing colors when I hear sounds. What first strikes me is the colour of someone's voice. Then it fades off... for it does interfere. If, say, a person says something, I see the word; but should another person's voice break in, blurs appear. These creep into the syllables of the words and I can't make out what is being said. (Record of June 1953.) (Luria, 1968, p. 24-25)

These descriptions suggest a sensory overload that Shereshevsky experienced due to his multiple and strong synaesthetic reactions. Another synaesthete reported that she had to leave the lecture hall while a new professor at her university spoke because the synaesthetic perceptions induced by his voice were too unpleasant to bear (personal communication). Rehme & Zedler (2010) tested a synaesthete's performance at solving Sudoku puzzles under time pressure. In addition to a flooding of colours, the participant even reported symptoms of vertigo, dysphoria and high irritability. Although other case studies have mentioned negative effects of synaesthesia in email lists and forums, there has only been one group study about these effects. Gimmestad (2011) analysed the negative effect that synaesthesia can have due to sensory overload on grapheme-colour synaesthetes. She did qualitative research on sensory sensitivity and personality, and quantitative research on the startle eye reflex with congruent and incongruent stimuli on grapheme-colour synaesthetes. Although results for synaesthetes were not significantly worse than those for non-synaesthetes when testing their startle reflex on congruently or incongruently coloured graphemes, tendencies of slower reaction times and reports of discomfort, sensory overload and sensation avoiding were found in synaesthetes.

In summary, evidence has been found both to support the claim that synaesthesia can be used as a memory aid, as well as to support the claim that synaesthesia can disturb memory functions as it distracts from focusing on the items to memorise. The latter could be aligned with the face overshadowing effect, and be given the name *synaesthetic concurrent overshadowing effect*: more attention is put on the synaesthetic concurrent(s) than the inducing stimulus. Hence, voice identification performance might be negatively influenced.

Because no clear picture emerges as to the positive and negative effects synaesthesia has on memory, no hypothesis is formulated here. If synaesthesia does have a positive effect on memorising voices, I do hypothesise, however, that participants whose synaesthesia is induced by the sound of voices perform better than participants whose synaesthesia is induced by other modalities, or those who do not have synaesthesia. Furthermore, all studies related to synaesthesia and memory have been conducted with letters, numbers, words or geometric figures as stimuli and grapheme-colour synaesthetes as participants. The two studies described below have been conducted to address the following issues: to test whether a memory advantage can be found in voice-induced synaesthetes, and to test whether results can be generalised and/or transferred to other modalities which are not active in the person's synaesthesia. Here, it was tested whether synaesthetes were also able to use their synaesthetic perceptions to remember voices when they do not have voice-induced synaesthesia.

6.2 ABX voice comparison task

The short voice comparison task described in this section was conducted along with the voice description experiment described in the previous chapter. Its purpose was to act as a pilot study for a fuller examination of voice identification skills by synaesthetes. It had an ABX design, which describes the following setup: participants are confronted with stimulus A, stimulus B, and finally stimulus X, and are asked to judge whether X is more similar to A or B. In the case of the ABX voice comparison task described here, participants listened to voice A, voice B and voice X and had to decide whether X was the same voice as in recording A or B. The same stimuli as in the voice description experiment were used, i.e. two speakers producing ten different voice qualities each.

6.2.1 Methods

Stimuli The same recordings as those for the voice description experiment outlined in Section 5.2.2 were used. The short sentence was: "People look, but no-one ever finds it." The long sentence was: "These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon." (See Section 5.2.2 for more details.)

The intensity of the recordings was equalised to 65 dB_{SPL} per ABX triad using Audacity. Audio samples of the triads are included on the supplementary CD under "Ch6_Voice_identification\ABX".

Participants There were three participant groups for this experiment: 14 synaesthetes, 10 phoneticians and 28 controls. For more details about the participants, see Section 5.2.5.

Procedure As previously described, the setup of the ABX voice comparison was as follows: participants listened to voice A, voice B and voice X and had to decide whether X was the same voice as in recording A or B. They were asked to choose whether the voice of the last sentence was identical to the voice in the (a) first sentence, or (b) second sentence. Ten sets of voices were used in a fixed order, starting with two easy comparisons so that participants could become acquainted with the task. The recordings used are listed in Table 22 in the order presented during the experiment. A and B were either the same speaker using different voice qualities, or different speakers using the same voice quality. X was always the same speaker using the same voice quality as either stimulus A or B, but reading a different sentence than was used in A and B.

There was approximately 1 second of silence between the recordings A, B and X. 4 seconds of silence followed before the next triad of stimuli automatically started to play. Thus, participants had 4 seconds plus the time while X was playing to make their decision.

6.2.2 Results

Performance of the three different participant groups in correctly identifying the voices is summarised in Table 23. Because group sizes were different, the results are given in percent. The column 'overall (incl. whisper)' shows the results of all ten ABX comparisons, the column 'voiced (excl. whisper)' lists the results of the nine ABX comparisons that comprised voiced stimuli, and the column 'whisper' displays results of the comparison between whispering voices only. The separation of voiced and voiceless stimuli was made because noticeable differences in correct performance were found for these two different phonation types across and within groups. Every participant group performed best in

A			B			X			Correct
VQ	Speaker	Sentence	VQ	Speaker	Sentence	VQ	Speaker	Sentence	
nasal	Laver	short	nasal	Nolan	short	nasal	Laver	long	A
modal	Nolan	long	harsh	Nolan	long	modal	Nolan	short	A
breathy	Nolan	short	modal	Nolan	short	breathy	Nolan	long	A
creak	Nolan	long	creak	Laver	long	creak	Nolan	short	A
falsetto	Nolan	short	falsetto	Laver	short	falsetto	Laver	long	B
lowered lx	Laver	long	creak	Laver	long	creak	Laver	short	B
lowered lx	Laver	long	modal	Laver	long	lowered lx	Laver	short	A
modal	Laver	long	denasal	Laver	long	denasal	Laver	short	B
raised lx	Laver	short	denasal	Laver	short	raised lx	Laver	long	A
whisper	Laver	long	whisper	Nolan	long	whisper	Laver	short	A

Table 22: Stimuli for the ABX voice comparison task. VQ = voice quality, lx = larynx.

	Overall (incl. whisper)	Voiced (excl. whisper)	Whisper
synaesthetes	74.3%	75.4%	64.3%
phoneticians	84.0%	86.7%	60.0%
controls	88.2%	93.7%	39.3%

Table 23: Participants' percentage correct identification performance in the ABX voice comparison task.

		Overall	Voiced Samples	Whispered Sample	Difference voiced−whisper
Mean Rank	Synaesthetes	20.07	18.86	30.21	19.04
	Phoneticians	28.00	27.30	29.10	24.25
	Controls	29.18	30.04	23.71	31.04
χ^2		3.70	6.24	2.77	6.46
p		0.158	0.041	0.258	0.037

Table 24: Kruskall-Wallis tests for measuring group differences in voice comparison performance for all as well as voiced and whispered stimuli separately.

the category 'voiced' and worst in the category 'whisper'. The comparison across groups within categories shows that synaesthetes are the best-performing group in 'whisper' with 64.3% correct responses, and controls are the best-performing group in 'voiced' with 93.7% correct and 'overall' with 88.2% correct responses.

Kruskall-Wallis tests were performed to test for significant differences of the performance between participant groups. This test was chosen over an ANOVA – a standard test to compare group differences – because data was neither normally distributed (Shapiro-Wilk for synaesthetes W=0.68, p=0.031, phoneticians W=0.74, p=0.002, controls W=0.67, p<0.001) nor homogenous (Levene statistic W=13.31, p<0.001). Thus, a non-parametric test had to be used. Statistical results are given in Table 24. Groups with a higher mean rank identified more voices correctly than other participant groups. The χ^2 value measures whether there is a difference in performance between any of the three groups. Tests were ran separately for all samples, voiced samples, whispered samples and the difference of voiced−whispered. The difference measure was introduced to test whether group performance differed not only generally but also in relation to different VQs, i.e. whether interactions between groups and VQs exist. As significant values were found for the voiced samples and the difference of voiced−whispered samples, additional Mann-Whitney U tests were conducted to find out which groups differ significantly in their voice identification performance. To avoid a misleading effect of multiple comparisons, the Bonferroni correction was applied to interpret the level of significance: As three group comparisons were made, the level of significance (p<0.05) was devided by 3, resulting in a new level of p<0.017. Synaesthetes' performance differed significantly from that of controls in voiced samples (U=112, z=-2.46, p=0.013) and the difference of voiced−whispered samples (U=106, z=-2.46, p=0.013). Controls had the most correct responses in the voiced samples, whereas synaesthetes had the lowest percentage of correct responses. The group scoring highest in the whisper condition were the synaesthetes, while the group of controls performed poorest here. This suggests an interaction between voice qualities and group for these two participant groups. There are no significant differences between groups in the 'whisper' condition. However, synaesthetes performed 25% better in this task than controls, as shown in Table 23. The reason why this result does not reach significance here is most likely the amount of data per condition: while 520 replies were included in the 'overall' condition, only 52 replies were included in the 'whisper' condition.

This pilot study revealed group differences in correct voice comparison performance between voice-induced synaesthetes and non-synaesthetes. Surprisingly, group differences were not consistent across stimuli: while controls outperformed synaesthetes when voiced stimuli were used, synaesthetes outperformed controls when whispered stimuli were used (although not significantly so). This phenomenon can be referred to as *double dissociation* (e.g. Ellis & Young, 1997): while group A performs well in task one but badly in task two, group B performs badly in task one but well in task two, indicating that not only does the difficulty of the task have an impact on performance, but also the condition synaesthesia. Despite the lack of significance in the whisper condition, it was felt that a performance difference of 25% justified a more elaborate follow-up study to further test this intriguing group difference, that reverses performance between groups depending on whether speech is voiced or not. Hence, a voice identification task with modal and whispered stimuli was conducted, which is described in the following section.

6.3 Voice parade

In this experiment, speaker identification performance among voice-induced synaesthetes, other synaesthetes (i.e. synaesthetes whose perceptions are *not* induced by the sound of voices) and non-synaesthetes was investigated. Two groups of synaesthetes were chosen to test whether any synaesthesia-related memory differences occur only when voices are synaesthetic inducers, or also for inducers unrelated to voice. For this, two voice parades following the setup introduced in Section 6.1.1 were used: one parade used modal voice quality in both the presentation of the target voice and in the voice line-up, whereas the other parade used whispered speech for both. Two voice parades were conducted because the pilot study described in the previous section suggested opposite performances of synaesthetes and controls when voiced vs. whispered speech was used. The aim of these two parades is to find out whether synaesthetes perform better or worse than non-synaesthetes when asked to identify a previously-heard unfamiliar voice. Previous studies have shown that synaesthetes have a memory advantage, especially in those areas that involve their synaesthesia (Rothen & Meier, 2009; Rothen et al., 2012; Yaro & Ward, 2007). Taking into account these findings on memory advantage among synaesthetes, while keeping in mind possible adverse influences such as sensory overload, this is an attempt to extend and refine our knowledge of how memory for voices among earwitnesses works. In addition to the two voice parades some qualitative data was collected to help interpret the results.

6.3.1 Methods

Experimental design

The design of the voice parades followed the McFarlane Guidelines (cf. Nolan, 2003) as closely as possible, with the exception that read rather than spontaneous speech was used. This decision was taken because many voice-induced synaesthetes have multiple types of synaesthesia – for example one that is induced by spoken words – and I wanted to avoid any confounding influence of speakers using different words.

Two different parades were conducted: one with modal voice, i.e an ordinary, neutral mode of phonation (see Laver, 1980, for an extensive description), the other with whisper, where no voicing is present throughout the samples. The two parades each consisted ultimately of seven speakers (target voice + six foils), six of whom were present in both parades.

Participants were allowed to choose to participate in one or both parades. This compromise was necessary for two reasons: first, to avoid participants dropping out (or not even taking part at all) because of the experiment's lengthy design, and second, to collect as much data as possible from synaesthetes who are willing and available to do the tests.

Stimuli

Recordings from the CHAINS corpus (Cummins et al., 2006) were used as, unlike most corpora, it includes whispered speech. All speakers in the corpus who were used for the stimuli in this experiment were men from the Dublin area in their twenties and early thirties. Recordings of two short stories were used. Disfluencies while reading the short stories were edited out by the creators of the corpus. All recordings were made in mono with a bit rate of 705 kbps, 16 bit sample size and a sampling rate of 44 kHz. Whispered speech was normalised to an intensity of 60 dB and the modal speech to 70 dB using Praat (Boersma & Weenink, 2012). When taking measurements of the stimuli, however, I found that the modal speech recordings varied around 70 dB with a range of approximately 8 dB.

Stimuli were selected from the corpus as follows. For the target voice samples (both modal phonation and whisper), a part of the story 'Cinderella' (see Appendix H) read by two different speakers was used (each approximately 50 seconds in duration). For the two parades, eight speakers (target voice + seven foils) were initially chosen, reading 'The North Wind and the Sun' (each approximately 30 seconds in duration, see Appendix I).

Since, according to McFarlane's guidelines, a voice parade needs to undergo a mock test with "mock witnesses" unrelated to the crime in order to investigate whether certain voices are more likely to be picked than others, the original selection of eight voice recordings was presented to three phoneticians. The recordings of the voice parade, i.e. 'The North Wind and the Sun' were used. The phoneticians were asked to listen to the samples imagining they were witnesses in a voice parade procedure and to comment generally on whether one or more of the voices stood out from the rest and should be excluded from the parade. Later, they were asked to make pairwise comparisons of the speakers. For this, they were asked to give a general similarity rating from 1-7, where 1 is very similar and 7 very different, with the option to specify or comment on their ratings. All three judges suggested excluding one particular speaker from the modal line-up, whose voice quality was too different from the others'. Two of the phoneticians then judged the whispered stimuli as well, and both suggested excluding one particular speaker from the whisper line-up for similar reasons. Because no suitable alternatives could be found, both parades were reduced by one speaker, resulting in seven speakers per parade. Audio samples of the two parades including target recordings can be found on the supplementary CD under "Ch6_Voice_identification\Voice_parade".

Speaker characteristics

Speaking rate and other acoustic measurements were taken from the initial target recording and all recordings in the line-up, for both the modal and whisper condition, to analyse influences of speaker differences on the identification performance. The speaking rate was measured by dividing the total time by the number of syllables. Further, the LTF values of the first four formants were taken for all recordings. For this, sound files were downsampled to 11025 Hz with a precision of 200 samples for the interpolation method in Praat (Boersma & Weenink, 2012). See Section 5.3.2 on page 124 for further methodology. For the modal voice recordings, additional measures of f0 were taken: mean f0 in Hz and pitch range in semitones. Data of these measurements per speaker are given in Table 25 for modal and in Table 26 for whisper. Although six out of the seven speakers were present in both parades, their numbers in the line-ups (1-7) do not necessarily correspond between the modal and the whisper parade.

Participants

Participants were 12 voice-induced synaesthetes, 20 other synaesthetes, and 34 controls. They were paid for their participation. Group sizes differed owing to the availability of the individuals. Whereas 33 participants took part in both voice parades, 12 participants took part in the modal voice parade only and 22 took part in whisper only. All participants were native speakers of English; all but 2 were monolingual. Information about participants' age and gender is in Table 27.

	Target	1	2	3	4	5*	6	7
mean f0 (Hz)	122	102	103	114	107	120	112	118
pitch range (semitones)	16.9	14.6	13.4	11.6	17.8	14.6	14.7	14.7
LTF1 (Hz)	448	449	472	455	429	458	410	461
LTF2 (Hz)	1410	1406	1390	1355	1340	1361	1346	1462
LTF3 (Hz)	2415	2429	2486	2367	2324	2406	2343	2444
LTF4 (Hz)	3402	3314	3688	3406	3483	3377	3364	3576
speaking rate (syll/sec)	4.6	4.7	5.1	5.1	5.1	5.2	5.1	4.5

Table 25: Measures of speaker characteristics for the modal voices for the initial target recordings and the seven voice parade recordings. Speaker 5* is the target speaker.

	Target	1	2	3	4	5	6*	7
LTF1 (Hz)	705	646	643	673	657	734	711	547
LTF2 (Hz)	1583	1430	1600	1611	1501	1557	1570	1430
LTF3 (Hz)	2530	2270	2558	2565	2478	2591	2487	2385
LTF4 (Hz)	3539	3337	3617	3632	3450	3418	3369	3368
speaking rate (syll/sec)	4.9	5.3	4.8	4.8	4.4	4.6	5.0	5.5

Table 26: Measures of speaker characteristics for the whispering voices for the initial target recordings and the seven voice parade recordings. Speaker 6* is the target speaker.

47% of the participants gave information about their educational background. Of these, all but 2 had at least a bachelor degree or were studying at university at the time they took part in the experiment. Information about occupation was given by 46% of the participants. As it has been found that synaesthetes work in the fields of arts or music significantly more often than non-synaesthetes (Rich et al., 2005), it might be noteworthy that the only 2 people working in music were voice-induced synaesthetes. 1 voice-induced synaesthete works in the field of arts, 2 elsewhere and for the rest no data is available. Of the other synaesthetes, 4 work in the field of arts, 2 elsewhere and for the rest it is unknown. None of the controls reported working in the field of arts or music.

Voice-induced synaesthetes Nationalities are as follows: 6 English, 2 Scottish, 2 US American, 1 Canadian, 1 Ugandan. They all reported additional types of synaesthesia. These are, in order of frequency: general sound - colour, general sound - texture, letter-colour, music-colour, music-shape, number lines, number-colour, smell-colour, days-colour, taste-colour, pain-colour, names-colour, months-colour, ordinal linguistic personification, words-taste/smell, words-texture, temperature-visual, all types of sensation-visual, pain-taste, pain-shape, words-shape, taste-shape, smell-shape, pain-texture, taste-texture and "ticker tape". 6 out of the 7 people responding to the question claimed they were musical.

Other synaesthetes Nationalities are as follows: 8 US American, 5 Scottish, 4 English, 1 Australian, 1 Canadian, 1 Northern Irish. Synaesthetes in this group had

		Voice syns	Other syns	Controls
total	n	12	20	34
	female	11	20	28
	male	1	0	6
	mean age	40	37	29
	age SD	17.5	14.9	13.4
modal	n	9	13	22
	female	8	13	18
	male	1	0	4
	mean age	36	38	29
	age SD	16.1	13.7	14.8
whisper	n	12	19	25
	female	11	19	20
	male	1	0	5
	mean age	40	36	30
	age SD	17.5	15.3	13.0

Table 27: Number of participants per group, their average age, and standard deviation (SD) of their age. The top columns describe the demographics of all participants, the middle columns those of the participants taking part in the modal voice parade, and the bottom columns of those taking part in the whisper voice parade.

the following types of synaesthesia, in order of frequency: letter-colour, number-colour, general sound - colour, words-colour, number lines, days-colour, months-colour, smell-colour, words-texture, taste-colour, words-taste/smell, sound-texture, sound-taste/smell, ordinal linguistic personification, music-colour, pain-colour, names-colour, sound-pain, mirror touch, people-colour. 6 out of the 8 people responding to the question claimed they were musical.

Controls Nationalities are as follows: 13 Scottish, 11 English, 3 Northern Irish, 3 US American, 1 Australian, 1 French (bilingual), 1 Irish, 1 Welsh. 9 out of the 14 people responding to the question claimed they were musical.

Procedure

The experiment was conducted online to reach as many synaesthetes as possible. The two voice parades could be carried out independently. If both were done, the participant could choose the order, and could do them in one session or in separate sessions. The procedure for each was the same, as follows.

Before being presented with instructions, participants were asked about sex, age and nationality, discriminating between English, Scottish, Northern Irish and Welsh as well, to account for differences in accents. After a test sound, they were presented with the initial target voice recording, followed by two filler tasks: a facial expression reading task and a "risk behaviour" test. These filler tasks were chosen because no additional voices would be heard, i.e. no additional auditory linguistic stimulation was used. The first distractor task, the facial expression reading task, used the visual stimuli of the *Cambridge Mindreading Face-Voice Battery* (Golan et al., 2006). Short silent video clips (a few seconds in length) of individuals expressing a mental state or emotion were played and participants were asked to choose one of four given words to describe which emotion or state had been presented. The risk behaviour task was designed by Marie-Hélène Grosbras and her group (unpublished). It comprised a computer game in which participants had to inflate 30 balloons to an arbitrary size with the aim of collecting as many points as possible. Following the filler tasks, the voice parade was played. Participants were asked to listen to the stimuli in the given order 1-7. They were informed that the target speaker might or might not be present in the voice parade, although in both parades the speaker was indeed present. After having listened to all speakers once, participants could replay individual speakers an unlimited number of times, and finally respond by pressing the appropriate number of the voice recording or the "not present" button. At the end, some more qualitative data was collected. The following mandatory questions were asked:

1. Are you a synaesthete? If so, can you please list all kinds of synaesthetic associations you have, e.g. environmental sound to colour, spoken words to texture, written letters to colour etc.

2. If you are not from the UK, please specify which country or state you are from.

3. Do/Did you have regular contact with Irish people?

4. What kind of headphones or speakers did you use? (E.g. in-ear, closed cup, make, model)

6.3.2 Results

Voice identification performance The overall percentage of correct identifications in the voice parade using modal voice quality was 36.2%, and using whispered samples 25%, as shown in Table 28. The chance level was 12.5%, as eight responses were possible: choosing one of the seven voices, or indicating that the target voice was not present in the line-up. Binomial tests were used to test whether participants identified the correct voice significantly above chance. The p value in Table 28 indicates whether the groups' performance in correct identifications differed significantly from chance. These cases (i.e. if $p<0.05$) are highlighted in bold.

In the modal voice condition, the group of non-synaesthetes performed significantly above chance. So too did participants overall, when group affiliation is ignored. In the whisper condition, voice synaesthetes and all participants in total differed significantly from chance. The column '% correct' shows that the best performance overall was achieved

Condition	Group	Correct		Total	p
		%	n	n	
modal	voice syn	33.3	3	9	0.092
	other syn	30.8	4	13	0.069
	non-syn	**39.1**	**9**	23	**0.001**
	all	**36.2**	**17**	47	<**0.001**
whisper	voice syn	**41.7**	**5**	12	**0.011**
	other syn	21.1	4	19	0.207
	non-syn	20.0	5	25	0.195
	all	**25.0**	**14**	56	**0.008**

Table 28: Correct identifications per group per voice condition in absolute numbers and percentages. p describes whether correct identification is different from chance using binomial tests with a significance level of $p < 0.05$; significant differences indicated in bold. The n of "all" can exceed the sum of n's of all groups because group membership of some participants could not be established.

Predictor	B	SE	Wald	df	p	95% CI for Odds ratio		
						Lower	Odds ratio	Upper
constant	-1.39	0.50	7.69	1	.006		0.25	
groups			2.17	2	.338			
voice synaesthetes	1.05	0.77	1.86	1	.173	0.63	2.86	12.92
other synaesthetes	0.07	0.75	0.01	1	.932	0.24	1.07	4.66
voice quality	0.94	0.66	2.06	1	.151	0.71	2.57	9.33
VQ*groups			1.33	2	.515			
VQ by voice synaesthetes	-1.30	1.13	1.33	1	.249	0.03	0.27	2.49
VQ by other synaesthetes	-0.43	1.05	0.17	1	.681	0.08	0.65	5.11

Note: R^2=0.00 (Hosmer & Lemeshow), 0.036 (Cox & Snell), 0.051 (Nagelkerke). Model $\chi^2(5)$=3.71, p=0.592

Table 29: Logistic regression showing the influence of participant group, VQ of the stimuli and their interaction on voice identification performance.

by voice synaesthetes identifying the target voice in whisper. This result is of striking importance, because identification performance usually drops severely with whispered speech (Evans & Foulkes, 2009; Orchard & Yarmey, 1995; Yarmey et al., 2001); this drop in performance shows for non-synaesthetes (39.1% correct in modal and 20% in whisper).

Voice-induced synaesthetes only identified whispered voices significantly better than chance, and non-synaesthetes only identified modal voices significantly better than chance. However, groups did not differ significantly from each other, as logistic regression revealed. Results of this test are presented in Table 29. The beta coefficient B reports the individual contribution of the predictors to the model, SE its standard error. The basic model only contains a constant, so predictors are added to the model. The predictor "groups" specifies that there are three different groups of participants who may behave differently.

