Speaker discrimination based on 'facewear speech'

Natalie Fecher and Dominic Watt
Department of Language and Linguistic Science, University of York, York, United Kingdom
{natalie.fecher|dominic.watt}@york.ac.uk

Introduction. Previous behavioural and neurological research has shown that speech content and speaker-specific properties of speech are processed in a mutually dependent way. It has been reported that the extraction of indexical information encoded in the speech signal – that which helps listeners to tell one speaker apart from another – depends to a significant extent upon the segmental content of an utterance (Mullennix & Pisoni, 1990; Andics *et al.*, 2007; Cutler *et al.*, 2011). Building on these findings, the present study investigates lay listeners' ability to distinguish between two unfamiliar speakers when all they have available for comparison are /Cɑː/ syllables. In this context, we examine whether some consonants possess greater speaker-discriminating potential than others. Moreover, we explore whether speaker discrimination is further complicated when the listeners' decisions are based on 'facewear speech', namely speech that has been produced while the speaker's face is disguised by a forensically-relevant face covering. The goal of this work is to extend previous research on the influence of the segmental content of an utterance on speaker discrimination, and to offer new insights into the likely effects of facial disguise on speaker discriminability.

Method. The task of 24 participants (13F, 11M, mean age 25.2) was to make timed decisions about which pair of speech samples – out of two pairs presented in each of 432 experimental trials – were produced by the same speaker ('two-interval forced-choice' procedure). The speech material was extracted from the 'Audio-Visual Face Cover' corpus (Fecher, 2012) and was highly controlled (e.g. for amplitude, interstimulus intervals, and the occurrence of a response bias). It consisted of /Cɑː/ syllables with a systematically varying onset (/p t f s m n/). These syllables were produced by four male speakers a) in a control (no facewear) condition, b) while wearing a motorcycle helmet, and c) with a piece of tape adhered across their mouths.

Results and discussion. In total, 78.2% (*SD* = 5.5) of all speaker discriminations were correct. The listeners were able to distinguish between the speakers significantly better than chance level (50%), even under the degraded listening conditions caused by the helmet and tape (*p*s <.001). Repeated-measures ANOVA revealed a significant main effect of facewear [$F(2,46) = 234.27$, $p < .001$, $\eta_p^2 = .91$] and consonant [$F(5,115) = 9.54$, $p < .001$, $\eta_p^2 = .29$] on response accuracy, as well as a significant main effect of facewear on response time [$F(1,31) = 32.75$, $p < .001$, $\eta_p^2 = .59$]. In comparison to the near-ceiling performance achieved by the listeners in the control condition (92.6%), response accuracy dropped by 18% in the helmet and 25% in the tape condition. The reduced proportion of correct responses in the two facewear conditions, along with the significant delay in response (*p*s < .001), indicate that speaker discrimination became more difficult for the perceiver – and correspondingly more error-prone – when facewear was involved in the task. Furthermore, the consonantal content of the test syllables was found to impact quite considerably on speaker discriminability. This implies that some consonants provided more speaker-specific cues that led to successful speaker discrimination than others. Further statistical evaluation and detailed auditory/acoustic analysis of the test material provided evidence that facewear modified the articulatory and acoustic properties of speech both on the segmental and suprasegmental levels. In addition, some of the facewear-induced changes to the perceptual properties of speech (see also Fecher & Watt, 2013) appeared to manifest themselves in a speaker-specific manner (i.e., some speakers seem to have been more resistant to 'facewear effects' than others).

References

Andics, A., McQueen, J. M. & van Turennout, M. (2007). Phonetic content influences voice discriminability. *Proceedings of the 16th International Congress of Phonetic Sciences* (ICPhS), Saarbrücken, Germany, August 6–10, 2007, pp. 1829–32.

Cutler, A., Andics, A. & Fang, Z. (2011). Inter-dependent categorization of voices and segments. *Proceedings of the 17th International Congress of Phonetic Sciences* (ICPhS), Hong Kong, China, August 17–21, 2011, pp. 552–555.

Fecher, N. & Watt, D. (2013). Effects of forensically-realistic facial concealment on auditory-visual consonant recognition in quiet and noise conditions. *Proceedings of the 12th International Conference on Auditory-Visual Speech Processing* (AVSP), Annecy, France, August 29–September 1, 2013, pp. 81–86.

Fecher, N. (2012). The 'Audio-Visual Face Cover Corpus': Investigations into audio-visual speech and speaker recognition when the speaker's face is occluded by facewear. *Proceedings of the 13th Annual Conference of the International Speech Communication Association* (Interspeech), Portland, Oregon, USA, September 9–13, 2012.

Mullennix, J. W. & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics* **47**(4), pp. 379–390.