On the link between glottal vibration and sonority

Megan Risdal, Ann Aly, Adam J. Chong, Patricia Keating & Jesse Zymet
Dept. of Linguistics, University of California, Los Angeles

I. Introduction

- Source-filter interactions: degree of oral constriction (filter) can affect ease of initiating and sustaining voicing source, and its amplitude (Bickley & Stevens, 1987; Fant, 1997; Halle & Stevens, 1967; Solé, 2015; Stevens, 2000)
  - Mittal et al (2014): examined differences in strength of glottal excitation across six voiced consonants
- Degree of oral constriction argued to correlate with phonological sonority
- But previous studies of physical manifestation/quantification of sonority either:
  - Make no connection to inherent source-filter dependencies (Parker 2002), or
  - Divorce glottal state (source) from aperture (filter) (e.g. Miller 2012)
- **Research Questions:**
  1. Does strength/degree of voicing differ across different segment types?
  2. How do source-filter interactions distinguish between voiced consonants and vowels of varying degrees of constriction?
     - ➢ Replication of Mittal et al. (2014) with more degrees of constriction and vowels
  3. How do these differences correlate with standard notions of sonority?
  4. Does voicing also change during a segmental constriction? If so, how?

II. Method

- 13 participants (6 M; 7 F): trained phoneticians, fluent/highly proficient English speakers (8 AmE speakers) – one F excluded from section IV due to lack of voicing in stops
- 5 out of 6 segments from Mittal et al. (2014), plus 16 others (total = 21):

14 Consonants: 3 reps. [aCa]
- Approximants: [j, w, l, ɹ]
- Trill & Tap: [r, ɾ]
- Nasal: [n]
- Fricatives: [ð, ɣ, ʁ, z]
- Affricates & Stop: [dʒ, gɣ, d] –
  - Affricates analyzed separately as stop and fricative

7 Vowels: 3 reps. [wV]
- Front unrounded: [i, e, a]
- Front rounded: [y, ø]
- Back rounded: [o ,u]
- Collapsed together in Figs. 1-2

- Simultaneous EGG and audio signal recorded (B&K microphone)
- Analysis intervals: Intervals with at least three glottal pulses during target constriction
  - Tokens excluded if lacked three glottal pulses (n = 112 out of 897)
- **Measures** (VoiceSauce: Shue et al., 2011; EggWorks: Tehrani, 2015):
  1. **Contact Quotient (CQ)**: Prop. of vibratory cycle where vocal fold contact is higher than specified threshold – here, CQ_H (Hybrid method):
     - Contacting moment: begins at the negative peak in the dEGG signal
     - Decontacting moment: ends when the EGG signal crosses a 37% threshold.
     - This version of CQ best reflects differences in phonation in modal-to-breathy range (Kuang, 2011; Kuang & Keating, 2012)
  2. **Strength of Excitation (SoE**; Mittal et al., 2014): strength of impulse-like excitation derived from the instant of significant excitation of the vocal tract (Murty & Yegnanarayana, 2008; Yegnanarayana & Murty, 2009). Related to RMS energy, but more sensitive.
     - Signal filtered with zero frequency resonators (ZFR)
     - Slope of negative-going ZFR signal = relative amplitude of impulse-like excitation
     - Depends on both source and filter; no equivalent EGG measure

III. Results: Mean CQ (left) and SoE (right)

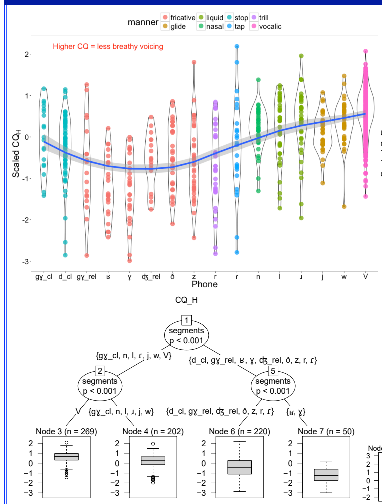


Figure 1. TOP: Scaled CQ_H across all consonants, collapsing vowels into one category. L → R = decreasing constriction degree.
BOTTOM: Conditional Inference tree partition of the data by CQ_H.

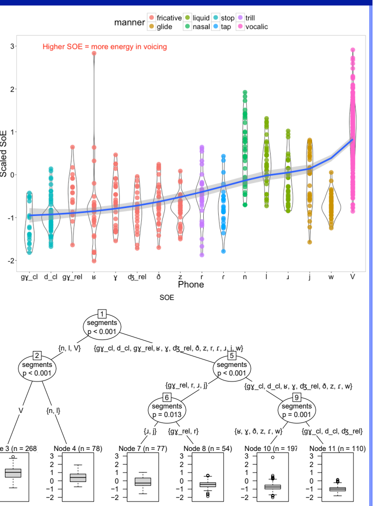


Figure 2. TOP: Scaled SoE across all consonants, collapsing vowels into one category. L → R = decreasing constriction degree.
BOTTOM: Conditional Inference tree partition of the data by SoE.

