PERCEPTION OF SPEAKER SOCIAL-INDEXICAL INFORMATION FROM LOCALISED PHONETIC VARIANTS

Ania Kubisz

University of York
ania.kubisz@york.ac.uk

ABSTRACT

The present paper investigates perceptions of speaker social-indexical information, including gender, age and social-class, from smaller phonetic segments such as gender-correlated phonetic variants. Since fundamental frequency (F0) is not the only cue to speaker gender identification, the perceptions are examined using gender-ambiguous sounding speech.

The results of the study show that while speaker social-indexical information is identifiable at the segmental level, listeners seemed to be more sensitive to certain types of indexical-information than others.

Keywords: perception, speaker social-indexical information, Tyneside English.

1. INTRODUCTION

Previous socioperceptual studies focus on identifying speaker-indexical information such as ethnicity [27, 34], geographic origin [9, 11] or personality traits [5, 8, 20]. Researchers have also investigated female and male voice identification [10, 26]. Even though it has been established that listeners are quite accurate at identifying adult female and male voices, it is still unclear how listeners identify gender in the speech signal [26]. Literature provides evidence that fundamental frequency impacts femininity and masculinity judgments [15, 26]. However, fundamental frequency is not always a decisive factor. There is an overlap of female and male pitch ranges, such that a lower-pitched female voice might be erroneously taken for a higher-pitched male voice and vice versa [10]. Furthermore, Johnson et al. [18] showed in their study that a voice judged as most stereotypically female had lower mean fundamental frequency than the non-stereotypical female voice. Also Klatt & Klatt [19] demonstrated that voices judged as typically female were not always characterised by high pitch.

Finally, it has been reported that listeners are able to distinguish male and female speakers in the absence of acoustic information normally found in speaker fundamental frequency [4, 12, 17, 21]. These findings imply that parameters of the vocal tract are not the only factors deciding whether a speaker sounds feminine or masculine, which further implies that gender-specific acoustic information does not rely heavily on fundamental frequency.

Because fundamental frequency is not the only cue to speakers' gender identification, it is hypothesised that when speaker-social information embedded in fundamental frequency is not accessible to the listener, this type of information can be identified from other cues, such as gender-correlated phonetic variants.

Therefore, this paper examines whether speaker social-indexical information can be identified at the segmental level.

This study builds on earlier research on perception of speaker-indexical information in child speech [15]. Following the findings of Foulkes et al. [15], it is hypothesised that listeners familiar with the dialect and particular variant realisations should be sensitive to speaker-indexical information carried by these variants.

A set of gender-correlated phonetic variants identified in Tyneside English were selected for the purpose of a broader study. Variants are sociolinguistically marked in terms of speaker gender, age and social class.

Perceptions of Tyneside-localised variants were compared and contrasted with perceptions of other localised variants from the wider North-East region or non-marked supra-local variants. While the broader study investigated perceptions of variants the FACE, GOAT and NURSE vowels, glottalised /p, t, k/, glottal and pre-aspirated /t/ and T-to-R, the present paper reports results for variants of the NURSE vowel.

It was decided to use Tyneside English phonetic variants in the study because Newcastle is considered to be the hub of the North East region, and as such, its dialect has been extensively researched and described [7, 13, 14, 23, 24, 30, 31, 32, 33]. Furthermore, Tyneside English is stereotypically perceived as the variety spoken in all of the North East.

2. METHOD

For the purpose of this study, speaker pitch was shifted to obtain the effect of a gender-ambiguous-sounding voice.

Single-word stimuli were used. The advantage of using single words over connected speech is that listeners can focus with greater ease on the specific type of information present in the acoustic signal [25]. At the same time, this approach allows the researcher to control for more parameters and therefore draw more reliable conclusions from the data when analysing which phonetic cues listeners rely on.

2.1. Stimuli

A total of four voices were used in this study. Two phoneticians were asked to record target stimuli using different Tyneside variants. Two other speakers recorded fillers used in the study. Speakers were in their forties and mid-twenties.

Stimuli selected for this study account for specific phonological contexts. Vowels occur in three phonological contexts: word-finally in open syllables, preceding a nasal, and preceding a fricative in one instance. For example, words in the NURSE group included: *nurse, turn, fur, blur* and *stir*.

