Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children

Joan A. Sereno, Shari R. Baum, Grant C. Marean, and Philip Lieberman

Citation: The Journal of the Acoustical Society of America **77**, S7 (1985);

View online: https://doi.org/10.1121/1.2022525

View Table of Contents: http://asa.scitation.org/toc/jas/77/S1

Published by the Acoustical Society of America

Session D. Speech Communication I: Vowel Perception, Lexical Access

Bruno H. Repp, Chairman
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695

Chairman's Introduction—8:45

Contributed Papers

8:50

D1. Categorical tendencies in vowel imitation: I. Synthetic vowel stimuli. David R. Williams and Bruno H. Repp (Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695)

The question whether isolated stationary vowels are imitated in a continuous or categorical fashion [R. D. Kent, Phonetica **28**, 1–25 (1973)] was pursued further by replicating Kent's study with slight modifications. The subjects (the two authors) imitated synthetic stimuli from [u] – [i] and [i] – [æ] continua at three different temporal delays. Acoustic analysis of the response vowels revealed very similar patterns across delays. Both subjects showed clear evidence of nonlinearities in the stimulus–response mapping and of preferred response formant frequencies, though strictly categorical responses were generally absent. The origin of these nonlinearities and their relation to the phonemic vowel categories of English are not fully understood at present. Vowel imitation responses presumably reflect the joint influences of perceptual and articulatory factors that need to be disentangled in future research. [Work supported by NICHD Grant HD-01994.]

9:02

D2. Categorical tendencies in vowel imitation: II. Natural vowel stimuli. Bruno H. Repp and David R. Williams (Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695)

From the vocal imitation responses recorded in our first experiment, we selected tokens approximating [u]-[i] and [i]-[æ] continua in each subject's own acoustic vowel production space. The imitation experiment was then repeated with these natural vowels as the stimuli. The results from the two experiments will be compared in this presentation. If the second experiment replicates the response patterns found in the first experiment, then the perceptual processing of synthetic vowels will have been ruled out as one possible source of nonlinear stimulus–response mapping in the imitation task. [Work supported by NICHD Grant HD-01994.]

9:14

D3. A perceptual-anchor interpretation of categorical phenomena on a vowel continuum. Rina Goldberg (Research Laboratory of Electronics, MIT, Cambridge, MA 02139 and A.T.& T. Bell Laboratories, Naperville, IL), Neil A. Macmillan (Department of Psychology, Brooklyn College, Brooklyn, NY 11210 and Research Laboratory of Electronics, MIT, Cambridge, MA 02139), and Louis D. Braida (Research Laboratory of Electronics, MIT, Cambridge, MA 02139)

The Perceptual Anchor Model of Context Coding [Braida *et al.*, J. Acoust. Soc. Am. **76**, 722–731 (1984)], developed for auditory intensity, considers resolutions to be determined by both sensory and memory processes, and postulates that tasks differ in the memorial coding they require. The theory was tested with steady-state synthetic vowels in the range /i/-/I/-/ɛ/. Resolution was measured in four discrimination conditions (two-interval forced-choice and same–different, fixed-, and roving-level), and in complete identification conditions with and without a standard. In identification and roving-level discrimination, resolution was poorer than in fixed-level discrimination, but not uniformly so: sensitivity differences among tasks were smallest near category boundaries. The data suggest that some categorical phenomena can be understood in terms of perceptual anchors. [Work supported by NSF.]

9:26

D4. Notes on the history of vowel theory. James D. Miller (Central Institute for the Deaf, St. Louis, MO.63110)

The history of metrics used to describe the vowels will be reviewed. The early arguments concerned whether the absolute or relative pitches of one or more resonances were crucial for their specification. Special attention will be given to the work of R. J. Lloyd of Liverpool who forcefully argued in 1890 that ratios of the pitches of audible resonant peaks characterized the vowels. Arguments and experiments relating to these issues will be traced from the late 1800s to the present. [Supported by NINCDS 03856.]

