VOICE QUALITY PERCEPTIONS BY SYNAESTHETES, PHONETICIANS AND CONTROLS

Anja Moos[a,b], *David Simmons*[a] *& Rachel Smith*[b]

[a]School of Psychology; [b]Glasgow University Laboratory of Phonetics, Glasgow University, UK
anjam@psy.gla.ac.uk

ABSTRACT

Voice-induced synaesthesia is an under-researched neurological condition which leads to multisensory perceptions while hearing someone's voice. This paper is the first attempt to analyse acoustically which aspects of voice qualities (VQs) trigger consistent colour perceptions and associations in synaesthetes, phoneticians and control participants. An online experiment revealed that f0 influenced brightness and colour associations with the voice for all groups and showed some idiosyncratic patterns for synaesthetes and others for phoneticians. Synaesthetes, for example, were less influenced by f0 in their colour associations than others and phoneticians used scales for high-low and tense-relaxed according to the perceived f0 and VQ more systematically than others. A steeper spectral tilt triggered smoother, softer and more relaxed associations with the voice. A short ABX voice comparison task illustrated that overall voice recognition was similar amongst groups but synaesthetes outperformed controls in recognising a speaker in whisper.

Keywords: synaesthesia, voice quality, colour associations

1. INTRODUCTION

Synaesthesia is a neurological condition in which stimulation in one sensory modality triggers perceptions in other modalities (most commonly colour associations with sequences such as days of the week, numbers, or letters of the alphabet; see [12]). Language is the most prevalent synaesthetic inducer, whether in the spoken or written modality. Whereas much research has been performed on grapheme-colour synaesthesia (e.g. [1, 3]), we are not aware of any scientific research on voice-induced synaesthesia. In this type of synaesthesia, synaesthetic perceptions include colour visualizations and/or texture sensations triggered by the voice itself, not (only) by the words spoken.

Previous work on other types of synaesthesia suggests colour associations with vowels based on formant frequencies [7] and positioning of the articulators [4], and empirical support for these proposals is given in [8]. [13] found correlations for musical pitch and timbre with luminance (brightness) in both synaesthetes and control subjects, with synaesthetes showing higher consistency in colour associations. We hypothesise that colour and/or texture associations will, likewise, bear systematic relations with the acoustic attributes of a voice, such as pitch and timbre. Further, we hypothesise that all groups will show similar patterns of associations but that synaesthetes will have more consistent associations.

As a first step towards assessing voice-induced synaesthesia from a phonetic perspective, quantitative data (alongside qualitative data) has been collected in an online experiment to establish how synaesthetes, phoneticians and others describe voices and what colours they associate with them. Distinguishable results for different acoustic features of VQs encourage further investigation into this new topic that has a potential application in forensic phonetics, given that synaesthetes can have memory advantages and use their additional synaesthetic perceptions as mnemonics [11, 14].

2. METHODS

2.1. Participants

Participants were divided into three groups (recruited via online advertisement):

7 synaesthetes: 6 female, 1 male; mean age 34, SD 19. Nationalities are 2 American, 3 British, 1 Indian, 1 Swedish. ***10 phoneticians:*** 7 female, 3 male; mean age 40, SD 16. Nationalities are 3 American, 5 British, 1 Irish, 1 Australian. ***28 controls:*** 17 female, 11 male; mean age 23, SD 4. Nationalities are 6 American, 18 British, 2 Singaporean, 1 New Zealander, 1 Swedish.

2.2. Auditory stimuli

Recordings of two male phoneticians reading two sentences in 10 different voice qualities were used (2x2x10 design): modal, raised larynx, lowered larynx, nasal, denasal, falsetto, breathy, whisper, harsh, creak. The two sentences were "These take the shape of a long round arch with its path high above and its two ends apparently beyond the horizon" and "People look but no one ever finds it". Original sound files were recorded onto reel-to-reel tape in 1976 and 1979. They were digitized to WAV format with 11025 Hz. For ease of presentation online they were converted to MP3 with 192 kbps.