Preliminary tests with Adobe Audition 3.0 [1] revealed that regarding the range of possible pitch manipulation and the final outcome in terms of voice naturalness, male voices gave better results than female voices. In other words, when working with male voices, it was possible to apply a wider range of pitch manipulations before the voice started to sound unnatural. The results were less optimistic for female voices which would lose their naturalness before they started to sound gender-ambiguous. Therefore, only male voices were used in this study.

The tokens were recorded in a recording studio to .wav sound files at a sampling rate of 44.1 kHz and 16 bit mono resolution. All tokens were manipulated in Adobe Audition 3.0 [1] using the Pitch Shifter function to raise pitch and obtain the effect of gender-ambiguous-sounding voice. In addition to preserving the tempo of the samples, high precision and default appropriate settings were selected. Pitch Shifter allows changes in fundamental frequency by semitones and cents, where 1 semitone is equal to 100 cents. Each token was manipulated individually between 1.0 and 4.0 semitones. Average F0 of target stimuli was 135 Hz.

The algorithm implemented by the Pitch Shifter allows the speech tempo to be preserved and the formant values to be adjusted to changes in pitch [1]. Because this study investigates perception of gender-correlated phonetic variables in the absence of gender-specific fundamental frequency, the aim was to manipulate only one of the phonetic cues, that is, fundamental frequency. Preserving tempo and adjusting formant values to changes in pitch sustained other acoustic features of the recordings. Furthermore, this approach allowed to control for pitch and draw more specific conclusions about the acoustic cues responsible for perceptions of speaker-indexical information.

All tokens were normalised for volume in Adobe Audition CS5.5 [2] using the Match Volume function. A single token was pre-selected and the remaining tokens were matched in volume to the pre-selected token using the file total root mean square power (RMS) function and limiting settings to ensure the output files were not clipped or overly loud.

Finally, stimuli were judged by sociophoneticians as being appropriate for the study, fulfilling the criteria of naturalness and gender-ambiguity.

2.2. Procedure

The experiment was conducted in laboratory conditions and administered in SurveyGizmo [28]. At the beginning of the experiment there was a short training session which made participants familiar with the types of scales used in the experiment. Participants heard four words, each with a different scale, and attempted to evaluate the speakers. A total of 531 single-word stimuli and fillers were presented via headphones at a comfortable hearing level, one at a time. Each stimulus was played once only. There were three breaks during the session, during which participants were asked to complete sudoku.

During the experiment, a visual representation of a stimulus was displayed on screen at the same time as the recording was played. Sound was played after an image and scale were loaded. The onset delay for audio was about a second.

In order to avoid visual priming, except for two instances referring to filler words, pictures excluded images of men or women. The images served as an additional element in the study, which alleviated a possible feeling of boredom.

As far as the words are concerned, these were not tested for gender bias.

Listeners were instructed to listen to each stimulus and evaluate it using a Visual Analogue Scale (VAS) slider with a 0 to 100 point scale, incrementing by 1 point and logging participant choices on the *x* axis. Listeners were also asked to go with their first impressions and to not overthink their choices.

The scales of choice were VAS, which are continuous, fine-grained scales. One of their major advantages over other types of scales is that they give the subjects more flexibility when providing subjective ratings. This, in turn, makes the analysis more precise [22].

Wording in each of the scales was colour-coded in a consistent manner for the benefit of the participant. Distinctive colours aimed to associate a particular colour with a particular scale.

Stimuli were presented in a fixed order and the slider was reset to a midpoint position on the scale after each evaluation. Additionally, the slider did not allow for stimuli to be left unrated and so, in order to proceed, participants had to move it.

Each stimulus was evaluated four times along four dimensions: perceived speaker gender – maleness and femaleness, perceived speaker age and social class. These alternatives were presented in a mixed order, in such a way, that every stimulus was rated along only one dimension per block and on all four of them in total.

