

This is a title and this is too

Stefano Coretta¹

The University of Manchester^{a)}

¹ Put your abstract here.

^{a)} stefano.coretta@manchester.ac.uk; other info

I. METHOD

A. Speakers

B. Equipment

The acquisition of the audio signal was achieved with the software Articulate Assistant Advanced™ (AAA, v2.17.2) running on a Hewlett-Packard ProBook 6750b laptop with Microsoft Windows 7, with a sample rate of 22050 MHz (16-bit) in a proprietary format. A FocusRight Scarlett Solo pre-amplifier and a Movo LV4-O2 Lavalier microphone were used for audio recording.

C. Materials

The target stimuli were disyllabic words with $C_1V_1C_2V_2$ structure, where $C_1 = /p/$, $V_1 = /a, o, u/$, $C_2 = /t, d, k, g/$, and $V_2 = V_1$ (e.g. /pata/, /pada/, /poto/, etc.). The lexical stress of the target words was placed by speakers of both Italian and Polish on V_1 , as intended. The make-up of the target words was constrained by the design of the experiment, which included ultrasound tongue imaging (UTI). Front vowels are difficult to image with UTI, since their articulation involves tongue positions which are particularly far from the ultrasonic probe, hence reducing the visibility of the tongue contour. For this reason, only central and back vowels were included. Since one of the variables of interest in the exploratory study was the closing gesture of C_2 , only lingual consonants were used. A labial stop was chosen as the first consonant to reduce possible coarticulation with the

following vowel (although see [Vazquez-Alvarez and Hewlett 2007](#)). The target words were embedded in a frame sentence, *Dico X lentamente* ‘I say X slowly’ in Italian, and *Mówię X teraz* ‘I say X now’ in Polish. These sentences were chosen in order to keep the placement of stress and emphasis similar across languages, so to ensure comparability of results.

D. Procedure

The participant was asked to read the sentences with the target words which were sequentially presented on the computer screen. The order of the sentence stimuli was randomised for each participant. Each participant read the list of randomised sentence stimuli six times. Due to software constraints, the order of the list was kept the same across the six repetitions within each participant. Each speaker read a total of 72 sentences, with a grand total of 576 tokens (288 per language). The reading task lasted between 15 and 20 minutes, with optional short breaks between one repetition and the other.

E. Data processing, measurements, and statistical analysis

The audio recordings were exported from AAA in `.wav` format for further processing. A forced aligned transcription was accomplished through the SPeech Phonetisation Alignment and Syllabification software (SPPAS) ([Bigi, 2015](#)). The outcome of the automatic annotation was manually corrected when necessary, according to the criteria in Table I. The releases of C1 and C2 were detected automatically by means of a Praat scripting implementation of the algorithm described in [Ananthapadmanabha et al. \(2014\)](#). The durations in milliseconds of the following intervals were extracted from the annotated acoustic landmarks with Praat

TABLE I. List of measurements as extracted from acoustics.

landmark		criteria
vowel onset	(V1 onset)	appearance of higher formants in the spectrogram following the burst of /p/ (C1)
vowel offset	(V1 offset)	disappearance of the higher formants in the spectrogram preceding the target consonant (C2)
consonant onset	(C2 onset)	corresponds to V1 offset
closure onset	(C2 closure onset)	corresponds to V1 offset
consonant offset	(C2 offset)	appearance of higher formants of the vowel following C2 (V2); corresponds to V2 onset
consonant release	(C1/C2 release)	automatic detection + manual correction (Ananthapadmanabha et al., 2014)

scripting: vowel duration (V1 onset to V1 offset), consonant closure duration (V1 offset to C2 burst), and C1 release to C2 release duration (Rel-Rel duration).

The durational measurements were analysed with linear mixed-effects models using `lme4` in R ([Bates et al., 2015](#); [R Core Team, 2017](#)).

II. RESULTS

Only the most relevant terms will be presented. For the others see tables and appendixes.

A. Vowel duration

A linear mixed-effects model was fitted with the following terms: vowel duration as the outcome variable; fixed effects for C2 voicing, C2 place of articulation, vowel identity, language, and speech rate (as syllables per second); by-speaker and by-word random intercept with by-speaker random slopes for C2 voicing. All logical interactions were included. According to t-tests with Satterthwaite's approximation of degrees of freedom, the following terms and interactions were significant: C2 voicing, C2 place of articulation, vowel identity, language, speech rate, the interaction between vowel and C2 voicing, and vowel and C2 place. Vowels are 14 ms longer when followed by a voiced stop, although vowel identity enters in an interaction with C2 voicing. The effect of voicing seems to be greater for /a/ and smaller for /u/, with /o/ having an intermediate effect. Polish has on average shorter vowels than Italian (-25.5 ms), although the effect of voicing is estimated to be the same in both languages. /u/ is 13.5 ms shorter when followed by a velar stop. The effect of C2 place on /a/ and /o/ is smaller. Speech rate has a negative effect on vowel duration, such that faster rates correlate with shorter vowel durations.

B. Consonant closure duration

The same maximally specified model as with vowel duration has been fitted to consonant closure durations as the outcome variable. The following terms and interactions were significant: C2 voicing, C2 place of articulation, vowel identity, language, and interactions between language and C2 place, language and vowel identity, C2 voicing and place, C2 voic-

ing and vowel, and a three-way interaction between C2 voicing, place and vowel identity. Stop closure is 15 ms shorter if the stop is voiced. Vowel identity has an effect on closure duration in voiced stops, but not in voiceless stops, and more so in voiced velar than in voiced coronal stops: closure after /a/ is the shortest, while after /u/ is the longest, with closure after /o/ in the middle.

C. Release to Release interval duration

The duration of the interval between the release of C1 and the release of C2 is not affected by C2 voicing.

III. DISCUSSION

ACKNOWLEDGMENTS

Thanks to...

APPENDIX A: OPTIONAL APPENDIX

Ananthapadmanabha, T. V., Prathosh, A. P., and Ramakrishnan, A. G. (2014). “Detection of the closure-burst transitions of stops and affricates in continuous speech using the plosion index,” The Journal of the Acoustical Society of America **135**(1), 460–471.

- 83 Bates, D., Mächler, M., Bolker, B., and Walker, S. **(2015)**. “Fitting linear mixed-effects
84 models using lme4,” *Journal of Statistical Software* **67**(1), 1–48.
- 85 Bigi, B. **(2015)**. “SPPAS - Multi-lingual approaches to the automatic annotation of speech,”
86 *The Phonetician* **111–112**, 54–69.
- 87 R Core Team **(2017)**. “R: A language and environment for statistical computing” [https:](https://www.R-project.org/)
88 [//www.R-project.org/](https://www.R-project.org/).
- 89 Vazquez-Alvarez, Y., and Hewlett, N. **(2007)**. “The ‘trough effect’: an ultrasound study,”
90 *Phonetica* **64**(2-3), 105–121.