The following acoustic measurements were taken to feed into statistical analysis: mean, minimum and maximum f0 in Hz, pitch range in semitones; long-term formant distributions LTF1, LTF2, LTF3 and their SDs in Hz (cf. [9]); harmonic-to-noise ratio in dB. Following [2], we used the ratios of the amplitude of the first harmonic to the second harmonic (H1*-H2*) and to the amplitude of F3 (H1*-A3*) for measures of spectral tilt. For information on the bandwidth of F1 we used H1*-A1 (cf. [2]). Spectral tilt and bandwidth measures were taken at three points 20 ms apart during a stable portion of the vowels in "apparently" and "ever".

2.3. Response display

Eight semantic differentials related to colour, texture and other vocal attributes were given in the form of continuous horizontal sliders: dry-fluid, low-high, light-dark, hard-soft, sharp-fuzzy, colourful-grey, smooth-rough, tense-relaxed. In addition, a set of 16 colours was displayed. They consisted of the 11 focal colours (white, black, blue, green, yellow, red, grey, brown, orange, pink, and purple) plus dark green, lime green, pale pink, cyan and dark blue to have a richer set to accommodate synaesthetes' needs. Colours were measured on ten different monitors using a chromameter; the mean values of these measurements were used for statistical analysis. 16 images of textures were also displayed (not discussed further here.)

2.4. Procedure

Given the rarity of voice-induced synaesthesia, the experiment had to be conducted online. LimeSurvey [6] was used. It took approximately 1.5h. Participants were asked to use the highest-fidelity audio equipment available to them. The first sentence was presented with speakers and VQs in random order. For each stimulus, participants had to describe the voice in their own words (text field), adjust sliders for eight semantic differentials, and pick a colour and a texture. They had the option to comment on the strength of their synaesthetic experiences or associations. After an optional break the second sentence was presented with speakers and VQs in random order.

At the end there was a short ABX voice comparison task. Here one sentence was played in two different VQs by the same speaker (6 times), or in the same VQ by two different speakers (4 times) (A and B). Then the other sentence was played (X) which had the same VQ and speaker as either A or B and participants judged whether it was more similar to A or B. Finally, a synaesthesia questionnaire was completed.

3. RESULTS

As this was a first exploratory study where many acoustic measures were taken that were potentially correlated, data reduction was necessary to aid interpretable analysis. Therefore a factor analysis with promax rotation was carried out on the acoustic properties of the stimuli. Four factors together accounted for 71.2% of the variability in the data and appeared to correspond with f0 measures, LTFs, tilt/bandwidth, and pitch range respectively, cf. image files 1 and 2. For each factor, the acoustic variable that correlated best with the factor was identified (mean f0, LTF2, H1*-A3* and pitch range respectively). Their mean values per VQ are shown in Table 1. These variables were entered, along with participant group, as predictors for mixed-effects modeling. Dependent variables were luminance, redness, greenness and blueness scales for the colour choices, and the eight semantic differentials. All reported results are significant at a level of $p<.05$.

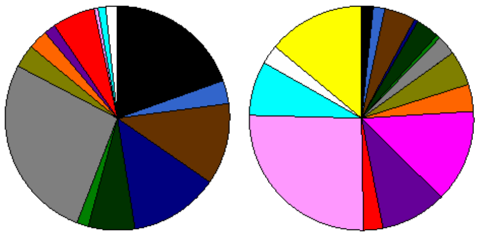
Table 1: List of four main acoustic features per voice quality. Formants for falsetto and f0 for whisper could not be measured but were replaced with the mean of all other VQs for statistical analyses.

VQ	f0 (Hz)	LTF2 (Hz)	H1*-A3* (dB)	pitch range (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

3.1. Colours and semantic differentials

Colour and luminance: On average, participants chose lighter colours for voices with higher f0, LTF2 and pitch range and a steeper spectral tilt. They also gave lighter ratings on the light-dark scale. Increasing pitch, LTF2 and pitch range also resulted in more colourful ratings on the colourful-grey scale. Higher pitch was associated with redder colours, lower pitch with greener and bluer colours. Colour choices of two sample VQs are given in Fig. 1.

