Adaptation in speech production to transformed auditory feedback

John F. Houde and Michael I. Jordan

Citation: The Journal of the Acoustical Society of America **97**, 3243 (1995); doi: 10.1121/1.411746
View online: https://doi.org/10.1121/1.411746
View Table of Contents: https://asa.scitation.org/toc/jas/97/5
Published by the Acoustical Society of America

a vowel and nasal sound [N] can be a mora like a CV syllable. For these stimuli, the RT was not influenced by the increase of the number of morae. These results suggests a possibility that a mora consisting of a CV syllable is the smallest segment of the processing unit in Japanese speech production. However, morae consisting of a single vowel or nasal sound cannot perform as independent processing units.

1aSC16. Perceptual interaction of *F*1 and *F*0. Jose R. Benki (Linguist. Dept., Univ. of Massachusetts, Amherst, MA 01003)

While $F1$ is a primary perceptual correlate of vowel height, other dimensions of vowels, such as $F0$, covary with $F1$ in natural speech. This study investigates the perceptual interaction of $F0$ and $F1$ in back vowels using the Garner paradigm. Two factors are hypothesized to determine the interaction: the magnitude of $F1-F0$ and the locations of the harmonics. $F1$ and $F0$ are predicted to integrate negatively for vowels in a threshold region (3.0 Bark $<F1-F0<3.5$ Bark), whereas subthreshold ($F1-F0$ <3.0 Bark) and suprathreshold vowels ($F1-F0 \gg 3.0$ Bark) are predicted to show no interaction. Previous work [Benki *et al.*, J. Acoust. Soc. Am. **95**, 2977(A) (1994)] suggests that $F1$ and $F0$ do not interact in the threshold region but integrate positively in the suprathreshold region. The predicted dependence on the locations of the harmonics follows from work by Hughes and Diehl [J. Acoust. Soc. Am. **95**, 2978(A) (1994)] showing that $F1$ discriminability is enhanced when a harmonic is near nominal $F1$. To assess these effects, CVC stimuli ranging in $F1$ from 300–600 Hz and $F0$ from 90–120 Hz were presented to listeners in the Garner baseline and correlated classification tasks. A perceptual $F0/F1$ space is inferred using a detection-theory analysis of the accuracy data. [Work supported by NSF and NIH.]

1aSC17. Word frequency effects on the acoustic duration of morphemes. Beth L. Losiewicz (Dept. of Psych., Colorado College, Colorado Springs, CO 80901)

The acoustic duration of the English past tense (ED) morpheme was measured for matched high- and low-frequency verbs (e.g., KNEADED/NEEDED). The ED on low-frequency verbs was of longer acoustic duration than the ED on matched high-frequency words. The rhyming portion of the matched verbs also showed a lengthening effect for the low-frequency words; in contrast to previous reports that word frequency does not affect word acoustic duration [Geffen and Luszcz, Mem. Cogn. **11**, 13–15; Wright Mem. Cogn. **7**, 411–419]. However, this effect was statistically independent of the ED length effect, and the final phonetic segment of a low-frequency monomorphemic verb stem was not longer in acoustic duration than a homophonous segment on a matched high-frequency verb (e.g., the /d/ in KNEAD/NEED). Further, the ED morpheme is of longer acoustic duration than a homophonous segment in a nonverb homophone (e.g., RAPPED/RAPT), as earlier reported for the morpheme /s/ [Walsh and Parker, J. Phon. **11**, 201–206]. This set of evidence corroborates a frequency-dependent dual-access processing theory of linguistic morphology: that high-frequency complex words are processed holistically and low-frequency complex words are processed componentially [cf. Bybee, *Morphology* (1985)].

1aSC18. Adaptation in speech production to transformed auditory feedback. John F. Houde and Michael I. Jordan (Dept. of Brain and Cognit. Sci., MIT, 79 Amherst St., Cambridge, MA 02139)

This study investigated the ability of the speech production system to learn to compensate for changes in auditory feedback. The setup used for this was a DSP system that transformed the immediate feedback a subject received when speaking. This system can analyze a subject's speech into a formantlike representation, possibly alter it, and then use it to resynthesize speech which is fed back to the subject with no noticeable delay (16 ms). The first of the experiments investigated whether subjects would learn to compensate for a change in vowel identity when producing CVC words. It was found that compensatory articulations were indeed learned, and that these persisted even when no auditory feedback was provided. The findings suggest similarities between speech and other sensorimotor tasks, such as reaching, which also show such adaptation. Other experiments characterizing the degree to which this effect generalizes across differing word and vowel environments will also be presented.

