

Voicing in Speech Production and Perception

Institute of Phonetics and Speech Processing, Munich, Germany

18th November 2015

Luis M.T. Jesus

Institute of Electronics and Informatics Engineering of Aveiro (IEETA) and School of Health Sciences (ESSUA), University of Aveiro, Aveiro, Portugal

lmj@ua.pt; <http://sweet.ua.pt/lmj>

Summary

Speech science has long tried to integrate a multitude of factors from physical sciences to formulate empirically based theories about the human communication system and data driven (and/or knowledge based) models of speech production and perception. This oral presentation attempts to bridge the knowledge gap between the information that we can derive from various measures of speech production (acoustic, aerodynamic and articulatory) and speech perception. The focus will be on the production and perception of voicing in speech produced by the interaction of two simultaneous sources (voicing and noise) which have a very different nature. The presence of both phonation and frication in these mixed-source sounds offers the possibility of mutual interaction effects, with variations across place of articulation. The acoustic and articulatory consequences of these interactions are seldom explored and few automatic techniques for finding parametric and statistical descriptions of these phenomena have been proposed so far. Recent results from research that attempts to link speech production and speech perception of fricatives and stops, conducted at University of Aveiro's Speech, Language and Hearing Laboratory (SLHlab) in Portugal, will be presented.

Motivation and Rationale

Current understanding of aerodynamic (Koenig, Fuchs, & Lucero, 2011; Shadle, 2010, 2012; Solé, 2010), articulatory (Loucks, Shosted, De Nil, Poletto, & King, 2010; Proctor, Shadle, & Iskarous, 2010) and acoustic interactions (Iskarous, 2012; Pape, Jesus, & Birkholz, 2015) that govern the production principles involved in voicing of speech sounds (Gobl & Chasaide, 2010; Shadle, 2010, 2012), particularly voicing during consonant production (Jesus & Jackson, 2008; Ohala & Solé, 2010; Pape, Jesus, & Perrier, 2012; Pape & Jesus, 2015; Shadle, 2010, 2012; Zygis, Fuchs, & Koenig, 2012), is still limited. The phonetic realisation of voicing in different languages is highly variable (Jesus & Jackson, 2008; Keating & Kuo, 2012; Pape & Jesus, 2014b, 2015; Recasens & Mira, 2012), and most definitions of voicing are based on properties of the acoustic signal and use articulatory terms (Pinho, Jesus, & Barney, 2012). Despite the fundamental interaction of voicing mechanisms with supralaryngeal configurations and airflow



(Proctor et al., 2010), the differences in aerodynamic behaviour have rarely been used to investigate voicing in continuous speech productions (Pinho et al., 2012). Measurements of this kind might lead, in the longer term, to a more in-depth understanding of the conditions required for the maintenance or cessation of glottal vibrations (Pinho et al., 2012). The use of stimuli in a rich variety of contexts (resulting in multiple within word and crossword interaction effects) in speech research (Shue et al., 2010), reveals details about production mechanisms resulting from real physiological conditions and demands placed upon the speech system, and extends voice measures beyond the common clinical focus of sustained vowels (Jesus, Martinez, Hall, & Ferreira, 2015; Pinho, Jesus, & Barney, 2013). Qualitatively and quantitatively defining non-modal voicing (Esposito, 2010b) based on factors more closely related to phone production (laryngeal behaviour) than to the acoustic signal (Garellek & Keating, 2011), could facilitate the exploration of relationship between laryngeal activity (Kreiman et al., 2012) and the observed electroglottographic (EGG) signal (Herbst, Fitch, & Švec, 2010; Mooshammer, 2010; Pinho et al., 2012, 2013; Recasens & Mira, 2012). New findings on acoustic correlates of prosody (Gobl & Chasaide, 2010; Shue et al., 2010) related to voice quality and the role of the subglottal system (Arsikere, Lulich, & Alwan, 2011; Lulich et al., 2012; Lulich, Alwan, Arsikere, Morton, & Sommers, 2011) in vocal fold vibration unveiled novel physiological and acoustical characteristics of the voice source.