Non-synaesthetes are defined as the reference category of participant groups, so voice synaesthetes and other synaesthetes are added separately. Additionally, the two different voice qualities and interactions between group and VQ are added to the model as predictors which may influence the outcome. The interactions are added because results from the ABX voice comparison task suggest that synaesthetes' performance is better for whispered stimuli and controls' performance is better for modal voice quality. The Wald statistic shows whether the beta coefficient is significantly different from zero for all the predictors. Neither participant groups nor VQ, nor in fact their interaction, give a significant contribution to the prediction of the outcome. The odds ratio supports this finding: a value above 1 indicates that there are more correct identifications for a predictor, and a value below 1 indicates that there are fewer correct identifications. If its confidence interval crosses 1, non-significance can be assumed. The finding is further confirmed by the non-significant χ^2 value and the small power indicated by the low R^2 values described in the note of Table 29.

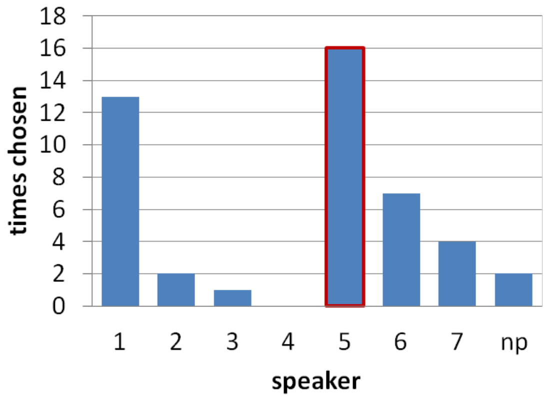
Thus, there is no significant difference between the performance of voice synaesthetes, other synaesthetes and non-synaesthetes in correct voice identification in modal or whispered speech. This lack of significant group differences – despite the large differences in the "% correct" results between voice-induced synaesthetes and non-synaesthetes in the whisper condition shown in Table 28– is most likely due to the low number of participants.

Influences of covariates A lot of factors other than group affiliation could possibly have influenced participants' performance. While most of these factors involve the listeners, e.g. accent and other demographics, both speaker characteristics and the experimental set up also play a role. These covariates were examined because they are of interest for finding influences on speaker recognition with respect to forensic phonetics, as introduced in Section 6.1.1.

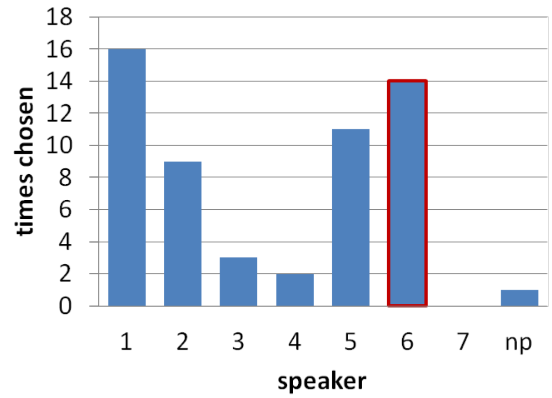
Speaker characteristics: Voice confusability and voice similarity

A crucial factor affecting identification performance in a voice parade is the similarity of the voices used. Even after controlling for speaker differences, such as age, accent and social background, speakers may still sound more or less different from each other. It is probably impossible to create a line-up of voices, which all sound equally different from, or similar to, each other. One would expect similar sounding voices to be more frequently confused with one another than non-similar sounding ones. However, other factors such as order of presentation can also influence patterns of confusability among voices.

Figure 39 displays how often each voice in the voice parades has been (correctly or incorrectly) identified to be the target voice. The target speaker in the modal voice parade is speaker 5 and in the whisper voice parade speaker 6. An eighth option ('np') means that target voice "not present" was chosen. In the modal voice parade, it seems

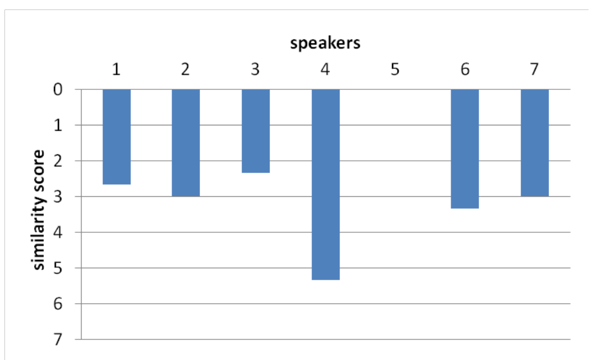


(a) Modal.

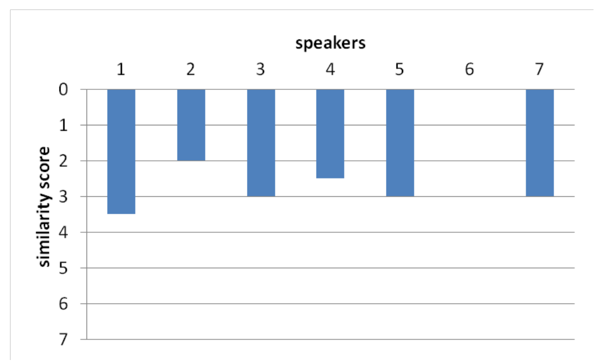


(b) Whisper.

Figure 39: Voice parade responses for all subjects. Numbers on the x-axis refer to speaker numbers in the voice parade, np = not present. The target voices are highlighted in red.



(a) Modal.



(b) Whisper.

Figure 40: Similarity ratings of foils and the target voices from 1 (very similar) to 7 (very different). Modal ratings based on three independent professional judges, whisper on two.

that speaker 1, who has been chosen the second most frequently, is very similar to target speaker 5, whereas speakers 3 and 4 seem very different from the target speaker. In the whisper voice parade, speaker 1 has been wrongly identified most often as the target speaker, whereas speaker 4 and 7 seem very different from target speaker 6 and speaker 1. Speaker similarities and differences are analysed in the following paragraphs.

As described in Section 6.3.1, trained forensic phoneticians rated the similarity of the speakers on a scale from 1 (very similar) to 7 (very different) before they were chosen for the voice parades. The average similarity ratings between each foil and the target voice are given in Figure 40. Comparing the results of the voice parade responses in Figure 39 and similarity ratings in Figure 40 shows that high similarity ratings do not necessarily imply misidentification and vice versa. The voice rated as most similar to the target voice in the modal condition is voice 3, which was rarely chosen as the target voice, whereas voice 1, second most similar to the target voice according to the professionals' ratings, was most often erroneously identified as the target voice. The voice rated as most different, voice 4, was not chosen by the participants at all. In general, the similarity ratings mostly support the idea that voices rated as more similar to the target are more often misidentified than voices rated as more different from the target for modal speech. However, similarity ratings can only be used as a rule of thumb.

This relationship between similarity ratings and (mis)identifications cannot be found for whisper. The similarity ratings for whisper only range from 2-3.5, indicating that all foils sounded more similar to the target voice than different from it. The voice which was rated as most different from the target voice is speaker 1, who was in fact chosen more often than target speaker 6. Hence, additional measurements were taken to search for an influence of speaker differences on participant responses. LTF values of the initial target recording and the seven line-up recordings are given in Table 25 on page 173 for modal and Table 26 on page 173 for whisper. Although different texts were read in the two different recordings, a comparison of the data is valid, because LTF values stabilise after about 6 seconds of pure vocalic stream (Moos, 2010). For the modal voices, the values of speaker 1 are very similar to those of the initial target recording. This could be one reason why participants, especially synaesthetes, often misidentified speaker 1 as the target. For whisper, the LTF values of the target speaker himself match his initial recording best. Speaker 1, who was incorrectly chosen as the target by participants more often than the correct target, on the other hand, has much lower LTF values than the target speaker.

Additionally, measures relating to the fundamental frequency were taken for the modal voice condition (cf. back to Table 25 on page 173). The target speaker had the highest mean f0 of all speakers, who all were in the normal range of fundamental frequencies for men (Jessen et al., 2005; Hudson et al., 2007). The pitch range of the target speaker differs between the initial and the line-up recording. The pitch ranges of the speakers seem to be lower than average, when comparing with other studies of English using read speech (cf. Traunmüller & Eriksson, 1995).

The average amount of syllables per second has been used as a measure of speaking rate. In both voice quality conditions, they varied between 4.5 and 5.5 syllables per second, as shown in Tables 25 and 26 – again a normal range for English speakers (Hewlett & Rendall, 1998; Dauer, 1983; Trouvain, 2003). Also this feature does not explain why some foils were misidentified as the target more often than others because there does not seem to be a systematic relation between speaking rate and participants' responses.

The high proportion of false identifications of speaker 1 in both line-ups might be due to the primacy effect. Serial position effects were first researched by Ebbinghaus (see translated republication of Ebbinghaus, 2004). It is assumed that directly after hearing a list, a recency effect takes place: those items mentioned last are still in the working memory and recall is therefore facilitated. When recall is delayed (as in the current experiment

Synaesthetes	Synaesthetic perceptions with voices		
	All	Some	None
voice-induced	4	2	1
other	3	0	4

Table 30: Number of synaesthetes who had synaesthetic perceptions with all, some or no voice stimuli in the modal and whisper parades.

owing to the use of filler tasks), the primacy effect is stronger. In this case, the items which are mentioned first are stored in the long-term memory, and are therefore easier to recall. Moreover, they are also easier to recall as there are no other items to remember simultaneously. Participants had to compare each voice with the previously heard target voice; a different task from plain memory recall. Still, after having heard the seven voices of the line-up, it will not only be difficult to remember each comparison, it will also be difficult to remember the voices. Maybe speaker 1 was chosen more often because his voice was remembered best, making it easier to compare with the target speaker. Alternatively, at least, the comparison between his voice and the target will be well remembered. Even when the comparison did not conclude in a perfect match, participants might feel more comfortable in choosing this voice rather than "guessing" some of the voices played in medial positions. The option to replay individual voices of the line-up might not have been used to a large extent; and if so, it might only have shifted the serial effect.

In summary, measures of speaker characteristics such as f0, LTFs, pitch range, speaking rate and similarity ratings, do not show a conclusive pattern for a systematic relation between speaker characteristics and their (mis)identification in voice parades. The most likely reasoning for misidentification is the primacy effect.

Synaesthetic perceptions from the stimuli

Synaesthetes were sent an additional short questionnaire (see Appendix J), asking whether they experienced synaesthetic perceptions during the task, whether they thought these perceptions helped or hindered the memorisation and identification process, and how they would describe the perceptions, in particular differences between those for whispered and modal speech. Unfortunately, only six voice-induced and nine other synaesthetes completed the questionnaire, so no statistical analysis was carried out on the data. Qualitative descriptions and interpretations are possible nonetheless.

Table 30 displays how many voice-induced and other synaesthetes had synaesthetic perceptions with the voice stimuli. The voice-induced synaesthete who did not have synaesthetic perceptions with the stimuli reported to only experience them in one-to-one conversations. For the 'other' synaesthetes reporting to have had synaesthetic perceptions, they were not induced by the sound of the voice, but rather by the words being said, or, in one case, by the volume of the sound. Descriptions by the voice-induced synaesthetes

Synaesthetes	Parade	Helped	Hindered	No effect
voice	modal	4	0	2
	whisper	2	1	3
other	modal	1	0	6
	whisper	1	0	6

Table 31: Synaesthetes' judgements of whether their additional perceptions helped, hindered or had no effect on speaker identification.

revealed that their perceptions were not the same for whisper and modal voicing. Two people had more vivid perceptions with the whisper, but three found whisper concurrents less intense than those induced by modal voices. It also became apparent that whisper seems to induce mostly achromatic colour perceptions, as the following comments from participants suggest:

> The whispered voices were only grey scale whereas normal voices have a variety of colours depending on the tone and timbre of the speaker. It is much easier to discern between red and yellow in your mind's eye than light grey and slightly lighter grey!

> Whispers are more grey/silver/smoky with a different texture to normal voicing, which is smoother and more colourful.

> The whispered ones were all very similar - I think it's the hissy element of whispering, they were all the same grainy white/grey colour. The normal voices had more notable differences, more distinctive colours.

Table 31 shows whether voice-induced or other synaesthetes thought that their synaesthetic perceptions helped, hindered or had no effect on memorising and identifying the voices, for both parades. The only 'other' synaesthete, who stated that perceptions helped for both whisper and modal, was the one whose perceptions changed with the volume of the sound. It seems that judgements about the effect of the synaesthetic perceptions do not have any systematic relationship to performance. In other words, only a few correct identifications were made by those stating a beneficial effect, and some correct identifications were made by those stating no effect. Only one participant reported a disadvantage and, indeed, she was unable to correctly identify the target voice. However, incorrect identification is the more likely outcome in any case.

Listener characteristics

Familiarity with the Irish accent 17 out of 44 participants of the modal voice parade claimed to have regular contact with at least one Irish person, and so did 18 out of 55 participants of the whisper voice parade. Contrary to expectation, the non-significant

		B (SE)	95% CI for Odds ratio		
			Lower	Odds Ratio	Upper
modal	Constant	-0.397 (0.73)			
	Age	-0.005 (0.02)	0.956	0.995	1.035
whisper	Constant	-0.230 (0.78)			
	Age	-0.026 (0.02)	0.931	0.974	1.018

Note modal: R^2=0.001 (Hosmer & Lemeshow), 0.001 (Cox & Snell), 0.002 (Nagelkerke). Model $\chi^2(1)$=0.066, p=0.797
Note whisper: R^2=0.023 (Hosmer & Lemeshow), 0.026 (Cox & Snell), 0.038 (Nagelkerke). Model $\chi^2(1)$=1.449, p=0.229

Table 32: Logistic regression of age as a predictor of correct identification performance: age does not influence performance of participants. B = beta, SE = standard error, CI = confidence interval.

trend showed reverse results: for modal, 9.2% of the participants knowing Irish people identified the target voice correctly, whereas 27.3% of those not knowing Irish people identified it correctly. For whisper, the contrast is 5.5% versus 18.2%, respectively. The Fisher's exact test did not reveal any significant influence of familiarity with the Irish accent on identification performance for whisper (χ^2=0.720, p(exact, 1-tailed)=0.311), nor did the χ^2 test for modal (χ^2=1.972, p(asymptotic)=0.160). However, group sizes might be too small to allow this result to be reliable.

Nationality In total, participants came from 10 different countries, with England, Wales, Scotland and Northern Ireland being treated as separate countries. To test whether nationality had an influence on performance in the voice parades, a χ^2 test was conducted with the three nations that most participants came from, namely England, Scotland and the US. This test did not reveal significant differences: χ^2=1.368, Fisher's=1.393, p(exact)=0.474 for modal, and χ^2=2.361, Fisher's=2.284, p(exact)=0.388 for whisper.

Age The age span of participants ranged from 18 to 72. As it was found that age influences performance in voice parades (e.g. Bull & Clifford, 1984), the influence of age was tested using a logistic regression. Here, age was entered as a continuous predictor for the dependent variable identification performance. Results of this test are captured in Table 32. The extremely low beta values for age in both voice quality conditions suggest that age did not influence participants' performance in the voice parades. The odds ratio supports this suggestion: a value above 1 indicates that, with increasing age, there are more correct identifications, and a value below 1 indicates that, with increasing age, there are fewer correct identifications. The value in this experiment is very close to 1 with its confidence interval crossing 1 for both modal and whisper – highlighting that age is not a good predictor for identification performance. This finding is confirmed by the non-significant χ^2 value and the small power indicated by the low R^2 values (see notes of Table 32).

Musicality Almost half the participants gave information about their musicality. Of these, 14.3% of the non-musical participants identified the whispering target voice correctly, and so did 21.1% of the musical ones. 20% of the non-musical participants identified the target voice correctly in the modal voice setting, and so did 41.2% of the musical ones. This trend was non-significant in Fisher's exact tests: χ^2=0.151, p(exact 1-sided)=0.589 for whisper, and χ^2=0.749, p(exact 1-sided)=0.380 for modal. With only nine participants reporting as not musical, however, the group size was too small to allow any firm conclusions from this sample.

Experimental set-up

First versus second voice parade Participants who took part in both the modal and the whisper voice parade could have had an advantage in the second voice parade, as they by then knew that it was important to remember the voice, rather than the content of what was being said.[13] Indeed, for the modal line-up, there were 24.6% correct identifications when taken as the first parade and 40% correct when it was the second. For whisper, there were 23.8% correct responses when it was the first parade and 28.6% correct ones when it was second. However, a χ^2 test did not reveal significant performance differences according to order: χ^2=0.141, p=0.708 for modal and χ^2=0.127, p=0.722 for whisper.

Sound quality Participants' reports on the audio devices they used varied in precision. Therefore, quality of the audio devices was split into three groups: headphones, external speakers and built-in speakers. It was assumed that headphones would carry best sound quality and built-in speaker worst. For the modal voice parade, 26 participants used headphones, 5 external speakers and 14 built-in speakers. For the whisper parade, 31 participants used headphones, 9 external speakers and 15 built-in speakers. Fisher's exact tests showed no significant influence of sound quality/equipment on correct identification performance: χ^2=0.624, Fisher's=0.568, p(exact)=0.903 for modal, and χ^2=0.860, Fisher's=1.105, p(exact)=0.686 for whisper.

[13]In the instructions, it said "We ask you to pay attention to this recording; you will be asked about it later on as an 'earwitness"', so their attention was not explicitly drawn to either the voice or the semantic content.

6.4 Discussion

6.4.1 Summary and discussion of voice memory

The two voice identification experiments presented here are the first studies to have tested an application of voice-induced synaesthesia. Results suggest that synaesthetes performed differently from controls in voice identification tasks. More specifically, voice-induced synaesthetes performed significantly worse in identifying speakers in the ABX voice comparison task when voiced stimuli were used. This effect was less strong in the voice parade experiment. But in both experiments, especially in the voice parade, voice-induced synaesthetes outperformed the control group in identification performance when whispered stimuli were used, although not significantly so. This means that voice-induced synaesthetes identified voices correctly more often in whisper than in voiced speech samples, to a degree that their identification rate exceeded that of control participants in the modal voice condition. This is a particularly striking result, because all publications to date have shown a drastic drop in identification performance from modal to whispered speech (Orchard & Yarmey, 1995; Evans & Foulkes, 2009; Brungart et al., 2001; Yarmey et al., 2001). This is explained by the increasing difficulty in identifying a speaker in whisper because many voice characteristics are lost when voicing is not present, most importantly f0.

The most comparable previous study was done by Orchard & Yarmey (1995), who offered a target-present 6-person line-up with non-distinctive speakers in both whisper and normal voicing, a target recording length of 30 seconds and line-up recording length of 20 seconds, amongst other designs. Their participants' correct identification performance dropped by 16% from modal (54%) to whisper (38%), and that of my non-synaesthetic participants by 19% (from 39% to 20%). Although my participants had the advantage of hearing the target voice for longer, they had the disadvantage of having an additional speaker in the line-up. Additionally, as the experiment was carried out online, they might have experienced difficulties in maintaining full concentration while "being their own experimenter" at the same time.

The crucial question about the main result is: why are voice-induced synaesthetes apparently better[14]at identifying whispering voices than others, even though this is a harder task? Two main approaches can be taken to address this question. First, there might be a difference in their synaesthetic perceptions between voiced and voiceless stimuli which

[14]Rothen et al. (2012) argued that enhanced performance found for grapheme-colour synaesthetes compared with controls usually lies within a standard deviation of the normal range of memory performance and is therefore rather ordinary than extraordinary; however, differences are stable and therefore of importance. Hence, I believe the difference in correct identifications of 41% by voice-induced synaesthetes and 20% by controls in whisper is of importance too, even though the group difference did not reach significance. All the same, voice-induced synaesthetes were the only group to perform significantly above chance in the whisper condition.

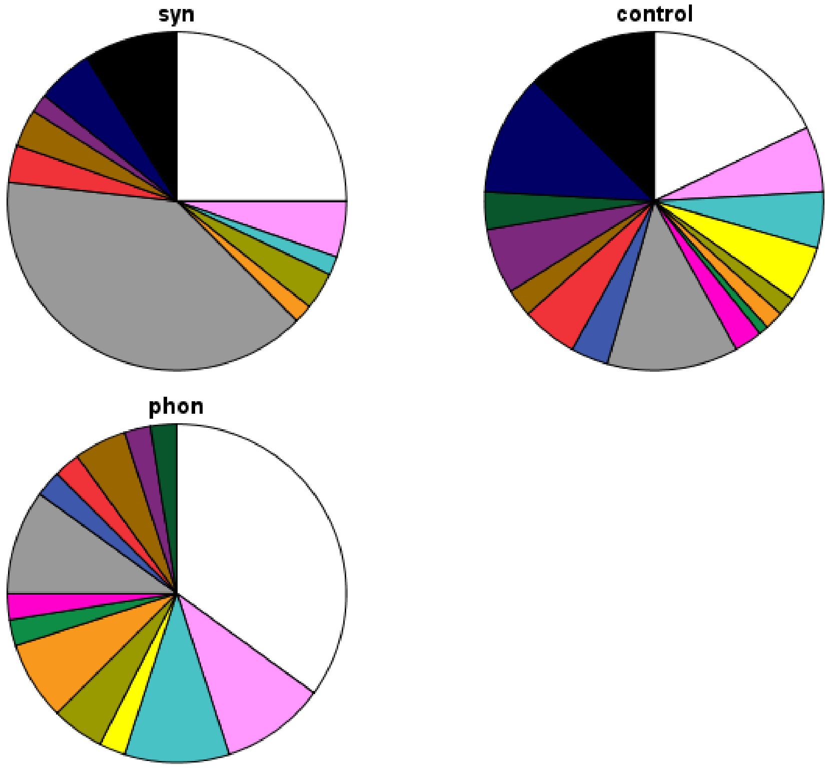


Figure 41: Colour associations with whispered stimuli per group show a high percentage of achromatic associations, especially for synaesthetes.

allows them to focus more on the voice when it is whispering; the additional perceptions could be more helpful in the case of whisper and more distracting or disturbing in the case of voiced speech. Second, the perception of voiced and voiceless stimuli may be underlain by different neurological processes, for example because neighbouring brain areas might be active for different acoustic frequency ranges (cf. Lauter et al., 1985; Talavage et al., 2004).

Descriptions by the synaesthetes are in favour of the first approach: most synaesthetes reported that whisper induced solely achromatic visual perceptions, such as grey, white and silver, whereas the modal voice induced a wider range of colours. Those taking part in the ABX voice comparison task also took part in the voice description experiment described in the previous chapter. Approximately 75% of them chose grey, white or black from the palette of 16 colours when hearing whispered stimuli, as depicted in Figure 41 alongside with the colour choices for whisper by phoneticians and controls. In contrast, synaesthetes associated grey, white and black less than 25% of the time with the remaining voiced stimuli. So there is indeed a difference in synaesthetic perceptions between voiced and voiceless stimuli. Are achromatic perceptions more useful, or less distracting, for voice identification than more colourful perceptions? A possible explanation for this could lie in a finding of the previous chapter: a positive LTF2-luminance correlation was found. As the LTF2 values of the target speakers' initial and line-up recording are very similar, they presumably induce the same luminance values. This, in combination with other factors, could have made it easier to identify the target speaker, as the luminance of the voice was inherently "known" by the voice-induced synaesthetes. As this association was not made by the controls – at least not consciously – they would not benefit from the stable LTF2 value in the same way.

An alternative interpretation to using the LTF2-luminance relationship as synaesthetic memory aid for an enhanced identification performance in whisper is the role of volume. Shereshevsky as well as one of the voice-induced participants reported that the volume of a sound influenced their synaesthetic perceptions. Shereshevsky was easily disturbed by loud sounds which induced synaesthetic visualisations that interfered with seeing the real world. One of my participants even suggested that the different volume settings of the participants' computers might affect the results. As she was the only participant reporting influences of volume, a negative impact of the absolute volume settings per participant is unlikely. Nonetheless, there was a relative difference in intensity between the modal and whisper recordings – assuming that participants did not change the volume between the first and second experiment, if they took part in both: the whisper recordings had been equalised to 60 dB by the creators of the corpus (Cummins et al., 2006), whereas the intensities of the modal voice recordings lay around 70 dB. If the voice-induced synaesthetes were as sensitive as Shereshevsky, the volume of the modal voice recordings might have interfered with their memory performance more than the volume of the whispered voice recordings.

