November 30, 1994

The Spectral Center of Gravity Effect and Auditory Filter Bandwidth

Marc Fagelson, *University of Texas at Austin*
Linda M. Thibodeau, *University of Texas at Dallas*

had either the same F_0 and spectrum envelope as one of the vowels ("voiced vowel" precursor), the same F_0 but a flat spectrum envelope ("buzz" precursor), or the same spectrum envelope as one of the vowels, but with noise excitation ("whispered vowel" precursor). All three types of precursors resulted in increased accuracy when the precursor was brief (200 ms). When the precursor was 1 s, only the voiced vowel precursor led to an increase in identification accuracy. These results suggest that both adaptation and perceptual grouping may play a role.

3aSP12. Modeling formant frequency discrimination for isolated English vowels using excitation patterns. Yijian Zheng and Diane Kewley-Port (Dept. of Speech and Hear. Sci., Indiana Univ., Bloomington, IN 47405)

Thresholds for formant discrimination across three sets of female and male vowels with different $F0$ were significantly different in a recent report [Kewley-Port *et al.*, J. Acoust. Soc. Am. **95**, 2978(A) (1994)]. This analysis examined whether excitation patterns could model these and other effects of stimulus parameters on formant thresholds. The goal was to determine if an "auditory metric" would be constant across the three stimulus sets when ΔF thresholds varied by 25 Hz. A separate discrimination study showed that listeners only attend to harmonic components within a restricted region near the formant [Sommers and Kewley-Port, J. Acoust. Soc. Am. **93**, 2422(A) (1993)]. Based on those results, four critical bands around the altered formant were selected, and the area between the critical-band spectra for the standard and just discriminable vowel was calculated. This spectral distance across formant frequency and gender was shown to be constant in three analyses: (1) ΔF threshold differences across the three sets of vowels were no longer significant; (2) slopes for ΔF thresholds (approximately 1.0) were flat for spectral distance; (3) variability of spectral distance across $F1$ and $F2$ is significantly smaller than that of ΔF thresholds. Results suggest that the auditory system has an inherent nonlinear transformation which changes threshold differences to be almost constant in the internal representation.

3aSP13. Modeling listeners' categorization of a large *F1-F2-F3* continuum. Terrance M. Nearey and Michael Kiefte (Dept. of Linguist., Univ. of Alberta, Edmonton, AB T6G 2E7, Canada)

A number of alternate spectral representations have been suggested for vowel spectra [see H. Hermansky, J. Acoust. Soc. Am. **87**, 1738–1752 (1990)]. To better evaluate the perceptual relevance of some of these, 972 vowels were synthesized. The stimuli were each 115 ms in duration with a falling $F0$ contour (125–100 Hz). $F1$ ranged (in 0.5 Bark steps) from 250 to 760, $F2$ from 750 to 2260, and $F3$ from 1360 to 3080 Hz. $F4$ and $F5$ were fixed at 3500 and 4500 Hz, respectively. (Constraints were placed on formant separations to ensure relatively natural stimuli.) Fifteen speakers of Western Canadian English categorized the stimuli as the vowels /i, ɪ, e, ɛ, æ, ʌ, ɒ, o, ʊ, u, ɝ/. Preliminary results indicate that while nominal synthesis formant frequencies can provide a relatively good fit to the data, alternate representations such as cepstral coefficients based on Hermansky's PLP analysis may provide moderate improvements of fit. However, linear transformations of the PLP cepstra show strong correlations with formant frequencies [similar to those noted by D. Broad and F. Clermont, J. Acoust. Soc. Am. **86**, 2013–2017 (1985)]. [Work supported by SSHRC.]

3aSP14. Formant movement and duration cues in the identification of vowels. Amy T. Neel and Diane Kewley-Port (Dept. of Speech and Hear. Sci., Indiana Univ., Bloomington, IN 47405)

Traditionally, target values of $F1$ and $F2$ are viewed as primary determinants of vowel identity. Several recent studies, however, have demonstrated the importance of dynamic formant information to vowel identification. The present study examines the contribution of both dynamic and durational information to vowel identification using sine-wave vowel analogs. Sine-wave stimuli consisting of two tones representing $F1$ and $F2$ were constructed using careful LPC measurements of ten vowels produced in /dVd/ context by male and female speakers. Four types of stimuli were constructed by varying two factors: (1) appropriate versus fixed vowel duration and (2) variation in tones representing formant movement throughout the token versus static target formant values. Listeners identified the sine-wave vowel analogs using a key-word response form. Results demonstrated that stimuli with appropriate vowel duration were identified with significantly greater accuracy than those with fixed length. For appropriate duration stimuli, there was no significant difference for dynamic versus static tokens. However, for fixed length stimuli listeners identified dynamic tokens with significantly greater accuracy than static stimuli. This suggests that intrinsic vowel duration, as expected, is an important cue to vowel identity and that dynamic formant information is used more by listeners when duration cues are unavailable.

