Exploiting Dutch speech collections for forensic speech science

Willemijn Heeren[1,2], Meike de Boer[1], Laura Smorenburg[1], David van der Vloed[2]
[1]*Leiden University Centre for Linguistics, Leiden University, Leiden*
[2]*Netherlands Forensic Institute, The Hague*
{w.f.l.heeren, m.m.de.boer, b.j.l.smorenburg}@hum.leidenuniv.nl, d.van.der.vloed@nfi.nl

Whereas linguistic research has mainly focused on capturing the behavior of language users in a general sense, more recently, individual variation has been taken up in speech perception and speech production modelling. In forensic speech science, individual variation is at the center of attention; how do individual speakers compare to the general population in their speaking behavior?

In forensic speaker comparisons (FSC), one or more disputed recordings by an unknown speaker (e.g. a threat, intercepted telephone recordings), are compared to one or more recordings from the suspect in the case to assess *similarity*. The speech features included in this comparative analysis are also related to other speakers' behavior, to estimate the *typicality* of the speech features. The 'other speakers' are known as the background population, which in casework often comes down to a representative group of other same-sex speakers with a comparable language background. Ideally, background populations are available for the modelling of all relevant speech features. In practice, however, such statistics are not yet widely available; this presentation explains the issue, and how it is being dealt with in current research through the use of large speech corpora.

Over the years, many speech features have been investigated in order to quantify their speaker discriminability, or strength-of-evidence. For instance, it has been assessed how well speakers are discriminated using vowel formant information (e.g., Nolan & Grigoras 2005; McDougall 2006), fundamental frequency (e.g. Gold 2014, ch. 6), or hesitation markers (e.g. Hughes et al. 2016). Earlier work on Dutch was carried out by Van den Heuvel (1996) and Kraayeveld (1997). Speech features show variability between speakers, but also within speakers due to variation in e.g. speech style, linguistic context, and language spoken. For instance, one's pitch is generally higher when excited than when sad, meaning that when emotions vary between recordings in an analysis, differences in pitch between recordings may *not* be indicative of different speakers. Also, between speech styles, specific speaker features may not be available to the same degree. In general, it is unclear how the amount of speaker-dependent information in speech features (and thus their usability in FSC) changes under different circumstances.

In order to shed more light on these issues, and for the Dutch language in particular, an ongoing research project[1] investigates the interaction of linguistic and speaker information in spontaneous, conversational speech. The study assesses various possible effects on the acoustic realization of features used in FSC and on the speaker-dependent information they contain: e.g. effects of phonetic context, of word class, and first versus second language use. Through this poster presentation we intend to demonstrate several examples of how Dutch speech collections (e.g. CGN, Oostdijk 2000; D-LUCEA, Orr & Quené, 2017; NFI-FRIDA, Van der Vloed et al. 2020) contribute valuable information to advance forensic speech science. This includes results on acoustic and speaker-dependent variation, and strength-of-evidence evaluations for several features. For instance (Fig. 1a), when labial sounds follow /x/ (as in Dutch "goed", meaning *good, well*), there is more between-speaker variation and better speaker classification (21.4% correct with multinomial logistic regression) than when non-labial sounds follow /x/ (as in "geen", meaning *no*, 17.6% correct). Also (Fig. 1b), pausing behavior depends on the language one speaks, with much individual variation. This shows that when comparing filled pauses in FSC one should not ignore a language mismatch.

[1] This work is being supported by an NWO VIDI grant (276-75-010).

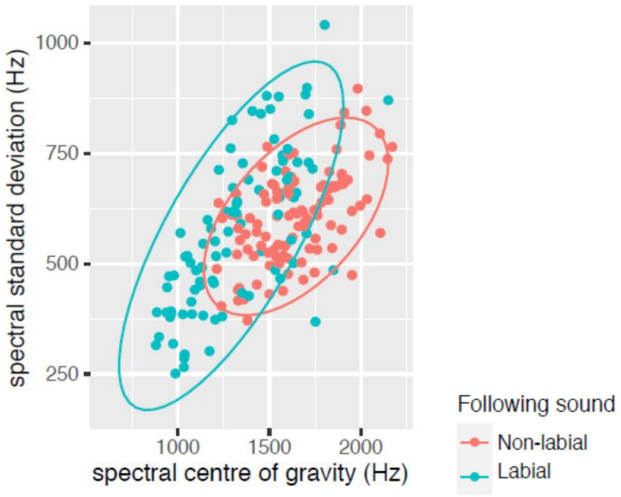

Fig. 1a. Speaker variation (N=43) for the "g"-sound /x/ (N=2,820) is affected by what type of speech sound follows /x/ (Smorenburg & Heeren, 2020).