The second interpretation as to why synaesthetes outperformed others in identifying whispering speakers is neurological in nature. Belin et al. (2000) found a voice-selective area in the auditory cortex that showed more activation when passively listening to vocal sounds compared to non-vocal sounds, namely the upper and central part of the superior temporal sulcus (STS). The authors found reduced activation in this area when vocal sounds were bandpass-filtered, so that only low-frequency (bandpass at 200 Hz with 50 Hz bandwidth) or higher-frequency components (bandpass at 1600 Hz with 200 Hz bandwidth) were audible. More precisely, activation was lower than in the unfiltered high-quality sound condition when sounds were manipulated with the higher bandpass-filter, but lowest with sounds manipulated with the lower bandpass-filter. Similarly, behavioural tests showed that participants performed worse in discriminating vocal vs. non-vocal sounds and the sex of the speaker in the higher bandpass condition, and worst in the lower bandpass condition. This indicates that low frequency components, including f0, are essential features for these tasks. As whisper lacks low frequency information, the fact that voice identification is generally poorer in whispered than voiced speech parallels the findings of Belin and colleagues. Unfortunately, Belin et al. (2000) only showed differences in brain activation induced by vocal sounds, excluding whisper, compared to non-vocal sounds, including for example wind. Assuming that wind and whisper share activation in a certain frequency range, a reverse research question – which brain areas are active when hearing frication-like sounds such as whisper and wind – could have possibly shown which brain regions may be structurally or functionally different in voice-induced synaesthetes. Although Talavage et al. (2004) found different brain areas within the auditory cortex activated with stimuli of different frequencies, the sound quality of these stimuli remained the same. Hence, comparisons between voiced and voiceless stimuli cannot be made. For now, it can only be speculated that voice-induced synaesthetes might have enhanced brain activation or structural differences in area(s) of the auditory cortex responsible for processing unvoiced sounds in a higher frequency range, and/or that this area shows stronger interconnectivity with areas in the visual cortex dealing with luminance perception.

The final question about the main results is: why do other synaesthetes show a different performance from voice-induced synaesthetes, especially in whisper? It has been shown that synaesthetes' memory is mostly enhanced with stimuli involved in their type of synaesthesia and only partly extends to other areas (Rothen et al., 2012; Rothen & Meier, 2010). Most of the other synaesthetes had grapheme-colour, music-induced and/or environmental sound-induced synaesthesia. It seems that memory for voices is not one of the areas the memory benefit extends to in other synaesthetes. This is to be expected in grapheme-colour synaesthetes, as the sound of voices is a very different domain from letters and colours. A transfer from memory of music to memory of voices, however, does not seem far fetched. But as there is currently no research into memory of musical sounds in synaesthesia, interpretations cannot be made in this respect. However, it is still a curiosity why the two groups of synaesthetes perform similarly in voice identification in modal voice quality, but differently in whisper.

Another group of synaesthetes who might show enhanced voice identification performance is that of time-space synaesthetes. Simner et al. (2009) tested time-space synaesthetes and controls on their autobiographical memory, memory for public or cultural events such as the death of Pope John Paul II or Oscar winners, and the ability to manipulate real or imagined objects in three-dimensional space. The authors found an enhanced memory performance related to the realm of the inducer and concurrent, e.g. memorising autobiographical and other important dates, but no advantage in other visual tasks not related to their synaesthesia. Could that mean that time-space synaesthetes are better witnesses because they remember events more accurately? This would probably have to be tested on a smaller time scale than the historical scale that Simner and colleagues used. But inferring from the results of the memory studies on grapheme-colour synaesthetes, and from the performance of the 'other' synaesthetes in the voice parade, it can be conjectured that the theory proposed by Rothen et al. (2012) holds true: a memory advantage for synaesthetes is only given in topics relating to their synaesthesia.

6.4.2 Discussion of covariate results

Correct identification performance in a voice parade can be influenced by many different factors (e.g. listeners' abilities, circumstances of exposure and identification process). All statistical analyses testing these factors returned negative results, although these factors were tested because previous publications had found, or at least discussed, influences of them. This disparity is briefly discussed, before turning to influences of speaker characteristics.

It has been previously found that the listener's age influenced performance in voice parades. Children and elderly people have on average a lower hit-rate than middle aged adults (Clifford, 1980; Künzel, 1990, but also see Öhman et al., 2011). An effect of age was not found in this study. The reason for not finding a decrease in performance with older participants could be twofold: first, ten participants older than 50 may not have been enough to make statistical tests reliable. Second, while Clifford and Künzel bin participants into age-groups and compare group results, age was used as a scalar predictor here. Note also that one has to be careful with implying that older people are less reliable as earwitnesses than younger adults because individual differences are large and performance of individuals in the age groups in Clifford's study did overlap.

Neither nationality nor familiarity with the Irish accent influenced results in this study, but was found to be relevant in other studies (Malik, 2010; Goldstein et al., 1981; Vanags et al., 2005). Although more familiarity with the Irish accent is to be expected by British participants rather than those living in other continents, only one participant was actually Irish and four Northern Irish, where familiarity can be assumed. Also the self-report of familiarity with the accent did not correlate with identification performance. This is probably due to the fact that familiarity of an accent by hearing it has a weaker influence on perceptual tasks than actively speaking it or growing up with it.

Having synaesthesia is another influencing factor of listener characteristics and its influences have been discussed in the previous Section (6.4.1) for the most part. An intriguing aspect of visual concurrents is that of possible audiovisual integration. In the field of forensic phonetics, phenomena such as the face overshadowing effect have been researched. Maybe a phenomenon similar to this exists for synaesthetic concurrents: a *concurrent overshadowing effect*. For an overshadowing effect, it is assumed that different processes take place for handling the audio stimulation and the visual stimulation, whereas priority is given to the visual stimulus at least for the first few seconds (Cook & Wilding, 2001). It has been found that voice memory is enhanced when auditory as well as visual information of the face is given over a longer period of time (Sheffert & Olson, 2004; von Kriegstein & Giraud, 2006). It could therefore be assumed that something similar happens in the concurrent overshadowing effect. However, restrictions to this parallel may apply. The beneficial influences of the face-information on voice memory could be due to an adaptation process to the visual input of the face in its whole, after which the focus shifts to lipreading to support the auditory input. A similar adaptation process may take place only in voice-induced synaesthetes when listening to whisper, as they adapt to achromatic visual stimuli more quickly than colourful stimuli (owing to a smaller degree of variation in achromatic stimuli). This would explain the enhanced performance of voice-induced synaesthetes in identifying speakers in whisper.

Synaesthetes were asked whether they thought their synaesthetic perception had any influence on memorising the voice. Although no statistical analysis was carried out, no systematic relation between judgement and performance was found. This finding is in agreement with similar research on the relationship between confidence ratings and accuracy in performance (Philippon et al., 2007; Watt, 2010; but see Künzel, 1990, for a conflicting finding). It is interesting to note that people often seem unable to estimate their abilities accurately – a phenomenon often observed in the context of psychological research (Lundeberg et al., 1994) and face recognition (Bothwell et al., 1987; Busey et al., 2000). This judgement varies between individuals as much as their identification skills. Although some individuals perform better than others, the enhanced performance cannot be generalised. Wilding, Cook and Davis say:

> However, there was no evidence in Cook's extensive data that participants who identified one voice correctly were more likely to be successful on the other voice than those who failed on the first voice. This finding suggests that there are no consistent individual differences in voice recognition ability other than those due to special experience. (Wilding et al., 2000, p. 560)

Künzel (1990) found the same. My data tentatively confirm this finding, because only two participants identified the target voice correctly in both voice parades. One, a control participant, unfortunately did not provide details about his professional background or musicality. The other participant, a voice-induced synaesthete, works in the film music industry, an expertise that needs a trained ear.

Many studies have shown a decline in correct identification performance, the more time elapsed between the initial exposure to the target voice and conducting of the voice parade (McGehee, 1937, 1944; Kerstholt et al., 2004, 2006; Bull & Clifford, 1984). I assume that results of this study do not contribute to these findings because the time range of 19 min to 20 hours is not large enough for memory decay to take place.

Previous research has reported contradictory results on the influence of incidental or deliberate exposure to the target recording. Cook & Wilding (2001) and Perfect et al. (2002) found no significant differences between the types of exposure, but Clifford (1980) and Saslove & Yarmey (1980) did. My participants were instructed to attend to the recording, because they would function as an earwitness later. It was not specified that they should attend to the voice specifically, so it is unknown whether participants set their focus of attention on the voice or the semantic content. As the experiment was conducted online, it is also possible – or even quite likely – that many participants did not read the instructions carefully, and hence did not know they would be asked about the recording again later on. Thus, an interpretation of the non-significant result to the question whether performance was increased in the second voice parade in those participants partaking in both experiments, is difficult. The results are probably most closely related to the work of Perfect et al. (2002), in which one group of participants was told that the authors were looking for a relationship between voice memory and mathematical skills to remember the voice, and the other group was told they needed to remember the voice. So their first group had to attend to both tasks, whereas deliberate attention to the voice could be expected from the second group. Similarly, my participants were made aware of attending to the voice recording in some respect when they took part in the first voice parade, but learned to memorise the target voice in their second run. As no significant differences in identification performance were found by Perfect et al. and myself, I conclude that it is not of major importance whether participants attended partly or fully to the voice.

6.4.3 Methodological considerations

As the aim was to have a voice parade close to forensic standards, the McFarlane guidelines published in Nolan (2003) were followed as closely as possible. However, some compromises had to be made to accommodate for the special case of synaesthetic participants. Foremost, the experiment was conducted online to reach as many synaesthetes as possible. Inevitably, control over certain factors was lost as a result. No control over sound quality across participants was possible. Although participants were asked to report on which type and make of audio device was used, with examples included in the question, descriptions varied widely in their accuracy. Despite the non-significant result of the influence of sound quality, it cannot be guaranteed that this had no bearing on participants' performance, because variation of sound quality within headphones alone may vary largely, and background noise of the environment might have been present as well. Another risk of online experiments is the absence of an experimenter, or in the case of a forensic voice parade, an inspector. The design did not allow the participant to skip any parts or do them in an order other than specified. However, there was no control over whether the participant was distracted from other things in their immediate environment, and whether instructions were read and understood.

Read speech was used for both the initial target recording and the voice parade recordings, although spontaneous speech is recommended in the McFarlane guidelines. This was done to avoid additional influences of word-induced synaesthesia, as this is a common type of synaesthesia. In this way, differences in performance between synaesthetes and controls can be traced to voice-induced synaesthesia. As a comment of a participant revealed, the volume of the recordings also influenced her synaesthetic reactions – something that could not be controlled for. Potentially, other features of the stimuli could have been inducers of certain types of synaesthesia which could not all possibly be accounted for. However, avoiding the use of different words ensured that the biggest ‘rival’ inducer was controlled for.

Further, the use of eight foils is suggested in the McFarlane guidelines. As time restrictions did not allow for recording the audio stimuli myself, they were drawn from a speech corpus. The CHAINS corpus (Cummins et al., 2006) was in fact the only corpus comprising whispered speech suitable for the purposes of this experiment. Although 36 speakers were recorded for this corpus, samples for the line-up could only be drawn from 16 speakers who had the same background respecting age, sex and dialect. The careful selection process restricted the choice to seven speakers in total; the target speaker plus six foils. This was judged to be sufficient, as Broeders & van Amelsvoort (1999) ask for a minimum of five foils in their guidelines.

Although the analysis of the results was largely comprehensive for the amount of data available, further comparisons in speaker characteristics could have been made with a Vocal Profile Analysis (VPA, cf. Beck, 2005; Laver, 1980). VPA is a perceptual scheme used in speech therapy and forensic phonetics. It is used to describe the global physiological configuration of laryngeal and supralaryngeal settings as voice quality features, but also prosodic and temporal organisation features. Time restrictions did not allow for an in-depth analysis. Nonetheless, two brief and preliminary VPA made by two forensic phoneticians detected some voice quality features that might have influenced the confusion of the voices in the parades, whereas other criteria seemed less illuminating in this respect.

Finally, it could be argued that a setup other than a voice parade would have led to clearer results with respect to the main research question. The disadvantage of using a voice parade for voice identification performance is that performance is tested on only one occasion. Multiple trials would overcome this disadvantage, but would lead to longer or multiple testing periods, which would most likely lead to a large drop-out rate – a risk in online experiments as it is. As the voice parade and ABX voice comparison task produced similar results by using two different methods, it is assumed that results are reliable and would be similar when using other methods of testing.

6.4.4 Grouping and classifying synaesthetes

Synaesthetes often have multiple types of synaesthesia (cf. Chapter 5.2.5, Novich et al., 2011). But even within a type, synaesthesia can express itself in different ways, especially in less common and thus less well defined types. Voice-induced synaesthesia, along with other types of synaesthesia involving sound are such types. Hence, it was not always straightforward to classify a synaesthetic participant as a *voice-induced* or *other* synaesthete. Originally, it was intended only to allow voice-induced synaesthetes, or those who do not have any auditory inducer, to take part in the experiment, to have clearly distinct groups of synaesthetes and ensure unambiguous discrimination of auditory and non-auditory inducers. However, it turned out to be nearly impossible to find synaesthetes who did not have any concurrents with auditory inducers, as for example a grapheme-colour synaesthete would most likely perceive their concurrents with a spoken grapheme as well. Therefore, 'other synaesthetes' also included those with auditory inducers. In many cases, it was clearly distinguishable whether the sound of a voice would be the inducer or not, based on self-reports of the synaesthetes. A few cases proved to be challenging in two respects. First, it was not always clear whether general sound - colour synaesthesia included the sound of voices – an issue that could be clarified after some discussion. Second, the strength of different types of synaesthesia varies. As a consequence, some types superimpose upon other types. Here is an example comment from a participant:

> My synaesthesia is not as strong in that way – I see colours when people speak, but it's not to do with the sound of their voices, but what they say. For instance, two people could both say June and it would still be goldish-white. However, if someone had a deep voice or strong, regional accent, it would change the depth/shade a bit.

Following from this description, she was put in the group of 'other synaesthetes', because her voice-induced synaesthesia would only occur in rare circumstances, was very weak and still dominated by word-induced colours.

The description highlights the complexity of synaesthetic perceptions. It shows how types of synaesthesia overlap and even interact, becoming "melded" into one perception. It also emphasises that different types have different intensities, some forms dominating while others are milder. From this, it can only be concluded that the groupings were based on the best of my knowledge and belief – as is the case for all other research on synaesthesia.

7 General discussion

In other words, the world as it is perceived should not be mistaken for a copy of the world as it is described by physics or biology.

(Sagiv, 2005, p. 7)

7.1 Main findings of the thesis

7.1.1 Study on vowel sound - colour associations

The first study reported here examined the colour and grey shade associations that synaesthetes and non-synaesthetes have with particular vowel sounds. Systematic vowel sound - colour and vowel sound - grey shade associations were found in both groups. Vowel formants F1 and F2 significantly influenced u* values – greenness to redness – of the colours chosen by participants, where changes in F1 played a more important role than those in F2. This effect was stronger for synaesthetes. Open vowels with a high F1 such as [a] were perceived as most red, whereas close front vowels with low F1 and F2 such as [i] were perceived as most green. These influences of the acoustics of vowel sounds were significant at a group level, although there were individual differences in colour associations. Similarly, systematic influences were found for luminance associations with vowel sounds. Higher vowel formants induced higher L* responses, i.e. lighter colour or grey shade choices. Here, F2 had a stronger influence on the results. Front and open vowels were perceived as lighter than back vowels, meaning that, for example, [i] and [ɛ] were perceived as lighter than [o]. Finally, it was found that synaesthetes' colour associations were significantly more consistent, measured through repeated exposure, than those of non-synaesthetes. However, there was no significant difference in consistency between the two groups when associating grey shades with the vowel sounds.

Data of the group average of vowel sound - colour associations by synaesthetes show very systematic shifts in colour space as formant values, or articulatory settings, change. As the first study of its kind, the findings reveal that the vowel space can be projected onto a colour space with impressive systematicity, as shown in Figure 19a on page 78. Although this finding indicates the influence of vowel *acoustics* on colour associations, a linguistic influence has to be acknowledged as well. In the subgroup of Scottish synaesthetes, the systematic acoustic influence was overlaid by graphemic influences: [e] embodied a "colour island" distinct from the systematic shifts from green/yellow for [i] to red for [a], because it was also associated with red, presumably by analogy to the Scottish pronunciation of <a> in isolation or in words such as *hate*. Thus, we can conclude that there is a co-existence of acoustic influences on colour associations and established effects such as that of graphemes and their frequencies of occurrence.

7.1.2 Voice description study

The study on voice descriptions was the first to explore descriptions of voices by synaesthetes, phoneticians and controls on a group level by collecting qualitative data (verbal descriptions of the voices) and quantitative data (colour and texture associations and the use of semantic differentials). Analysis of the verbal descriptions revealed how the participant groups have different foci for their descriptions: synaesthetes often used their synaesthetic perceptions to describe a voice alongside more general terms, whereas phoneticians used more technical terms, and controls focused on describing personal characteristics of the speaker as well as his voice. Acoustic analyses of the influence of differences in voice quality on colour and texture choices and ratings of the semantic differentials were made. Higher fundamental frequency, a larger pitch range, and higher formants were associated with lighter colours across groups. High f0, high vowel formants, larger pitch range and steeper spectral tilt resulted in associating textures that were 'fluid' and 'smooth', whereas a shallow spectral tilt resulted in 'rough' and 'line-like' texture choices. Additionally, consistency in people's associations was measured for textures, colours and semantic differentials. There was a non-significant tendency for synaesthetes to be more consistent in some of their associations than controls, mainly in colour associations. Regular patterns that show cross-modal usage of the expert knowledge of phoneticians were found as they used semantic differentials such as *smooth-rough* on a more extensive scale.

The acoustic analysis of voice-inducers in general, and the use of visual texture responses specifically, represent new departures in synaesthesia research, and contribute to the growing area of synaesthesia research using methodologies from a broad range of scientific fields.

7.1.3 Studies on voice identification performance

An ABX voice comparison task and two voice parades were conducted to test whether synaesthetes performed differently from non-synaesthetes in speaker identification. Voice-induced synaesthetes performed significantly worse than non-synaesthetes in identifying speakers in the ABX voice comparison task when voiced stimuli of various voice qualities were used. This effect was less strong and non-significant in the voice parade that used normal voice quality. But in both experiments, especially in the voice parade, voice-induced synaesthetes outperformed the control group in identification performance when whispered stimuli were used. Although group performance did not differ significantly here, voice-induced synaesthetes were the only group to identify speakers significantly above chance level when whispered stimuli were used.

This was the first series of experiments to study the practical use of voice-induced synaesthesia and at the same time the transferability of memory advantages of other synaesthetes, as have been found in other studies (c.f. Rothen et al., 2012). Results indicate no benefit in speaker identification for other synaesthetes and only some benefit in voice-induced synaesthetes, namely when speakers employed a whispered voice quality.

7.2 Shared findings and implications

Since all the experiments reported here shared the use of spoken stimuli, it is possible to generalise several of the findings. Whereas the first experiment studied aspects of vowel sounds as an inducer, the others tested voice quality. The focus of the analysis of spoken stimuli was acoustic-phonetic in nature since this is an under-researched area in synaesthesia research, and it offers one way to give a quantitative footing to observations regarding speech and colour/texture. The most surprising and revealing findings of the studies were not the extraordinary behaviour of synaesthetes (such as their within-subject consistency in colour associations) but rather the common underlying mechanisms that were found across participant groups regarding shared associations such as that [a] is red and that there are luminance associations with the intrinsic vowel pitch as well as the fundamental frequency. These common underlying mechanisms for luminance associations are not only shared between vowel sounds and f0, but also with musical pitch (Ward et al., 2006; de Thornley Head, 2006; Mudge, 1920). It seems that the cross-modal link of synaesthesia-like audio-visual processing is very common amongst human beings to an extent that it could be defined as 'default' or 'standard'. Ludwig et al. (2011) even went so far as to claim that this link is hard-wired in primates, as they found that chimpanzees match white with high pitch and black with low pitch: "Rather than being a culturally learned or a linguistic phenomenon, this mapping constitutes a basic feature of the primate sensory system." (Ludwig et al., 2011, p. 20661)

In this context, it is of interest to investigate the influence of sound symbolism. Although sound symbolism in its broadest scope has not yet been proven to be universal (Reay, 2006), no counter-evidence has been found regarding topics related to my research, such as the relationship between a luminance scale and the loudness or pitch scale, or in fact, spatial positioning (Spence, 2011). It could be hypothesised that a similar universality may be found for the relation between voice quality and texture. Indeed I found that a *harsh* voice, for example, was associated with 'rough' textures, whereas a *modal* or *breathy* voice was aligned with 'smooth' textures. Ludwig and her colleagues do not believe that audio-visual mappings are culturally or linguistically influenced. Could the cause and effect be reversed? How much do colour and texture associations with voices and speech sounds tell us about the origin of language? Could the use of metaphors have evolved out of cross-modal associations? Dark shades, for example, are associated with a low-pitched voice. Dark also connotates something sinister or evil. When asked to imitate a villain, it is common for people to lower their voice. As early as 1920, Mudge noted: "Although there is a wide variation of color-imagery, it is probable that the common term 'tone-color' is not a mere figure of speech based upon analogy." (Mudge, 1920, p. 345)

Synaesthesia research can help in investigating these shared cross-modal phenomena in human beings, since synaesthetes consciously perceive what others are not aware of but seem to unconsciously experience, nonetheless.

7.2.1 Synaesthesia: A complex phenomenon

The role of vowel sounds as an inducer

By revisiting the vowel sound - colour associations of synaesthetes and analysing the vowel sound - grapheme relation, it can be concluded that not all participants show plain vowel sound - colour synaesthesia. While previous research highlighted the influences of graphemes on colour associations even in spoken stimuli (Beeli et al., 2007; Paulesu et al., 1995), here a systematic influence of acoustic features of the vowel sounds was found. Irrespective of whether we believe that these two influences are intermingled in every synaesthete or whether one inducer dominates the other, it becomes apparent that the phonemes of a language play a smaller role in this influence than vowel sounds in general, because the relationships between colour associations and the vowel quadrilateral show an acoustic-phonetic and/or graphemic influence, rather than a phonemic influence of vowel-categories specific to a language.

A theory that has been proposed to explain why phonemic influences on colour associations are less common than graphemic influences is that of the adjacency of the involved brain areas. Simner (2007) suggests that tastes as concurrents are more often tied to phonemes and colours more often to graphemes because the "colour perception area" V4 is situated closer to the "grapheme processing area", and the primary gustatory cortex[15] closer to the regions involved in phonology and lexical semantics.[16] However, this theory does not explain the influence of acoustics on colour associations. Here, it can only be speculated that the underlying mechanism of acoustic influences is shared – for example, with music-induced synaesthesia and its influences of (tone or intrinsic vowel) pitch. Such a shared mechanism is also suggested by the positive correlation between increasing pitch and increasing luminance. Alternatively, Simner (2007) claims that high-frequency letters and numbers pair with high-frequency colour terms. According to Sereno (1994), 16% of the English vowel sounds occurring in verbs and nouns are open vowels (C3.5 and C5), i.e. they occur more often than other phonemes on average. Since open vowels were associated with red (a high-frequency colour term), an analogy to Simner's interpretation on grapheme-colour associations can be partly drawn for vowel sound - colour associations.

[15]The gustatory cortex is located partly on the insular lobe, a part of the cerebral cortex folded deep within the lateral sulcus between the temporal lobe and the frontal lobe, and on the frontal operculum on the inferior frontal gyrus of the frontal lobe.

[16]An area in the anterior insula is related to both speech perception and production and lies adjacent to the primary gustatory area (Ward et al., 2005).

