Fine-grained phonetics and developmental universals for glottal features

Eun Jong Kong (ejkong@ling.ohio-state.edu) and **Mary Beckman** (mbeckman@ling.ohio-state.edu)
Department of Linguistics, the Ohio State University

Background

- The best substantiated of Jakobson's (1941) claims about implicational universals in acquisition concerns phonation type: **Voiceless unaspirated are produced first, and the mastery of voiced or aspirated stops implies the mastery of voiceless unaspirated ones.**
- Claim is borne out in studies of many languages.
  1) unaspirated stops are produced in babbling at 6-7 months
  2) aspirated stops mastered at about 2 years in English (Macken & Barton, 1980), Cantonese (Clumeck et al., 1981), etc.,
  3) voiced stops are not mastered until 4-5 years in French (Allen, 1985), Thai (Gandour et al, 1986), Hindi (Davis, 1995), etc.
- Validation depends on accurate phonetic characterization: e.g., if English stops in word-initial position characterized as in traditional accounts (i.e., voiced [b, d, g] in opposition to voiceless [p, t, k]) rather than as suggested by Lisker & Abramson's (1964) voice onset time data (i.e., predominately voiceless [p, t, k] in opposition to aspirated [pʰ, tʰ, kʰ]), then English is exception to claim.
- Therefore, it is important to take language-specific phonetic characteristics into account in comparing the order of acquisition of the "same" phonation types across languages.

Questions for this study

- What are the implications for Jakobson's developmental universal of more recent phonetic studies documenting much finer-grained phonetic differences across languages having purportedly identical phonetic categories?
- Is VOT the only (or even the best) way of describing the phonation types of a given language?
- Can we use a comparison of VOT values and other measures in children's productions and adults' productions to explain the developmental universal?

Data

- Word-initial coronal and dorsal stops in four languages with a two-way phonation-type contrast, elicited in a word repetition task using an audio prompt only (for adults) or using an audio prompt accompanying a picture (for children).
- Three words were used for each target C in a variety of vowel contexts (aiming for five broad V contexts). For example:
  - Coronal unaspirated [t] vs. aspirated [tʰ] elicited in Cantonese /ti:p³⁵/ 'plates', /ta:n³⁵/ 'eggs', …, vs. /tʰi:n⁵⁵/ 'sky', /tʰa:i⁵⁵/ 'ties', ..., and English *dish*, *day*, ..., vs. *tickle*, *tail*, …
  - Dorsal voiced [g] vs. voiceless [k] elicited in Greek /gazi/ 'gas', /gol/ 'goal', …, vs. /karta/ 'card', /kokalo/ 'bone', …, and Japanese /gasu/ 'gas', /gomi/ 'trash', …, vs. /karasu/ 'crow', /koara/ 'koala', … [see demo on laptop]
- Subjects were about 20 children aged 2-3 years and 3 female adult speakers for each language.

Analysis 1 — VOT

- We measured voice onset time (VOT), the latency from the burst that marked the release of the lingual closure to the onset of voicing in the closure (negative VOT = "voicing lead") or in the following vowel (positive VOT = "voicing lag").
- Histograms of all of the adult values (Fig. 1, left panels) showed expected patterns of (1) short vs. long lag values for Cantonese and (2) mostly short vs. long lag values for English.

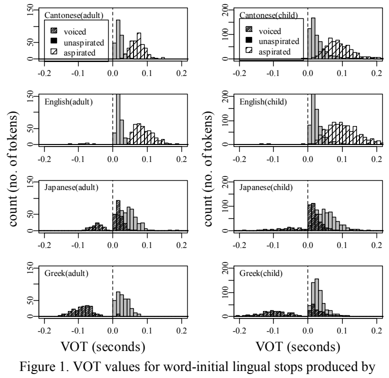

Figure 1. VOT values for word-initial lingual stops produced by adult (left panels) and child (right panels) speakers.

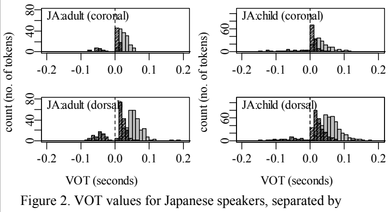

Figure 2. VOT values for Japanese speakers, separated by coronal versus dorsal place of articulation.

- However, (3) the voiceless stops of Japanese had "short" lag values that are intermediate between the truly short lag values for the unaspirated stops and the long lag values of the aspirated stops of Cantonese. Moreover, (4) the voiced stops of Japanese showed a bimodal distribution, with many short lag values that overlapped with the values for the voiceless stops. This was particularly true of the dorsal stops (Fig. 2, left panels).
- By contrast, (5) the voiced stops of Greek showed much larger negative (voicing lead) values which did not overlap at all with the values for the voiceless stops, even though (6) the voiceless stops showed prototypically short lag values.
- The children's VOT patterns (Figs. 1-2, right panels) showed distributions very much like the adult patterns, even for minutely language-specific details such as (3) the intermediate lag values of the Japanese voiceless stops and (5) the particularly long voicing lead values of Greek.

Q1) What is the acoustic cue that separates the Japanese voiced from voiceless stops in the region of overlapping VOT values?
- Note that this pattern of overlapping values is superficially like the VOT distribution in English children with "covert contrast" (Macken & Barton, 1980).

Q2) Why are the Greek children capable of producing the very long voicing lead values characteristic of the Greek voiced stops?
- Recall that these children are all much younger than the ages reported for mastery of voiced stops in French, Thai, Hindi, etc.