The literature is sparse concerning the differing contributions (across languages) of acoustic parameters or auditory features for voicing distinction (Broersma, 2010). It is generally agreed that voice onset time (VOT) is the most dominant perceptual cue for voicing distinction in stops. However, analyses of real speech data also show that for a significant number of productions there is no audible release (Lousada, Jesus, & Hall, 2010; Lousada, Jesus, & Pape, 2012; Pape & Jesus, 2015). The missing burst forces the perceptual system to rely on other voicing parameters/cues to extract and perform the given voicing distinction task (Alwan, Jiang, & Chen, 2011; Pape & Jesus, 2014a, 2014b). Thus, the following questions could be raised, especially when seen in the light of cross-linguistic perceptual research: How does the perceptual system choose important voicing cues, and how is weighting among the available cues mediated? For vowel perception, it has been shown that the human perceptual system is not only able to perform certain weighting techniques between cues (i.e., to apply cue-trading) in order to achieve a robust perceptual outcome, but, in addition, this weighting differs across different dialects and languages (Esposito, 2010a; Garellek, 2012). For the perception of obstruent voicing (Li, Menon, & Allen, 2010; Li, Trevino, Menon, & Allen, 2012; Pape et al., 2015; Pape & Jesus, 2014a; Silbert, 2012), this cue weighting is assumed to be highly language-dependent (Smith & Hayes-Harb, 2011; Smith & Peterson, 2012; Weber, Broersma, & Aoyagi, 2011). While some languages merely rely on the strong cues like VOT, other languages rely on voicing maintenance or closure and vowel duration cues instead (Pape et al., 2015; Pape & Jesus, 2014a, 2014b). Thus, when comparing different languages, a number of different acoustic parameters have to be taken into account when examining the cue mediation for voicing distinction (Broersma, 2010; Shultz, Francis, & Llanos, 2012) and the perception of voice quality (Bishop & Keating, 2012; Kreiman, Gerratt, & Khan, 2010; Kreiman & Gerratt, 2010, 2012).



References

- Alwan, A., Jiang, J., & Chen, W. (2011). Perception of place of articulation for plosives and fricatives in noise. *Speech Communication*, 53(2), 195–209. <http://doi.org/10.1016/j.specom.2010.09.001>
- Arsikere, H., Lulich, S. M., & Alwan, A. (2011). Automatic estimation of the first subglottal resonance. *The Journal of the Acoustical Society of America*, 129(5), EL197. <http://doi.org/10.1121/1.3567004>
- Bishop, J., & Keating, P. (2012). Perception of pitch location within a speaker's range: Fundamental frequency, voice quality and speaker sex. *The Journal of the Acoustical Society of America*, 132(2), 1100. <http://doi.org/10.1121/1.4714351>
- Broersma, M. (2010). Perception of final fricative voicing: Native and nonnative listeners' use of vowel duration. *The Journal of the Acoustical Society of America*, 127(3), 1636. <http://doi.org/10.1121/1.3292996>
- Esposito, C. M. (2010a). The effects of linguistic experience on the perception of phonation. *Journal of Phonetics*, 38(2), 306–316. <http://doi.org/10.1016/j.wocn.2010.02.002>
- Esposito, C. M. (2010b). Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association*, 40(02), 181–198. <http://doi.org/10.1017/S0025100310000046>
- Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics*, 40(1), 152–161. <http://doi.org/10.1016/j.wocn.2011.10.003>
- Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, 41(02), 185–205. <http://doi.org/10.1017/S0025100311000193>
- Gobl, C., & Chasaide, A. N. (2010). Voice source variation and its communicative functions. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 378–423). Chichester: Blackwell.
- Herbst, C. T., Fitch, W. T. S., & Švec, J. G. (2010). Electroglottographic wavegrams: A technique for visualizing vocal fold dynamics noninvasively. *The Journal of the Acoustical Society of America*, 128(5), 3070. <http://doi.org/10.1121/1.3493423>
- Iskarous, K. (2012). Articulatory to Acoustic Modeling. In A. Cohn, C. Fougerson, & M. Hoffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 472–483). Oxford: Oxford University Press.
- Jesus, L. M. T., & Jackson, P. J. B. (2008). Frication and voicing classification. In A. Teixeira, V. Lima, L. Oliveira, & P. Quaresma (Eds.), *Computational Processing of the Portuguese Language* (pp. 11–20). Berlin: Springer -Verlag.