Figure 1: Colour associations for creak (left) and falsetto (right), data pooled across all participants.



Textural aspects of semantic differentials: It was found that higher f0, LTF2, pitch range and a steeper tilt resulted in softer, smoother and more fluid ratings on the respective scales. However, increasing pitch and pitch range and a shallower tilt resulted in sharper ratings on sharp-fuzzy.

Other attributes: Results for the low-high scale were similar to luminance results: Increasing f0, LTF2 and pitch range triggered higher choices on the scale. For tenseness increasing pitch and LTF2, decreasing pitch range and a shallower tilt were rated as more tense on the tense-relaxed scale.

3.2. Group differences

Semantic differentials revealed a complex pattern of results. Two differential scales (sharp-fuzzy and smooth-rough) did not show any group differences at all. The other six scales showed a variety of group differences. The most important unifying theme seems to be that synaesthetes are less influenced by f0 than others.

Colour and luminance: While phoneticians and controls rated voices with low f0 greyer and with high f0 more colourful, synaesthetes were not affected by f0 in this respect. Synaesthetes rated voices greyer, the higher their LTF2; phoneticians and controls showed the reverse behaviour. Phoneticians chose darker on the light-dark scale the higher LTF2 became, whereas synaesthetes and controls showed the reverse behaviour. Synaesthetes and controls chose lighter values, the steeper the spectral tilt; phoneticians did not show an effect here.

Textural aspects of semantic differentials: Synaesthetes rated voices overall as more fluid than the other participant groups. Also, the higher LTF2, the harder synaesthetes rated the voice. The lower LTF2, the more fluid synaesthetes perceived it as, whereas controls showed the reverse pattern.

Other attributes: Overall, phoneticians rated voices as higher and tenser than other groups. Although every group associated higher values on the low-high scale with higher f0, phoneticians did this to a more extreme degree than others. There was hardly any effect of f0 on the tense-relaxed scale for synaesthetes and controls; but phoneticians rated the voice more relaxed for lower f0.

3.3. ABX voice comparison task

Results for the 10 voice comparisons show that phoneticians identified the correct voice 84% of the time, synaesthetes 77.1%, and controls 88.2%. Synaesthetes performed significantly worse than controls overall ($\chi^2(1)=5.70$, $p<.02$), but outperformed controls in identifying the whispered voice (86% vs. 39% correct; $\chi^2(1)=4.83$, $p<.03$; phoneticians 60%, n.s.). Synaesthetes' performance is much better than found in previous whisper research where recognition rate of unfamiliar voices is at 38% [10].

4. DISCUSSION

This is a first approach to analysing associations between voice qualities and descriptions of these using colours and semantic differentials at a group level and overall. Overall, an increase in f0 resulted in lighter, higher, more colourful, redder, less blue and green and sharper associations. Higher f0 being perceived as lighter is in line with previous findings on musical pitch [13]. The positive correlation with redness and negative one with blueness suggests a warm-cold perception of the voices ("warm" voices having a higher f0).

A bigger pitch range triggered lighter, higher, more colourful, more relaxed, more fluid, smoother, softer and sharper associations, suggesting mostly pleasant connotations. An increase in LTF2 meant higher, tenser and smoother responses. For synaesthetes and controls it also resulted in lighter responses, as has also been shown for vowel sounds [8]. A steeper tilt resulted in softer, smoother, more fluid, fuzzy and relaxed ratings. Spectral tilt and harmonic-to-noise ratio both correlated highly with the same factor, indicating that smooth, soft, fluid, fuzzy and relaxed ratings stand for increased harmonicity, whereas rough, hard, sharp, dry and tense stand for noisy, more turbulent speech.