2.3. Participants

Listeners who participated in the study were from Tyneside, and so they fulfilled the criterion of being familiar with the dialect under investigation. They were volunteers recruited from the undergraduate and graduate student bodies at the University of York and Newcastle University. Twenty-four female and seven male listeners participated in the study. Although the aim was to obtain a balanced sample of male and female participants, this proved difficult in practice. However, an imbalanced sample is not a problem. With the exception of three persons, whose age ranges were 25-34, the rest of the participants were ages 18-24. Each participant was paid £12 for a completed experiment.

3. RESULTS

Table 1 presents patterns of use of the NURSE vowel variants in Tyneside English for comparison with findings of the perceptual study.

Table 1: Usage patterns of the NURSE vowel variants in Tyneside English [7, 29, 33].

NURSE variant	**Most frequently used by**
[ɔː] - localised retracted variant	older WC males
[øː] - localised fronted variant	younger MC & WC females but also older WC
[ɜː] - supra-local centralised variant	used across all speaker groups

Statistical tests were carried out using the lme4 library in the software package R [6]. Linear regression mixed effects models with random slopes were performed [4]. This method allowed to account for differences between respondents and to normalise data. For all of the analyses individual participants were modelled as random effects while phonetic variants were modelled as fixed effects.

Plots presenting evaluations of perceived speaker maleness and femaleness were mirror images of one another. Statistical results reported for evaluations of perceived maleness and femaleness of the speaker were also similar. When looking at Figure 1 it can be noticed that the fronted variant [øː] was evaluated as more female sounding than the retracted variant [ɔː]. However, statistical analysis showed a significant difference only between evaluations of localised retracted and supra-local centralised variants ($p < 0.01$).

Figure 1: NURSE localised [ɔː] (■1), [øː] (■2) and supra-local [ɜː] (■3) variants -- evaluation of perceived speaker femaleness.

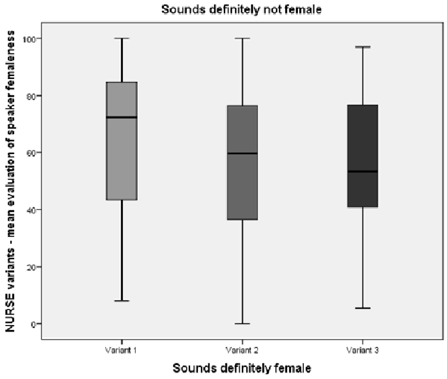
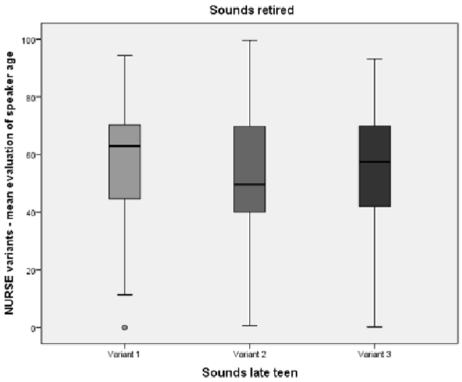
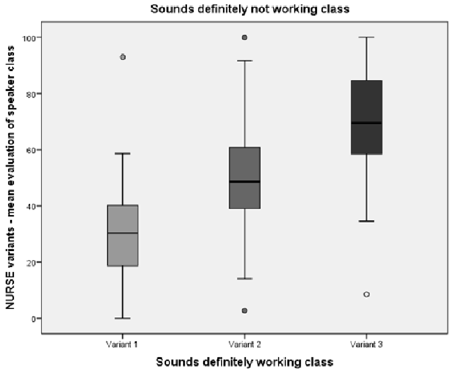


Figure 2: NURSE localised [ɔː] (■1), [øː] (■2) and supra-local [ɜː] (■3) variants -- evaluation of perceived speaker age.



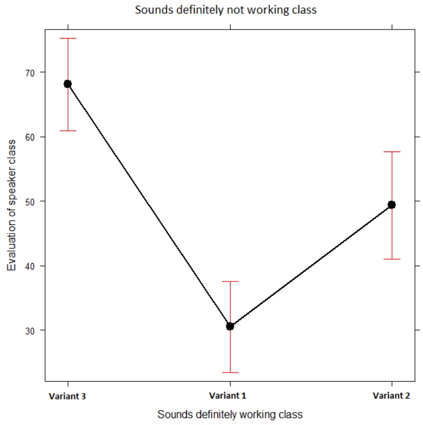
As far as evaluations of speaker age are concerned, only the medians for the male [ɔː] and female [øː] illustrate a difference in perception. While the fronted variant [øː] was evaluated as somewhat less old sounding, the retracted variant [ɔː] was found to be slightly older sounding in comparison. The supra-local variant, on the other hand, was evaluated similarly to the female variant. As could be expected from the spread of evaluations, no statistically significant differences were reported.