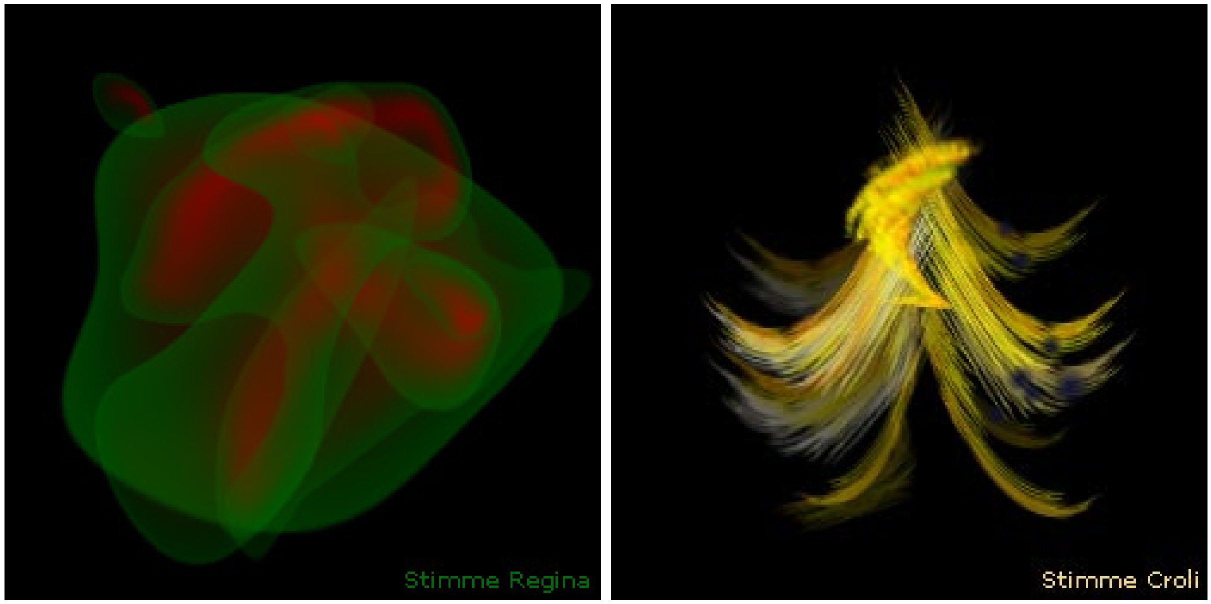


Figure 42: Voice portraits of two friends of a synaesthete. Courtesy of Dr. Alexandra Dittmar.

The role of voices as an inducer

The studies on voice-induced synaesthesia explored which role voices play when inducing synaesthesia and whether a memory benefit might occur with voices. By eliciting verbal descriptions of voices I was aiming to find the concurrents that synaesthetes perceive with voices, because so far it had been unknown which concurrents are most common in voice-induced synaesthesia. With colour being the most common concurrent followed by texture, shape, movement and arrangement in space, it became apparent that most concurrents are of a visual nature. The complexity of concurrents is demonstrated in verbal descriptions such as "Mustard yellow, outlined in dark brown, with touches of drippy or slimy green. Dark background. The voice moves upward and curls like a snail. There are 'hairs' around some of the words." Similarly, a set of voice portraits shown in Figure 42 illustrate the complexity. The images capture the synaesthetic perceptions of the voices of two friends of a voice-induced synaesthete.[17]

It was difficult to disentangle which aspects of the voice sound trigger concurrents in a systematic way. The fundamental frequency was the strongest aspect influencing luminance values of colour associations; and f0, vowel formants, pitch range and spectral tilt influenced texture and colour associations. So ultimately, it can be concluded that some acoustic features of voice qualities have a systematic influence on synaesthetic perceptions. Nonetheless, it was found that voice is only one type of inducer that was, in all but one case, intermingled with other inducers such as music, words and letters. The fact that

[17]Neither the voices nor the synaesthete were part of my experiments. The graphs are an illustration of synaesthetic voice perceptions similar to the way my participants might have displayed/shown/captured them if they were asked to paint rather than describe the voices.

systematic acoustic influences of the voice were found highlights the role of a voice as an inducer, but the lack of consistency also suggests that it either does not play a major role, or is labile and context-sensitive, in most synaesthetes who have additional types of synaesthesia. Possibly voices play a minor role in synaesthesia because they are mainly used as an instrument to send the semantic message. Although emotion and affect also carry meaning – and change the voice quality – they might still play a minor role because most inducers seem to be overlearned items such as sequences (alphabet, numbers, words, months or musical notes). A method for testing the influence of a voice disregarding other influences has yet to be designed. Ideas such as removing the linguistic content from speech seem the only solution, but this hardly seems to approximate a realistic setting, as discussed in Section 5.5.

It remains a puzzle why enhanced memory performance in voice-induced synaesthetes should only be found with whispering voices. Perhaps it owes to the fact that no easily definable indicator has been found showing which aspects of a voice induce synaesthetic perceptions. The inducer rather seems to be an interplay of various acoustic criteria which make it too complicated to be used as a mnemonic, unless the acoustic information is reduced, as in the case of whisper by the absence of a pitch.

The co-existence of types of synaesthesia

The voice description experiment in particular highlighted the issue of co-existing types of synaesthesia. Some important questions need to be addressed when speaking about co-existing types: to what extent are concurrents influenced by multiple inducers? To what extent would the concurrents differ when triggered by a single inducer? For example, if one voice was mostly blue and one voice mostly yellow and they both said a red word, would the concurrent be purple in one case and orange in the other? Would the stronger one dominate, or would they somehow blend into each other differently? Multiple extensive case studies will have to be conducted with a multitude of stimuli with a controlled variation of the possible co-inducers to address these questions. Should it be concluded that inducers are sometimes inseparable, i.e. that there are hybrid types? Perhaps synaesthesiae cannot be compartmentalised; they might be like internal organs: although we can identify them separately and know their individual functions, they only work in conjunction with each other. Similar to the idea that there is a continuum of being more or less synaesthetic, a continuum could be hypothesised that allows more or fewer inducers to act in a manifestation of synaesthesia of various strengths: an *inducer continuum*. Since certain types of synaesthesia seem to cluster more often than others, and clusters mostly contain various inducers but only one or few concurrents (Novich et al., 2011), it can be assumed that types within a cluster can be placed on the higher end of the suggested inducer continuum. This would, again, face us with the difficulty of giving suitable names to the combinations of inducers. These names would be more general at the high end of the inducer continuum (e.g., 'sound-visual'), and more specific at the low end (e.g. 'days-colour'). In the case of types of synaesthesia examined in this thesis, the umbrella-term 'speech sound' could be used for the inducer. Whichever term is used for different types of synaesthesia, it should be defined as closely as possible, i.e. listing all known inducers with the degree of certainty of their existence and strength, thus leading to both an inducer continuum and a concurrent intensity continuum.

It is evident that synaesthesia expresses itself in very heterogeneous ways (see also Novich et al., 2011). "As Simner's article emphasizes, we must treat the heterogeneity of the condition as an interesting clue rather than an inconvenience to be swept under the rug." (Eagleman, 2012, p. 18) In testing under-researched types of synaesthesia, I embraced this heterogeneity. Results made us aware of the multifaceted features that cannot be found in straightforward types such as grapheme-colour synaesthesia – or, more precisely, the simplified version of it that is proffered following standardised tests. Heterogeneity triggers a dilemma in research, namely how to find the balance between single case and group studies, and between qualitative and quantitative data analysis. Although I generally prefer group studies over case studies, because they have more power when referring to the general population, the individual/idiosyncratic nature of synaesthesia presents a challenge in this respect. Here, this challenge was overcome by conducting qualitative as well as quantitative analyses on a group level, but also explaining within-group differences in synaesthetes regarding their vowel sound - colour associations and verbal descriptions of voices.

One explanation as to why there are different types of synaesthesia in co-existence is that multiple neurological "systems" are present at once: functional or structural differences in different parts of the brain could be responsible for different aspects of synaesthetic perceptions; for example, grapheme-colour associations could be processed in the V4 area for a synaesthete (e.g. Hubbard et al., 2005), who may process music-colour synaesthesia in the superior temporal sulcus in the temporal lobe at the same time (Calvert et al., 2001). Alternatively, a non-localised increased activity or connectivity could explain the phenomenon of synaesthesia more generally. These ideas are discussed in the following section.

Findings in the context of theories of synaesthesia

Researchers have constructed different theoretical models with the aim of explaining the neurological basis, or neurological manifestation, of synaesthesia, as described in Section 1.6. My results suggest a co-existence of different kinds of synaesthesia having an impact on the processing of a single stimulus, or perhaps the existence of different influences from graphemes, speech sounds, voices and other inducers, on a single kind of synaesthesia.

How can this co-existence be neurologically explained? Many neuroimaging studies on grapheme-colour synaesthetes have found functional or structural differences in brain areas usually related to the concurrent (Rouw & Scholte, 2007; Hubbard et al., 2005; Sperling et al., 2006; Nunn et al., 2002; Aleman et al., 2001). It is assumed that these differences in brain activity or connectivity cause the synaesthetic perceptions and therefore the differences in processing stimuli from non-synaesthetes.

In the case of vowel sound - colour associations, the data probably cannot be accounted for solely by hyper-connection between adjacent brain regions dedicated exclusively to processing visual graphemes and colour, because an acoustic-phonetic influence has been proven to exist. Moreover, there are studies on grapheme-colour synaesthesia which showed neither functional nor structural differences between synaesthetes and controls in concurrent-related brain areas (Rouw et al., 2011; Hupé et al., 2011). If these contradictory findings cannot be explained by methodological weaknesses of the studies, they suggest that multiple different synaesthesia-related neurological systems co-exist. Rouw et al. (2011) found that grapheme-colour synaesthetes show hyper-connectivity in the parietal lobe implicated in binding, suggesting that their unusual abilities might relate to letter-sound integration in general (with the letter representing a concept), rather than *visual* letter processing specifically. Hubbard et al. (2011) and Weiss et al. (2005) also assign a key role to the parietal cortex for synaesthetic binding.

It was found that synaesthetes and non-synaesthetes at least partly share similar mechanisms underlying their associations. Simner et al. (2005) stated:

> The letter-colour pairings of our control groups differ from those of synaesthetes in terms of consistency, automaticity, and phenomenology, but certain patterns of responses appear to be common to both synaesthetes and non-synaesthetes alike. Such similarities between populations suggested that grapheme-colour synaesthesia, like the music-colour variant, may stem from an exaggeration of mechanisms for cross-modal association that are common to us all. Both populations are influenced by linguistic priming, in that colours tend to be paired with the initial letter of the colour name (e.g., b-blue), although this effect was more dominant in non-synaesthetes. (Simner et al., 2005, p. 1081)

Assuming that the parietal lobe is also active in non-synaesthetic binding tasks (such as pairing <b> with blue), the interpretation of Hubbard et al. (2011) and Weiss et al. (2005) on the role of the parietal lobe in synaesthesia should be extended: additional functions of this brain area other than binding are the dorsal stream of the visual system and the use of symbolic functions in language. The dorsal stream is mainly active in tasks involving spatial relationships and coordination in body space. This could play a role for projector synaesthetes who perceive their synaesthetic associations outside their body and would therefore relocate their focus onto the point in space where the synaesthetic perception takes place. The symbolic functions in language could also play a larger role in synaesthetes than non-synaesthetes, as the symbolic function might be redefined in synaesthetes: letters or words might become symbols for colours, or vice versa. Thus, the additional activation in the parietal lobe in synaesthetes might be a neurological marker for synaesthesia after all, but the reasons might differ from some interpretations found in the literature.

In addition to the role of the parietal lobe, activation in the retrosplenial cortex (RSC) in the limbic lobe may also be important. Hupé et al. (2011) found increased white matter in the RSC in synaesthetes, an area involved in processing emotion, memory, relating objects to their contexts, and switching from the external world to internal mentation/contemplation. All these functions are likely to be involved in synaesthetic perceptions. The authors interpret their finding thus: "[...] the structural differences we observed may be related to the complex construction of meaning by the brain, involving not only perception but certainly at least language, memory, and emotion" (Hupé et al., 2011, p.8). Consequently, they take the focus off of colour-related areas and approach the neurology of synaesthesia with a more holistic view, covering complex mechanisms of dealing with multi-modal perception not only on the level of sensory perception, but including more abstract entities. Although this was a case study done with only one rather extraordinary case, Bor et al. (2008) found that the "typical" colour areas in the extrastriate cortex like V4 were inactive during synaesthetic perceptions for Daniel Tammet, while areas in the lateral prefrontal cortex were active. The authors attribute this to the fact "that he has an unusual and more abstract and conceptual form of synaesthesia" (Bor et al., 2008, p. 311). Maybe this abstract or conceptual form is more common than it has been thought to be, and the interplay of different inducers found in my participants parallels this "more abstract" form, especially in the case of voice-induced synaesthetes. Assuming a genetic predisposition towards being synaesthetic (Tomson et al., 2011; Barnett et al., 2008; Baron-Cohen et al., 1996), differences in the parietal lobe and the RSC could explain why different types of synaesthesia run in families: the same brain areas exhibit structural particularities leading to different "phenotypes" of synaesthesia.

The thoughts of Hupé and colleagues and Bor and colleagues about the abstractness and inclusion of language and emotion into synaesthesia research bring to mind the discussion about re-defining synaesthesia, which was initiated by Simner (2012a). Simner speculated about the definition of synaesthesia being due to cross-talk between brain areas, what could then, strictly speaking, mean that exceptionally good speakers or writers might have synaesthesia caused by hyper-associations in language-related brain areas. Although this is only a hypothesised scenario, this idea in combination with the interpretations of Hupé et al. (2011) and Bor et al. (2008) could be an explanation as to why it is common for artists and musicians to be found amongst synaesthetes (Rich et al., 2005). The tendency of synaesthetes to be more musically inclined than others has been found in my synaesthetic population as well.

In summary, some studies have found an increase in brain activation across brain areas that are not directly or primarily involved in the processing of the stimulus or synaesthetic perception. As integration of sensory information is a key characteristic of synaesthesia, it is not surprising to find activation in brain areas active in binding and sensory integration. Unfortunately, it remains unclear whether this activation is, in the case of grapheme-colour synaesthesia, due to the binding of the visual presentation of the letter with the synaesthetically perceived colour or due to the integration of the speech sound associated with the (concept of the) letter, which may trigger the synaesthetic association as a result of a chain reaction. For more complex and less researched types such as voice-induced synaesthesia, it is even more speculative to apply interpretations of findings on grapheme-colour synaesthesia: only direct neuroscientific investigations will be able to resolve this issue. In conclusion, it seems that even within a type, synaesthesia expresses itself in different neurological ways. This could lead to the assumptions that some synaesthetes' neurological processes are similar to those of controls – or vice versa, that some controls process stimuli similar to synaesthetes. This potential variation across the whole population, once again, adds to the debate regarding whether synaesthesia is a continuous spectrum rather than a binary condition – or whether, at least, the underlying neurological reasons might lie on such a continuum.

The (re)definition of synaesthesia and its types

The classical definition of synaesthesia can be considered rather rigid, taking into account the multifariousness of the condition. Two main areas are addressed in recent challenges to the current definition: to discriminate between perceptual and conceptual inducers, and to allow the idea of a synaesthesia spectrum.

Distinguishing between concepts and percepts inducing synaesthesia has usually been addressed by differentiating between 'high-level' and 'low-level' synaesthesia. The concept or semantic nature of a stimulus, for example the meaning of the number 5, no matter whether presented as '5', 'V', 'five' or simply thought of, is the inducer of high-level synaesthesia, whereas only the physically printed representation '5' is the inducer of low-level synaesthesia. Danko Nicolić, however, thinks that the term 'synaesthesia' is not appropriate to describe perceptions induced by a concept or an idea. Therefore, a new term is suggested: *Ideaesthesia*, "a phenomenon in which a mental activation of a certain concept or idea is associated consistently with a certain perception-like experience" (Nikolić, 2009, p.4). Although the author, in fact, used this definition to describe the word 'synaesthesia' in his paper, it becomes apparent that he means to define ideaesthesia, which should be separated from *synaesthesia literal*, where sensory perceptions induce concurrents. He explains the introduction of the new term:

> One alternative is the word ideaesthesia, which is a combination of two ancient Greek words, one for concept, “idea”, and the other for sensation, “aisthesis”. In translation, ideaesthesia means sensing concepts or perceiving meaning. This sends a considerably different message than does union of senses [a translation for synaesthesia]. Thus, given the available empirical evidence, the word ideaesthesia describes the discussed phenomenon much more accurately than the word synaesthesia.” (Nikolić, 2009, p.4)

While it is understandable to want to distinguish between sensory and semantic induction of the phenomenon, I would suggest, instead of assigning people into two groups – ideaesthetes and synaesthetes – to take the meaning of ‘synaesthesia’ less literally. The phenomenon has proven to be very multifaceted, and moreover, it was found that types of synaesthesia are intertwined to an extent that makes them inseparable. The attempt to clarify this entanglement by splitting the phenomenon into two presents itself as, most likely, impossible to achieve. Data from the voice description experiment suggests that this phenomenon is synaesthetic rather than ideaesthetic, as the sound of different voices is not a concept but an auditory percept (although I did not ask my participants what they perceive when they *think* of a particular voice). The shared influence of graphemes and vowel sounds found in the vowel sound - colour association experiment highlights the influence of both conceptual (graphemic) and perceptual (phonetic-acoustic) influences at once. Thus, I think the introduction of an additional term, instead of simplifying, might complicate matters of definition, because, again, it may not be possible to separate the influence of various inducers, which may well be both conceptual and sensory, or, as Sidoroff-Dorso (2010) calls them, quasi-sensory.

Considering the idea of a synaesthesia spectrum (Simner, 2011; Eagleman, 2012), it should be pointed out that there was a commonality of underlying mechanisms in cross-modal associations for my participants in general, as well as a spread of within-subject consistency across participants, independent of their group affiliation. The issue of cross-modal binding has been addressed in various studies (Spence, 2011; Wrembel, 2010; Marks, 1974; Ludwig et al., 2011) and verified in my experiments, as statistical analyses resulted in shared pitch-luminance associations (for both vowel sounds and voices), pitch-colourfulness associations and voice acoustics-texture associations across all participants. On one hand, the distribution of within-subject consistency across participants for the vowel sound - colour associations suggests grouping participants into synaesthetes and non-synaesthetes. For the vowel sound - grey shade associations and the results on voice-colour, voice-texture and voice-semantic differential associations, on the other hand, results suggest a synaesthetic spectrum spanning across participant groups.

These opposing findings, in turn, highlight the different manifestations of synaesthetic perceptions and suggest that we should not only think of a synaesthesia spectrum, but also place types of synaesthesia on a spectrum rather than assuming that all types function in a similar way – as has been proposed with the inducer continuum in Section 7.2.1. This view is supported by some synaesthetic individuals who call themselves 'semi-synaesthetes' or discriminate between 'mini'- and 'maxi'-synaesthetes, because their type(s) of synaesthesia seem(s) less obvious or strong than 'proper' synaesthesia; e.g. when 'only' perceiving grey shades rather than a variety of colours or having a timeline as their 'only' kind of synaesthesia (communication via the German synaesthesia email list hosted by Regina Pautzke at http://Synaesthesieforum.de). Verbal descriptions and the painting in Figure 38 on page 153 indicate that, for several individuals, concurrents are evoked by an interplay of the sound of the voice and the letters or words being said. A participant in the voice identification experiment also mentioned that the volume of a sound changes her perception of a voice. While these descriptions indicate an interaction of inducers, other findings (Hubbard et al., 2005; Grossenbacher & Lovelace, 2001; Smilek et al., 2001) point at only one inducer.

It is therefore not surprising that findings in synaesthesia research are somewhat heterogeneous in parts – as is the condition itself. The research area with least agreement in the outcomes of studies on synaesthesia so far seems to be that of neurological research, in which it seems unclear whether synaesthesia stems from structural or functional differences, and from direct or indirect ones (see e.g. Bargary & Mitchell, 2008; Hubbard et al., 2011). Following the suggestion of multiple spectra rather than clear-cut classifications around the phenomenon, as just described, I share the view of Cohen Kadosh & Terhune (2012), hypothesising that different types of synaesthesia may result from different neurological mechanisms such as enhanced connectivity, disinhibition or a lack of cortical specialisation. Or to put it in Eagleman's words: "Likewise, the end result of synaesthesia may spring from several fundamentally different neural processes (e.g., neuronal overgrowth, under-pruning, imbalanced inhibition, and excitation) all of which happen to converge on the similar result of unusual perceptual or cognitive pairings" (Eagleman, 2012, p. 18).

7.3 Challenges and limitations

When formulating research questions and designing experiments, every researcher is faced with the challenge to decide which aspects to focus on, whether to conduct broad or in-depth studies, which methods to use, etc. While studies can be designed in a more or less self-contained fashion (depending on their topic), one must always look at ways to explore further. This section deals with the challenges and limitations found in my studies.

Possibly the biggest challenge that I faced when conducting the experiments was the difficulty of recruiting voice-induced synaesthetes, as this is a rare type of synaesthesia. Although the international mailing list hosted by Sean Day, the British mailing list hosted by the UK Synaesthesia Association, and the subject pool of Julia Simner and Jamie Ward were used to recruit synaesthetes, only a relatively small number of participants engaged in the experiments. As this was anticipated, experiments were conducted online – a setup that allowed the inclusion of more participants but offers less control over the conduct of the experiment. Further difficulties are discussed in the following two sections.

7.3.1 Stimulus presentation

Colour concurrents are usually very specific; instead of "green", a description of a synaesthetically induced colour by a synaesthete could be "green of an unripe banana". Many experiments in which colour responses are required, employ a large set of colours such as colour wheels or other colour sliders to facilitate responses. It could therefore be argued that the colour palette used in the vowel sound - colour and the voice description study did not cover a large enough range of colours to satisfy synaesthetes' needs, or that results are less precise with only 16 colour choices. However, three major points have to be taken into account: (1) It is possible that descriptions of synaesthetic colours are only so specific because synaesthetes are asked to give a concrete description. (2) Some studies reduce the colour responses to the 11 basic colour terms of English before doing analyses (e.g. Day, 2004; Simner et al., 2005; Wrembel, 2009). (3) Psychophysical constraints were taken into account: to allow the participant to react to the perception of the sound rather than the memory of it, fast response was facilitated, which would not have been possible using a colour-picker system with (near-)unlimited options.

Another criticism, similar to the limited set of colour choices, can be made related to the texture response displays, where again a set of 16 categories was given to choose from. This method cannot be compared to other forms of presentation, as there has not been any research conducted that included visual texture concurrents. Despite this lack of alternatives with which to compare my method with, one shortcoming was that the samples were not controlled for their luminance – only colour information was removed. It also remains unclear to what extent "typical" synaesthetic texture perceptions were covered. Since participants did not complain about an unsuitable range of texture choices, I assume the variation in the set was large enough to suit its purpose of a first exploratory study. In addition, the texture analysis was run with nominal data description rather than quantitative measures, as introduced by Petrou et al. (2007), for example. Thus, the measures do not parallel colorimetric analyses, but rather share the methodology of analysing the semantic differentials.

Another methodological question in the voice description experiment is about the presentation of the voice stimuli. By choosing different voice qualities produced by two trained phoneticians, I aimed at controlling for variations in the voice that might otherwise have been difficult to specify without the help of extensive and differentiated analyses by speech therapists or trained phoneticians. Because the differences in voice qualities only influenced synaesthetic perceptions to a limited extent, one could argue that a limited set of two speakers, even though they produced different voice qualities, was a suboptimal choice. It might have been better to use different speakers using their natural voice qualities, and to additionally introduce female voices. Only a follow-up study could resolve that uncertainty. Similarly, one could argue that English vowel phonemes might have been a better choice than cardinal vowels for the stimuli in the vowel sound - colour experiment. However, the main interest of this experiment was the influence of acoustic vowel characteristics, not the categorical influence of phonemes, especially as not all synaesthetes were native speakers of English.

Since both experiments on voices were carried out online, the limitations of online experiments need to be addressed as well. In general, with online experiments the experimenter runs the risk of receiving incorrect details given by participants such as nationality, native language and in this case also types of synaesthesia etc.[18] Experience shows that participants often do not fully attend to the instructions. Furthermore, there may be uncontrolled distractions, such as background noise, children crying, telephone calls etc. Finally and foremost, if audio samples are used, the sound quality of the audio equipment in unknown. These disadvantages had to be accepted in order to reach a wider audience of synaesthetes.

