Psychoacoustic measures of stop production in Cantonese, Greek, English, Japanese, and Korean

Tim Arbisi-Kelm*, Mary E. Beckman**, Eunjong Kong**, and Jan Edwards*

*University of Wisconsin-Madison, Madison, WI; **Ohio State University, Columbus, OH

INTRODUCTION AND RATIONALE

- Spectral analyses of stop bursts have revealed that place of articulation can be discriminated, using both recurrent and non-recurrent cues present within the acoustic signal (e.g., Wesson and Blumstein, 1979; Forrest et al., 1988).
- While prior studies have had some success in uncovering such cues for American English, it is not clear whether these parameters are equally pertinent to stop consonant classification in other languages.
- Measures of burst frequency may be appropriate for both the compact and potentially bimodal peaks of velar stops and the diffuse bursts of alveolar stops, as well as capture the variability found across vowel contexts.
- Furthermore, the use of multilinear methods in linear acoustic analysis is that it imposes different scales of loudness and frequency on the acoustic signal than does the human ear, thus generating power spectra with different frequency distributions than are predicted by the auditory system (e.g., Zipser 1991, Kuczyński-Piotr 1983).
- This study evaluates whether a spectral model analysis would be more translatable if the open mouth is first "closed" by putting a piece of tissue into the mouth, so that peaks will correspond to audible combinations of energy in the representation of broad-specific boundaries in the auditory pathway.

METHOD

Materials

- Languages: Cantonese, English, Greek, Japanese, Korean
- All data recorded in each country with a native speaker as the experimenter.
- Participants: 10 adult speakers of each language.
- Stimuli:
  - Velar and alveolar stop consonants placed in word-initial position in familiar words in the following vowel contexts: /a, i, o/.
  - These were generated by taking a 20-ms Hamming window, centered at the burst, in order to obtain a frequency distribution of the burst energy.
- The resulting long-term averaged spectra were then normalized and converted into probability distributions, in order to compute the linear frequency scale spectral moments.
- The first four spectral moments—centroid, standard deviation, skewness, and kurtosis—were then calculated, using the formulas defined in Forrest et al. (1988).
- Spectral slices were generated across a 10-ms Hamming window, centered at the burst, in order to obtain a frequency distribution of the burst energy.
- The very small window was used to effectively isolate the front cavity resonances from the lateral resonances of the following vowel.
- We first designed a method of calculating specific loudness (SL) against equal rectangular bandwidths (ERB) as an initial step in developing psychoacoustic measures of loudness.
- We then constructed a set of transform long-term averaged spectra (LTS) derived from the LTS (specific loudness), using program modified after those used in Moore, Glasberg, and Bregel (1997).

Acoustical analysis

- Using fast Fourier transforms (FFT), we first downsampled the audio files from 44.1 to 2048 Hz in order to compute the spectra of the bursts as defined in Forrest et al. (1988) and high-pass filtered the output to 70 Hz to minimize aliasing noise.
- We then generated 2048-point FFTs for each burst, centered on a 20-ms Hamming window, centered at the burst, in order to obtain a frequency distribution of the burst energy.
- The resulting long-term averaged spectra were then normalized and converted into probability distributions, in order to compute the linear frequency scale spectral moments.
- The first four spectral moments—centroid, standard deviation, skewness, and kurtosis—were then calculated, using the formulas defined in Forrest et al. (1988).

RESULTS

Linear discriminant analyses

- Spectral moment analyses: for the English stops, it was found that the three measures correctly predicted the place of articulation category 71% of the time. The mean measures were less successful in discriminating place of articulation for stops in the other four languages.
- Auditory-based analysis: the measures peak ERB, CI and range ERB distinguished alveolar and velar stops for all four languages with a higher degree of accuracy overall than did the linear spectral moment measures.

Differences across languages and vowel contexts

- For English, peak ERB + CI outperformed spectral moments measures in correctly categorizing alveolars in each vowel context (including alveolars overall, 56 vs. 76), as well as overall for velars (73 vs. 65), but was not more successful in classifying velars (85 vs. 71) or alveolars (85 vs. 81).
- This was not the case for Cantonese, Greek, and Japanese, where peak ERB + CI performed equal to or better than spectral moments measures in all vowel contexts (including front vowels), with the exception of Cantonese /k/ (Both auditory- based and spectral measures both poorly discriminated Korean fricatives.)
- While velars on the whole were predicted with greater accuracy by the auditory-based measures, the alveolar results were even stronger. Peak ERB + CI classified alveolar stops with considerably greater success than did spectral moments in Korean (91 vs. 64), Japanese (78 vs. 37), Greek (83 vs. 61), and Cantonese (77 vs. 65).

DISCUSSION AND CONCLUSION

- Traditional spectral analyses based on linear model scales successfully distinguished back velar stops from alveolar stops in English, but were less successful in discriminating velar from alveolar stops in front vowel contexts in English, Greek, Cantonese, and Korean.
- Measures in all languages except English were unable to demonstrate good segmentation of tongue backness in velar versus alveolar stops, as indicated by the systematic variation in the ERB of the highest peak.

ACKNOWLEDGEMENTS

Supported by NIDCD grant R01 DC02932 to Jan Edwards.

Thanks also to the children who participated in the study, the parents who gave their consent, and the schools who let us use their facilities for our experiments.  

the modulating project