9:38

D5. A shift in formant frequencies is not the same as a shift in the center of gravity of a multiformant energy concentration. Dennis H. Klatt (Room 36-523, MIT, Cambridge, MA 02139)

Chistovich *et al.* [*Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, New York, 1979), pp. 143–158] observed that when two formants are within about three critical bands, the auditory percept is like that of a single equivalent formant, in that a change in relative formant amplitude has about the same perceptual effect as a change in formant frequencies (shifting the center of gravity of the collective energy concentration). Klatt [*Proc. ICASSP* (Paris, 1982), pp. 1278–1281], on the other hand, was unable to replicate these findings. A small change in formant amplitude did not produce any measurable change in phonetic quality of a vowel. This experiment attempts to reconcile these two results by examining the various differences in stimuli and experimental design in the two studies. It is concluded that large changes in the relative levels of $F1$ and $F2$ for [ɔ] can change the perceived phonetic quality of the vowel, but the change is not the same as that induced by moving the formant frequencies because opposing changes in these two variables do not tend to cancel one another in the percept. [Work supported in part by an NIH grant.]

9:50

D6. Acoustic analyses and perceptual data on anticipatory labial coarticulation in adults and children. Joan A. Sereno, Shari R. Baum, Grant C. Marean, and Philip Lieberman (Department of Linguistics, Box E, Brown University, Providence, RI 02912)

The present study investigated anticipatory labial coarticulation in the speech of adults and children. CV syllables, [s], [t], and [d] before [i] and [u], were produced by an adult male speaker and a female child speaker age 3 years 6 months. Each syllable was computer-edited to include only the noise-excited portion of fricative-vowel stimuli and only the aperiodic portion of stop-vowel stimuli. LPC spectra were computed for each excised segment. Analyses of the effect of the following vowel on the spectral

peak associated with the second formant frequency and on the characteristic spectral prominence for each consonant were performed. Perceptual data were obtained by presenting the aperiodic consonantal segments to subjects who were instructed to identify in a forced choice paradigm the following vowel [i] or [u]. Both the acoustic and perceptual data show strong coarticulatory effects for adults and the absence of such coarticulations in the speech stimuli of the child. The results are discussed in terms of the articulatory and perceptual aspects of coarticulation in language learning.

10:02

D7. Influence of spectral prominence on perceived vowel height. S. Hawkins and P. S. Beddor (Haskins Laboratories, 270 Crown Street, New Haven, CT 06511)

Earlier experiments on the perception of vowel height comparing oral with nasal vowels, and one- with two-formant vowels, indicated (1) center of gravity (measured in terms of a simple weighted average of spectral components in the $F1$–$F2$ region) does not closely predict perceived vowel height and (2) $F1$ influences the perceived height of oral vowels more than that of nasal vowels. This experiment examined the influence of the relative prominence of formant peaks on perceived height. Two-formant vowels, /a/ and /o/, were synthesized with steady $F0$ and equal formant amplitudes. Each had narrow (45 Hz), medium (70–75 Hz), and wide (120–150 Hz) bandwidth versions. Continua were made from the medium-bandwidth vowels by systematically manipulating formant frequencies and amplitudes. For both /a/ and /o/, the medium-bandwidth vowels were paired with (a) the narrow-bandwidth vowel and (b) the wide-bandwidth vowel. Subjects selected the "best-match" pair for each set. Narrow-bandwidths had no effect on perceived height, i.e., subjects matched the narrow-bandwidth vowel with the medium-bandwidth vowel of the same formant frequencies. Wide bandwidths (characteristic of nasal vowels) affected perceived height: "medium-wide" matches were relatively variable and were biased away from pairs with the same formant frequencies. The direction of these biases will be discussed in terms of nasal vowel perception. [Supported by NIH grants NS-07237 and F32-NS-07196.]

10:14

D8. Effects of contextual and noncontextual nasalization on perceived vowel height. Rena A. Krakow, Patrice S. Beddor, Louis M. Goldstein, and Carol A. Fowler (Haskins Laboratories, 270 Crown Street, New Haven, CT 06511)

Listeners' ability to separate the spectral effects of nasal coupling from those of tongue position was examined by comparing labeling responses to three continua: oral /bɛd-bæd/; contextually-nasalized /bɛ̃nd-bæ̃nd/; and noncontextually nasalized /bɛ̃d-bæ̃d/. The continua were generated by articulatory synthesis and were identical to one another in tongue position. Results showed the same percent /ɛ/ responses to the oral and contextually nasalized vowels, but significantly fewer /ɛ/ responses to the noncontextually nasalized vowels. That is, the noncontextually nasalized vowels were perceived as lower than the oral and contextually nasalized vowels. Perceptual lowering of the nasalized vowels relative to the oral vowels is consistent with upward shifts in $F1$ due to nasal coupling. However, the absence of a lowering effect in the contextual nasal condition needs to be explained. We propose that subjects' vowel judgments are influenced by spectral effects of nasal coupling when no conditioning environment is present. When a nasal consonant is present, these spectral effects are perceptually analyzed as due to the consonant and do not influence vowel height judgments. [Work supported by NIH grants HD-01994 and F32-NS-07196, NSF grant BNS-81-11470.]