7.3.2 Data interpretation

One of the challenges to data interpretation was classifying participants according to their types of synaesthesia. As mentioned above, types of synaesthesia can blend into one another. When these types are closely related to each other, it may be difficult to assign one specific type. Sometimes the synaesthetic perceptions may actually be triggered by a mix of inducers, meaning that it is impossible to define the potentially multiple inducers correctly. This became apparent in the descriptions of some participants who, for example, claimed to perceive colours induced by the words being said, but the colours might change slightly if the voice quality was extraordinary; or those who reported that the perceptions were induced by the sound of the voice, but the perception may change when volume is manipulated. Additionally, the inducers may vary in how strongly they affect the concurrents. Also, some types of synaesthesia are simply less clearly defined than others. A lot of individual variation exists in synaesthetes in general and in under-researched types such as voice-induced synaesthesia in particular. Thus, the grouping of synaesthetes, for

[18]In fact, I excluded one participant after detecting by further examination using internet resources that he was not a native speaker of English.

example into voice-induced and others in the voice description experiment, is difficult to pursue and runs the risk of being somewhat imprecise.

Another factor that may have influenced the results is that of participants' nationalities and dialect backgrounds. As shown with the analysis of a subgroup of Scottish synaesthetes in the vowel sound - colour associations, language background can influence synaesthetic perceptions, and this may have played a role in voice-induced synaesthesia as well. It is known that voice quality is a feature of accents of a language (Stuart-Smith, 1999; Laver & Trudgill, 1979), so participants of a shared dialectal background might have more agreement in their descriptions of voice qualities than a group of individuals of non-shared dialectal background, such as my group of participants was. Or, since the speakers in the voice parades were Irish, Irish participants might have performed better at the speaker identification task then the heterogeneous group of synaesthetes tested.

Finally, it was sometimes difficult to interpret results because of a lack of numbers. Many statistical tests are only reliable when a certain amount of data is entered, which could not always be reached. Thus, for the time being, results have to be interpreted with caution.

7.4 Ideas for the future

To reach a more detailed understanding of voice-induced synaesthesia, more research involving more participants is needed. Stimuli presentation could be altered for this purpose: different speakers using their natural voices rather than just a few speakers producing different voice qualities could function as inducers. A detailed Vocal Profile Analysis (VPA) could aid the analysis of results in finding differing speaker characteristics. The influence of speaker sex should be examined as well. With a behavioural outline and proof of voice-induced synaesthesia, neuroimaging studies on this kind of synaesthesia could be conducted. We could investigate to what extent voice-selective brain regions found by Belin et al. (2000) in the upper bank of the superior temporal sulcus (STS) act or interact differently with other regions which are potentially active in the binding or processing of the concurrents in voice-induced synaesthetes but not other synaesthetes.

In order to obtain more knowledge about the development of synaesthesia and related neural structural or functional differences, it would be desirable to conduct longitudinal studies. Since crucial aspects of the development happen in early childhood, it is obviously difficult to do so. One would have to find young children whose synaesthetic perceptions became apparent to their knowledgeable parents, who would have to agree to their child taking part in repeated behavioural and neuroimaging studies up into adulthood. Although the feasibility of such an endeavour seems hard to reach, results would be invaluable to the research community studying synaesthesia. It might be easier to accomplish in families with synaesthetic parents.

To reach a broader understanding of synaesthesia and its complexity, we need to meet the challenge of finding appropriate ways to measure it. Section 7.2.1 "Synaesthesia – A complex phenomenon" suggests that this is a difficult task. A combination of in-depth case studies of people with varying degrees of synaesthesia, including both qualitative and quantitative measures, and large group studies focusing on differences and similarities in cross-modal associations and within- and between-subject consistency are potential methods to approach this task. When enough studies covering a large spectrum of synaesthesia-related research have been conducted, it may be the right time for a meta-analysis, trying to relate findings of different disciplines about different aspects of synaesthesia to each other, with the aim of telling a comprehensive and conclusive story about the phenomenon.

7.5 Conclusion

This series of experiments was the first study on synaesthesia with a strong acoustic-phonetic background in the analysis of inducers and the first to introduce visual texture response displays. It therefore illustrates how synaesthesia research can be and is being broadened, despite having the limitations that all synaesthesia research encounters.

The vowel sound - colour associations of synaesthetes showed systematic acoustic-phonetic influences of vowel sounds on colour choices next to graphemic influences and their vowel sound - colour associations are highly consistent in contrast to those of non-synaesthetes. The vowel sound - grey shade associations of both synaesthetes and non-synaesthetes, on the contrary, revealed a common underlying mechanism of associating acoustic parameters such as intrinsic vowel pitch with a luminance scale and a similar consistency in vowel sound - grey shade associations across groups. In the voice description experiment it became apparent that primary associations with and descriptions of voices differ between groups: while synaesthetes use mostly synaesthetic descriptions (most of them referring to colours), controls use mainly personal or phonetic descriptions and phoneticians mostly phonetic descriptions. Quantitative analyses of their colour and texture associations and their consistency scores in a retest, however, highlighted once again shared underlying mechanisms across participant groups. In the voice identification task it was tested whether voice-induced and other synaesthetes and controls perform differently from each other. The puzzling result showed that controls tend to outperform synaesthetes in identifying a voice in modal voice quality, but voice-induced synaesthetes are the only group performing significantly above chance when asked to identify a whispering voice.

All in all, my studies in under-researched areas of synaesthesia show that the traditional definition of synaesthesia only holds true under certain circumstances and with certain manifestations of synaesthesia. However, it is much more common to find the phenomenon equipped with a colourful multifacetedness which is difficult to break into analysable pieces

– and perhaps cannot and should not be cut into bite-sized pieces for scientific purposes that try to define a phenomenon based on its parts rather than as a whole.

A Southern British English vowel phonemes and corresponding graphemes

Symbols for the phonemes of Southern British English, and examples of words in which these phonemes occur (courtesy of Prof. Francis Nolan).

Symbol	*Example*	*Alternative spellings*
/iː/	bead	**ee** (beet), **e** (be), **ie** (thief), **ei** (receive), **e.e** (delete), **ey** (key), **y** (pity)
/ɪ/	bid	**e** (decided), **o** (women), **a** (village)
/e/	bed	**ea** (head), **a** (any)
/æ/	bad	**ai** (plaid)
/ɑː/	bard	**al** (balm), **a** (father), **au** (aunt), **a** followed by /s/, /θ/, /f/ (pass, path, graph), **er** (clerk)
/ɒ/	body	**ou** (cough), **au** (sausage), **a** after /w/ (watch)
/ɔː/	bawdy	**ou** (bought), **au** (sauce), **al** (talk), **or** (for), **our** (four), **ore** (more), **oor** (door), **a(r)** after /w/ (water, war)
/ʊ/	book	**u** (full), **oul** (should), **o** (woman)
/uː/	boot	**ou** (soup), **u** (flute), **o** (do)
/ʌ/	bud	**ou** (trouble), **oo** (flood), **o** (son), **oe** (does)
/ɜː/	bird	**er** (herd), **ear** (heard), **ur** (fur), **or** (word), **our** (journey)
/ə/	banana	**e** (ageless), **o** (objection), **u** (campus), **er** (better), **ou** (famous), **ure** (nature), etc
/eɪ/	bay	**ai** (wait), **ey** (whey), **ei(gh)** (sleigh), **a.e** (hate), **ea** (great)
/aɪ/	buy	**y** (sky), **i(gh)** (sigh), **i.e** (time)
/aʊ/	bout	**ow** (how)
/ɔɪ/	boy	**oi** (groin)
/əʊ/	boat	**o** (no), **oe** (foe), **o.e** (note), **ou** (soul), **ow** (know)
/ɪə/	beer	**ear** (ear), **ere** (here), **ea** (idea), **eir** (weird), **ier** (tier)
/eə/	bear	**are** (care), **air** (hair)
/ʊə/	poor	**ure** (sure), **our** (tour), **ur** (alluring)

B Vowel sounds in different English accents

The vowel phonemic systems of certain accents of English (reproduced from Prof. Mike MacMahon, 2011).

		SCOTTISH	RP	N IRISH	N ENGLISH	GEN AM
1	SEAT	i	i	i	i	i
2	SIT	ɪ	ɪ	ɪ	ɪ	ɪ
3	WAY	e	eɪ	ɛɪ	eɪ	eɪ
4	WEIGH	e	eɪ	ɛɪ	eɪ *or* ɛɪ	eɪ
5	SEVER	ɛ	ɛ	ɛ	ɛ	ɛ
6 *	SEVEN	ɛ̈ *or* ɛ *or* ɪ	ɛ	ɛ	ɛ	ɛ
7 *	PAM	a	a	a	a	a
8 *	PALM	a *or* ɑ	ɑ	a	ɑ	ɑ
9 *	COT	ɔ	ɒ	ɒ	ɒ	ɑ
10 *	CAUGHT	ɔ	ɔ	ɔ	ɔ	ɔ
11	NOSE	o	əʊ	o	əʊ	oʊ
12	KNOWS	o	əʊ	o	əʊ *or* oʊ	oʊ
13 *	CUT	ʌ	ʌ	ʌ	ʌ *or* ə *or* ʊ	ʌ
14 *	FULL	u	ʊ	u	ʊ	ʊ
15	FOOL	u	u	u	u	u
16 *	BIRTH	ɪr	ɜ	ɜr	ɜ	ɜr
17 *	BERTH	ɛr	ɜ	ɜr	ɜ	ɜr
18	WORTH	ʌr	ɜ	ʌr *or* ɜr	ɜ	ɜr
19	HORSE	ɔr	ɔ	ɔr	ɔ	ɔr
20	HOARSE	or	ɔ	or	ɔ	ɔr
21	TIDE	ʌi *or* ɪi	aɪ	aɪ	aɪ	aɪ
22	TIED	ae	aɪ	aɪ	aɪ	aɪ
23	HOUSE	ʌu	aʊ *or* ɑʊ	ɑu	aʊ *or* ɑʊ	aʊ
24	BOY	ɔe	ɔɪ	ɔɪ	ɔɪ	ɔɪ
25	DEAR	ir	ɪə	ir	ɪə *or* iə	ɪr
26	HAIR	er *or* ɛr	ɛə	ɛr	ɛə	ɛr
27	POOR	ur	ʊə	ur	ʊə *or* uə	ʊr
		SCOTTISH	RP	N IRISH	N ENGLISH	GEN AM

C Short subject pool questionnaire

Synaesthetes experience a fascinating phenomenon called synaesthesia, a "joining together" of two senses that are normally experienced separately. There are many, many types of synaesthesia. For example, synaesthetes may perceive colours when they read numbers, letters, or words; the colours may be superimposed on the graphemes themselves or merely perceived in their "mind's eye." Synaesthetes may "hear colours," "see sounds," or feel touch sensations on their own bodies when they observe others being touched. Likewise, the days of the week or even numbers may appear in a specific location, form, or shape.

This questionnaire takes no more than five minutes. The info you give us about yourself will be treated completely confidentially.

Name:

Contact (Email and/or phone number):

Take a look at the following questions to help determine whether you are a synaesthete. If so, you may be able to help us in better understanding how this extraordinary phenomenon works.

Please answer YES, NO, or SOMETIMES to the following statements:

1. I experience colours when I look at written numbers.
2. I experience colours when I look at written letters.
3. I experience colours when I look at written words.
4. I experience colours when I hear people say numbers.
5. I experience colours when I hear people say letters.
6. I experience colours when I hear people say words.
7. I experience colours when I hear people's voices.
8. Each number/ letter/ word has a specific colour.
9. I associate numbers to colours.
10. I associate letters to colours.
11. I associate words to colours.
12. I experience touch on my own body when I look at someone else being touched (i.e., I feel touch sensations on my own body when I observe them on another person's body).

13. I experience touch on my own body when I look at something else being touched (i.e., I feel touch sensations on my own body when I observe them on objects).

14. I experience touch in response to body postures.

15. Do these experiences have specific locations (i.e., on your body, on words or objects, in front of your eyes) or not (i.e., you just "know" or they feel as though they are in your "mind's eye")? Please describe.

16. Do you think about ANY of the following being arranged in a specific pattern in space (i.e., in a line, a circle, etc.)?
    ALPHABET
    CALENDAR YEAR
    DAYS OF THE WEEK
    WEEKS
    TIME
    NUMBERS (NUMBER LINE)

17. Do you think about numbers/ letters/ words as having personalities or genders?

18. Do you experience colours in response to:
    SOUNDS
    MUSIC
    VOICES
    TOUCH

19. Please match the triggers on the left with the experiences on the right IF you ever experience the two together. For example, if you experience colours in response to numbers, then write "Numbers – Colour" below, OR draw a line between them. (We assume that everyone experiences e.g. music when listening to music, so please don't connect the same triggers with the same experiences.)

TRIGGERS	EXPERIENCES
Letters of the alphabet	Colours
Words	Shapes
Numbers	Touch
Days of the week	Taste
Months of the year	Smell
Pain	Sounds
Touch	Music
Body postures	Pain
Voices	
Music	
Sounds	
Colours	
Shapes	
Taste	
Smell	
Emotions	
Fingers	
Faces	
Places	
Other	

Further comments:

D Synaesthesia questionnaire (by Jamie Ward and Julia Simner)

Dr J. Ward, Dept. of Psychology, University of Sussex, Falmer, Brighton, BN1 9QH, UK.
Tel: +44 (0)1273 876598; Fax: +44 (0)1273 678058; E-mail: jamiew@sussex.ac.uk
Dr J. Simner, School of Psychology, Edinburgh University, 7 George Square, Edinburgh. EH8 9JZ, UK.
Tel: +44 (0)1331 650-3450; E-mail: j.simner@ed.ac.uk

Synaesthesia Research Programme:
Information for Participants

What is synaesthesia?

People with synaesthesia experience unusual sensations (e.g., of colour, of taste) when doing things that wouldn't usually trigger those sensations for non-synaesthetic people. In some cases, this means that a synaesthete might experience a sensation in one of the 5 senses (hearing, vision, taste, touch, smell) that is triggered by a different sense (e.g. sounds trigger tastes, smells trigger colours etc.).

What are the aims of the research?

The aim of the research is to understand the cognitive, developmental and biological basis of synaesthesia. This might also tell us more about ordinary perceptual experiences and its relationship to thinking, memory and language.

What is involved with taking part?

First of all, you will be asked to fill in a general questionnaire (enclosed) and to describe any synaesthetic experiences that you have in response to a list of letters and words. You do not have to answer all the questions if you feel uncomfortable about it. However, it is useful for our research to gain as complete a picture as possible and all information you give will be treated in confidence. Following this, we may contact you again (by either phone, e-mail or letter) to invite you to take part in further studies. These will involve basic tests of memory and perception. None of the tasks are harmful or stressful.

Will my data be kept confidential?

Your personal details (name, address, etc.) will not be passed on to anybody else outside of our research group without first gaining your written consent. A current list of people in our research group can be found on our website (www.syn.ucl.ac.uk). You will be referred to in our records and in any publications by your initials (or other code), in accordance with the data protection act.

How long will the research go on for and can I drop out?

The amount of times we may ask you to take part in an experiment is likely to differ from person to person. As a general rule of thumb, we may contact you a couple of times over the course of one year. You are under no obligation to take part, and you may refuse to take part for whatever reason and without giving any explanation.

Please fill in the following :

Name of participant : ______________________________

Address : ______________________________

____

Telephone number : ______________________________

E-mail : ______________________________

I have read the information above and I agree to take part in the study. I understand that I may withdraw at any point in the future.

Signed (by participant) : ______________ Date : ______________

SYNAESTHESIA QUESTIONNAIRE

Section A1. ABOUT YOURSELF

Initials: ________________________

Date of Birth:__________________

Sex: Male / Female

Are you left or right handed? LEFT RIGHT AMBIDEXTROUS

Is English your first language? YES NO (please state ____________)

Please indicate any other languages that you have learned to speak proficiently. At what age did you start to learn?

Language 1 ____
____
____

Language 2 ____
____
____

(If you learnt more please give details separately)

Does anyone in your family experience synaesthesia?

YES NO DON'T KNOW

If YES - please describe their relationship to you (father, cousin etc.) and the effects that they experience?

____

____

Section A2. TYPES OF SYNAESTHESIA

Please match the triggers on the left with experiences on the right. For instance, if you experience colours in response to numbers then draw a line in between 'numbers' (left) and 'colours' (on right), and so on. There is no need to draw lines between the same things (e.g. colours - colours) as this is assumed to be true of everyone.

TRIGGERS	**EXPERIENCES**
Letters of alphabet	Colours
English words	Shapes
Foreign words	Tastes
Peoples names	Smells
Numbers	Noises
Days of week	Music
Months of year	Pain
Voices	Touch
Pains	
Touch	
Body postures	
Music (instrumental)	
Noises	
Smells	
Tastes	
Colours	
Shapes	
Emotions	
Fingers	
Faces	
Places	

Other (e.g., musical notation, chess pieces, cars, animals, plants); please specify: ________

Do these experiences have specific locations (e.g., they feel as if they are on your body, on words or objects in the environment, in front of your eyes) or not (e.g. they feel as if they are in 'your minds eye')? Please describe.

____

____

____

Section A3. COLOURED SPEECH & COLOURED READING

I experience colours when I listen to someone speaking... YES NO

I experience colours when I look at written material (reading silently)... YES NO

(If you answered NO to both questions then please move on to Section A4. If you answered YES to either question then please continue.)

Is your synaesthesia more intense when hearing speech or reading silently?

Hearing more intense Reading more intense Equally intense Not applicable

Is your synaesthesia more automatic when hearing speech or reading silently?

Hearing more automatic Reading more automatic Equally automatic Not applicable

When looking at written material, i.e. silent reading, which of the following statements applies to you... (please tick)

- I see a coloured copy of the letters in my mind's eye and black and white on the page
- I see a block of colour (but not letters) in my mind's eye and black and white on the page
- I see colour that appears to be coming out from underneath the (black and white) text on the page
- I see colour that appears to float above the (black and white) text on the page
- I have a strong sense of 'knowing' the colour of the letters but I do not 'see' them in any of the ways described above (please describe if you can)
- I don't experience any colours
- Other. Please explain

____

When listening to someone speaking which of the following statements applies to you... (please tick)

- I see a coloured copy of the letters in my mind's eye (like coloured subtitles)
- I see a block of colour (but not letters) in my mind's eye
- I see colour that appears to be located in the space outside of my body
- I see colour that appears to come from the speaker's mouth
- I have a strong sense of 'knowing' the colour but I do not 'see' them in any of the ways described above (please describe if you can)
- I don't experience any colours
- Other. Please explain

____

Section A4. ABILITIES and DIFFICULTIES

Do you find that you often get left and right confused?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

Do you have problems navigating or finding your way around places?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

Do you have (or have you ever had) problems with understanding numbers and/or calculation?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

Do you have (or have you ever had) problems with reading and spelling?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

If YES, then were you ever formally assessed for dyslexia? ______________________________

Do you have a problem recognizing or remembering familiar faces, including people you see often, close friends or relatives?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

When you are given a telephone number to remember, which of the following best describes the way you would remember it?

- ☐ Say it to myself over and over in my head
- ☐ Take a 'mental picture' of what it looks like
- ☐ Other (please explain) ____

Can you read music? YES (fluently) YES (nonfluent) NO

If yes, at what age did you start learning? ______________________

Do you play any instruments? If yes, to what level? At what age did you start?

Instrument 1____

____

Instrument 2____

____

(If you learnt more please give details separately)

Section B1. THE ALPHABET

↓ In the column marked '0-9', please indicate how intense your synaesthetic experience is on a 0 to 9 scale (where 0 = nothing at all, and 9 = a very stong experience). This might relate to how confident you are that you are experiencing something. You can use the entire range of numbers (if some sensations are stronger than others) or repeat the same numbers (if the intensity doesn't vary much).
↓ In the column marked '**Description**', we would like you to describe succinctly your synaesthetic experience (e.g. deep blue, fried onion). If you don't experience anything at all then just put a dash in the column.

	0-9	**Description**
a		
b		
c		
d		
e		
f		
g		
h		
i		
j		
k		
l		
m		

	0-9	**Description**
n		
o		
p		
q		
r		
s		
t		
u		
v		
w		
x		
y		
z		

Do capital letters (ABC…) have the same synaesthetic experience as lower case letters (abc…)? (please circle)?

TRUE for all letters TRUE for some letters FALSE

Is the synaesthetic experience more intense for…? (please circle)

Capitals Lowercase No difference Varies from letter to letter

Do you think about the alphabet being arranged in a specific pattern in space (e.g. in a line, or circle)?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

If you answered 'agree' or 'strongly agree' then please try to draw this in the space below

Section B2. NUMBERS

↓ In the column marked '0-9', please indicate how intense your synaesthetic experience is on a 0 to 9 scale (where 0 = nothing at all, and 9 = a very stong experience). This might relate to how confident you are that you are experiencing something. You can use the entire range of numbers (if some sensations are stronger than others) or repeat the same numbers (if the intensity doesn't vary much).
↓ In the column marked '**Description**', we would like you to describe succinctly your synaesthetic experience (e.g. deep blue, fried onion). If you don't experience anything at all then just put a dash in the column.

	0-9	Description
0		
1		
2		
3		
4		
5		
6		
7		
8		
9		

	0-9	Description (same as 0,1,2,3...?)
zero		
one		
two		
three		
four		
five		
six		
seven		
eight		
nine		

Do you think about the numbers being arranged in a specific pattern in space (e.g. in a line, or circle)?

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

If you answered 'agree' or 'strongly agree' then please try to draw this in the space below and answer the additional questions (if not, then move to the next section).

Does your number line appear to be in the space outside of your body or your mind's eye? (e.g. '1' is about 10 inches from my left shoulder or the number line is not defined relative to your body)____

Does your number line appear in a fixed position or do you 'zoom' in on certain parts of the number line when thinking of specific numbers? FIXED ZOOM IN

Do you actively use your number line when carrying out calculation tasks? YES NO

Section B3. DAYS AND MONTHS

↓ In the column marked '0-9', please indicate how intense your synaesthetic experience is on a 0 to 9 scale (where 0 = nothing at all, and 9 = a very stong experience). This might relate to how confident you are that you are experiencing something. You can use the entire range of numbers (if some sensations are stronger than others) or repeat the same numbers (if the intensity doesn't vary much).
↓ In the column marked '**Description**', we would like you to describe succinctly your synaesthetic experience (e.g. deep blue, fried onion). If you don't experience anything at all then just put a dash in the column.

	0-9	Description
Monday		
Tuesday		
Wednesday		
Thursday		
Friday		
Saturday		
Sunday		
January		
February		
March		

	0-9	Description
April		
May		
June		
July		
August		
September		
October		
November		
December		

Do you think about the DAYS being arranged in a specific pattern in space (e.g. in a line, or circle?)

Strongly disagree Disagree Neither agree nor disagree Agree Strongly agree

If you answered 'agree' or 'strongly agree' then please try to draw this in the space below and answer the additional questions (if not, then move to the next page).

Do your days appear to be in the space outside of your body or your mind's eye? (e.g. 'today' is about 10 inches from my left shoulder) or elsewhere?____________________________

Do your days appear in a fixed position or can you 'zoom' in on certain parts when thinking of specific days? FIXED ZOOM IN

Do you actively picture your days when planning your time? YES NO

Section B3. DAYS AND MONTHS (continued)

Do you think about the MONTHS being arranged in a specific pattern in space (e.g. in a line, or circle?)

Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly agree

If you answered 'agree' or 'strongly agree' then please try to draw this in the space below and answer the additional questions (if not, then move to the final question in this section).

Do your months appear to be in the space outside of your body or your mind's eye? (e.g. 'today' is about 10 inches from my left shoulder or are they not defined relative to your body)?

____

Do your months appear in a fixed position or can you 'zoom' in on certain parts when thinking of specific days? FIXED ZOOM IN

Do you actively picture your months when planning your time? YES NO

Do you think about anything else not mentioned earlier (e.g., temperature, show sizes, years, weight, height, salaries) as being arranged in a specific pattern in space?

Strongly disagree | Disagree | Neither agree nor disagree | Agree | Strongly agree

If you answered 'agree' or 'strongly agree' then please try to draw this in the space below

Section B4. WORDS

↓ In the column marked '0-9', please indicate how intense your synaesthetic experience is on a 0 to 9 scale (where 0 = nothing at all, and 9 = a very stong experience). This might relate to how confident you are that you are experiencing something. You can use the entire range of numbers (if some sensations are stronger than others) or repeat the same numbers (if the intensity doesn't vary much).
↓ In the column marked '**Description**', we would like you to describe succinctly your synaesthetic experience (e.g. deep blue, fried onion). If you don't experience anything at all then just put a dash in the column.
NOTE: Some of the words are deliberately made-up!