- Jesus, L. M. T., Martinez, J., Hall, A., & Ferreira, A. (2015). Acoustic Correlates of Compensatory Adjustments to the Glottic and Supraglottic Structures in Patients with Unilateral Vocal Fold Paralysis. *BioMed Research International*, Article ID, 1–9.
- Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2), 1050. <http://doi.org/10.1121/1.4730893>
- Koenig, L. L., Fuchs, S., & Lucero, J. C. (2011). Effects of consonant manner and vowel height on intraoral pressure and articulatory contact at voicing offset and onset for voiceless obstruents. *The Journal of the Acoustical Society of America*, 129(5), 3233. <http://doi.org/10.1121/1.3561658>
- Kreiman, J., & Gerratt, B. R. (2010). Perceptual sensitivity to first harmonic amplitude in the voice source. *The Journal of the Acoustical Society of America*, 128(4), 2085–2089. <http://doi.org/10.1121/1.3478784>
- Kreiman, J., & Gerratt, B. R. (2012). Perceptual interaction of the harmonic source and noise in voice. *The Journal of the Acoustical Society of America*, 131(1), 492–500. <http://doi.org/10.1121/1.3665997>
- Kreiman, J., Gerratt, B. R., & Khan, S. ud D. (2010). Effects of native language on perception of voice quality. *Journal of Phonetics*, 38(4), 588–593. <http://doi.org/10.1016/j.wocn.2010.08.004>
- Kreiman, J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J., & Alwan, A. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *The Journal of the Acoustical Society of America*, 132(4), 2625. <http://doi.org/10.1121/1.4747007>
- Li, F., Menon, A., & Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *The Journal of the Acoustical Society of America*, 127(4), 2599–2610. <http://doi.org/10.1121/1.3295689>
- Li, F., Trevino, A., Menon, A., & Allen, J. B. (2012). A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise. *The Journal of the Acoustical Society of America*, 132(4), 2663–2675. <http://doi.org/10.1121/1.4747008>
- Loucks, T. M. J., Shosted, R. K., De Nil, L. F., Poletto, C. J., & King, A. (2010). Coordinating Voicing Onset with Articulation: A Potential Role for Sensory Cues in Shaping Phonological Distinctions. *Phonetica*, 67(1-2), 47–62. <http://doi.org/10.1159/000319378>
- Lousada, M., Jesus, L. M. T., & Hall, A. (2010). Temporal Acoustic Correlates of the Voicing Contrast in European Portuguese Stops. *Journal of the International Phonetic Association*, 40(3), 261–275. <http://doi.org/10.1017/S0025100310000186>

- Lousada, M., Jesus, L. M. T., & Pape, D. (2012). Estimation of stops' spectral place cues using multitaper techniques. *DELTA*, 28(1), 1–26. <http://doi.org/10.1590/S0102-44502012000100001>
- Lulich, S. M., Alwan, A., Arsikere, H., Morton, J. R., & Sommers, M. S. (2011). Resonances and wave propagation velocity in the subglottal airways. *The Journal of the Acoustical Society of America*, 130(4), 2108–2115. <http://doi.org/10.1121/1.3632091>
- Lulich, S. M., Morton, J. R., Arsikere, H., Sommers, M. S., Leung, G. K. F., & Alwan, A. (2012). Subglottal resonances of adult male and female native speakers of American English. *The Journal of the Acoustical Society of America*, 132(4), 2592–2602. <http://doi.org/10.1121/1.4748582>
- Mooshammer, C. (2010). Acoustic and laryngographic measures of the laryngeal reflexes of linguistic prominence and vocal effort in German. *Journal of the Acoustical Society of America*, 127, 1047–1058.
- Ohala, J. J., & Solé, M. J. (2010). Turbulence and Phonology. In S. Fuch, M. Toda, & M. Zygis (Eds.), *Turbulent Sounds: An Interdisciplinary Guide* (pp. 37–101). Berlin: De Gruyter Mouton.
- Pape, D., & Jesus, L. M. T. (2014a). Cue-weighting in the perception of intervocalic stop voicing in European Portuguese. *The Journal of the Acoustical Society of America*, 136(3), 1334–1343. <http://doi.org/10.1121/1.4890639>
- Pape, D., & Jesus, L. M. T. (2014b). Production and perception of velar stop (de)voicing in European Portuguese and Italian. *EURASIP Journal on Audio, Speech, and Music Processing*, 2014(1), 6. <http://doi.org/10.1186/1687-4722-2014-6>
- Pape, D., & Jesus, L. M. T. (2015). Stop and Fricative Devoicing in European Portuguese, Italian and German. *Language and Speech*, 58(2), 224–246. <http://doi.org/10.1177/0023830914530604>
- Pape, D., Jesus, L. M. T., & Birkholz, P. (2015). Intervocalic fricative perception in European Portuguese: An articulatory synthesis study. *Speech Communication*. <http://doi.org/10.1016/j.specom.2015.09.001>
- Pape, D., Jesus, L. M. T., & Perrier, P. (2012). Constructing physically realistic VCV stimuli for the perception of stop voicing in European Portuguese. . In H. Caseli, A. Villavicencio, A. Teixeira, & F. Perdigão (Eds.), *Computational Processing of the Portuguese Language* (pp. 338–349). Berlin: Springer -Verlag.
- Pinho, C. M. R., Jesus, L. M. T., & Barney, A. (2012). Weak Voicing in Fricative Production. *Journal of Phonetics*, 40(5), 625–638. <http://doi.org/10.1016/j.wocn.2012.06.002>

