Perceptual and Motor Skills, 1981, 53, 515-519.

INFLUENCE OF PERSONAL CHARACTERISTICS OF THE SPEAKER ON PHONETIC QUALITY IN PERCEPTION OF VOWELS[1]

ROBERT ALLEN FOX

Ohio State University

Summary.—This study was designed to investigate the effect of speaker's voice quality on perceived phonetic quality in vowels using a perceptual scaling task. 16 subjects scaled sets of vowels produced by six different speakers (three women, three men). Analysis indicated no significant influence of speakers' quality on judgments of similarity of vowels. Results suggest that judgments of similarity were made on the basis of linguistically categorized images.

Perception of vowels can be characterized as the identification of the phonetic quality of a vowel based on its acoustic characteristics—the most important acoustic parameters being the first two or three formants (Peterson, 1952, 1961; Ladefoged, 1975). Phonetic quality, however, cannot be determined in a straightforward manner from absolute frequency values of formants. For any particular speaker these absolute values depend upon length of the vocal tract as well as the shape of the articulators, although the over-all formant patterns of particular phonetic qualities will be similar from speaker to speaker (Joos, 1948). The speech wave thus not only contains information relevant to phonetic quality but includes information about the physical (voice) quality of each individual speaker. The latter is sometimes termed speaker's quality[2] and refers to the acoustic characteristics of a vowel which serve to differentiate individual speakers but are not necessarily related to any linguistic distinctions. This would include not only the absolute range of frequencies of formants for a given speaker but factors such as clarity, roughness, intensity, fundamental frequency, and speaking rate (Carterette & Barnebey, 1975; Holmgren, 1963; Voiers, 1964).

To what extent does the speaker's quality of a vowel affect its perceived phonetic quality, or to what degree are the two qualities extracted from the speech wave in an independent manner? Although these questions have received relatively little attention, they are very important to the construction of models of vowel perception. Many perceptual models (Massaro, 1975) incorporate a "normalization" process wherein speakers' variations are separated

[1]Reprints of this article may be obtained from Robert A. Fox, Linguistics Department, Ohio State University, Columbus, OH 43210. Data discussed though not presented here are on file in Document NAPS-03883 with Microfiche Publications, P.O. Box 3513, Grand Central Station, New York, NY 10017. Remit $4.00 for fiche or $14.05 for photocopy.

[2]There is potential terminological confusion in the literature concerning these two qualities. One can use the terms *phonetic quality* and *vowel quality* interchangeably in reference to that aspect of vowels responsible for linguistic distinctions. Similarly, both *quality of a speaker* and *personal quality* refer to individual, nonlinguistic acoustic variations in vowels.

from linguistically significant information. This normalization process could utilize acoustic properties within the vowel itself such as fundamental frequency (Slawson, 1968) or the higher formants (Fujisaki & Kawashima, 1968). Alternatively, normalization for a particular vowel could be done on the basis of the formant structure of other vowels from that speaker (Gerstman, 1968) although it is clear that vowels can be identified without necessary reference to other sounds since subjects are fairly accurate in identifying isolated words from different speakers when they are presented in random order (Peterson & Barney, 1952).

Most research directed at the influence of speakers' quality on perception of speech has concentrated on the effect which the speakers' quality of one segment (or phrase) has upon the perception of a following token. For example, using synthetic speech, Ladefoged and Broadbent (1957) demonstrated that varying the frequencies of formants of a precursor phrase could influence the identification of an unvarying token. Although this effect has, however, been replicated for synthetic speech (Ainsworth, 1974; Thompson & Hollien, 1970), efforts to produce the same effect using natural speech have been unsuccessful (Dechovitz, 1977; Fox, unpublished data). Further evidence suggesting that differences in personal quality of speakers affect identification of vowels can be found in reports by Cole, Coltheart, and Allard (1974) and by Summerfield and Haggard (1975) who demonstrated that choice reaction times to pairs of stimuli were increased when they were produced by different speakers rather than by the same speaker. However, these results do not necessarily indicate that information on quality of the speaker directly influences the decision about phonetic quality, merely that extra cognitive processing is necessary when there is a shift in frame of reference for a formant.

Rather than examining how the quality of a speaker for one segment can affect the perceived phonetic quality of a second segment, the present study concerned the extent to which subjects are affected by personal quality when making perceptual judgments between vowels from a single speaker. Subjects performed paired comparisons on six different sets of vowel-pairs, each stimulus set varying in terms of quality of the speaker but not in phonetic quality. These judgments of similarity were then analyzed to determine whether a significant effect due to quality of the speaker was present. Such a result would obtain only if listeners take into consideration the absolute physical characteristics of the vowels being scaled in addition to the linguistically significant information.