	0-9	Description		0-9	Description
ked			cud		
bem			wholemeal		
shid			prince		
boak			peg		
snite			hair		
hance			crow		
dringe			graze		
squate			gnome		
doop			pet		
nar			cool		
serial			cinder		
died			hint		
bat			wrung		
symbols			tab		
pod			lute		
cent			pat		
yew			sent		
gender			bruise		
sheik			Neil		
wrap			sloth		
hour			sells		
bind			closer		
wreck			cereal		
link			choke		
pint			pier		
kneel			plumb		
hertz			whirlpool		
loot			dyed		
knights			peer		
brakes			psalm		
cells			froth		
pot			ewe		
bind			tasty		
tub			phrase		
bark			rap		

Section B4. WORDS (continued)

	0-9	Description		0-9	Description
beg			shed		
knot			shake		
hurts			breaks		
hay			circle		
rye			fir		
brews			tank		
grumble			mint		
kelp			fur		
pig			brow		
wharf			pub		
hey			choir		
cymbals			foot		
vale			thorny		
cod			prey		
pit			kit		
nights			rung		
wry			covert		
veil					

Section B5. PERSONALITIES & GENDERS

Do you think about letters as having personalities or genders? (please circle)

Genders? YES NO Personalities? YES NO

Do you think about numbers as having personalities or genders? (please circle)

Genders? YES NO Personalities? YES NO

(If you answered NO to all four questions then please move on to Section B6. If you answered YES to any question then please continue.)

↓ In the column marked 'm/f?, we would like you write the gender of the letter, as either m (= male) or f (= female) or leave a dash if you don't feel strongly either way.
↓ In the column marked '0-9', please indicate confident you feel about each letter's letter on a 0 to 9 scale (where 0 = no feelings, and 9 = a very stong feeling). You can use the entire range of numbers (if some feelings are stronger than others) or repeat the same numbers (if the intensity doesn't vary much).
↓ In the column marked 'Description', we would like you to describe succinctly and to the best of your ability the personality of any of the letters below (e.g. bossy). If you don't experience anything at all then just put a dash in the column.

	m/f	0-9	personality	0-9
a				
b				
c				
d				
e				
f				
g				
h				
i				
j				
k				
l				
m				

	m/f	0-9	personality	0-9
n				
o				
p				
q				
r				
s				
t				
u				
v				
w				
x				
y				
z				

	m/f	0-9	personality	0-9
0				
1				
2				
3				
4				

	m/f	0-9	personality	0-9
5				
6				
7				
8				
9				

Section B6. MISCELLANEOUS

Do you experience touch sensations on your own body when you observe them on another person's body?

Strongly disagree   Disagree   Neither agree nor disagree   Agree   Strongly agree

Do punctuation symbols give synaesthetic experiences? YES NO
If YES, please write your associations to the following symbols...

	0-9	Description		0-9	Description
!			@		
?			£		
.			$		
,			%		
:			&		
;			*		
"			#		
'			+		
(			-		
)			x		
{			÷		
}			=		
/			→		

MANY THANKS FOR YOUR TIME!

(Please return to the address on the front sheet)

E Formant values of vowel stimuli

	F1	F2	F3
C1	225	2290	3400
C1.5	270	2280	3350
C2	360	2360	3130
C2.5	482	2185	2923
C3	630	2030	2650
C3.5	692	1830	2575
C4	790	1665	2440
C4.5	705	1170	2570
C5	590	960	2660
C5.5	480	750	2580
C6	380	800	2550
C6.5	430	820	2470
C7	351	875	2476
C7.5	240	600	2270
C8	220	800	2320
C8.5	229	1672	2970

F Synaesthesia questionnaire used in voice description experiment

The following information was given to the participant groups of phoneticians and controls to test whether they are synaesthetes or not:

Synaesthesia is a neurological condition with cross-modal sensations. Here we examine a rare kind of synaesthesia where people's voices (i.e. "human sound") trigger colour and texture perceptions. These synaesthetes actually see or associate colours/textures with voices.

Does anything from the list below trigger spontaneous, involuntary and persistent colour and/or texture perceptions or associations for you? Please select and specify if necessary. E.g. if you have colour perceptions for English words but not French words (although you speak both languages), tick "Words" and add "English words only".

Check any that apply

* Letters
* Music
* Words

* Days of week
* Months of Year
* Pains
* Touch
* Body postures
* Noises
* Smells
* Tastes
* Colours
* Shapes
* Emotions
* Fingers
* Faces
* Places
* Names
* Numbers
* Voices
* Other:

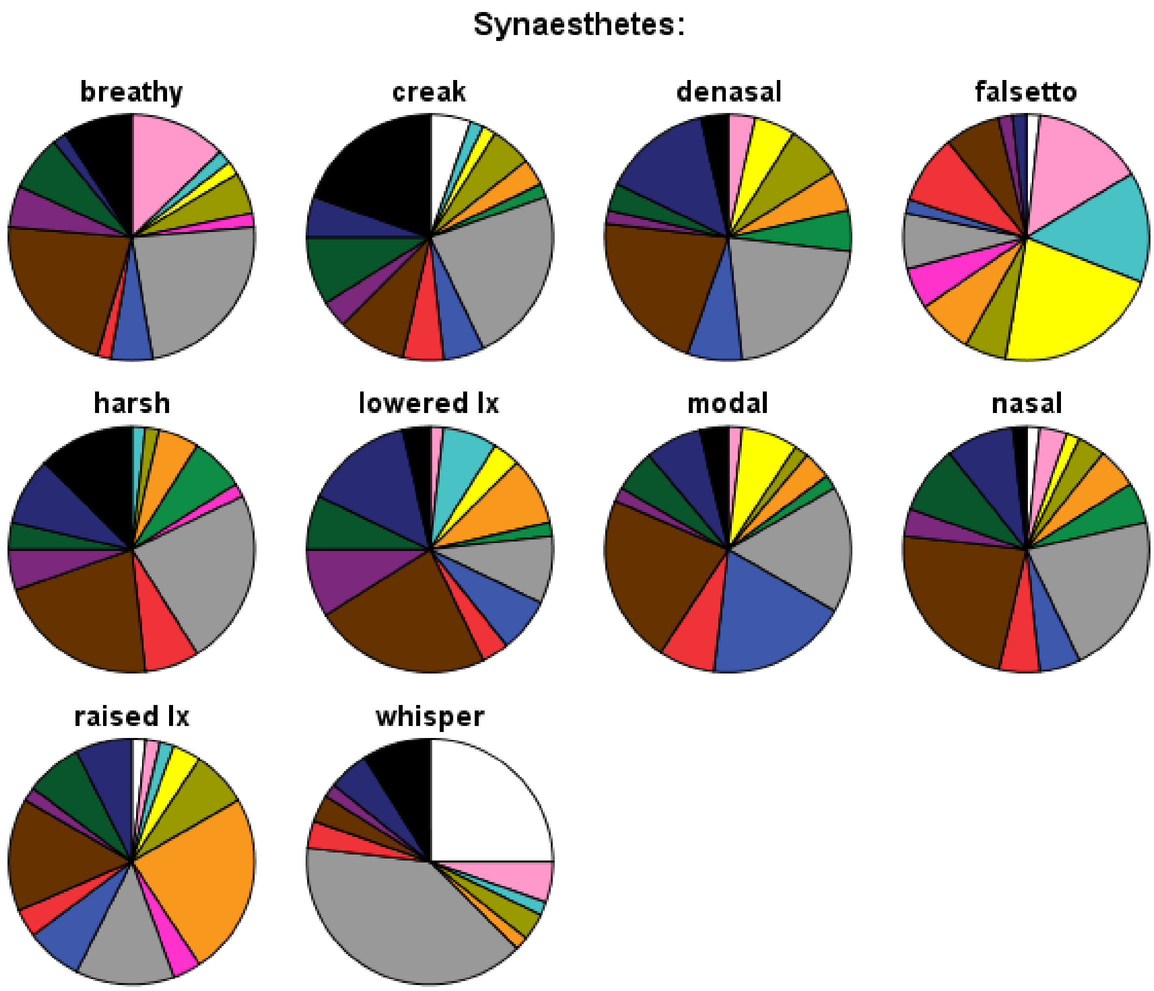
Do any of the triggers listed above - or indeed anything else - lead to other sensations than colours or textures? Please specify.

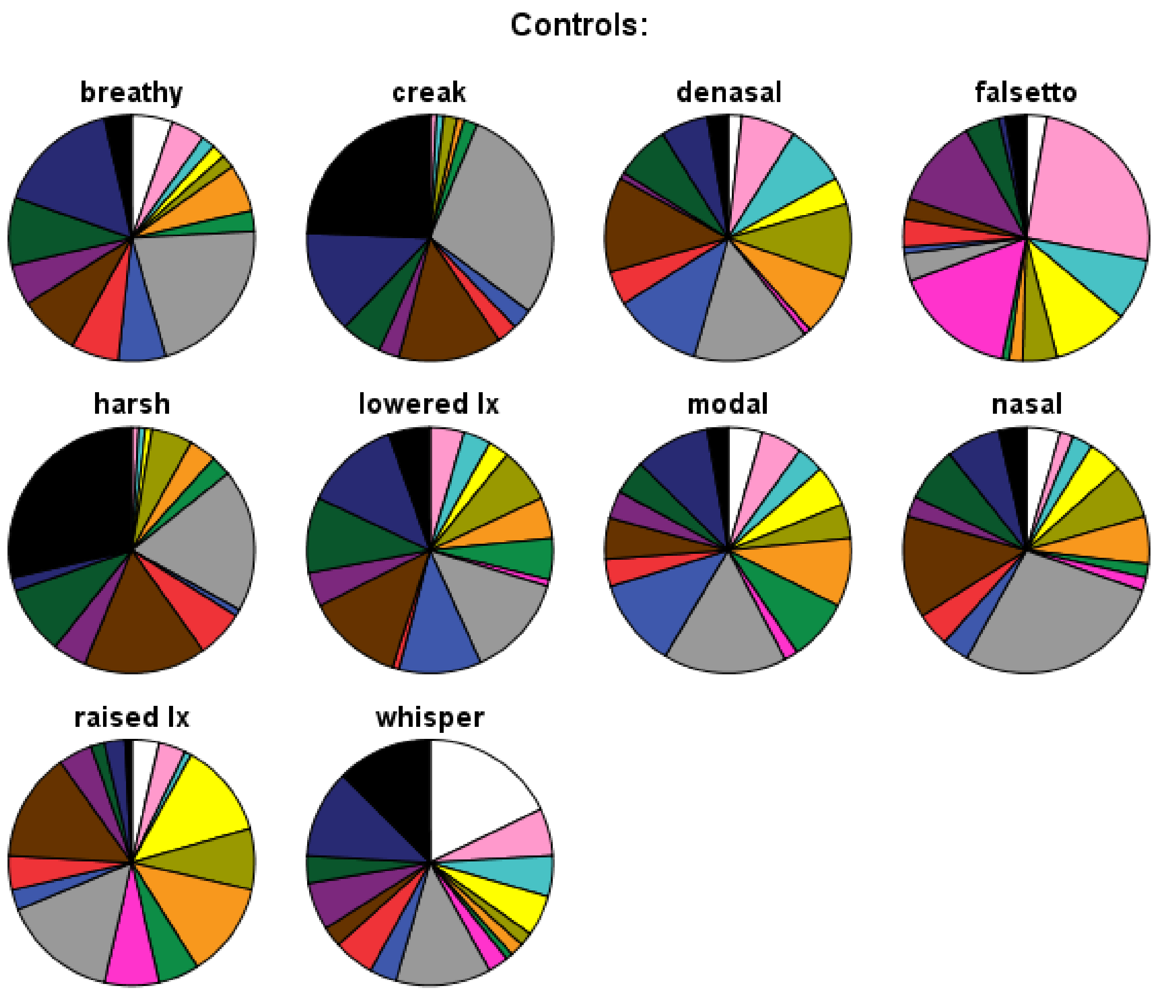
The same list was given to synaesthetes with the following introductory text instead:

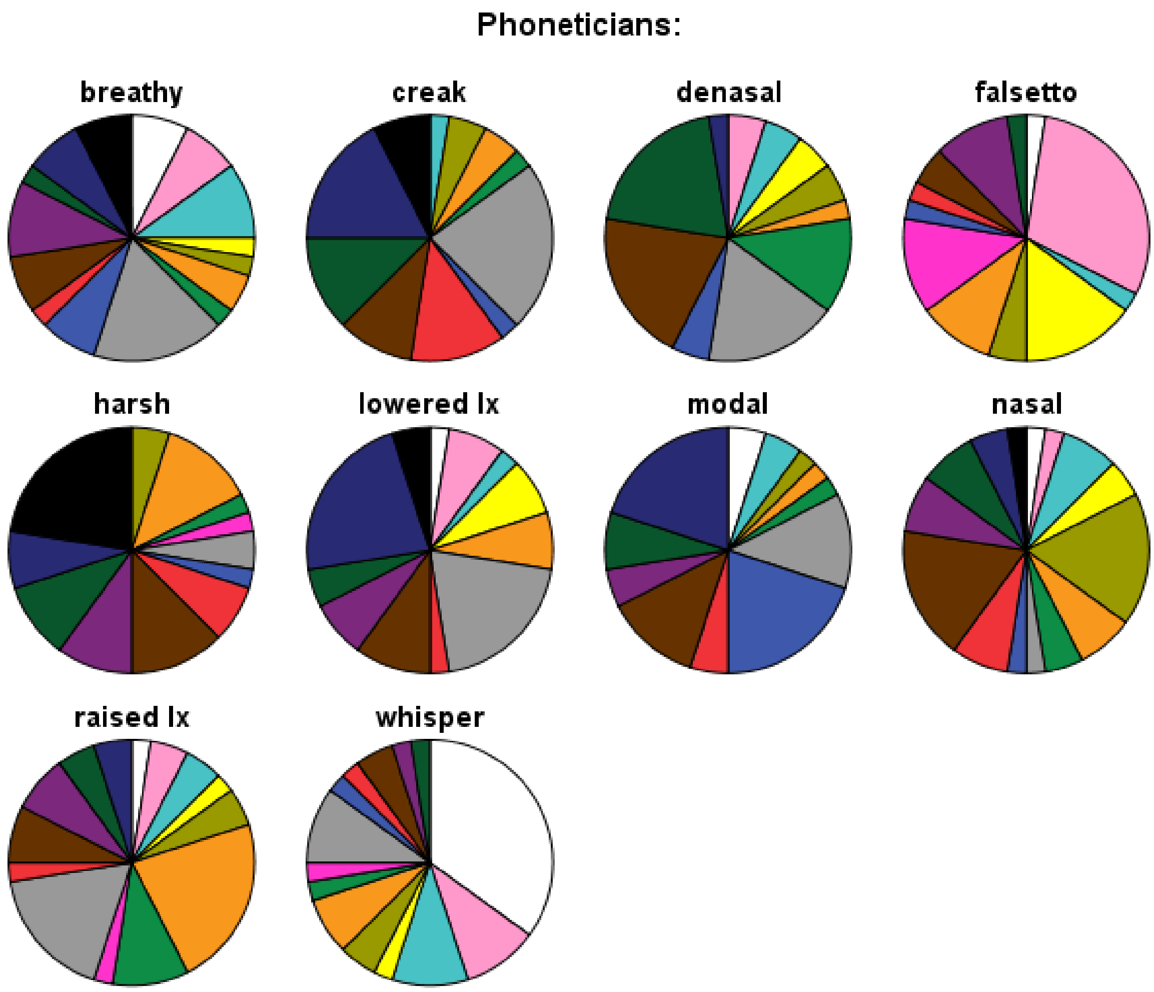
Does anything else than people's voices trigger colour and/or texture perceptions? Please select and specify if necessary (for example tick 'words' and add "French words only" if applicable).

G Colour associations per participant group per voice quality in the voice description study

The colours used in these pie charts are approximations of the colours used in the experiment.




Controls:
breathy
creak
denasal
falsetto
harsh
lowered lx
modal
nasal
raised lx
whisper



H Cinderella

Once upon a time there was a girl called Cinderella. But everyone called her Cinders. Cinders lived with her mother and two stepsisters called Lily and Rosa. Lily and Rosa were very unfriendly and they were lazy girls. They spent all their time buying new clothes and going to parties. Poor Cinders had to wear all their old hand-me-downs! And she had to do the cleaning! One day, a royal messenger came to announce a ball. The ball would be held at the Royal Palace, in honour of the Queen's only son, Prince William. Lily and Rosa thought this was divine. Prince William was gorgeous, and he was looking for a bride! They dreamed of wedding bells! When the evening of the ball arrived, Cinders had to help her sisters get ready. They were in a bad mood. They'd wanted to buy some new gowns, but their mother said that they had enough gowns. So they started shouting at Cinders. "Find my jewels!" yelled one. "Find my hat!" howled the other. They wanted hairbrushes, hairpins and hair spray.

I The north wind and the sun

The North Wind and the Sun were arguing one day about which of them was stronger, when a traveller came along wrapped up in an overcoat. They agreed that the one who could make the traveller take his coat off would be considered stronger than the other one. Then the North Wind blew as hard as he could, but the harder he blew, the tighter the traveller wrapped his coat around him; and at last the North Wind gave up trying. Then the Sun began to shine hot, and right away the traveler took his coat off. And so the North Wind had to admit that the Sun was stronger than he was.

J Synaesthesia questionnaire for voice identification experiment

Questions to help interpret the data of the voice identification experiment:

(Please delete the wrong answers when a choice is given. Otherwise just type away.)

1. Did you have any synaesthetic perceptions with the voice stimuli, i.e. the story at the beginning and the voice line-up at the end of the experiment?

[yes/ no]

2. Did you take part in both the whisper and the normal voice experiment?

[yes/ no]

3. Were the synaesthetic perceptions present for some or all of the whispered and/or normal voices?

[whisper: for all/ some/ none of the voices/ didn't do that test]

[normal: for all/ some/ none of the voices/ didn't do that test]

4. If you had synaesthetic perceptions with both the whispered and the normal voices, were they different or similar in quality and intensity? Can you briefly describe the perceptions, especially the differences (if present) between those induced by whisper and normal voicing?

5. Do you think your synaesthesia had any effect on memorising and identifying the voices? If so, did it help or hinder?

[whisper: no effect/ helped/ hindered / didn't do that test]

[normal: no effect/ helped/ hindered / didn't do that test]

6. Did you have synaesthetic perceptions during the facial expression task and balloon task as well?

[yes both/ yes face/ yes balloon/ no]

7. What is your occupation and highest level of education?

8. Have you ever regularly played a musical instrument? Or have you ever had singing lessons? To what level did your training go?

Thank you very much for your help again!

K Glossary

Words printed in bold in a glossary entry have their own entry.

ABX Experimental design for a perception experiment, where stimuli A and B are presented followed by stimulus X. The participant then has to judge whether X is more similar to A or B.

allophone An audibly distinct variant of a **phoneme**, e.g. the 'clear' [l] of *leaf* and the 'dark' [ɫ] of *feel*.

ANCOVA (analysis of covariance) A statistical procedure that uses the F-ratio to test the overall fit of a linear model controlling for the effect that one or more covariates have on the outcome/dependent variable. In experimental research, this linear model tends to be defined in terms of group means and the resulting **ANOVA** is therefore an overall test of whether group means differ, after the variance in the outcome/dependent variable explained by any covariates has been removed. (Field, 2009)

ANOVA (analysis of variance) A statistical procedure that uses the F-ratio to test the overall fit of a linear model. In experimental research, this linear model tends to be defined in terms of group means and the resulting ANOVA is therefore an overall test of whether group means differ. (Field, 2009)

articulators Speech organs which are used to produce speech sounds, such as the tongue, the lips, the soft and hard palate, etc.

associator A **synaesthete** who associates their **concurrents** with their **inducers** by simply thinking of the concurrent or seeing it in front of their mind's eye, as opposed to a **projector**.

brightness Attribute of a visual sensation according to which an area appears to emit more or less light. (Fairchild, 1998)

cardinal vowels Reference vowels referring to points on the periphery of the vowel quadrilateral as shown in Figure 4 on page 42. They are independent of any language as they are established on theoretical grounds and represent possibilities of vowel production with the human vocal tract.

Chi-square-test (χ^2 test) Although this term can apply to any test statistic having a χ^2 distribution, it generally refers to Pearson's χ^2 test of the independence of two categorical variables. Essentially, it tests whether two categorical variables forming a contingency table are associated. (Field, 2009)

chroma The **saturation** or intensity of a colour (or purity of a **hue** according to Munsell), for example pale pink has lower chroma than bright pink.

chromaticity In lights, a measure of the combination of **hue** and **saturation** in a colour; the quality of colour, independent of its brightness.

CIELUV (CIE 1976 (L* u* v*)) A perceptually uniform colour space used for self-luminous colours such as those displayed on a computer screen. The three co-ordinates are L* for **lightness** or **luminance**, u* which approximates the redness-greenness axis, and v* which approximates the yellowness-blueness axis. (Fairchild, 1998)

colourfulness Attribute of a visual sensation according to which the perceived colour of an area appears to be more or less **chromatic**. (Fairchild, 1998)

colorimetry The science and technology used to quantify human colour perception and to describe it physically.

concurrent The synaesthetic perception which occurs in addition to the perception of the inducing stimulus, e.g. colour in grapheme-colour **synaesthesia**. The modality in which synaesthetic experiences take place. (Ward & Simner, 2003)

dB (decibel) A logarithmic unit of intensity of a sound, usually a ratio of the measured sound to a reference sound (where the reference sound marks the common threshold of human hearing, usually sound pressure level (SPL)).

diphthong A vowel whose quality changes perceptibly in one direction within a single syllable: e.g. [aʊ] in house, whose articulation changes from relatively open to relatively close and back. (Matthews, 1997)

DTI (diffusion tensor imaging) A magnetic resonance imaging (MRI) technique that enables the measurement of the restricted diffusion of water in tissue in order to produce neural tract images. It also provides useful structural information of tissue, e.g. about connectivity in the brain.

EEG (electroencephalography) The recording of the brain's electrical activity (ionic current flows within the neurons of the brain), measured via electrodes attached to the scalp, to show brain activity with a relatively high temporal resolution, while having relatively poor spatial resolution.

f0 (fundamental frequency) The lowest component frequency of speech; it stands for the rate of vibration of the vocal folds (usually measured in Hz). It is the physical correlate of the perception that is referred to as 'pitch'. (Ogden, 2009)

face overshadowing effect The presentation of a face during encoding of a voice supposedly impairs later voice identification, most likely because priority is given to processing visual information.

FFT (fast Fourier transform) In speech analysis, the FFT is used as a narrowband amplitude spectrum, offering an analysis of the distribution of acoustic energy. When analysing a vowel sound, **harmonics** (multiples of the fundamental frequency) can be detected.

Fisher's exact test A test which computes the exact probability of a statistic. It was originally designed to overcome the problem that, with small samples, the sampling distribution of the **chi-square** statistic deviates substantially from a chi-square distribution. (Field, 2009)

fMRI (functional magnetic resonance imaging) A procedure measuring brain activity by detecting changes in oxygen levels in the different brain areas. These oxygen levels vary with brain activity, as the blood flow changes accordingly.

foil In the context of **forensic phonetics**, a foil is a person other than the suspect who is used in a **voice parade**. A foil is not involved in the crime but their sex is the same as that of the suspect and their voice features are similar with regards to age, social and educational background, nationality, accent etc.

forensic phonetics The application of the knowledge, theories and methods of general phonetics to practical tasks that arise in the context of police work or the presentation of evidence in court. The most central aspect of forensic phonetics is speaker identification, including voice comparison and voice profiling.

formants Acoustic features of a vowel; spectral peaks (i.e. higher-intensity regions) of the sound spectrum, corresponding to resonant frequencies of the **vocal tract** during vowel production. For any given vowel the first formant, F1, is the lowest-frequency peak. F1 and F2 are the main formants that define the vowel quality, i.e. that enable us to differentiate vowels such as/i/ and /a/ perceptually. Normally measured in Hertz (**Hz**).

H/N (harmonic-to-noise ratio) Describes how much aperiodic noise (such as frication) can be found in a signal. The more aperiodic noise is present, the lower the ratio is; whisper, for example, has a low ratio because there is no voicing, or periodicity, in the signal.

harmonics In a periodic sound, such as a vowel, a harmonic (H) is a component frequency of the signal, namely a whole-number multiple of the fundamental frequency **(f0)**.