Analysis 2 — Burst intensity

- To answer Question 1, we took a measure of burst intensity that was the sum of energy in the region above 2000 Hz in a spectrum calculated over a 5 ms window beginning at the stop release.
- This measure was inspired by Sundara (2005), who used a peak RMS intensity measure to discriminate the voiced vs. voiceless stops of Canadian French (a dialect which is like Japanese in having a pattern of overlapping VOT values).

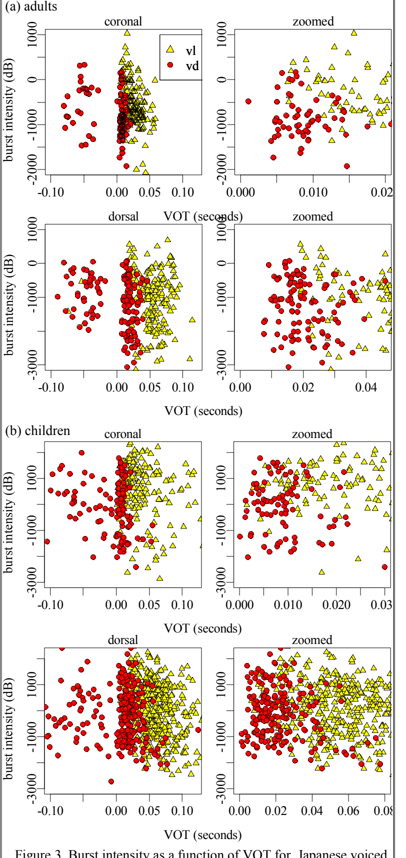

Figure 3. Burst intensity as a function of VOT for Japanese voiced (red circles) and voiceless (yellow triangles) stops. Left panels show the entire distribution, right panels zoom to the region of overlapping VOT values.

- MANOVA analyses showed a significant ($p<0.001$) difference between the voiced and voiceless stops in the region of VOT overlap for both places of articulation for both the adult productions (Fig. 3a) and the child productions (Fig. 3b).
- Given the greater intra-oral air pressure build-up in voiceless stops, we predicted greater burst intensity in the voiceless stops. The mean difference in burst intensity was in the predicted direction.

Analysis 3 — Peak amplitude during closure

- To answer Question 2, we took a measure of peak intensity of the voice bar during the closure (i.e., amplitude of the first spectral peak in a spectrum calculated over a 6 ms window centered at the highest intensity glottal pulse during closure) relative to the intensity of the voice bar in the region just before the burst (Fig. 4).

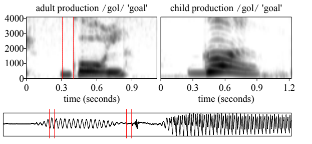

Figure 4. Spectrograms of adult (left) and child (right) productions of the Greek word /gol/ and zoomed-in waveform showing the windows over which the spectra were calculated for the peak closure amplitude measure in the adult production (red cursors).

- This measure was inspired by Burton, Blumstein, & Stevens's (1992) analysis of the contrast among oral voiced stops, pre-nasalized stops, and nasals in Moru, showing that the two types of stop are distinguished from nasals by a falling off of intensity just before the burst, and the pre-nasalized stops are distinguished from the oral stops by the intensity of the nasal murmer during closure.
- Note that the voiced stops of Greek developed relatively recently from clusters of nasal followed by homorganic voiceless stops, and voiced stops alternate with pre-nasalized stops, depending on the prosodic environment and the social/stylistic context (Arvaniti & Joseph 2000).

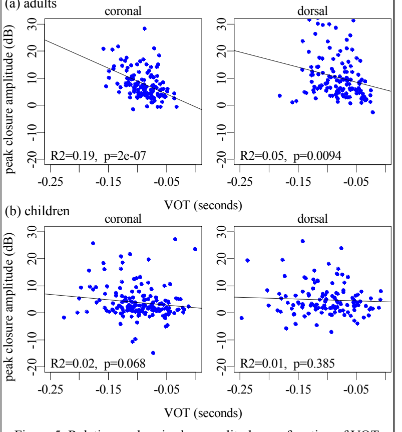

Figure 5. Relative peak voice bar amplitude as a function of VOT.

- The adult's voiced stops with longer lead VOT tend to show a bigger relative peak closure amplitude (Fig. 5a), suggesting that the stops with longer lead values are pre-nasalized to some extent.
- The correlation between relative peak voice bar amplitude and VOT is less evident in the children's voiced stops (Fig. 5b), possibly because many of the children's stops are more like Moru nasals than they are like Moru pre-nasalized stops.
- Compare the higher intensity of the voice bar just before the release in the child's production of /g/ in Fig. 4.

Discussion and conclusion

- The four languages that we examined in this study all have been described in terms of a simple two-way phonation-type contrast between either short lag and long lag (Cantonese and English) or between voicing lead and voicing lag (Greek and Japanese).
- However, a closer examination of the acoustics of the voiced-voiceless contrasts in Greek and Japanese suggests that more is involved than control of VOT, and that the other parameters involved facilitate the mastery of this difficult or "marked" contrast.
- Japanese "voiced" stops with short lag values have less intense bursts than voiceless stops with identical VOT values, and Japanese-acquiring children seem to master this aspect of production relatively early.
- Greek "voiced" stops seem to be partially pre-nasalized, and Greek-acquiring children can use (pre-)nasalization to vent air pressure during closure, making for early mastery of voicing lead.
- These minor variations on the general tendency for voiced stops to be mastered late lend further support to an interpretation Jakobson's developmental universal in terms of the universal constraints imposed by the more or less stringent articulatory and aerodynamic requirements of the more or less "marked" sounds.

Acknowledgements

- This work was supported by NIDCD grant 02932 to Jan Edwards. We thank the speakers who let us record them.