Figure 3: NURSE localised [ɔː] (■1), [øː] (■2) and supra-local [ɜː] (■3) variants -- evaluation of perceived speaker social class.



Results in Figure 3 show that listeners were quite sensitive to social-class information carried by the three variants. A statistically significant difference between evaluations of the two localised variants and the supra-local variant was reported ($p < 0.001$) (See Fig. 4). Furthermore, the results corroborate with findings of the production studies.

Figure 4: Results of the linear regression mixed effects models with random slopes for variants of the NURSE vowel. Variant 1 - [ɔː], Variant 2 - [øː] and Variant 3 [ɜː] -- evaluation of perceived speaker social class.



6. CONCLUSIONS

The results indicate that listeners familiar with the dialect under investigation were sensitive to speaker-indexical information on the segmental level. In fact, speaker-indexical information was extracted from phonetic segments alone. However, listeners seemed to be more sensitive to certain types of indexical-information than others.

As far as evaluations of gender of the speaker are concerned, it should be reminded that even though the voices sounded gender-ambiguous, it is often the case that localised variants are attributed to male speakers. This could explain why listeners did not evaluated the variant most frequently used by female speakers as female.

Evaluating age of the speaker seemed to pose a similar problem to the listeners. Overall, variants were evaluated as middle aged sounding.

However, listeners seemed to be quite sensitive to information about speaker social class and correctly identified patterns conditioned by social class of the speaker. It seems that after removing the cue of gender specific fundamental frequency, social class of the speaker became the most salient social-indexical feature. It may be worth mentioning at this point that a number of participants reported feeling uncomfortable having to evaluate social class of the speaker.

Perhaps different results would be obtained from participants of older age than the current group. Especially perception of speaker age could differ.

7. REFERENCES

[1] Adobe Audition 3.0 User Guide. 2007. *Adobe Systems Incorporated.*

[2] Adobe Audition CS 5.5. 2012. *Adobe Systems Incorporated.*

[3] Assmann, P. F., Nearey, T. M. 2007. Relationship between fundamental and formant frequencies in voice preference. *Acoustical Society of America* 122 (2), 35-43.

[4] Baayen, R., Davidson, D., Bates, D. 2008. Mixed-effects modeling with crossedrandom effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.

[5] Ball, P., Giles, H. 1988. Speech Style and Employment Selection: The Matched Guise Technique. In: Breakwell, G. M., Foot, H., Gilmour, R. (eds.), *Doing Social Psychology: Laboratory and Field Exercises*. Cambridge: Cambridge Univ. Press, 121–49.

[6] Bates, D., Maechler., M. 2008. lme4: Linear mixed effects models using S4 classes. R package version 3.0.1. 2008. Retrieved from http//CRAN.R-project.org.

[7] Beal, J., Burbano Elizondo, L., Llamas, C. 2012. Urban North-Eastern English: Tyneside to Teesside. *Edinburgh: Edinburgh University Press.*

[8] Bezooijen, R. Van. 1988. The Relative Importance of Pronunciation, Prosody and Voice Quality for the Attribution of Social Status and Personality Characteristics. In: Hout, R. Van, Knops, U. (eds), *Language Attitudes in the Dutch Language Area*. Dordrecht: Foris, 85–103.

[9] Bezooijen, R. Van, Gooskens, C. 1999. Identification of Language Varieties: The – Contribution of Different Linguistic Levels. *Journal of Language and Social Psychology,* 18 (1). 31-48.

[10] Biemans, M. 2000. *Gender Variation in Voice Quality.* PhD dissertation, University of Utrecht. Utrecht: LOT.

[11] Clopper, C., Conrey, B., Pisoni, D. B. 2005. Effects of Talker Gender on Dialect Categorization. *Journal of Language and Social Psychology* 24, 182-206.