With an f0 of 100 Hz, H1 (the first harmonic) is also 100 Hz, H2 is 200 Hz, H3 is 300 Hz, etc. (An alternative numbering system gives H1 as 2*f0, H2 as 3*f0, etc.)

hue The identification or name of a colour. (Anderson Feisner, 2001)

Hz (Hertz) A unit of frequency, defined as the number of cycles per second of a periodic phenomenon, e.g. a sound wave.

inducer The perceptual or conceptual stimulus triggering a synaesthetic perception, in grapheme-colour **synaesthesia** for example the graphemes.

intonation Changes in fundamental frequency **(f0)** during the course of an utterance (which do not reflect lexical tone). (Borden et al., 2003)

LTF (long-term formant distribution) A method used to determine average **formant** values of a speaker for vocalic parts of their speech.

LCD (liquid crystal display) A flat panel display, electronic visual display, or video display that uses the light modulating properties of liquid crystals.

lightness The property of a colour that pertains to the subjective **brightness** perception of a colour along a lightness-darkness axis (called 'value' in the **Munsell colour system**).

linear mixed effect modelling A statistical model that describes a relationship between a response variable (e.g. a colour association) and predictors. It takes into account both fixed effects (e.g. stimulus acoustics, participant groups) and random effects (e.g. different individual participants, who act in different ways).

logistic regression A version of multiple regression in which the outcome/dependent variable is a categorical variable. If the categorical variable has exactly two categories the analysis is called binary logistic regression, and when the outcome has more than two categories, it is called multinomial logistic regression. (Field, 2009)

LPC (linear predictive coding) A tool used in speech processing for representing the spectral envelope of a speech signal, i.e. the approximate energy distribution across the frequency range, usually with a wideband amplitude spectrum. When analysing a vowel sound, **formants** are detected, as the LPC algorithm 'picks the peaks'.

luminance A measure of **lightness** of a colour. Light grey or bright pink have a high luminance, dark grey or dark brown have a low luminance. Usually measured in cd/m^2 (candela per square metre).

MANOVA (multivariate analysis of variance) A family of tests that extend the basic **ANOVA** to situations in which more than one outcome/dependent variable has been measured. (Field, 2009)

Mann-Whitney U test A non-parametric test that looks for differences between two independent samples. That is, it tests whether the populations from which two samples are drawn have the same location. (It is a non-parametric equivalent to an independent t-test.) (Field, 2009)

meta-analysis Analysis of data from a number of independent studies of the same subject (published or unpublished), especially in order to determine overall trends and significance. A meta-analysis contrasts and combines results from different studies, with the aim of finding commonalities, disagreements and relationships between the studies that may lead to new findings and/or greater statistical power.

MEG (magnetoencephalography) A technique for mapping brain activity by recording magnetic fields produced by electrical currents occurring naturally in the brain, to show brain activity with a relatively high temporal resolution, while having relatively poor spatial resolution.

mirror touch A type of **synaesthesia** in which, whenever the person sees someone else being touched, the synaesthete feels the sensation on their own body.

mnemonic A memory aid, or a learning technique that aids information retention.

monophthong Vowel sound of unchanging vowel quality and with steady **formants** throughout, for example in the words *mask* (/ɑ/) or *feet* (/i/).

Munsell colour system Albert Munsell (1858-1918) was the first to separate **hue**, value (**lightness**), and **chroma** into perceptually uniform and independent dimensions, and was the first to systematically illustrate the colours in three-dimensional space in the early 20th century.

OLP (ordinal linguistic personification) A type of **synaesthesia** in which ordered sequences, such as ordinal numbers, days, months and letters, are associated with personalities.

opponent colours The **hues** at the opposite ends of colour scales, such as red-green and blue-yellow. Under normal circumstances, there is no hue that could be described as a mixture of opponent hues, e.g. a greenish red.

PET (positron emission tomography) An imaging technique using computer analysis of gamma-rays produced after introduction of a radioactive isotope (Borden et al., 2003). It is used to image tissue concentration by increasing the metabolic activity in the tissue (often in the brain).

phone A speech segment representing the smallest discrete segment of sound in a stream of speech (Oxford Dictionaries, 2010), or a speech sound which is identified as the realization of a single **phoneme** (Matthews, 1997).

phoneme The smallest unit of speech that gives meaningful contrasts between two words. We can distinguish between *tree* and *free*, so /t/ and /f/ are phonemes in English (and many other languages, although they are language specific).

phonetician A speech scientist studying the sounds of human speech in the fields of physiological production, auditory perception and acoustic properties.

primacy effect A cognitive bias that suggests that information presented first in a list is recalled better than information presented later on. For example, a person who reads a sufficiently long list of words is more likely to remember words toward the beginning than words in the middle.

projector A **synaesthete** who sees their **concurrents** placed somewhere outside their body, as opposed to an **associator**. For example, for a projector, the colour of a letter might be projected onto the letter itself, or it might be presented as a "screen" in front of their eyes.

quartertones In music, a pitch aurally and logarithmically halfway between the usual notes of a chromatic scale (i.e. semitones).

RGB The RGB (red, green, blue) colour space is an additive colour model of light defined by the three **chromaticities** of red, green and blue. It is used as the basis for displaying colours on a screen.

RP (Received Pronunciation) The form of British pronunciation that many educated people in Britain use, and that is thought of as the standard form; spoken mainly in the south of England.

RSC (retrosplenial cortex) An area of the cingulate cortex (in the limbic lobe) of the brain, which is sometimes associated with recall of episodic information, emotion processing, memory, relating objects to their contexts, switching from the external world to internal mentation, and is generally heavily interconnected with other parts of the posterior cingulate region.

saturation The **colourfulness** of an area judged in proportion to its **brightness**. (Fairchild, 1998)

semantic differentials A type of rating scale designed to measure the connotative meaning of objects, events, and concepts (here the sound of a voice). Opposite ends of the scale are meant to have opposite meanings, for example light-dark.

SSBE (Standard Southern British English) The modern equivalent of what has been called 'Received Pronunciation' ('**RP**') where 'Standard' should not be taken as implying a value judgment of 'correctness'. It is an accent of the south east of England which operates as a prestige norm there and (to varying degrees) in other parts of the British Isles and beyond. (International Phonetic Association, 1999)

schwa The mid-central vowel that can usually be found in unaccented syllables of English such as in the second syllable of *father*, transcribed as /ə/. (Matthews, 1997)

SD (standard deviation) Shows how much variation or dispersion exists from the mean of a set of data points, with a low value showing little variation and a high value showing a lot of variation.

SE (standard error) The standard deviation (**SD**) of the sampling distribution of a statistic, e.g. of the mean. It tells us how much variability there is in this mean across samples from the same population. Large values, therefore, indicate that a statistic from a given sample may not be an accurate reflection of the population from which the sample came. (Field, 2009)

sound symbolism The study of the relationship between the sound of an utterance and its meaning. (Hinton et al., 1994)

spectral tilt Energy distribution across the frequency range of an acoustic signal at a specified time point or period: a steep slope indicates high energy in low frequencies and low energy in high frequencies. A flat slope indicates even energy distribution across the frequency range.

Stroop effect In a Stroop task (named after John Ridley Stroop (1897-1973)), reaction time is measured to a task involving incongruent stimuli. Stroop's original task involved naming the ink colour of a word while ignoring the meaning of the word. If the word *blue* was printed in red, red was the correct answer. The incongruency of word meaning and ink colour slows down reaction time.

supralaryngeal Refers to the speech organs above the larynx, i.e. the **vocal tract**. It is usually used to describe settings of the **articulators** and how they may change the speech sound or **voice quality**.

synaesthesia (pl. synaesthesiae) A phenomenon in which additional (sensory) perceptions are triggered by apparently unrelated sensory or conceptual stimuli. For example, people with grapheme-colour synaesthesia see colours while reading letters or numbers, or associate colours with those graphemes.

synaesthete A person with the condition called **synaesthesia**.

ticker tape synaesthesia Ticker tape synaesthetes see words spelled out in front of them when they hear speech, as if a ticker tape appeared.

timbre The character or quality of a musical or vocal sound (distinct from its pitch and intensity) depending on the particular voice or instrument producing it, and distinguishing it from sounds proceeding from other sources. It is caused by the proportion in which the fundamental tone is combined with the **harmonics** or overtones (= German Klangfarbe, i.e. "colour of sound"). (Simpson, 1989)

TMS (transcranial magnetic stimulation) A non-invasive technique to stimulate a restricted part of the cortex based on electromagnetic induction. A coil generates a magnetic field stimulating the brain, both to elicit responses directly and to modify excitability and plasticity.

Tonotopy A fundamental organising principle of the auditory system. There is tonotopy in the cochlea, the small snail-like structure in the inner ear that sends information about sound to the brain. There is also tonotopy in the human auditory cortex, the part of the brain that receives and interprets sound information. Different frequencies are processed at different places in these areas (cochlea and auditory cortex).

V4 The visual cortex of the brain is the part of the cerebral cortex responsible for processing visual information. Its subarea V4 is supposedly tuned for colour, orientation, spatial frequency and simple geometric shapes.

verbal overshadowing effect The verbal description of a voice between initial exposure and a **voice parade** by the witness supposedly impairs later voice identification performance, probably because cognitive processes are different for description and retrieval.

vocal tract The passages above the larynx through which air passes in the production of speech: the throat, the mouth and the nasal tract.

voice parade (or **voice line-up)** A procedure in **forensic phonetics** which serves to identify a suspect of a crime. Similar to a visual line-up, the (ear)witness is auditorily presented with a series of recordings of different voices.

voxel-based morphometry A neuroimaging analysis technique that allows investigation of focal differences in brain anatomy, using the approach of statistical parametric mapping. It usually involves a comparison of the local concentration of gray matter between two groups of subjects.

VPA(Vocal Profile Analysis) An auditory perceptual assessment of voice quality **(VQ)** providing a systematic phonetic framework for describing speakers' habitual voice quality. Laryngeal, **supralaryngeal**, prosodic and temporal organisation features are assessed. (Beck, 2005)

VQ (voice quality) The features that habitually or permanently characterise a speaker's voice; in particular, those involving a fixed articulatory setting such as a nasalized voice quality (with incomplete raising of the soft palate), whisper (without voicing), etc.

References

Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh University Press.

Aguirre, G.K. (2007). Continuous carry-over designs for fMRI. *NeuroImage* 35: 1480–1494.

Aleman, A., Rutten, G.J.M., Sitskoorn, M.M., Dautzenberg, G. & Ramsey, N.F. (2001). Activation of striate cortex in the absence of visual stimulation: An fMRI study of synesthesia. *NeuroReport* 12(13): 2827–2830.

Anderson Feisner, E. (2001). *Colour. How to Use Colour in Art and Design.* Laurence King Publishing.

Asher, J.E., Aitken, M.R., Farooqi, N., Kurmani, S. & Baron-Cohen, S. (2006). Diagnosing and phenotyping visual synaesthesia: A preliminary evaluation of the revised test of genuineness (TOG-R). *Cortex* 42: 137–146.

Audacity Team (2010). Audacity: Free audio editor and recorder. [Computer program]. URL `audacity.sourceforge.net`.

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R.* Cambridge University Press.

Bargary, G., Barnett, K.J., Mitchell, K.J. & Newell, F.N. (2009). Colored-speech synaesthesia is triggered by multisensory, not unisensory, perception. *Psychological Science* 20(5): 529–533.

Bargary, G. & Mitchell, K.J. (2008). Synaesthesia and cortical connectivity. *Trends in Neurosciences* 31(7): 335–342.

Barnett, K.J., Feeney, J., Gormley, M. & Newell, F.N. (2009). An exploratory study of linguistic-colour associations across languages in multilingual synaesthetes. *The Quarterly Journal of Experimental Psychology* 62(7): 1343–1355.

Barnett, K.J., Finucane, C., Asher, J.E., Bargary, G., Corvin, A.P., Newell, F.N. & Mitchell, K.J. (2008). Familial patterns and the origins of individual differences in synaesthesia. *Cognition* 106: 871–893.

Baron-Cohen, S., Bor, D., Billington, J., Asher, J., Wheelwright, S. & Ashwin, C. (2007). Savant memory in a man with colour form-number synaesthesia and Asperger syndrome. *Journal of Consciousness Studies* 14(9-10): 237–251.

Baron-Cohen, S., Burt, L., Smith-Laittan, F., Harrison, J. & Bolton, P. (1996). Synaesthesia: Prevalence and familiality. *Perception* 25: 1073–1079.

Baron-Cohen, S., Harrison, J., Goldstein, L.H. & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception* 22: 419–426.

Baron-Cohen, S., Wyke, M.A. & Binnie, C. (1987). Hearing words and seeing colours: An experimental investigation of a case of synaesthesia. *Perception* 16: 761–767.

Beck, J.M. (2005). Perceptual analysis of voice quality: The place of vocal profile analysis. In: W.J. Hardcastle & J.M. Beck (eds.), *A Figure of Speech. A Festschrift for John Laver*, 285–322. Lawrence Erlbaum Associates.

Beeli, G., Esslen, M. & Jäncke, L. (2007). Frequency correlates in grapheme-color synaesthesia. *Psychological Science* 18: 788–792.

Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P. & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403: 309–312.

Ben-Tal, O. & Sagiv, N. (2010). Auditory-visual synaesthesia: Investigations using dynamic sounds and spontaneous reports between sleep and wakefulness. In: *2010 Meeting of the UK Synaesthesia Association*, 18.

Bendor, D. & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436: 1161–1165.

Berlin, B. & Kay, P. (1999). *Basic Color Terms. Their Universality and Evolution.* CSLI Publications.

Boersma, P. & Weenink, D. (2012). Praat: Doing phonetics by computer. [Computer program]. URL `http://www.praat.org/`.

Bor, D., Billington, J. & Baron-Cohen, S. (2008). Savant memory for digits in a case of synaesthesia and Asperger syndrome is related to hyperactivity in the lateral prefrontal cortex. *Neurocase* 13(5-6): 311–319.

Borden, G.J., Harris, K.S. & Raphael, L.J. (2003). *Speech Science Primer. Physiology, Acoustics, and Perception of Speech.* Lippincott Williams & Wilkins, 4th edition.

Bothwell, R.K., Deffenbacher, K.A. & Brigham, J.C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology* 72(4): 691–695.

Brodatz, P. (1966). *Textures: A Photographic Album for Artists and Designers.* Dover Publications.

Broeders, A. & van Amelsvoort, A. (1999). Lineup construction for forensic earwitness identification: A practical approach. *Proceedings of the 14th ICPhS* 1373–1376.

Broeders, A. & van Amelsvoort, A. (2001). A practical approach to forensic earwitness identification: Constructing a voice line-up. *Problems of Forensic Sciences* 47: 237–245.

Broeders, A., Combier-Langeveld, T. & Vermeulen, J. (2002). Arranging a voice line-up in a foreign language. *Forensic Linguistics* 9(1): 104–112.

Brungart, D., Scott, K. & Simpson, B. (2001). The influence of vocal effort on human speaker identification. *Eurospeech* 747–750.

Bull, R. & Clifford, B.R. (1984). Earwitness voice recognition accuracy. In: G.L. Wells & E.F. Loftus (eds.), *Eyewitness testimony. Pychological perspectives*, 92–123. Cambridge University Press.

Burns, J. (2010). Voice quality: Labelling by untrained listeners. Unpublished student essay, University of York.

Busey, T.A., Tunnicliff, J., Loftus, G.R. & Loftus, E.F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review* 7(1): 26–48.

Calvert, G.A., Hansen, P.C., Iversen, S.D. & Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the bold effect. *NeuroImage* 14(2): 427–438.

van Campen, C. (2009). How does the memory of synaesthetes work? In: *Proceedings of 3rd International Congress on Synaesthesia, Science & Art*.

Catford, J.C. (1994). *A Practical Introduction to Phonetics*. Oxford University Press.

Charpentier, F. & Moulines, E. (1989). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Proceedings of Eurospeech* 2: 13–19.

Clark, J., Yallop, C. & Fletcher, J. (2007). *An Introduction to Phonetics and Phonology*. Blackwell Publishing, 3rd edition.

Clarke, A., Halley, F., Newell, A., Griffin, L. & Chantler, M. (2011). Perceptual similarity: A texture challenge. In: J. Hoey, S. McKenna & E. Trucco (eds.), *Proceedings of the British Machine Vision Conference*, 120.1–120.10.

Clifford, B.R. (1980). Voice identification by human listeners: On earwitness reliability. *Law and Human Behavior* 4(4): 373–394.

Cohen Kadosh, R., Cohen Kadosh, K. & Henik, A. (2007). The neuronal correlate of bidirectional synesthesia: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience* 19(12): 2050–2059.

Cohen Kadosh, R., Henik, A., Catena, A., Walsh, V. & Fuentes, L.J. (2009). Induced cross-modal synaesthetic experience without abnormal neuronal connections. *Psychological Science* 20(2): 258–265.

Cohen Kadosh, R. & Terhune, D.B. (2012). Redefining synaesthesia? *British Journal of Psychology* 103(1): 20–23.

Cook, S. & Wilding, J. (1997a). Earwitness testimony 2: Voices, faces and context. *Applied Cognitive Psychology* 11: 527–541.

Cook, S. & Wilding, J. (1997b). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology* 11: 95–111.

Cook, S. & Wilding, J. (2001). Earwitness testimony: Effects of exposure and attention on the face overshadowing effect. *British Journal of Psychology* 92: 617–629.

Cummins, F., Grimaldi, M., Leonard, T. & Simko, J. (2006). The CHAINS corpus: CHAracterizing INdividual Speakers. In: *Proceedings of The International Conference on Speech and Computer*, 431–435. St. Petersburg, Russia.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1): 51–62.

Day, S. (2004). Trends in synesthetically colored graphemes and phonemes – 2004 revision. Unpublished.

Day, S. (2005). Some demographic and socio-cultural aspects of synesthesia. In: L.C. Robertson & N. Sagiv (eds.), *Synesthesia. Perspectives from Cognitive Neuroscience*, 11–33. Oxford University Press.

Dixon, M.J., Smilek, D. & Merikle, P.M. (2004). Not all synaesthetes are created equal: Projector versus associator synaesthetes. *Cognitive, Affective, & Behavioral Neuroscience* 4(3): 335–343.

Djonov, E. & Leeuwen, T.V. (2011). The semiotics of texture: From tactile to visual. *Visual Communication* 10(4): 541–564.

Eagleman, D.M. (2012). Synaesthesia in its protean guises. *British Journal of Psychology* 103(1): 16–19.

Eagleman, D.M. & Goodale, M.A. (2009). Why color synesthesia involves more than color. *TRENDS in Cognitive Sciences* 13(7): 288–292.

Eagleman, D.M., Kagan, A.D., Nelson, S.S., Sagaram, D. & Sarma, A.K. (2007). A standardized test battery for the study of synesthesia. *Journal of Neuroscience Methods* 159: 139–145.

Edquist, J., Rich, A.N., Brinkman, C. & Mattingley, J.B. (2006). Do synaesthetic colours act as unique features in visual search? *Cortex* 42: 222–231.

Ellis, A.W. & Young, A.W. (1997). *Human Cognitive Neuropsychology. A Textbook with Readings*. Psychology Press.

Esterman, M., Verstynen, T., Ivry, R.B. & Robertson, L.C. (2006). Coming unbound: Disrupting automatic integration of synesthetic color and graphemes by transcranial magnetic stimulation of the right parietal lobe. *Journal of Cognitive Neuroscience* 18(9): 1570–1576.

Evans, I. & Foulkes, P. (2009). Speaker identification in whisper. In: *Proceedings of the IAFPA*.

Fairbanks, G., House, A.S. & Stevens, E.L. (1950). An experimental study on vowel intensities. *Journal of the Acoustical Society of America* 22(4): 457–459.

Fairchild, M.D. (1998). *Color Appearance Models*. Addison Wesley Longman.

Fernay, L., Reby, D. & Ward, J. (2012). Visualized voices: A case study of audio-visual synesthesia. *Neurocase* 18(1): 50–56.

Field, A. (2009). *Discovering statistics using SPSS*. SAGE Publications, 3rd edition.

Ford, A. & Roberts, A. (1998). Colour space conversions. Unpublished.

Forster, K.I. & Forster, J.C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35(1): 116–124.

Gheri, C., Chopping, S. & Morgan, M. (2008). Synaesthetic colours do not camouflage form in visual search. *Proceedings of the Royal Society B* 275: 841–846.

Gimmestad, K.D. (2011). *Assessment of Transient Negative Affect in Synesthesia*. Ph.D. thesis, University of Missouri-Kansas City.

Goggin, J.P., Thompson, C.P., Strube, G. & Simental, L.R. (1991). The role of language familiarity in voice identification. *Memory & Cognition* 19(5): 448–458.

Golan, O., Baron-Cohen, S. & Hill, J. (2006). The Cambridge Mindreading (CAM) Face-Voice Battery: Testing complex emotion recognition in adults with and without Asperger syndrome. *Journal of Autism and Developmental Disorders* 36(2): 169–183.

Goldstein, A.G., Knight, P., Bailis, K. & Conover, J. (1981). Recognition memory for accented and unaccented voices. *Bulletin of the Psychonomic Society* 17(5): 217–220.

Gomez, P. & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology* 53: 91–103.

Grossenbacher, P.G. (1997). Perception and sensory information in synaesthetic experience. In: S. Baren-Cohen & J.E. Harrison (eds.), *Synaesthesia. Classic and Contemporary Readings*, 148–172. Blackwell Publishers Ltd.

Grossenbacher, P.G. & Lovelace, C.T. (2001). Mechanisms of synesthesia: Cognitive and physiological constraints. *TRENDS in Cognitive Sciences* 5(1): 36–41.

Gurnsey, R. & Fleet, D.J. (2001). Texture space. *Vision Research* 41(6): 745–757.

Halley, F. (2011). Pertex v1.0. URL `http://www.macs.hw.ac.uk/texturelab/resources/databases/pertex/`.

Hammersley, R. & Read, J.D. (1985). The effect of participation in a conversation on recognition and identification of the speakers' voices. *Law and Human Behavior* 9(1): 71–81.

Hancock, P. (2006). Monozygotic twins' colour-number association: A case study. *Cortex* 42: 147–150.

Hanson, H.M. (1997). Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101(1): 466–481.

Harris, T.R., Fitch, W.T., Goldstein, L.M. & Fashing, P.J. (2006). Black and white colobus monkey (Colobus guereza) roars as a source of both honest and exaggerated information about body mass. *Ethology* 112(9): 911–920.

Heath, A.J. & Moore, K. (2011). Earwitness memory: Effects of facial concealment on the face overshadowing effect. *International Journal of Advanced Science and Technology* 33: 131–140.

von Helmholtz, H. (1863). *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Theorie der Musik*. F. Vieweg.

Henderson, L. (1985). On the use of the term 'grapheme'. *Language and Cognitive Processes* 1(2): 135–148.

Hewlett, N. & Rendall, M. (1998). Rural versus urban accent as an influence on the rate of speech. *Journal of the International Phonetic Association* 28: 63–71.

Higashikawa, M., Nakai, K., Sakakura, A. & Takahashi, H. (1996). Perceived pitch of whispered vowels – Relationship with formant frequencies: A preliminary study. *Journal of Voice* 10(2): 155–158.

Hinton, L., Nichols, J. & Ohala, J.J. (eds.) (1994). *Sound symbolism.* Cambridge University Press.

Hollien, H. (1996). Consideration of guidelines for earwitness lineups. *Forensic Linguistics* 3(1): 14–23.

Hollien, H., Huntley, R., Künzel, H. & Hollien, P.A. (1995). Criteria for earwitness lineups. *Forensic Linguistics* 2(2): 143–153.

Hubbard, E.M., Arman, A.C., Ramachandran, V.S. & Boynton, G.M. (2005). Individual differences among grapheme-color synesthetes: Brain-behavior correlations. *Neuron* 45: 975–985.

Hubbard, E.M., Brang, D. & Ramachandran, V.S. (2011). The cross-activation theory at 10. *Journal of Neuropsychology* 5(2): 152–177.

Hubbard, T.L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *American Journal of Psychology* 109(2): 219–238.

Hudson, T., de Jong, G., McDougall, K., Harrison, P. & Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16th ICPhS* 1809–1812.