10:26

D9. Perceptual correlates of the tense/lax distinction in general American English. C. B. Huang (Research Laboratory of Electronics and Department of Electrical Engineering and Computer Science, MIT, Room 36-512, Cambridge, MA 02139)

Data from perceptual tests in which subjects were asked to identify tense and lax vowels are used to suggest correlates for the tense/lax distinction in general American English. Four vowel parameters were manipulated in synthetic nonsense words of the form [dVs]: formant frequency, duration, breathiness, and first-formant bandwidth. The tense/lax pairs included [ɪ], [æɛ], [uʊ], and [ɑʌ]. Three separate identification tests are created in which the stimuli vary only in formant frequency and one of the other three parameters. Each test was presented to five naive native speakers of general American English. Results so far indicate that shortening the vowel shifts judgments toward the lax vowel, and that this effect is greater in the low vowel pairs than in the high vowel pairs. Increased breathiness shifts judgments toward the tense vowel for high vowels. The effect of breathiness on low vowels and the effect of bandwidth of the first formant on all vowels is as yet unclear. [Supported in part by NINCDS and a NSF Fellowship.]

10:38

D10. Reduction of formant bandwidth improves vowel identification with sensorineural impairment. Donald G. Jamieson, Curtis W. Ponton, and Blas Espinoza-Varas (Department of Psychology, University of Calgary, Calgary, Alberta, T2N 1N4, Canada)

Accurate identification of front–back vowels depends on the ability to locate the peak of $F2$. Since hearing impairment reduces frequency resolution, sensorineural hearing impaired listeners often have difficulty with certain vowel distinctions. Espinoza-Varas, Jamieson, and Wahn [J. Acoust. Soc. Am. Suppl. 1 **76**, S80 (1984)] showed that (a) for normal listeners, equal changes in $F2$ frequency in critical-band units produce equal changes in the judged "goodness" of vowels on front–back vowel continua, cued by changes in $F2$; and (b) sensorineural impaired listeners are less sensitive to these $F2$ changes and less accurate in their identification of vowels. The present paper continues this work, examining whether vowel sounds can be altered to reduce masking between adjacent spectral frequencies, while preserving vowel quality. Following Assmann [J. Acoust. Soc. Am. Suppl. 1 **76**, S81 (1984)] we examined identification for synthesized vowels with normal versus narrow formant bandwidths. We also examined vowel identification performance for natural vowels which had been filtered to reduce formant bandwidths, while preserving the locations of the spectral peaks. Some hearing impaired listeners show significantly improved performance with these vowel sounds (for example, categorization functions for the synthesized vowel continuum /u/-/i/ become steeper). No listener showed poorer performance with the processed sounds. We conclude that (a) masking between adjacent spectral frequencies may cause hearing impaired listeners to make errors in vowel identification and (b) this masking can be reduced by processing speech signals to reduce formant bandwidth. [Work supported by grants from the AHFMR, HWCRDP, and NSERC.]

10:50

D11. A theory toward discontinuous perception of speech. Kazunari J. M. Koike (Speech-Language Pathology and Audiology, University of Utah, Salt Lake City, UT 84112)

An experiment by Guberina (1963) was replicated for perceptual and acoustical analyses. The vowel /i/ spoken by a female voice and recorded on a tape was presented via a loudspeaker to normal hearing listeners through the following conditions: (1) no filter; (2) the band with cutoff frequencies of 200–400 Hz at most confortable level (MCL); (3) the band with cutoff frequencies of 2400–4800 Hz at MCL; and (4) two bands just *below* threshold of intelligibility simultaneously. The results were summarized as follows: (1) the vowel /i/ was perceived in no filter condition; (2) the vowel /u/ was perceived through the 200–400 Hz. Only the first formant was observed; (3) the vowel /i/ was perceived through the 2400–4800 Hz. Only the second and third formants were observed; and (4) when the two discrete bands were combined, the vowel /i/ was perceived. Only the first formant was observable. It was speculated that the suprathreshold low frequency speech energy enhanced the perception of the subthreshold high frequency speech energy and as a result the vowel /i/ was perceived. A theory toward discontinuous perception of speech will be discussed especially as it is applied to the rehabilitation of hearing-impaired listeners.