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
lmtj@ua.pt; http://sweet.ua.pt/lmtj

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, Ariskere, 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


http://doi.org/10.1121/1.3632091


http://doi.org/10.1016/j.wocn.2012.06.002