Humphrey, N. (1976). The colour currency of nature. In: T. Porter & B. Mikellides (eds.), *Colour for Architecture*, 95–98. Studio Vista.

Hupé, J.M., Bordier, C. & Dojat, M. (2011). The neural bases of grapheme-color synesthesia are not localized in real color-sensitive areas. *Cerebral Cortex* 1–12.

International Phonetic Association (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet.* Cambridge: Cambridge University Press.

Ishihara, S. (1978). *Tests for Colour-Blindness.* Kanehara Shuppan & Co., Ltd.

Jakobson, R. (1962). *Selected Writings I. Phonological Studies.* Mouton & Co.

Jakobson, R. & Waugh, L.R. (1987). *The Sound Shape of Language*. Mouton de Gruyter, 2nd edition.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass* 2(4): 671–711.

Jessen, M., Köster, O. & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *Journal of Speech, Language and the Law* 12(2): 174–213.

Jewanksi, J., Simner, J., Day, S.A. & Ward, J. (2011). The development of a scientific understanding of synesthesia from early case studies (1849–1873). *Journal of the History of the Neurosciences* 20: 284–305.

Kawahara, H. (2003). Exemplar-based voice quality analysis and control using a high quality auditory morphing procedure based on STRAIGHT. In: *ISCA workshop VOQUAL'03*, 109–114. Geneva, Switzerland.

Kay, C. & Mulvenna, C. (2006). Synaesthesia, neurology and language. In: N. Pitchford & C. Biggam (eds.), *Progress in Colour Studies*, volume 2, 203–224. John Benjamins Publishing Company.

Kay, P., Berlin, B., Maffi, L., Merrifield, W.R. & Cook, R. (2009). *The World Color Survey*. SCLI Publications.

Kerstholt, J.H., Jansen, N.J.M., van Amelsvoort, A.G. & Broeders, A.P.A. (2004). Earwitnesses: Effects of speech duration, retention interval and acoustic environment. *Applied Cognitive Psychology* 18(3): 327–336.

Kerstholt, J.H., Jansen, N.J.M., van Amelsvoort, A.G. & Broeders, A.P.A. (2006). Earwitnesses: Effects of accent, retention and telephone. *Applied Cognitive Psychology* 20: 187–197.

Köhler, W. (1924). *Tonpsychologie*. Urban & Schwarzenberg.

von Kriegstein, K. & Giraud, A.L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology* 1809-1820(10): 4.

Künzel, H.J. (1990). Phonetische Untersuchungen zur Sprecher-Erkennung durch linguistisch naive Personen. In: J. Göschel (ed.), *Zeitschrift für Dialektologie und Linguistik*, volume 69. Franz Steiner Verlag.

Kuwabara, H. & Takagi, T. (1991). Acoustic parameters of voice individuality and voice-quality control by analysis-synthesis method. *Speech Communication* 10: 491–495.

Ladefoged, P. (1993). *A Course in Phonetics*. Harcourt Brace College Publishers.

Ladefoged, P. & Ladefoged, J. (1980). The ability of listeners to identify voices. *UCLA Working Papers in Phonetics* 49: 43–51.

Landy, M.S. & Graham, N. (2004). Visual perception of texture. In: L.M. Chalupa & J.S. Werner (eds.), *The Visual Neurosciences*, 1106–1118. MIT Press.

Lauter, J.L., Herscovitch, P., Formby, C. & Raichle, M.E. (1985). Tonotopic organization in human auditory cortex revealed by positron emission tomography. *Hearing Research* 20(3): 199–205.

Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press.

Laver, J. (1991). *The Gift of Speech*. Edinburgh University Press.

Laver, J. & Trudgill, P. (1979). Phonetic and linguistic markers in speech. In: K.R. Scherer & H. Giles (eds.), *Social Markers in Speech*, 1–32. Cambridge University Press.

Liljencrants, J. & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4): 839–862.

LimeSurvey project team (2009). Limesurvey: An online survey application. URL `www.limesurvey.org`.

Lucassen, M.P., Gevers, T. & Gijsenij, A. (2011). Texture affects color emotion. *Color Research & Application* 36: 426–436.

Ludwig, V.U., Adachi, I. & Matsuzawa, T. (2011). Visuoauditory mappings between high luminance and high pitch are shared by chimpanzees (Pan troglodytes) and humans. *Proceedings National Academy of Sciences of the USA* 108(51): 20661–20665.

Lundeberg, M.A., Fox, P.W. & Punccohar, J. (1994). Highly confident but wrong: Gender differences and similarities in confidence judgments. *Journal of Educational Psychology* 86(1): 114–121.

Luria, A.R. (1968). *The Mind of a Mnemonist*. Jonathan Cape Ltd.

MacMahon, M.K.C. (2011). *Basic Phonetics*. University of Glasgow, 15th edition.

Malacara, D. (2011). *Color Vision and Colorimetry: Theory and Applications*. SPIE Press, 2nd edition.

Malik, K. (2010). *The effect of language familiarity on speaker identification. The study of Italian speech*. Master's thesis, University of Cambridge.

Marks, L.E. (1974). On associations of light and sound: The mediation of brightness, pitch, and loudness. *American Journal of Psychology* 87(2): 173–188.

Marks, L.E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin* 82(3): 303–331.

Marks, L.E. (1987). On cross-modal similarity: Auditory-visual interactions in speeded discrimination. *Journal of Experimental Psychology* 13(3): 384–394.

Matthews, P.H. (1997). *The Concise Oxford Dictionary of Linguistics*. Oxford University Press.

Mattingley, J.B., Rich, A.N., Yelland, G. & Bradshaw, J.L. (2001). Unconscious priming eliminates automatic binding of colour and alphanumeric form in synaesthesia. *Nature* 410: 580–582.

Maurer, D. & Mondloch, C.J. (2005). Neonatal synaesthesia: A reevaluation. In: L.C. Robertson & N. Sagiv (eds.), *Synesthesia: Perspectives from Cognitive Neuroscience*, 193–213. Oxford University Press.

McAllister, H.A., Dale, R.H., Bregman, N.J., McCabe, A. & Cotton, C.R. (1993). When eyewitnesses are also earwitnesses: Effects on visual and voice identifications. *Basic and Applied Social Psychology* 14(2): 161–170.

McGehee, F. (1937). The reliability of the identification of the human voice. *Journal of General Psychology* 17: 249–271.

McGehee, F. (1944). An experimental study of voice recognition. *Journal of General Psychology* 31: 53–65.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264: 746–748.

Meissner, C.A., Brigham, J.C. & Butz, D.A. (2005). Memory for own- and other-race faces: A dual-process approach. *Applied Cognitive Psychology* 19: 545–567.

Melara, R.D., DeWitt-Rickards, T.S. & O'Brien, T.P. (1989). Enhancing lineup identification accuracy: Two codes are better than one. *Journal of Applied Psychology* 74(5): 706–713.

Mills, C.B., Innis, J., Westendorf, T., Owsianiecki, L. & McDonald, A. (2006). Effect of a synesthete's photisms on name recall. *Cortex* 42: 155–163.

Moos, A. (2010). Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. *The Phonetician* 101/102: 7–24.

Moos, A., Smith, R., Miller, S. & Simmons, D. (in preparation). Cross-modal associations in synaesthesia: Vowel colours in the ear of the beholder .

Morrison, G.S. (2009). Forensic voice comparison and the paradigm shift. *Science and Justice* 49: 298–308.

Mudge, E.L. (1920). The common synaesthesia of music. *Journal of Applied Psychology* 4(4): 342–345.

Nikolić, D. (2009). Is synaesthesia actually ideaesthesia? An inquiry into the nature of the phenomenon. In: *Proceedings of the Third International Congress on Synaesthesia, Science & Art*. Granada, Spain.

Nikolić, D., Lichti, P. & Singer, W. (2007). Color opponency in synaesthetic experiences. *Psychological Science* 18(6): 481–486.

Nolan, F. (1983). *The phonetic bases of speaker recognition*. Cambridge University Press.

Nolan, F. (2003). A recent voice parade. *Forensic Linguistics* 10(2): 277–291.

Nolan, F. & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *Journal of Speech, Language and the Law* 12(2): 143–173.

Nonyane, B.A.S. & Theobald, C.M. (2007). Design sequences for sensory studies: Achieving balance for carry-over and position effects. *British Journal of Mathematical and Statistical Psychology* 60: 339–349.

Novich, S., Cheng, S. & Eagleman, D.M. (2011). Is synaesthesia one condition or many? A large-scale analysis reveals subgroups. *Journal of Neuropsychology* 5(2): 353–371.

Nunn, J., Gregory, L., Brammer, M., Williams, S., Parslow, D., Morgan, M., Morris, R., Bullmore, E., Baron-Cohen, S. & Gray, J. (2002). Functional magnetic resonance imaging of synesthesia: Activation of V4/V8 by spoken words. *Nature Neuroscience* 5(4): 371–375.

Ogden, R. (2009). *An Introduction to English Phonetics*. Edinburgh University Press.

Öhman, L., Eriksson, A. & Granhag, P.A. (2011). Overhearing the planning of a crime: Do adults outperform children as earwitnesses? *Journal of Police and Criminal Psychology* 26(2): 118–127.

Orchard, T.L. & Yarmey, A.D. (1995). The effects of whispers, voice-sample duration, and voice distinctiveness on criminal speaker identification. *Applied Cognitive Psychology* 9(3): 249–260.

Ou, L.C., Luo, M.R., Woodcock, A. & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. *COLOR research and application* 29: 232–240.

Oxford Dictionaries (2010). URL `http://oxforddictionaries.com`.

Palmeri, T.J., Blake, R., Marois, R., Flanery, M.A. & Whetsell Jr., W. (2002). The perceptual reality of synesthetic colors. *Proceedings National Academy of Sciences of the USA* 99(6): 4127–4131.

Paulesu, E., Harrison, J., Baron-Cohen, S., Watson, J.D.G., Goldstein, L., Heather, J., Frackowiak, R.S.J. & Frith, C.D. (1995). The physiology of coloured hearing. A PET activation study of colour-word synaesthesia. *Brain* 118: 661–676.

Perfect, T., Hunt, L.J. & Harris, C.M. (2002). Verbal overshadowing in voice recognition. *Applied Cognitive Psychology* 16: 973–980.

Petrou, M., Talebpour, A. & Kadyrov, A. (2007). Reverse engineering the way humans rank textures. *Pattern Analysis and Applications* 10: 101–114.

Philippon, A.C., Cherryman, J., Bull, R. & Vrij, A. (2007). Earwitness identification performance: The effect of language, target, deliberate strategies and indirect measures. *Applied Cognitive Psychology* 21(4): 539–550.

Pisoni, D. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics* 13(2): 253–260.

Pisoni, D. (1975). Auditory short-term memory and vowel perception. *Memory & Cognition* 3(1): 7–18.

R Foundation for Statistical Computing (2010). R: A free software environment for statistical computing and graphics. [Computer program]. URL `http://www.r-project.org/index.html`.

Ramachandran, V. & Hubbard, E. (2001a). Psychophysical investigations into the neural basis of synaesthesia. *Proceedings of the Royal Society B* 268: 979–983.

Ramachandran, V. & Hubbard, E. (2001b). Synaesthesia – A window into perception, thought and language. *Journal of Consciousness Studies* 8(12): 3–34.

Ramachandran, V. & Hubbard, E. (2003). The phenomenology of synaesthesia. *Journal of Consciousness Studies* 10(8): 49–57.

Rao, A.R., Bhushan, N. & Lohse, G.L. (1996). The relationship between texture terms and texture images: A study in human texture perception. *Proceedings of SPIE* 2670(206): 206–214.

Rao, A.R. & Lohse, G. (1996). Towards a texture naming system: Identifying relevant dimension of texture. *Vision Research* 36(11): 1649–1669.

Rao, A.R. & Lohse, G.L. (1993). Identifying high level features of texture perception. *CVGIP: Graphical Models and Image Processing* 55(3): 218–233.

Reay, I.E. (2006). Sound symbolism. In: K. Brown (ed.), *Encyclopedia of language & linguistics*, 531–539. Elsevier, 2nd edition.

Rehme, M. & Zedler, M. (2010). To be spoilt for choice: Self-sensory overload produced by synaesthesia in stress situations. In: *Meeting of the UK Synaesthesia Association*, 20–21.

Reich, A.R. & Lerman, J.W. (1978). Teflon laryngoplasty: An acoustical and perceptual study. *Journal of Speech and Hearing Disorders* 43: 496–505.

Rich, A., Bradshaw, J. & Mattingley, J. (2005). A systematic, large-scale study of synaesthesia: Implications for the role of early experience in lexical-colour associations. *Cognition* 98: 53–84.

Richer, F., Beaufils, G.A. & Poirier, S. (2011). Bidirectional lexical-gustatory synesthesia. *Consciousness and Cognition* 20(4): 1738–1743.

Rolls, E., O'Doherty, J., Kringelbach, M., Francis, S., Bowtell, R. & McGlone, F. (2003). Representations of pleasant and painful touch in the human orbitofrontal and cingulate cortices. *Cerebral Cortex* 13(3): 308–317.

Rosch Heider, E. (1972). Universals in color naming and memory. *Journal of Experimental Psychology* 93: 10–20.

Rose, P. (2002). *Forensic Speaker Identification*. Taylor & Francis.

Rothen, N. & Meier, B. (2009). Do synesthetes have a general advantage in visual search and episodic memory? A case for group studies. *PLoS ONE* 4(4): 1–9.

Rothen, N. & Meier, B. (2010). Grapheme-colour synaesthesia yields an ordinary rather than extraordinary memory advantage: Evidence from a group study. *Memory* 18(3): 258–264.

Rothen, N., Meier, B. & Ward, J. (2012). Enhanced memory ability: Insights from synaesthesia. *Neuroscience and Biobehavioral Reviews* 36(8): 1952–1963.

Rouw, R. & Scholte, H.S. (2007). Increased structural connectivity in grapheme-colour synesthesia. *Nature Neuroscience* 10(6): 792–797.

Rouw, R., Scholte, H.S. & Colizoli, O. (2011). Brain areas involved in synaesthesia: A review. *Journal of Neuropsychology* 5(2): 214–242.

Russell, J.A. & Pratt, G. (1980). A description of the affective quality attributed to enviroments. *Journal of Personality and Social Psychology* 38(2): 311–322.

Sagiv, N. (2005). Synesthesia in perspective. In: L.C. Robertson & N. Sagiv (eds.), *Synesthesia. Perspectives from Cognitive Neuroscience*, 3–10. Oxford University Press.

Saslove, H. & Yarmey, A.D. (1980). Long-term auditory memory: Speaker identification. *Journal of Applied Psychology* 65(1): 111–116.

Schiller, N.O. & Köster, O. (1998). The ability of expert witnesses to identify voices: A comparison between trained and untrained listeners. *Forensic Linguistics* 5(1): 1350–1771.

Sereno, J.A. (1994). Phonosyntactics. In: L. Hinton, J. Nichols & J. Ohala (eds.), *Sound symbolism*, 263–275. Cambridge University Press.

Sheffert, S.M. & Olson, E. (2004). Audiovisual speech facilitates voice learning. *Perception & Psychophysics* 66(2): 352–362.

Sidoroff-Dorso, A.V. (2010). Synaesthesia factor: General neurocognitive aspects in large-scale survey and their methodological implications. In: *Proceedings of the 4th Congress Synaesthesia in Arts and Science*. Almeria, Spain.

Simmons, D.R. (2011). Colour and emotion. In: C. Biggam, C. Kay, C. Hough & D. Simmons (eds.), *New Directions in Colour Studies*, 395–414. John Benjamins Publishing Company.

Simner, J. (2007). Beyond perception: Synaesthesia as a psycholinguistic phenomenon. *TRENDS in Cognitive Sciences* 11(1): 23–29.

Simner, J. (2011). Synaesthesia in colour. In: C.P. Biggam, C. Hough, C.J. Kay & D.R. Simmons (eds.), *New Directions in Colour Studies*, 309–318. John Benjamins Publishing Company.

Simner, J. (2012a). Defining synaesthesia. *British Journal of Psychology* 103(1): 1–15.

Simner, J. (2012b). Defining synaesthesia: A response to two excellent commentaries. *British Journal of Psychology* 103(1): 24–27.

Simner, J., Mayo, N. & Spiller, M.J. (2009). A foundation for savantism? Visuo-spatial synaesthetes present with cognitive benefits. *Cortex* 45: 1246–1260.

Simner, J., Mulvenna, C., Sagiv, N., Tsakanikos, E., Witherby, S.A., Fraser, C., Scott, K. & Ward, J. (2006). Synaesthesia: The prevalence of atypical cross-modal experiences. *Perception* 35: 1024–1033.

Simner, J., Ward, J., Lanz, M., Jansari, A., Noonan, K., Glover, L. & Oakley, D.A. (2005). Non-random associations of graphemes to colours in synaesthetic and non-synaesthetic populations. *Cognitive Neuropsychology* 22(8): 1069–1085.

Simpson, J. (ed.) (1989). *Oxford English Dictionary.* Oxford University Press, 2nd edition.

Simpson, L. & McKellar, P. (1955). Types of synaesthesia. *The British Journal of Psychiatry* 101: 141–147.

Sinke, C., Neufeld, J., Emrich, H., Bleich, S., Szycik, G. & Zedler, M. (2012). Reduced multimodal integration in synaesthetes: Testing auditory-visual integration in synaesthesia with the McGurck illusion. In: *Proceedings of the fourth International Congress on Synaesthesia, Science & Art*. Almeria, Spain.

Sjölander, K. & Beskow, J. (2005). Wavesurfer 1.8.5. [Computer program]. URL `http://www.speech.kth.se/wavesurfer/index.html`.

Skelton, R., Ludwig, C. & Mohr, C. (2009). A novel, illustrated questionnaire to distinguish projector and associator synaesthetes. *Cortex* 45: 721–729.

Smilek, D., Carriere, J.S., Dixon, M.J. & Merikle, P.M. (2007). Grapheme frequency and color luminance in grapheme-color synaesthesia. *Psychological Science* 18(9): 793–795.

Smilek, D., Dixon, M.J., Cudahy, C. & Merikle, P.M. (2001). Synaesthetic photisms influence visual perception. *Journal of Cognitive Neuroscience* 13(7): 930–936.

Smilek, D., Dixon, M.J., Cudahy, C. & Merikle, P.M. (2002). Synesthetic color experiences influence memory. *Psychological Science* 13(6): 548–552.

Smith, R., Moos, A., Cartwright-Hignett, W. & Simmons, D. (2011). Towards a phonetically-rich account of speech-colour synaesthesia. In: D.S. C. Biggam C. Hough & C. Kay (eds.), *New Directions in Colour Studies*, 319–328. John Benjamins Publishing Company.

van Son, R. (2002). Can standard analysis tools be used on decompressed speech? In: *COCOSDA2002 meeting, Denver*.

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception & Psychophysics* 73: 971–995.

Sperling, J.M., Prvulovic, D., Linden, D.E., Singer, W. & Stirn, A. (2006). Neuronal correlates of colour-graphemic synaesthesia: A fMRI study. *Cortex* 42: 295–303.

Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research.* SAGE Publications, 2nd edition.

Stroop, J.R. (1992). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General* 121(1): 15–23. (This reprinted article originally appeared in the Journal of Experimental Psychology, 1935, Vol 18, 643–662).

Stuart-Smith, J. (1999). Glasgow: Accent and voice quality. In: P. Foulkes & G. Docherty (eds.), *Urban Voices. Accent Studies in the British Isles*, 203–222. Arnold.

Stumpf, C. (1926). *Die Sprachlaute: Experimentell-phonetische Untersuchungen nebst einem Anhang über Instrumentalklänge*. Verlag von Julius Springer.

Talavage, T.M., Sereno, M.I., Melcher, J.R., Ledden, P.J., Rosen, B.R. & Dale, A.M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology* 91: 1282–1296.

Tamura, H., Mori, S. & Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on systems, man, and cybernetics* 8(6): 460–473.

Tartter, V.C. (1991). Identifiability of vowels and speakers from whispered syllables. *Perception & Psychophysics* 49(4): 365–372.

Thompson, C.P. (1985). Voice identification: Speaker identifiability and a correction of the record regarding sex effects. *Human Learning: Journal of Practical Research & Applications* 4(1): 19–27.

de Thornley Head, P. (2006). Synaesthesia: Pitch-colour isomorphism in RGB-space? *Cortex* 42: 164–174.

Tomson, S.N., Avidan, N., Lee, K., Sarma, A.K., Tushe, R., Milewicz, D.M., Bray, M., Leal, S.M. & Eagleman, D.M. (2011). The genetics of colored sequence synesthesia: Suggestive evidence of linkage to 16q and genetic heterogeneity for the condition. *Behavioural Brain Research* 223: 48–52.

Traunmüller, H. & Eriksson, A. (1995). The frequency range of the voice fundamental in the speech of male and female adults. Unpublished.

Trouvain, J. (2003). *Tempo Variation in Speech Production. Implications for Speech Synthesis*. Ph.D. thesis, Saarland University.

Vanags, T., Carroll, M. & Perfect, T.J. (2005). Verbal overshadowing: A sound theory in voice recognition? *Applied Cognitive Psychology* 19: 1127–1144.

Ward, J., Huckstep, B. & Tsakanikos, E. (2006). Sound-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all? *Cortex* 42: 264–280.

Ward, J., Jonas, C., Dienes, Z. & Seth, A. (2009). Grapheme-colour synaesthesia improves detection of embedded shapes, but without pre-attentive 'pop-out' of synaesthetic colour. *Proceedings of the Royal Society B* 277(1684): 1021–1026.

Ward, J. & Mattingley, J.B. (2006). Synaesthesia: An overview of contemporary findings and controversies. *Cortex* 42: 129–136.

Ward, J. & Simner, J. (2003). Lexical-gustatory synaesthesia: Linguistic and conceptual factors. *Cognition* 89: 237–261.

Ward, J., Simner, J. & Auyeung, V. (2005). A comparison of lexical-gustatory and grapheme-color synaesthesia. *Cognitive Neuropsychology* 22(1): 28–41.

Ward, J., Thompson-Lake, D., Ely, R. & Kaminski, F. (2008). Synaesthesia, creativity and art: What is the link? *British Journal of Psychology* 99: 127–141.

Warth, M. (2007). *Aspects of the Pronunciation of Two Idiolects: Polish and Scottish English*. Master's thesis, University of Glasgow.

Watt, D. (2010). The identification of the individual through speech. In: C. Llamas & D. Watt (eds.), *Language and identities*, 76–85. Edinburgh University Press.

Weiss, P.H., Zilles, K. & Fink, G.R. (2005). When visual perception causes feeling: Enhanced cross-modal processing in grapheme-color synesthesia. *NeuroImage* 28: 859–868.

Westland, S. & Ripamonti, C. (2004). *Computational colour science using MatLab*. Wiley.

Wilding, J., Cook, S. & Davis, J. (2000). Sound familiar? *The Psychologist* 13(11): 558–562.

Witthoft, N. & Winawer, J. (2006). Synesthetic colors determined by having colored refrigerator magnets in childhood. *Cortex* 42(2): 175–183.

Wrembel, M. (2007). "still sounds like a rainbow" – A proposal for a coloured vowel chart. In: *Phonetics Teaching & Learning Conference (PTLC)*.

Wrembel, M. (2009). On hearing colours – Cross-modal associations in vowel perception in a non-synaesthetic population. *Poznan Studies in Contemporary Linguistics* 45(4): 581–598.

Wrembel, M. (2010). Sound symbolism in foreign language phonological acquisition. *Research in Language* 8: 175–188.

Wrembel, M. & Rataj, K. (2008). Sounds like a rainbow – Sound-colour mappings in vowel perception. In: A. Botinis (ed.), *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics*, 237–240. Athens: University of Athens.

Yarmey, A.D. (1991). Descriptions of distinctive and non-distinctive voices over time. *Journal of the Forensic Science Society* 31(4): 421–428.

Yarmey, A.D., Yarmey, A.L., Yarmey, M.J. & Parliament, L. (2001). Commonsense beliefs and the identification of familiar voices. *Applied Cognitive Psychology* 15(3): 283–299.

Yaro, C. & Ward, J. (2007). Searching for shereshevskii: What is superior about the memory of synaesthetes? *The Quarterly Journal of Experimental Psychology* 60(5): 681–695.