Are temporal features of voice identity influenced by jaw size?

Daniel Friedrichs, Volker Dellwo

Department of Computational Linguistics, University of Zurich, Switzerland
daniel.friedrichs@uzh.ch, volker.dellwo@uzh.ch

Preferred presentation format: poster

Anatomical properties of the articulators impact the qualitative-spectral characteristics of speech (Fant, 1960). Here, we asked whether jaw anatomy can explain temporal dynamics of the speech amplitude envelope, which have been shown in prior research to vary between talkers (He & Dellwo, 2017). Two groups of adult talkers (N=7 each) with relatively short or long mandibles (mean length group 1: 17cm, s.d.=0.5; group 2: 18.5cm, s.d.=0.4cm) produced consonant-vowel (CV) repetitions for at least five seconds at a comfortable and maximum speech rate. CV combinations consisted of a labial consonant and the back vowel /a/ (/ma/, /ba/, /fa/) to ensure a reasonable degree of mandibular movement. Electromagnetic articulography (EMA) was used to record jaw movements during mouth opening and closure. Articulatory kinematics and the speech amplitude envelope were analyzed within a time window of four seconds starting at production onset. Results showed that talkers with shorter mandibles produced syllables at a significantly higher rate than those with longer mandibles in the fast speech condition, while no differences were found between groups for the comfortable speech rate. Reconstruction of jaw kinematics revealed that talkers with longer mandibles needed considerably more time to raise their mandibles (mouth closure), which resulted in a significantly longer amplitude fall time. These findings suggest that the jaw anatomy influences the syllabic and supra-segmental timing of speech. Further analyses are required to determine whether motor control changes after reaching an articulatory target, but it seems likely that raising the mandible is affected by other factors such as its resonant frequency in fast speech (cf. He & Dellwo, 2017). Knowledge about anatomical properties encoded in the speech signal is relevant in explaining speaker-specific timing characteristics but may also play a role in understanding how segmental timing evolved in the world's languages (cf. Blasi et al., 2019).

[Abstract Word count: 297]

References

Blasi, D. E., Moran, S., Moisik, S. R., Widmer, P., Dediu, D., & Bickel, B. (2019). Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363(6432), eaav3218. https://doi.org/10.1126/science.aav3218

Fant, G. (1960). Acoustic theory of speech production. Mouton, The Hague, The Netherlands.

He, L. & Dellwo, V. (2017). Between-speaker variability in temporal organizations of intensity contours. *The Journal of the Acoustical Society of America*, 141(5), EL488–EL494. https://doi.org/10.1121/1.4983398