3aSP15. A linear model of boundary shifts in /ʊ–ɪ/ and /ɔ–ɛ/ continua. Anna K. Nabelek and Alexandra Ovchinnikov (Dept. of Audiol. and Speech Pathol., The Univ. of Tennessee, Knoxville, TN 3796-0740)

Boundary locations were tested for /ʊ–ɪ/ and /ɔ–ɛ/ continua with steady-state or linearly changing formants in which $F2$ was varied. $F1$ and $F2$ trajectories had upward and downward directions. Boundary shifts were calculated for changing formant stimuli relative to the boundary for steady-state stimuli. The directions of boundary shifts indicated perceptual emphasis of the final segments of $F2$ trajectory which might be a consequence of low-frequency spread of masking from $F1$ to $F2$. A linear model was developed in which boundary shifts were related to spectral distance between $F1$ and $F2$ trajectories. Parameters were initial and final frequencies of $F1$ and boundary $F2$ established for each continuum with steady-state stimuli. When the distance was constant in time shifts depended on directions of $F1$ and $F2$ trajectories, described by two model terms containing differences of initial and final frequencies of $F1$ and $F2$. When the distance was changing in time additional shift was toward the greater spectral distance, described by two model terms containing ratios of initial and final frequencies. [Work supported by NIH.]

3aSP16. The spectral center of gravity effect and auditory filter bandwidth. Marc Fagelson (Dept. of Speech Commun., Prog. in Commun. Sci. and Disord., Univ. of Texas—Austin, CMA 2.200, Austin, TX 78712) and Linda M. Thibodeau (Univ. of Texas, Austin, TX 78712)

The spectral center of gravity refers to a listener's averaging of frequency and intensity components when formant peaks in a speechlike signal are separated by 3.5 Bark units or less. In this paper a total of 18 synthetic vowels whose spectra approximated /ae/ or /ʌ/ were generated digitally; each stimulus contained the first 40 harmonics of a 100-Hz fundamental. Nine spectra contained three formants, while the balance contained only two. Subjects with normal hearing and mild high-frequency hearing loss above 3000 Hz were instructed to identify synthetic vowels as either /ae/ or /ʌ/ as $F2$ frequency was varied between nine different values in 100-Hz steps for both the two-formant and three-formant stimuli. Probit analysis indicated that the normal-hearing subjects identified stimuli more consistently than the mildly hearing-impaired listeners across $F2$ frequencies for three-formant than for two-formant spectra. The $F2$ frequency corresponding to the perceived increase in vowel frontness occurred at a lower frequency for normal-hearing listeners. Auditory filter bandwidth was negatively correlated with the $F3-F2$ Bark difference. Results suggest that spectral averaging may help listeners disambiguate confusing speech signals.

3aSP17. The effects of pitch changes on the perception of vowel sequences. Benita K. Nordenstrom, Magdalene H. Chalikia, and Elizabeth M. Ebsen (Dept. of Psychol., Moorhead State Univ., Moorhead, MN 56563)

When listeners hear a repeated sequence of steady-state vowels (of the same duration and pitch) phonemic transformations occur, and they report hearing words and phrases absent in the original stimulus. A previous study [M. H. Chalikia, R. Meyer, and R. Lindemann, 34th Psychon. Society Meet. (1993)] investigated the possible effects of variations in vowel duration, pitch, or both. Listeners successfully matched these vowel sequences to the verbal forms heard with vowels of equal duration and pitch (base-line stimuli). In this study a broader variation in pitch was employed, and the vowel sounds were computer-generated rather than naturally produced. Six base-line sequences of six 60-ms vowels (at 100 Hz), followed by a 300-ms silent gap, were used. Four variations of these were created by randomly changing the pitch of individual vowels. Most listeners were able to match the modified vowel sequences to the verbal forms heard with the base-line stimuli, thus confirming the robust, stable nature of the verbal