1. CQ_H (LEFT; FIG. 1)
- Voiced fricatives in general have the lowest CQ_H (lower than voiced stops)
  - Most-breathy voicing
  - Independently expected: vocal folds in fricatives are somewhat spread to maintain continued airflow needed for fricative noise (e.g. Keyser & Stevens, 2006)
- Voiced stops have lower CQ_H than nasals, liquids, glides and vowels
- Vowels as a whole have highest CQ_H
  - Least breathy voicing/most glottal contact

2. SoE (RIGHT; FIG. 2)
- Voiced fricatives > Voiced stops
  - Stronger voicing energy, despite breathier voicing
- Vowels have highest SoE
  - Strongest voicing
- Voiced stops have lowest SoE
  - Weakest voicing

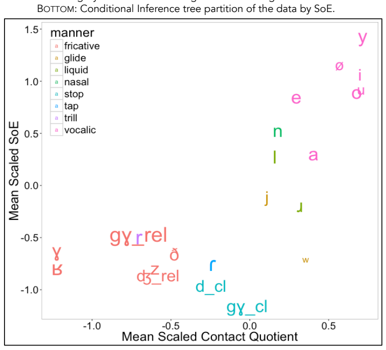


Figure 3. Two dimensional space of Scaled SoE by Scaled CQ_H by segment. Size of symbol indicates standard deviation.

- ➢ Together, CQ_H and SoE form a 2-D space (FIG. 3)
  - Accurately capture the ends of the scale
    - Vowels: Most sonorous (Highest CQ_H & SoE)
    - Voiced stops: Least sonorous (Low CQ_H & SoE)
- ➢ CQ_H and SoE make distinctions within different segmental categories
  - CQ_H makes distinctions amongst obstruents
  - SoE makes distinctions amongst sonorants

IV. Results: Timecourse of voicing (SoE)

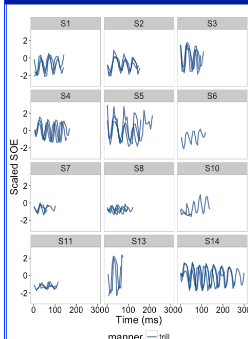


Figure 4. Scaled SoE across time by speaker and token for trills

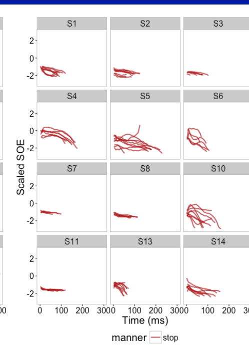


Figure 5. Scaled SoE across time by speaker and token for stops

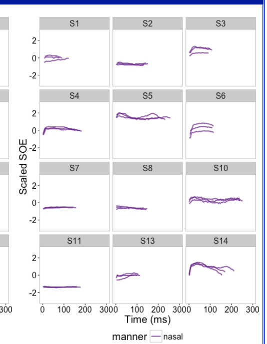


Figure 6. Scaled SoE across time by speaker and token for nasals

- Same differences seen within segments with changing conditions for voicing:
- In trills (LEFT; FIG. 4): open and close phases involve different degrees of glottal contact
  - SoE oscillations across an entire trill reflect this
  - Open phases more sonorant like – despite repeated tongue tip contact
- In stops (MIDDLE; FIG. 5): full closure gradually impedes voicing
  - SoE can drop throughout duration of voicing – as voicing becomes more difficult due to increase in supraglottal pressure
  - Voicing becomes weaker and breathier before dying out
- In contrast, in nasals (RIGHT; FIG. 6) strength of voicing is relatively stable throughout

V. Summary

1. Does voicing differ across different segment types? – Yes!
2. How do source-filter interaction distinguish between voiced consonants and vowels?
   - In general, in accord with previous work and predictions (e.g. Bickley & Stevens, 1987): the tighter the constriction, the breathier the voicing (CQ_H).
   - Also: the tighter the constriction, the weaker the voicing (SoE)
3. How do these differences correlate with standard notions of sonority?
   - At the broadest level: Vowels > approximants > obstruents
   - However, a number of reversals within each class (e.g. liquids have lower SoE than nasals)
4. Does voicing change during a segmental constriction? – Yes! How?
   - Trill: Strength of voicing oscillates with changing oral constriction
   - Stops: Voicing becomes weaker before extinguishing

Acknowledgments

We would like to thank students of the Winter 2015 Speech Production graduate course at UCLA for participating in the design and execution of the study, as well as Soo Jin Park for implementing the SoE measure in VoiceSauce. Thank you also to members of the UCLA Phonetics Lab, the audience at the 2016 CUNY Phonology Forum, and four anonymous reviewers for LabPhon 2016 for comments and discussion.