- Pinho, C. M. R., Jesus, L. M. T., & Barney, A. (2013). Aerodynamic measures of speech in unilateral vocal fold paralysis (UVFP) patients. *Logopedics, Phoniatrics, Vocology*, 38(1), 19–34. <http://doi.org/10.3109/14015439.2012.696138>
- Proctor, M. I., Shadle, C. H., & Iskarous, K. (2010). Pharyngeal articulation in the production of voiced and voiceless fricatives. *The Journal of the Acoustical Society of America*, 127(3), 1507. <http://doi.org/10.1121/1.3299199>
- Recasens, D., & Mira, M. (2012). Voicing assimilation in Catalan two-consonant clusters. *Journal of Phonetics*, 40(5), 639–654. <http://doi.org/10.1016/j.wocn.2012.06.001>
- Shadle, C. H. (2010). The Aerodynamics of Speech. In W. J. Hardcastle, J. Laver, & F. E. Gibbon (Eds.), *The Handbook of Phonetic Sciences* (2nd ed., pp. 39–80). Chichester: Blackwell.
- Shadle, C. H. (2012). The Acoustics and Aerodynamics of Fricatives. In A. Cohn, C. Fougerson, & M. Huffman (Eds.), *The Oxford Handbook of Laboratory Phonology* (pp. 511–526). Oxford: Oxford University Press.
- Shue, Y.-L., Shattuck-Hufnagel, S., Iseli, M., Jun, S.-A., Veilleux, N., & Alwan, A. (2010). On the acoustic correlates of high and low nuclear pitch accents in American English. *Speech Communication*, 52(2), 106–122. <http://doi.org/10.1016/j.specom.2009.08.005>
- Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, 132(2), EL95–EL101. <http://doi.org/10.1121/1.4736711>
- Silbert, N. H. (2012). Syllable structure and integration of voicing and manner of articulation information in labial consonant identification. *The Journal of the Acoustical Society of America*, 131(5), 4076–4086. <http://doi.org/10.1121/1.3699209>
- Smith, B. L., & Hayes-Harb, R. (2011). Individual differences in the perception of final consonant voicing among native and non-native speakers of English. *Journal of Phonetics*, 39(1), 115–120. <http://doi.org/10.1016/j.wocn.2010.11.005>
- Smith, B. L., & Peterson, E. A. (2012). Native English speakers learning German as a second language: Devoicing of final voiced stop targets. *Journal of Phonetics*, 40(1), 129–140. <http://doi.org/10.1016/j.wocn.2011.09.004>
- Solé, M. J. (2010). Effects of syllable position on sound change: An aerodynamic study of final fricative weakening. *Journal of Phonetics*, 38(2), 289–305.
- Weber, A., Broersma, M., & Aoyagi, M. (2011). Spoken-word recognition in foreign-accented speech by L2 listeners. *Journal of Phonetics*, 39(4), 479–491. <http://doi.org/10.1016/j.wocn.2010.12.004>

Zygis, M., Fuchs, S., & Koenig, L. L. (2012). Phonetic explanations for the infrequency of voiced sibilant affricates across languages. *Laboratory Phonology*, 3(2). <http://doi.org/10.1515/lp-2012-0016>