Vowel duration as a cue for consonant voicing?

C. W. Fruin and D. M. Bischoff

Citation: The Journal of the Acoustical Society of America **60**, S91 (1976); doi: 10.1121/1.2003606
View online: https://doi.org/10.1121/1.2003606
View Table of Contents: https://asa.scitation.org/toc/jas/60/S1
Published by the Acoustical Society of America

one category but not in the presence of a speech sound from a second category. The infants were trained with a single synthetically produced /a/ and /i/ whose formant frequencies and fundamental frequency were appropriate for a male voice (rise—fall pitch contour) and then tested with synthetic /a/ and /i/ tokens whose formant frequencies and fundamental frequencies were appropriate for female and child talkers, with both rising and rise-fall pitch contours. Generalization to novel tokens was often times immediate. The six-month old's behavior is suggestive of either an inherent, or easily learned, perceptual constancy for vowel categories when both the critical (formant frequencies) and noncritical (fundamental frequency, pitch contour) acoustic dimensions are randomly varied.

2:50

LL6. Discrimination of synthetic prevoiced labial stops by infants and adults. Rebecca E. Eilers (Mailman Center for Child Development, P.O. Box 520006, Biscayne Annex, Miami, FL 33152), Wesley R. Wilson and John M. Moore (Department of Speech, University of Washington, Seattle, WA 98195)

Individual data were obtained from eight normal 6-month-old infants on discrimination of eight synthetic labial stop pairs differentiable on the basis of voice onset time. All infants were presented with the pairs p^ha (+40) versus p^ha (+70), p^ha (+40) versus pa (+10), pa (+10) versus ba (−20), ba (−20) versus ba (−50), ba (−20) versus ba (−60), ba (−30) versus ba (−60), pa (0) versus ba (−30) and pa (+10) versus ba (−60). Discrimination was assessed for each infant on each pair using the VRISD paradigm (Visually Reinforced Infant Speech Discrimination). Adult discrimination was assessed for the same pairs in the same manner. With few exceptions, infants presented no evidence of discrimination of pairs other than the pair p^ha (+40) versus pa (+10) despite the fact that these infants were successful in the same paradigm with a variety of subtle natural speech contrasts. Contrary to the hypothesis that age restricts discrimination capacities, adults were far more successful than infants in discriminating the non-English contrasts. Individual data will be presented to illustrate these points and the use of synthetic speech stimuli in infant research will be discussed in light of these data. [Works supported by NIH-NICHD, HD-3-2793.]

3:00

LL7. Stimulus correlates in the perception of voice onset time (VOT): I. Discrimination of the time interval between tone bursts of different intensities and frequencies. P. L. Divenyi,* R. M. Sachs, and K. W. Grant (Central Institute for the Deaf, 818 S. Euclid, St. Louis, MO 63110)

In a four-level 2AFC paradigm, trained subjects had to discriminate time intervals that were marked by tone bursts of unequal duration (10 and 100 msec). Time differences discriminable at the $d'=1$ level, ΔT_{Thresh}, were determined in the 10—100-msec range of time intervals (onset to onset). When the frequencies and intensities of all markers were held constant (1 kHz and 86 dB SPL), ΔT_{Thresh} was a monotonic function of T. This relation became non-monotonic when the level of the first marker was decreased to 36 dB SPL or as the two marker frequencies were made different from each other by two octaves. Thus, a discrepancy in either the level or the spectrum of two time-marker sounds may constitute an acoustic basis for the nonmonotonic VOT discrimination functions, despite the fact that the depth of the local minimum found in the present data is less than that observed in VOT experiments. [Supported by a grant from NINCDS.]

*Presently at Neurophysiology—Biophysics Laboratories, Veterans Administration Hospital, Martinez, CA 94553.

3:10

LL8. Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. R. M. Sachs and K. W. Grant (Central Institute for the Deaf, 818 S. Euclid, St. Louis, MO 63110)

The three Lisker—Abramson CV series (ba-pa, da-ta, ga-ka) were computer synthesized at C.I.D. in 1-msec steps. For each 84-trial block of a high-uncertainty, same—different task (no feedback), one of the three continua was chosen at random. On each trial, the ability of three naive listeners to resolve for differences among VOT pairs ranging from VOT = 10—80 msec, in 10-msec steps was tested. As expected, there was a local maximum performance for each of the three speech continua, thus satisfying one criterion for categorical perception. A minimum-uncertainty AX task (with feedback) was run with three highly trained subjects for the ga-ka continuum. Only one standard VOT was presented in a block of trials. Results were not categorical: the minimum detectable ΔVOT was less than 2 msec at VOT = 10 msec and increased as VOT increased. Our results indicate that trained listeners can utilize subtle stimulus cues in synthetic speech if stimulus uncertainty is minimized. [Supported by a grant from NINCDS.]

