Voice quality as a function of information density and prosodic factors

Erika Brandt, Bistra Andreeva, Bernd Möbius
[1]Language Science and Technology, Saarland University, Saarbrücken
ebrandt@coli.uni-saarland.de

This study investigated the influence of information density (ID) on cepstral peak prominence (CPP) and CPP-smoothed (CPPS) in German content words. CPP measures the difference in amplitude (in dB) between the cepstral peak and the corresponding fundamental frequency. CPP(S) correlate well with perceived breathiness and hoarseness [3, 5]. Speech signals with well-defined harmonic structure show prominent peaks, and thus higher CPP(S) values than signals with less well-defined harmonic structure [4]. We expected to find higher CPP values in vowels that were difficult to predict from the context, and that appeared in low-frequency words. As controls, primary lexical stress, prosodic boundary, articulation rate, average vowel duration, and sentence position were used.

Vocalic segments (n = 40,203) of the Siemens Synthesis corpus (SI1000P) [7] were fed into the CPPS analysis tool [5] using the default settings for sustained vowels. CPP is calculated every 10 ms, and then averaged for every signal. CPPS is measured every 2 ms and then averaged and smoothed in cepstral magnitude across quefrency bins.

ID factors were surprisal ($S(unit_i) = -log_2 P(unit_i|context)$) and word frequency. Surprisal values for the preceding and following context were calculated from a n-phone language model using SRILM [8]. As a text corpus for language modeling and word frequency counting SDeWaC was preprocessed using the g2p tool in German-Festival [1]. Articulation rate (phonemes / second) was calculated excluding pauses on the sentence (global) and word level (local). Primary lexical stress (stressed vs. unstressed) was based on the canonical transcription of the SI1000P corpus. Boundary was defined as word, phrase or no boundary. Statistical analysis was performed with lme4 [2] and lmerTest [6].

We found a significant positive effect of biphone surprisal of the preceding context on CPP(S), and a significant effect of triphone surprisal of the following context for CPP. There was only a tendency for a negative effect of word frequency. Vowels immediately preceding both boundary positions showed significantly lower values in both CPP(S). Primary lexical stress was not significant, however, in interaction with triphone surprisal it had a positive effect on both metrics. Vowels in sentences at fast global speech rate showed lower CPP(S) values than at slow tempo. The opposite effect was observed for local speech rate. Average vowel duration had a significant strong positive effect on CPP. For CPPS, however, we only found a tendency for this effect. This result was due to the durational averaging that was part of the smoothing procedure. As expected, vowels in the last word of a sentence showed less well-defined harmonic structure than vowels in words with non-final position. To conclude, ID and voice quality were related, while controlling for other variables: Vowels that were difficult to predict showed less breathiness and hoarseness than easily predictable vowels.

References

[1] Marco Baroni and Adam Kilgarriff. “Large Linguistically-processed Web Corpora for Multiple Languages”. In: *Proceedings of the Eleventh Conference of the European Chapter of the Association for Computational Linguistics*. EACL ’06. Trento, Italy: Association for Computational Linguistics, 2006, pp. 87–90.

[2] Douglas Bates et al. “Fitting Linear Mixed-Effects Models Using lme4”. In: *Journal of Statistical Software* 67.1 (2015), pp. 1–48.

[3] Yolanda D. Heman-Ackah, Deirdre D. Michael, and George S. Goding. “The relationship between cepstral peak prominence and selected parameters of dysphonia”. In: *Journal of Voice* 16.1 (2002), pp. 20–27. ISSN: 08921997.

[4] James Hillenbrand, Ronald A. Cleveland, and Robert L. Erickson. “Acoustic Correlates of Breathy Vocal Quality”. In: *Journal of Speech Language and Hearing Research* 37.4 (1994), p. 769. ISSN: 1092-4388.

[5] James Hillenbrand and Robert A. Houde. “Acoustic Correlates of Breathy Vocal QualityDysphonic Voices and Continuous Speech”. In: *Journal of Speech, Language, and Hearing Research* 39.2 (1996), pp. 311–321. ISSN: 1092-4388.

[6] Alexandra Kuznetsova, Per B. Brockhoff, and Rune H. B. Christensen. “lmerTest Package: Tests in Linear Mixed Effects Models”. In: *Journal of Statistical Software* 82.13 (2017), pp. 1–26.

[7] Florian Schiel. *Siemens Synthesis Corpus - SI1000P*. 1997. URL: `https://www.phonetik.uni-muenchen.de/Bas/BasSI1000Peng.html`.

[8] Andreas Stolcke. “Srilm — an Extensible Language Modeling Toolkit”. In: *Interspeech* 2.Denver, Colorado (2002), pp. 901–904.