Group differences were less prominent than expected. However a planned retest may reveal synaesthetes' associations to be more consistent over time (cf. [3]). Finding significant results for all three groups suggests that there are mechanisms used by everyone to associate colours with voices; the question arises whether there may be a synaesthetic continuum, as proposed e.g. in [13]. Further individual analysis can address this question. Some interesting group differences are to be found nonetheless. The expertise of phoneticians is reflected in two interactions: (1) Although every group associated higher values on the low-high scale with higher f0, phoneticians did this to a more extreme degree, showing the ability to judge f0 correctly by using this scale accordingly. (2) There was hardly any effect of f0 on the tense-relaxed scale for synaesthetes and controls; but phoneticians rated the voice as more relaxed the lower the f0. This likely relates to articulatory settings as the speakers probably have vocal tract tension for higher f0 and relaxed muscles to widen the pharynx for lower f0.

While others rated voices with low f0 as greyer and with high f0 as more colourful, the colourfulness rating of synaesthetes was not affected by f0. This suggests idiosyncratic colour associations with a different set of voice "features" independent of f0 for synaesthetes. Their good performance in distinguishing the two speakers in whisper also suggests that they use other features of the voice and their additional synaesthetic perceptions to characterize it. Future work will investigate whether these perceptions afford voice recognition advantages to synaesthetes, cf. [11].

There are many more subtle and fine-grained group differences in the acoustic data which need further processing for clear interpretation. But as stated by [5], there are no spaces or scales to define voice quality that are appropriate for use by everyone, because people use different internal standards and templates, so it makes it a challenging task to research this psychoacoustic phenomenon, especially including the multisensory aspect of synaesthesia. The complexity of voices is also mirrored in the complexity of the verbal responses; synaesthetes usually report more than a monochrome image for a voice. One synaesthete, for example, describes a voice as "white, yellow, drippy. still with browns and greens, they are just less apparent this time. […]" Qualitative data analysis of the verbal responses will help to interpret individual and group differences in more detail.

5. REFERENCES

[1] Beeli, G., Esslen, M., Jäncke, L. 2007. Frequency correlates in grapheme-colour synaesthesia. *Psychological Science* 18, 788-792.

[2] Hanson, H. 1997. Glottal characteristics of female speakers: Acoustic correlates. *J. Acoust. Soc. Am.* 101, 466-481.

[3] Hubbard, E., Arman, A., Ramachandran, V., Boynton, G. 2005. Individual differences among grapheme-color synesthetes: Brain-behavior correlations. *Neuron* 45, 975-985.

[4] Jakobson, R. 1962. *Selected Writings I. Phonological Studies*. The Hague: Mouton & Co.

[5] Kreiman, J., Vanlancker-Sidtis, D., Gerratt, B. 2005. Perception of voice quality. In Pisoni, D., Remez, R. (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell, 338-362.

[6] LimeSurvey. *http://www.limesurvey.org/*

[7] Marks, L. 1975. On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin* 82, 303-331.

[8] Moos, A., Smith, R., Simmons, D. In preparation. Colour and luminance associations with vowel sounds in synaesthetes and non-synaesthetes.

[9] Nolan, F., Grigoras, C. 2005. A case for formant analysis in forensic speaker identification. *Journal of Speech, Language and the Law* 12, 143-173.

[10] 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, 249-260.

[11] Rothen, N., Meier, B. 2010. Grapheme-colour synaesthesia yields an ordinary rather than extraordinary memory advantage: Evidence from a group study. *Memory* 18, 258-264.

[12] 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.

[13] Ward, J., Huckstep, B., Tsakanikos, E. 2006. Sound-colour synaesthesia: To what extend does it use cross-modal mechanisms common to us all? *Journal Cortex* 42, 264-280.

[14] Yaro, C., Ward, J. 2007. Searching for Shereshevskii: What is superior about the memory of synaesthetes? *The Quarterly Journal of Experimental Psychology* 60, 681-695.