Between-speaker variability in segmental F1 dynamics in spontaneous speech

Willemijn Heeren [1] *and Lei He* [2]
[1] *Leiden University Centre for Linguistics, Leiden University, The Netherlands*
w.f.l.heeren@hum.leidenuniv.nl
[2] *Department of Computational Linguistics, University of Zürich, Switzerland*
lei.he@uzh.ch

It has repeatedly been argued that speakers differ in articulatory movements, and that the articulatory idiosyncrasy should be reflected in the speech signal (e.g. Dellwo et al. 2015, McDougall 2006), measurable in terms of formant curvature (McDougall 2006), speech rhythm (e.g. Dellwo et al. 2015, He et al. 2016), intensity dynamics (He and Dellwo 2017) and F1 dynamics (He et al. 2019). The rationale for intensity and F1 dynamics, in particular, is that intensity and F1 co-vary with the mouth opening and closing movements; calculating the speeds of intensity or F1 increases or decreases (positive or negative dynamics) may approximate such articulatory movements.

It has been shown that ~70% between-speaker variability was explained by supra-segmental measures of both intensity and F1 dynamics using Zürich German read speech (He and Dellwo 2017, He et al. 2019), suggesting that more speaker idiosyncrasy may be encoded in mouth closing movements. However, using spontaneous speech, it has been demonstrated that the differences between supra-segmental positive/negative intensity/F1 dynamics in explaining between-speaker differences seem to dwindle in Dutch, English and German (Lins Machado 2021, He and Heeren 2021). Since articulatory gestures are less controlled in spontaneous speech, supra-segmental dynamics measures calculated across utterances (mean, stdev, PVI) may be hard to interpret. Therefore, we calculated positive and negative F1 dynamics at the segmental level, including [±open] and [±rounded] vowels.

Method

Vocalic nuclei were manually annotated using Praat (www.praat.org), in spontaneous data:

1. Between 26 and 40 sentences (M = 33) were annotated from 14 speakers in English telephone conversations (DyVis corpus, Nolan 2011, task 2)
2. Between 25 and 43 sentences (M = 34) from Dutch face-to-face conversations were included from 16 gender-balanced speakers (Spoken Dutch Corpus, Oostdijk 2000)
3. 16 Zürich German sentences by each of 16 gender-balanced speakers were used from an interview with an experimenter (TEVOID corpus, Dellwo et al. 2015).

Moreover, each annotated vowel was classified as being an open/closed and unrounded/rounded vowel. F1 effects were expected to be clearest in open unrounded vowels.

The F1 trajectory of each vocalic nucleus was extracted automatically using Praat, and the F1 dynamics (F1[+] and F1[–]) were calculated following the basic procedure described in He et al. (2019). However, because different speakers each contributed a different set of vowels, mouth opening and closing speeds were computed as the raw F1 difference in Hertz over time, without F1 normalization. Only those vowels that contributed both F1[+] and F1[–] were maintained for analysis, and speeds were limited to a maximum of 1,000 Hz change per 25 ms (as speeds beyond this value were considered due to measurement errors). Thus, 74% of the Dutch and German data, and 59% of the English data were kept.

To assess between-speaker variability in F1[+] relative to that in F1[–], the variances from either measure were combined as $s^2(F1[+])/s^2(F1[–])$. This was done across vowels and per vowel class, for each of the languages. A value below 1 indicates that between-speaker variability is larger in negative dynamics. In order to assess speaker discriminability, per language, one-way ANOVAs were used to generate F-ratios with log-transformed F1[+] or

F1[–] as a dependent variable and Speaker as a random independent variable. The resulting F-ratios express the between-to-within speaker variation.

Results

Table 1 presents the ratio of the variances of F1[+] relative to F1[–] as well as the F-ratios. This is done for each of the languages, and within language, by vowel class. These preliminary results show that the result pattern in Zürich German spontaneous speech follows that in Zürich German read speech (He et al. 2019); between-speaker variability is larger in F1[–]. Also, speaker discriminability is higher in F1[–]. Dutch and English tend to not show higher between-speaker variability in F1[–], and speaker discriminability results vary; in Dutch there seems more information in F1[+], whereas for English, F1[–] and F1[+] contain comparable amounts of speaker information. When looking at the results by vowel class, F1 dynamics across languages contain most speaker discriminatory information in open-unrounded vowels, such as [a]. F1 dynamics of rounded vowels contain the least information.

Table 1. The ratios of variances and F-ratios by vowel class for F1[+] and F1[–] by language.

	Closed-rounded	Closed-unrounded	Open-rounded	Open-unrounded	*All*
Zürich German	N = 345	N = 533	N = 295	N = 508	*N = 1,681*
$s^2(F1[+])/s^2(F1[-])$	0.92	0.69	1.37	0.87	*0.85*
F1[+]	2.05	2.74*	1.70	4.90*	*4.23**
F1[–]	1.21	2.97*	3.50*	5.29*	*4.84**
Dutch	N = 486	N = 1,013	N = 218	N = 1,405	*N = 3,122*
$s^2(F1[+])/s^2(F1[-])$	0.58	1.03	1.64	1.20	*1.09*
F1[+]	2.21	3.65*	1.47	12.65*	*13.37**
F1[–]	2.02	3.21*	1.75	8.07*	*9.87**
English	N = 1,774	N= 3,769	N = 1,033	N = 4,824	*N = 11,400*
$s^2(F1[+])/s^2(F1[-])$	1.13	1.37	1.18	1.06	*1.17*
F1[+]	6.99*	10.54*	3.47*	20.11*	*33.73**
F1[–]	4.25*	11.24*	3.60*	19.63*	*32.79**

Acknowledgements This work is being supported by an IAFPA research grant to both authors, an NWO VIDI grant (276-75-010) to WH, and a UZH Forschungskredit grant (FK-20-078) to LH.

References

Dellwo, V; Leemann, A; Kolly, M-J (2015) Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *Journal of the Acoustical Society of America* 137: 1513–1528.

He, L; Dellwo, V (2016) The role of syllable intensity in between-speaker rhythmic variability. *International Journal of Speech, Language and the Law* 23: 243–273

He, L; Dellwo, V (2017) Between-speaker variability in temporal organizations of intensity contours. *Journal of the Acoustical Society of America* 141: EL488–EL494.

He, L; Heeren, W (2021) Between-speaker variability in dynamic formant characteristics in spontaneous speech. Paper presented at AISV 2021.

He, L; Zhang, Y; Dellwo, V (2019). Between-speaker variability and temporal organization of the first formant. *Journal of the Acoustical Society of America* 145: EL209–EL214.

Lins Machado, C (2021) A cross-linguistic study of between-speaker variability in intensity dynamics in L1 and L2 spontaneous speech. Paper presented at AISV 2021.

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

Nolan, F (2011) Dynamic Variability in Speech: a Forensic Phonetic Study of British English, 2006–2007, UK Data Service.

Oostdijk, N. (2000) The Spoken Dutch Corpus: Overview and First Evaluation. In LREC (pp. 887–894).