Perceptual Differences In Categorization Of Speech Sounds By Norwegian/English Bilinguals

Audny T. Dypvik and Elzbieta B. Slawinski
Dept. of Psychology, University of Calgary, 2500 University Drive, N.W., Alberta, Canada, T2N 1N4

I. INTRODUCTION

The ability to process auditory information is critical for speech perception. Deficits directly affect an individual's well-being and ability to participate in society. When a second language is introduced, bilinguals, and especially late bilinguals, perceptual abilities change with the degree and intensity of exposure to the second language.

Various acoustic cues serve to precipitate the perceptual distinction between phonemic contrasts. Spectral and temporal cues are used to discriminate the phonemic contrast between, for example, prevocalic /r/ and /l/ for English listeners (Polka and Strange, 1985). When consideration is taken for differing perceptual abilities of bilinguals, it has been found that bilingual Japanese/ English listeners with early and extensive exposure to English had improved perceptual ability in a phonemic distinction (Slawinski and MacNeil, 1994).

Perceptual categories are revealed in stop consonants. These are separated by well-defined phonemic boundaries and can be defined by listeners presented with a series of stops varying in equal physical steps relative to voicing. One such study (Elman, 1977) focused on perceptual switching in English and Spanish bilinguals. When the listeners heard the test stimuli in either an English or a Spanish context, the placement of category boundaries changed. Not only did the induction of a particular "language set" serve to alter the identification of stimuli by the listeners, but the degree of bilingualism interacted with the extent of perceptual switching effect (eg. Elman, 1977; Bohn and Flege, 1993).

Synthetic stimuli as well as natural speech stimuli have been implemented in speech studies. When Elman (1977) attemped to replicate his findings using synthetic stimuli, a reliable tendency toward a shifted perceptual boundary was not observed. Nonetheless, the noted absence of secondary voicing cues in their synthetic stimuli (such as spectral characteristics of the release burst and changes in the fundamental frequency following articulatory release) may have increased the difficulty for the bilinguals who conceivably place greater reliance on such cues. Since that time, synthetic stimuli have been improved.

Norwegian is one of the few European languages that is not an intonation language but a tone language (Moen, 1993). Norwegian has a distinct phonemic categorization for certain syllables, prominent among these being /v/ and /w/. This and a shortage of available data for the Norwegian language instigated the conceptualization of the current study. A revolutionary departure for speech synthesis was utilized - an articulatory synthesizer.

2. OVERVIEW OF APPROACH

In the past, the primary means of synthesizing speech was spectral reconstruction as exemplified by the Klatt synthesizer (Klatt, 1980). Recently, a real-time articulatory-based speech synthesizer has been developed and is currently in the process of being refined (Hill, Manzara and Taube-Schock, 1995). The "tube synthesizer", as it is called, together with supporting tools and components, allows one to create precisely controlled stimuli.

Energy in the form of pressure variations is injected into one end of the tube model. This corresponds to air being forced through the vocal folds causing them to vibrate. The energy is then spectrally modified by the resonant characteristics of the oropharyngeal and nasal cavities using an acoustic waveguide (in Hill, Manzara and Taube-Schock, 1995). The Distinctive Region Model (DRM) (Carre, 1992) was used as a basis to model the oropharyngeal cavity.

The tube model has 8 regions (vocal tract sections), which closely relate to human articulators, primarily the lips, tongue, and teeth. In contrast to older synthesizers, the control of speech is achieved by directly controlling the articulators as opposed to the formants.

A tool called MONET was developed to edit speech postures (roughly related to phonemes), interpolation rules, and timing data. MONET was used to create rules that combine basic postures into articulatory sequences. Once developed, these rules were used to create parameter tracks that drive the synthesizer. MONET can handle diphone combinations of postures as well as triphone combinations. Triphone combinations enable one to model co-articulatory effects.

In the current study, the tube synthesizer was used to create a series of preliminary stimuli for the phonemic distinction between /v/ and /w/. These phonemes have particular articulatory postures. We created 9 postures which included the 2 endpoints of /v/ and /w/ plus 7 intermediate

postures equally spaced and linearly interpolated.

Using these postures, we generated 4 series of stimuli. The first series used diphone transitions describing silence to the interpolated sequence of /w/ to /v/ postures to the following vowel /i/. This series also included frication energy injected at the tube section representing the teeth. Frication amplitude in this series depended upon the size of the orifice at the teeth. The second series was similar to the first except that no frication energy was injected. The third series used a triphone transition describing silence to the interpolated sequence of /w/ to /v/ postures to the following vowel /i/. Frication was included in the series. The fourth series was similar to the third except that frication was not injected. These stimuli will be presented to participants via headphones in an anechoic chamber (Industrial Acoustics Company, Inc.). Analysis of speech production will follow analysis of perception.

3. DISCUSSION

It is anticipated that the combination of examining categorization of English phonemic perception in the little-studied language of Norwegian with the use of real-time articulatory speech synthesis will lead to some exciting results.

Bilingual immigrants do indeed face unique challenges in every-day living. They perceive acoustically identical stimuli differently depending on the language being processed at the time, or "language set", and the extent of this perceptual difference is related to the timing and extent of exposure to the second language. We propose that because /v/ and /w/ are allophones in the Norwegian language, English phonemic boundaries of bilingual listeners between /v/ and /w/ will vary from those of English speakers.

To date, use of the "tube synthesizer" has revealed a very real clarity of articulation. As well, spectrograms already created from these synthesized-by-rules speech stimuli appear similar to those created from natural speech stimuli. A more natural approach to speech production is introduced when articulatory postures are used to dictate sound synthesis rather than formant manipulation. Validation of this approach will require substantiation of formerly-achieved results illustrating perceptual discrimination differences of certain phonemic boundaries by bilinguals.

4. REFERENCES

Bohn, O-S. and Flege, J.E. (1993). Perceptual switching in Spanish/English bilinguals. Journal of Phonetics, 21, 267-290.

Carre, R. (1992). Distinctive regions in acoustic tubes. Speech production modeling. Journal d'Acoustique, 5, 141-159.

Elman, J.L., Diehl, R.L. and Buchwald, S.E. (1977). Perceptual switching in bilinguals. Journal of Acoustical Society of America, 62(4), 971-974.

Hill, D.R., Manzara, L. and Taube-Schock, C-R. (1995). Real-time articulatory speech-synthesis-by-rules. Presented at AVIOS, San Jose, CA, USA.

Klatt, D.H. (1980). "Software for a cascade/parallel formant synthesizer". Journal of Acoustical Society of America, 67, 971-995.

Moen, I. (1993). Functional Lateralization of the Perception of Norwegian Word Tones – Evidence from a Dichotic Listening Experiment. Brain and Language, 44, 400-413.

Polka, L. and Strange, W. (1985). Perceptual equivalence of acoustic cues that differentiate /r/ and /l/. Journal of Acoustical Society of America, 78(4), 1187-1197.

Slawinski, E.B., and MacNeil, J.F. (1994). The relationship between perceptual and productive discrimination of English /r/ and /l/ sounds by Japanese speakers. International Journal of Psycholinguistics, 10(3), 1-23.

ACKNOWLEDGEMENTS

Thank to Dr. David R. Hill, Dr. Leonard Manzara and Craig-Richard Taube-Schock, MSc., for their expertise, generosity of time, and use of their speech synthesizer.

Dr. Greg Shaw added valuable support in programming.