Formant variability in five Hungarian vowels with regard to speaker Discriminability

Anna Huszár[1], Valéria Krepsz[1], Alexandra Markó[2,3] and Tekla Etelka Gráczi[1,3]

[1]Hungarian Research Institute for Linguistics, [2]ELTE, Department of Applied Linguistics and Phonetics, [3]MTA-ELTE Lendület Lingual Articulation Research Group

Introduction. Formant frequencies are often used in speaker identification [e.g. 1, 2, 3], and higher (F3, F5) formants were hypothesised and found to be more distinctive among speakers [e.g. 2, 3]. [4] tested three vowels' first five formants separated and in combination of 2 and 3 of them. When one formant was included in the test F4 or F5 gave the best classification rate depending on the vowel, when two or three formants were included the combination giving the best classification rate always included one or two of F2, F4 and F5 [4]. The test and train speech materials are often of the same speech style, as their combination may raise further questions, like the possible difference on the scale of hypo-hyperarticulation [5] and speech accommodation [e.g. 6]. The present study's main questions are how sentence repetition and reading aloud may differ with regard to vowel realisations and how speaker identification is affected by the speech style difference. Our question was how the vowel formants differ across the speech styles, and if the possible difference affects the across speech styles speaker classification rate. We hypothesised that reading aloud will result in more distinct formant values. We also hypothesised that varying the vowel and formant values used for LDA will impact the classification rates.

Methods. 30 male and 30 female speakers aged between 20 and 50 years were selected from the Hungarian "BEA" database that includes the recordings of various speech styles of 461 speakers [7]. The recordings were carried out in a sound attenuated room by an AT4040 microphone. Two speech styles were selected: (i) repeated and (ii) read sentences. The sentences were the same in the two tasks. In the first task, the interviewer reads these one-by-one and the speaker repeats them by ear. In the second task, the speaker reads the sentences aloud. The same 20-20 CVC-realisations of the 5 most frequent Hungarian vowel phonemes /ɒ a ɛ e i/ were analysed in the two tasks (the 60% of the V-realisaions appeared in opened syllables and 50% of them in the first syllable that is the position of the word stress). Altogether 200 vowel realisations per speaker. The first five formants were measured at the mid 50% of the vowels' duration automatically by a Praat script [8]. The basic settings were 5 formants in 5500 Hz for women, and 5 formants in 5000 Hz for men. However, in the cases where more than 2 false values appear in a vowel quality in the same formant the settings were changed to the most fitting found by manual measurement probes. The possible differences between the speech styles were tested by linear mixed models [9, 10], the speaker discriminability was tested by LDA [11] in R [12]. The train set was one of the tasks, the test set the other task. All 5 vowels, one-one vowel was selected for training and testing in each model.

Results. The interaction of the task and the vowel was significant in the case of F1, F2 and F3 in women with showing smaller difference between the vowels, thus centralisation, hypoarticulation in repeated sentences than in reading. F4 and F5 were significantly higher in repeated sentences in the women's speech. F2, and F3 were significantly higher in the repeated sentences in men's pronunciation. F1, F4 and F5 did only show significant differences across vowels. Each formant was significantly different across vowels in both groups. The variability of F1 of /ɛ/ was considerably low among the speakers in both gender groups. The LDA were built with all 5 formants, and also with selecting 4, 3, 2 and 1 formant(s).

The classification rates of the LDA-models were higher for the vowels tested separately,as expected, but especially when including F1 and F2. The formant combination resulting in the best classification rate was vowel dependent, and was higher in male speakers in general. F5 did rarely contribute to the highest classification rates in contradiction to the results of [4].

Discussion & conclusions. The first three formants were less distinct among the vowels in women's pronunciation than in men's speech that may mean that women tend to hypoarticulate and centralise their vowels in repeated speech. The higher classification rates in general found testing men's vowel production may result from this difference. The higher classification rates in models including only one out of the analysed vowels evidently arises from the of the formant values' dependence on vowels. The lower classification rates for F5 than for lower vowels may appear due to the less accurate measurement and higher number of missing values due to automatic measurements. The results show in general that the speech style is an important factor in speaker identification.

References
[1] Jessen, M. 2008. Forensic Phonetics. Lang Linguist Compass 2(4), 671–711.
[2] Jessen, M. 1997. Speaker-specific information in voice quality parameters. Forensic Linguistics 4(1), 84–103.
[3] Nolan, F. 1983. The phonetic bases of speaker recognition. Cambridge: Cambridge University Press.
[4] Cao, H., Dellwo, V. 2019. The role of the first five formants in three vowels of mandarin for forensic voice analysis. In: International Congress of Phonetic Sciences, Melbourne, 5 August 2019 - 9 August 2019, 617-621.
[5] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the h&h theory. In: Hardcastle, W. J., Marchal, A. (eds.). Speech Production and Speech Modelling. 403-439.
[6] Eskénazi, M. 1993. Trends in speaking style research. Keynote speech. In: Proceedings of Eurospeech'93, Berlin. 501–509. http://www.cs.cmu.edu/~max/#_-_Publications.
[7] Neuberger, T., Gyarmathy, D., Gráczi, T. E., Horváth, V., Gósy, M., Beke, A. 2014 Development of a large spontaneous speech database of agglutinative Hungarian language. In: 17th International Conference, TSD 2014, September 8-12, 2014., Brno.
[8] Boersma, P., Weenink, D. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.1.24, retrieved 29 September 2020 from http://www.praat.org/
[9] Bates, D., Mächler, M., Bolker, B., Walker, S. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67, 1–48.
[10] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software 82(13), 1–26. http://doi.org/10.18637/jss.v082.i13.
[11] Venables W. N., Ripley, B. D. 2002. Modern Applied Statistics with S, Fourth edition. Springer, New York. ISBN 0-387-95457-0, http://www.stats.ox.ac.uk/pub/MASS4/.
[12] R Core Team 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.