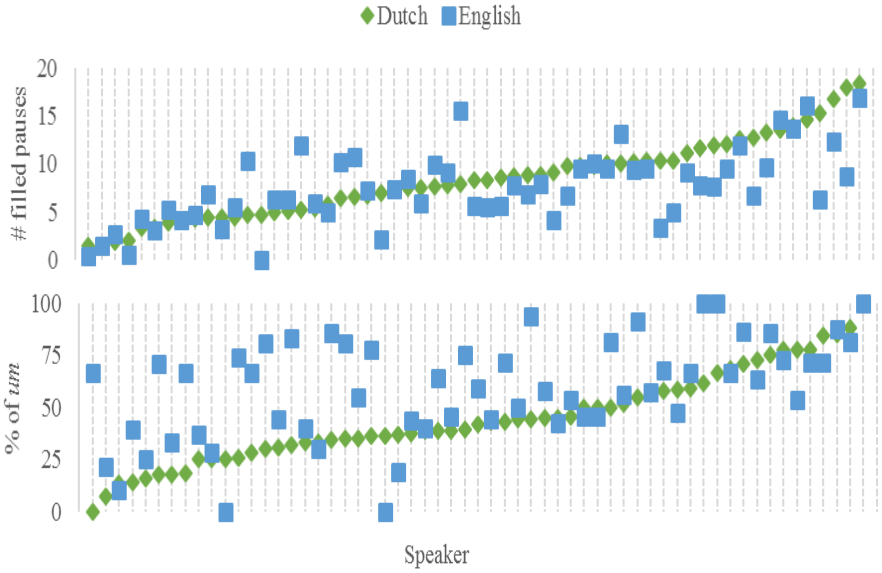

Fig. 1b. Speaker variation (N=59) for (*upper*) the number of filled pauses per minute, and (*lower*) the percentage of *um* (from the total *um + uh* tokens, n=2,101) in L1 Dutch and L2 English (speakers represented by vertical lines; sorted on Dutch), (De Boer & Heeren, 2020).

References

de Boer, M. M., & Heeren, W. F. (2020). Cross-linguistic filled pause realization: The acoustics of uh and um in native Dutch and non-native English. *The Journal of the Acoustical Society of America*, 148(6), 3612-3622.

Gold, E. (2014). *Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters.* PhD dissertation, University of York.

Hughes, V., Wood, S., & Foulkes, P. (2016). Strength of forensic voice comparison evidence from the acoustics of filled pauses. *Journal of Speech, Language and the Law*, 23(1), 99-132.

Kraayeveld, H. (1997). *Idiosyncrasy in prosody. Speaker and speaker group identification in Dutch using melodic and temporal information*. PhD dissertation, Radboud University Nijmegen.

McDougall, K. (2006) Dynamic features of speech and the characterization of speakers: towards a new approach using formant frequencies. *International Journal of Speech, Language and the Law* 13(1): 89–126.

Nolan, F., & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law Speech, Language and the Law* 12(2): 143–173.

Oostdijk, N. H. J. (2000) Het Corpus Gesproken Nederlands [The Spoken Dutch corpus]. *Nederlandse Taalkunde* 5: 280–284.

Orr, R., & Quené, H. (2017). D-LUCEA: Curation of the UCU Accent Project data. In J. Odijk & A. van Hessen (Eds.), *CLARIN in the Low Countries*, Berkeley: Ubiquity Press (pp. 177–190).

Smorenburg B.J.L & Heeren W.F.L. (2020), The distribution of speaker information in Dutch fricatives /s/ and /x/ from telephone dialogues, *Journal of the Acoustical Society of America* 147(2): 949-960.

Van den Heuvel, H. (1996) *Speaker variability in acoustic properties of Dutch phoneme realisations*. PhD dissertation, Radboud University Nijmegen.

van der Vloed, D., Kelly, F., & Alexander, A. (2020). Exploring the effects of device variability on forensic speaker comparison using VOCALISE and NFI-FRIDA: A forensically realistic database. In *Proceedings of the Odyssey Speaker and Lang. Recogn. Workshop* (pp. 402-407).