1aSC19. Differential masking of individual words within a sentence. Theodore S. Bell (UCLA School of Medicine, Head & Neck Surgery, 31-24 Rehabilitation Ctr., Los Angeles, CA 90024-1794) and Richard Wright (UCLA, Los Angeles, CA 90024-1543)

Three experiments tested the effectiveness of adjusting the amplitude of a noise masker to selectively alter intelligibility of individual words within sentences. Stimulus sentences contained three key words; all were commonly used and phonetically unique. A speech-shaped noise was presented at 65 dB SPL for all experiments. The initial study tested 12 listeners under two conditions: the noise level remained flat, or was attenuated 6 dB under the middle key word. The speech was presented at 54 and 58 dB SPL. The middle key word improved by as much as 30% (p <0.001), while other key words were unaffected. In the next two experiments, the noise was either increased or decreased by 3 dB at a key word. In one, the manipulation was at the first key word, and in the other was at the last key word. In both experiments, the speech was presented to ten listeners at 56 dB SPL. The 3-dB increment in noise significantly decreased the word's intelligibility, and the 3-dB decrease increased the intelligibility ($p<0.01$) without affecting other key words in the sentence. This technique has application in reducing variability in sentence intelligibility and greatly reducing perceptual dynamic range. Implications for "all-or-none" contextual recognition in adverse conditions are discussed.

1aSC20. Pitch perception physiology and psychophysics as a basis for the design of pitch detection algorithms. Robert A. Houde (RIT Res. Corp., 75 Highpower Rd., Rochester, NY 14623) and James Hillenbrand (Western Michigan Univ., Kalamazoo, MI 49008)

The generation of high-quality speech with a source-filter vocoder depends to a very great extent on accurate analysis of source parameters. After decades of research, even state-of-the-art pitch detection algorithms tend to make gross errors in the analysis of signals that present no difficulty for the human listener. In this study a review of a broad range of pitch detection algorithms was undertaken, with particular attention to the plausibility of those algorithms in relation to what is currently known about the psychophysics of pitch perception and the neural coding of speech signals. Our principal conclusion from this review is that the most plausible model is a time-domain pitch perception scheme proposed more than 4 decades ago by Licklider [J. C. R. Licklider, Experientia **7**, 128–133 (1951)], and extended in more recent studies. The implications of these findings for source-filter vocoders will be discussed, and an implementation of the Licklider model using level-crossing interval histograms will be described.

1aSC21. Contribution of different frequency regions to detection of additive noise in vowels. Muralidhar R. Kudumala (Electr. Eng., Univ. of Oklahoma, Oklahoma City, OK 73170) and Blas Espinoza-Varas (Univ. Oklahoma Health Sciences Ctr., Oklahoma City, OK 73170)

An important issue in audio coding is the detection of quantization noise masked by speech. This investigation examined how different frequency regions contribute to the detection of broadband noise masked by a vowel. The vowel was synthesized by addition of the first 32 harmonics of a 200-Hz fundamental, with amplitudes appropriate to /i/. Broadband noise was synthesized by random-phase addition of the harmonics of a 10-Hz fundamental that fall within the vowel bandwidth. Noise detection thresholds were measured in a 2IFC task with an adaptive procedure (Levitt, 1971). Detection thresholds were obtained for noise with spectrum envelope parallel to that of the vowel, and for the same noise containing -6.0-dB spectral notches in either of the following bands: (a) 0.2–0.5 kHz; (b) 0.2–1.0 kHz; (c) 4.2–5.0 kHz; and (d) 4.7–5.0 kHz. Thresholds obtained in three highly trained listeners ranged from -27 to -32 dB (expressed in terms of noise-to-vowel power ratio). Threshold differences between spectral notch conditions were small. The results were compared to predictions of the excitation pattern model proposed by Moore and Glasberg [Hear. Res. **28**, 209–225 (1987)]. [Work supported by OCAST-HR4-064.]