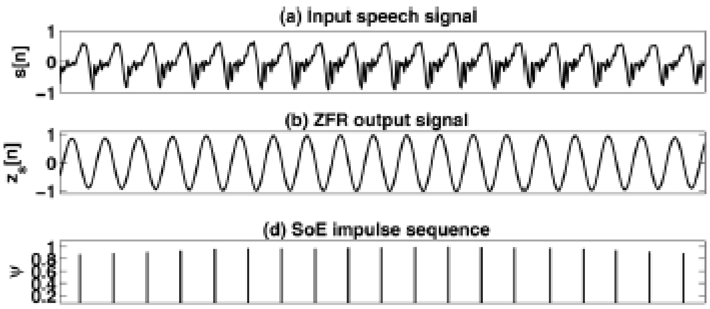
<u>Information about Strength of Excitation (SoE)</u>

Epoch: the instant of significant, impulse-like excitation of the vocal-tract system during speech production (Murty & Yegnanarayana 2008); also defined as moment of glottal closure.

- Takes place during closing phase of the glottal cycle, due to abrupt closure of vocal folds
- Groundtruth for epochs (or actual epochs) is taken to be *d*EGG closing peak, or positive-to-negative zero-crossings of the ZFR signal (signal filtered with zero-frequency resonators – see (b) below)

Strength of Excitation (SoE) (Mittal et al. 2014): the relative amplitude of impulse-like excitation at an epoch (Murty & Yegnanarayana, 2008; Yegnanarayana & Murty 2009). See (c) below.

- Measured as the slope of the ZFR signal around the epoch
- Related to closing peak in *d*EGG, except SoE reflects changes in both source and vocal-tract system characteristics (as shown by differences across segment types)
- Mittal et al. show values for a low vowel of .5 - .83, and for consonants as low as .06



<u>References</u>

Bickley, C. A. and Stevens, K. (1987). Effects of a vocal tract constriction on the glottal source: Data from voiced consonants. In Baer, T., Sasaki, C., and Harris, K., editors, *Laryngeal Function in Phonation and Respiration*, pp. 239–253. College-Hill Press, Boston, MA.

Fant, G. (1997). The voice source in connected speech. *Speech Communication*, 22:125–139.

Halle. M. and Stevens, K. (1967). On the mechanism of glottal vibration for vowels and consonants. *Quarterly Progress Report of the Research Laboratory of Electronics, MIT*, 85: 267-271.

Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, (3):651–674.

Keyser & Stevens, K. (2006). Enhancement and overlap in the speech chain. *Language*, 82: 33-63.

Kuang, J. (2011). *Production and perception of phonation contrast in Yi*. Masters thesis, UCLA.

Kuang, J. and Keating, P. (2012). Glottal articulations of phonation contrasts and their acoustic and perceptual consequences. *UCLA Working Papers in Phonetics*, 111: 123-161.

Miller, B. (2012). Sonority and the larynx. In Parker, S., editor, *The Sonority Controversy*, pp. 257–288. Walter de Gruyter.

Mittal, V. K., Yegnanarayana, B., and Bhaskararao, P. (2014). Study of the effects of vocal tract constriction on glottal vibration. *Journal of the Acoustical Society of America*, 136(4):1932–1941.

Murty, K. S. R. and Yegnanarayana, B. (2008). Epoch extraction from speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1602–1613.

Parker, S. (2002). *Quantifying the Sonority Hierarchy*. PhD thesis, UMass Amherst.

Shue, Y. L., Keating, P. A., Vicenik, C., and Yu, K. (2011). VoiceSauce: A program for voice analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences, Hong Kong*, pp. 1846–1849.

Solé, M.-J. (2015). Acoustic evidence of articulatory adjustments to sustain voicing during voiced stops. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow.

Stevens, K. (2000). *Acoustic Phonetics*. Current Studies in Linguistics Series. MIT Press.

Tehrani, H. (2012). Eggworks. http://www.linguistics.ucla.edu/faciliti/sales/software.htm

Yegnanarayana, B., and Murty, K. S. R. (2009). Event-based instantaneous fundamental frequency estimation from speech signals. *IEEE Trans. Audio, Speech, Lang. Process*. 17, 614–624.