3:20

LL9. The voicing boundary as a function of F_2 and F_3 transitions and fundamental frequency. Bruno H. Repp (Haskins Laboratories, 270 Crown St., New Haven, CT 06510)

It is well known that the voice-onset-time (VOT) boundary for synthetic stop consonants shifts towards longer VOT's as place of articulation shifts from front to back, even if all parameters of the signal except VOT and the starting frequencies of F_2 and F_3 are held constant. Is this a phonetic effect contingent on a decision about the place feature, or is it an auditory effect that is somehow dependent on the spectrum at syllable onset? In order to decide this question, F_2 and F_3 transitions were varied in a number of steps along a "place continuum," and the VOT boundary was determined at each step. The function relating the VOT boundary to the place continuum should be flat within place categories and show jumps at the category boundaries if the effect is phonetic, but it should be monotonically increasing if the effect is auditory. Neither hypothesis has been supported by the data so far, which show the function to be generally increasing but strongly nonmonotonic, with a number of seemingly haphazard but statistically reliable peaks and valleys. A radically different pattern (with the same general characteristics) is obtained if the fundamental frequency of the stimuli is changed. Further research on the acoustic basis of the effect and its perceptual and methodological implications will be reported.

3:30

LL10. Perception of voicing in final stops. Catherine G. Wolf (36-549, Massachusetts Institute of Technology, Cambridge, MA 02139)

The perception of voicing in final stops was investigated. Syllables which had been truncated at various points were presented to subjects for identification under two response conditions. In one condition, the response set consisted of two consonants (one voiced, one voiceless). In the other condition, subjects could choose either of the two consonants, or report that no final consonant was heard. The results indicated that the formant transitions, closure, burst, and vowel duration are important in determining whether a stimulus is heard as voiced or voiceless. In addition, there are coarticulation effects occurring as early as the first 50 msec of the preceding vowel which appear to affect the pattern of voiced/voiceless judgments. A common basis for the perception of voicing in initial and final stops is also considered. [Work supported by NINCDS.]

3:30

LL11. Vowel duration as a cue for consonant voicing? C. W. Fruin and D. M. Bischoff (Department of Hearing and Speech Sciences, Stanford University, Stanford, CA 94305)

Several studies state that the duration of the vowel preceding the consonant is a significant cue to the voicing characteristic of that consonant. The CVC's used in these studies have generally been prepared on the Pattern Playback. The present

study was designed to examine vowel duration as a consonant voicing cue in naturally produced speech. Several CVC's were recorded by both authors as stressed citation forms. Vowel durations were measured and parts were removed from: the endmost section of the vowel, the middle of the vowel, or the end *and* the middle of the vowel. At least for these stimuli produced by these speakers, no clear relation is seen between the length of the preceding vowel and the perception of the voicing characteristic of the final consonant.

3:50

LL12. Perception of vowel duration. William S-Y. Wang (Project on Linguistic Analysis, 2222 Piedmont Ave., University of California, Berkeley, CA 94720), Ilse Lehiste (Department of Linguistics, Ohio State University, Columbus, OH 43210), Chin-Kuang Chuang, and Nancy Darnovsky (Project on Linguistic Analysis, 2222 Piedmont Ave., University of California, Berkeley, CA 94720)

Lehiste and Pisoni have independently shown that the perception of vowel duration is influenced by the F_0 pattern that accompanies the vowel. Here we report some related results. Sixteen subjects are asked to judge the relative duration of pairs of synthesized stimuli. The stimulus pairs are made up of the vowels /i/, /a/, /a^i/, and nonspeech, where nonspeech is a single formant at 1500 Hz. The F_0 patterns are: level = 120 Hz, rising = 105—135 Hz, and falling = 135 to 105 Hz. The durations are 220, 240, 260, and 280 msec. The orthogonal combination of these factors makes up a total of 512 stimulus pairs. In comparing duration judgements of stimuli of rising (R), falling (F), and level (L) F_0's, we found that the following relation obtains: R > F > L, where " > " denotes "perceived as having greater duration." This relation holds to the same extent for both speech and nonspeech stimuli. There is no signficant difference in the perceived duration between /i/ and /a^i/. The most consistent effect, found also earlier in a pilot experiment we performed with 58 subjects, is that /i/ is heard as longer than /a/, in a ratio of 60% : 40%. Since /i/ is normally shorter than /a/ in natural speech, we posit a compensation mechanism mediating between perception and production that is responsible for this negative correlation between vowel height and duration. It is interesting to note that although the processes of duration judgment and pitch judgement (cf. the abstract by Chuang and Wang) are different, they seem to share a common compensation mechanism between production and perception. [Work supported by NSF Grant No. BNS 76-00017-Wang-2/78J34.2.]

