Why are vowels and tones perceived less categorically than stop consonants? – Categorical perception and speech sound types revisited

Since Liberman et al.'s (1957) study, copious research has been conducted to test the relationship between discrimination of speech sounds and their phonemic status in a particular language. Researchers share the consensus that discrimination of stimuli from stop consonant continua (e.g. the VOT continuum constructed by Abramson & Lisker, 1970) is the most categorical, while the perception of vowel stimuli exhibits a mainly continuous pattern. Studies on lexical tone perception (e.g. Abramson, 1979) show that tones are perceived in a fashion similar to vowels (but Cf. Wang et al., 1975). Such perceptual differences in speech sound types have been attributed to differences in acoustic cues: rapid CV or VC transitions and transient release bursts in stops versus fairly long and steady formant structures in vowels (see, e.g. Pisoni, 1975).

It is well recognized that the absolute formant values for any vowel vary a great deal from talker to talker (Peterson & Barney, 1952; Lehiste & Meltzer, 1973). Take /i/ or example. The typical acoustic characteristic of this vowel is a low first formant (F1) and a high second formant (F2). Nevertheless, the F1 values in different talkers may range from 180Hz to 400Hz, while the F2 can be anywhere between 2000Hz and 3500Hz. Similarly for tones, talker variation in pitch is also great (see, e.g., Ladefoged, 2001). For example, the absolute F0 value for the high level tone T55 in Mandarin Chinese can be 140Hz for a male speaker but 280Hz in a female voice. One can imagine any phonetic value between 140Hz and 280Hz (and some even outside this range) at which T55 can be realized. Because of these ranges in both the /i/ and the T55 cases, there are overlaps across talkers in formant or pitch values with other vowels or other tones. Thus, the listener has to be sensitive to all frequency values within such ranges, as any of them may represent a phonemic category for a particular speaker. Consequently, discrimination may be good even within the possible ranges of F1 and F2 in /i/ or that of F0 in T55. For this reason, perception of steady state vowels and level tone contrasts will be predictably more continuous, unless enough information is provided for talker normalization (Johnson, 2005).

Based on the observations made above, it is proposed here that stop consonants are perceived categorically because a contrast such as VOT is largely determined by the common physiological function of aspiration rather than being constrained by individual anatomical differences (length of vocal tract in vowels, size and mass of vocal folds in tones, etc.). There is little variation in this type of consonantal cues, leading to better defined category boundaries in plosives than in steady-state vowels or level tones. Also in the case of diphthongs or contour tones, formant movements or tonal contours are relatively independent of talker physiology. Just as in VOT contrast, the perceptual boundary between one diphthong and another (or a monophthong), or the boundary between one contour tone and another contour (or a level) tone, will also be better defined than in steady-state vowels or level tones. Thus, perception of vowel stimuli from a diphthong-diphthong (or a diphthong-monophthong) continuum and perception of tone stimuli from a contour tone-contour tone (or a contour tone-level tone) continuum are predicted to be more categorical than perception of stimuli from a steady state vowel continuum or perception of stimuli from a level tone continuum. Results from studies on tone perception by Wang et al. (1975), Abramson (1979) and Sun & Huang (under review) support this hypothesis.

References:

Abramson, A., & Lisker, L. (1970). Discriminability along the voicing continuum: Cross-language tests. *Proceedings of the 6th International Congress of Phonetic Sciences* (Academia, Prague). pp. 569-573.

Abramson, A. (1979). Noncategorical perception of tone categories in Thai. In *Journal of the Acoustical Society of America*, 61, S66.

Johnson, K. A. (2005). Speaker normalization in speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The Handbook of Speech Perception* (pp. 363-389). Malden, MA: Blackwell Publishing.

Ladefoged, P. (2001). A Course in Phonetics (4th Edition). Boston: Heinle & Heinle.

Lehiste, I. & D. Meltzer (1973). Vowel and speaker identification in natural synthetic speech. In *Language and Speech*, 16, 356-64.

Liberman, A., Haris, K., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.

Peterson, G. E. & H. L. Barney (1952). Control methods used in the study of vowels. In *Journal of the acoustical Society of America*, 24, 175-84.

Pisoni, D.B. (1975). Auditory short-term memory and vowel perception. In *Memory & Cognition*, 3, 7-18.

Sun, K.-C. & T. Huang (under review). A Cross-linguistic Study of Taiwanese Tone Perception by Taiwanese and English listeners. *Journal of East Asian Linguistics*.

Wang, S.Y., Chan, S. W., & Chuang, C. K. (1975). Cross-linguistic study of categorical perception for lexical tone. *Journal of the Acoustical Society of America*, 58, S119.