[12] Coleman, R. O. 1971. Male and female voice quality and its relationship to vowel formant frequencies. *Journal of Speech and Hearing Research* 14, 565-577.

[13] Docherty, G. J., Foulkes, P. 1999. Derby and Newcastle: instrumental phonetics and variationist studies. In: Foulkes, P., Docherty, G. (eds), *Urban Voices: Accent Studies in the British Isles*. London: Arnold, 46–71.

[14] Foulkes, P., Docherty, G. J., Watt, D. 2005. Phonological Variation in Child-Directed Speech. *Language* 81 (1), 177-206.

[15] Foulkes, P., Docherty, G. J., Khattab, G., Yaeger-Dror, M. 2010. Sound judgements: perception of indexical features in children's speech. In: Preston, D., Niedzielski, N. (eds), *A Reader in Sociophonetics*. Berlin: de Gruyter, 327-356.

[16] Galbraith, S., Daniel, J. A., Vissel, B. 2010. A Study of Clustered Data and Approaches to Its Analysis. *The Journal of Neuroscience* 30 (32), 10601–10608.

[17] Hubbard, D. J., Assmann, P. F. 2013. Perceptual adaptation to gender and expressive properties in speech: The role of fundamental frequency. *Acoustical Society of America* 133 (4), 2367–2376.

[18] Johnson, K., Strand, E. A., D'Imperio, M. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27, 359-384.

[19] Klatt, D. H., Klatt, L. C. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America* 87 (2). 820-857.

[20] Lambert, W. E., Hodgsen, R. C., Gardner, R. D., Fillenbaum, S. 1960. Evaluational Reaction to Spoken Language. *Journal of Abnormal and Social Psychology* 60, 44–51.

[21] Lass, N. J., Hughes, K. R., Bowyer, M. D., Waters, L. T., Bourne, V. T. 1975. Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Acoustical Society of America* 59 (3), 675-678.

[22] Llamas, C., Watt, D. 2014. Scottish, English, British?: Innovations in attitude measurement. *Language and Linguistics Compass* 8 (11), 610-617.

[23] Milroy, J., Milroy, L., Hartley, S., Walshaw, D. 1994. Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change* 6, 327-357.

[24] Milroy, J., Milroy L., Hartley, S. 1994. Local and supra-local change in British English: the case of glottalisation. *English World-Wide* 15, 1-33.

[25] Munson, B. 2007. The Acoustic Correlates of Perceived Masculinity, Perceived Femininity, and Perceived Sexual Orientation. *Language and Speech* 50 (1), 125 – 142.

[26] Munson, B., Babel, M. 2007. Loose Lips and Silver Tongues, or, Projecting Sexual Orientation Through Speech. *Language and Linguistics Compass* 1 (5), 416–449.

[27] Purnell, T., Idsardi, W., Baugh, J. 1999. Perceptual and phonetic experiments in American English dialect identification. *Journal of Language and Social Psychology* 18, 10–30.

[28] Surveygizmo. 2014. [Online] Available at: http://www.surveygizmo.com/

[29] Watt, D. 1998. *Variation and Change in the Vowel System of Tyneside English*. PhD thesis, the University of Newcastle upon Tyne.

[30] Watt, D. 2000. Phonetic parallels between the close-mid vowels of Tyneside English: Are they internally or externally motivated? *Language Variation and Change* 12, 69–101.

[31] Watt, D. 2002. 'I don't speak with a Geordie accent, I speak, like, the Northern accent': Contact-induced levelling in the Tyneside vowel system. *Journal of Sociolinguistics* 6 (1), 44-63.7

[32] Watt, D., Allen, W. 2003. Tyneside English. *Journal of the International Phonetic Association* 33, 267 - 271.

[33] Watt, D., Milroy, L. 1999. Patterns of variation and change in three Newcastle vowels: is this dialect levelling? In: P. Foulkes and G. Docherty (eds), *Urban Voices.* London: Arnold, 25–46.

[34] Wolfram, W. 2000. On Constructing Vernacular Dialect Norms. *Chicago Linguistic Society* 36, 335–58.