4:00

LL13. Influence of vowel height, intensity, and temporal order on pitch perception. Chiu-Kuang Chuang and William S-Y. Wang (Project on Linguistic Analysis, 2222 Piedmont Ave., University of California, Berkeley, CA 94720)

Recently, it has been observed that high vowels are produced with a higher F_0, even in tone languages. Earlier, H. Fletcher reported that louder pure tones are perceived as having a lower pitch, but only in the range 50 to 500 Hz, which is the range of F_0 in speech. Here we report some experiments which investigate these relations. Fifteen subjects are asked to judge the relative pitch of pairs of synthesized vowels. The vowels are /i, e, u, a/ with intensity differences (ΔI) of 0, 10, 20, and 30 dB SPL, and at frequency differences (ΔF_0) of 0, 1, 2, and 3 Hz, with a base frequency of 100 Hz. Each vowel is 300 msec long and the inter stimulus interval is 400 msec. The orthogonal combination of these factors, i.e., four vowel types, four ΔI's, four ΔF_0's, and the two positions within each pair, makes up a total of 512 stimulus pairs. The interaction between pitch perception and vowel height is highly significant statistically. Pooled across all subjects, /a/ is heard as 2.2 Hz higher than /u/, 0.8 Hz higher than /i/, and 0.2 Hz higher than /e/. The intensity factor by itself appears to have no effect on pitch perception, but interacts with the postion factor in influencing pitch perception. The negative correlation between production and perception in the relation between vowel height and pitch is interpreted as the result of a compensation mechanism mediating between production and perception. [Work supported by NSF Grant No. BNS 76-00017-Wang-2/78J34.2.]

4:10

LL14. An acoustic correlate of syllabicity in english. H. Semiloff (Speech Communications Research Laboratory, Santa Barbara, CA 93109)

The purpose of this research was to test three hypotheses (1) Increasing the duration of a nonsyllabic segment will cause the segment to be perceived as syllabic. (2) The perception of syllabicity is categorical. (3) Listeners will differ in their labeling responses to stimuli when told to judge them in different speech styles. Seven words were synthesized: "blow," "plight," "dress," "crest," "prayed," "broke," and "sport." The steady-state portion of the /l/ or the /r/ (for "sport," aspiration of /p/) was lengthened in 10-msec increments. Subjects were asked to decide whether the word was monosyllabic or disyllabic: "blow-below," "plight-polite," "dress-duress," "crest-caressed," "prayed-parade," "broke-baroque," and "sport-support." Five groups of 15 listeners each participated in the labeling task. Four groups heard the stimuli with a precursor frame, "The word you will hear next is ___" in one of four speech styles: formal-slow, formal-fast, casual-slow, casual-fast. Each group was told to judge the words according to the criterion of being spoken in one of the designated styles. The fifth group heard the stimuli with no precursor and received no style instructions. Results showed that (1) increased durations of /r/ and /l/ did result in perception of the words as disyllabic, but increasing the duration of aspiration in "sport" did not; (2) perception of syllabicity appears to be categorical; and (3) there were no statistical differences among the four speech styles. [Work supported by NSF SOC75-10043.]

4:20

LL15. Stimulus range as a determinant of phoneme boundaries along synthetic consonant continua. Michael Studdert—Kennedy (Queens College, CUNY, Flushing, NY 11367 and Haskins Laboratories, 270 Crown St., New Haven, CT 06510)

Brady and Darwin (see Darwin, C.J. The Perception of Speech, in E.C. Carterette, and M.P. Friedman, *Handbook of Perception*, Academic, New York (in press), Vol. 7, report shifts in the phoneme boundary along a synthetic voicing continuum as a function of the range of stimuli presented within a test. The present study reports comparable shifts along synthetic voicing and place of articulation continua. Implications for the interpretation of adaptation boundary shifts are considered.

4:30

LL16. Perceptual adaptation to the duration of vowels preceding stop consonants. Paul D. Williams and Donald J. Sharf (Section of Speech and Hearing Sciences, The University of Michigan, 1111 E. Catherine St., Ann Arbor, MI 48104)

The affect of selective adaptation on the ability to identify stop consonants as being either voiced or voiceless solely on the basis of preceding vowel duration was tested. Stimuli consisted of computer-modified, real-speech sounds edited from a single utterance of /æd/ which varied in duration in 25-msec steps. Each subject classified recorded, random orderings of the stimuli as either AT or AD before and after periods of repetitive listening to long and short adapting stimuli. Calculated phonetic boundaries showed a mean shift toward the category of the adapting stimulus in both conditions, indicating a reduction in sensitivity for that category. The findings of independent, adaptable feature detectors for the voiced and voiceless categories and a greater resistance to adaptation for the voiced feature were comparable to results reported for adaptation studies employing voice onset time. [P.D. Eimas and J.D. Corbit, Cognitive Psychology 4, 99—109 (1973); P.D. Eimas *et al.*, Perception and Psychophysics 13, 247—253 (1973)].