METHOD

Stimulus Materials

The stimulus tokens were constructed using nine different vowels /i ɪ ɛ æ a ʌ o ʊ u/ which appear in the phonetic environment [h—d]. This set

includes most of the relatively monophthongal vowels in American English. There were 36 different possible vowel dyads (excluding vowel-pairs with themselves and neglecting order). Six different sets of vowels were recorded, each set produced by a different speaker (a vowel-pair × speaker quality design). Each separate set of vowels utilized a different randomized order of stimulus pairs and consisted of four blocks of stimuli—each block containing all 36 vowel-pairs plus buffer-pairs.

The stimulus materials were designed to maximize differences in speakers' quality among the various vowel-sets. The speakers producing the stimuli included three women and three men. Within each of the sex subgroups there was a speaker with a relatively low fundamental frequency (F_0), one with a relatively medium F_0, and one with a relatively high F_0 ("relatively" because what would be a low F_0 for a woman might be considered medium for a man). The sex differentiation in the group of speakers produced a systematic variation among the vowels sets in terms of relative height of formants—women tend to have shorter vocal tracts than men and thus exhibit higher frequencies of formants in general.

Speakers producing the stimuli were all from the same dialect group to help ensure that the phonetic quality of corresponding vowels across different qualities of speaker was uniform. In addition, the vowels were produced under the direction of the experimenter who closely monitored the phonetic quality of all vowel stimuli produced.[3] Vowel-pairs appeared on the stimulus tape at 6-sec. intervals and were recorded using a Tandberg Series 9100X tape recorder with a Realistic "Highball" microphone. Care was taken to achieve similar levels of loudness and durations of vowel across all vowel-sets.

Experimental Task

Subjects were required to listen to two sequentially presented vowels and to judge the similarity or dissimilarity of the vowel-pair on a 9-point scale. Typed instructions describing the experimental task were read by all subjects. An introductory set of 10 practice vowel-pairs preceded the experimental stimuli. Each subject heard the stimuli as reproduced by a Tandberg 9100X Series tape recorder through Nova Pro earphones while sitting in an IAC sound-attenuated room. Each subject scaled the six separate vowel-sets in three 1-hr. sessions on consecutive days. Order of presentation of the stimulus-sets was counterbalanced across subjects.

[3]Slight idiolectal variation in the phonetic quality of corresponding vowels across different vowel sets, however, may be present since, as Ladefoged (1967) suggests, it may be the case that no single group of speakers, even those from the same dialect, will have vowels of *identical* phonetic quality. This fact would tend to increase the likelihood that a significant result due to the speaker's quality would emerge from the analysis but, as we shall see, since no significant variation was found due to speaker's quality, our conclusions concerning a prototypic comparison are even more strongly supported.

Subjects

Sixteen subjects, all nonlinguists from the same dialect group as the speakers, took part in the scaling experiment and were evenly divided into women and men. Each subject was paid $10.00 for participating. In order that the perceptual distances obtained be as free of error as possible and that each subject's data represent reliable and replicable scaling judgments, statistics for consistency were applied to the obtained perceptual data. For each of the 16 subjects correlations were calculated among the four blocks within each of the six stimulus-sets. Eleven of the 16 subjects demonstrated criterial consistency, i.e., all correlations were significant at the .01 level or beyond, and were used in subsequent analyses.

Results and Discussion

A three-way analysis of variance was performed on the data (speaker quality $\times$ vowel-pair $\times$ trial). A significant main effect for vowel-pair ($F = 23.09$, $df = 35/350, p < .001$) accounted for 34.9% of the variance but no significant effect for trial ($F = .221$, $df = 3/30, p > .50$) or quality of speaker ($F = .416$, $df = 5/50, p > .50$) was obtained. All interactions were at least marginally significant but significance stemmed from the large degrees of freedom involved and did not represent important effects (none accounted for more than 2.5% of the variance).

These results do not support the hypothesis that quality of the speaker of a vowel can affect its perceived phonetic quality, but this generalization may be limited to the particular experimental design used. Specifically, these results indicate that, when subjects evaluate the similarity of vowels from a particular speaker, nonlinguistic information does not play a role in the decision. For example, though the physical differences (seen in terms of frequency values for formants, F_0, etc.) between Speaker 1's [i] and [u] may be greater than between Speaker 6's [i] and [u], the perceptual distance between the two vowels (as reflected in judged similarity) in each case is the same. These results are compatible with the hypothesis that comparisons in a task judging similarity are done not on the basis of precategorical acoustic images (Crowder & Morton, 1969) but rather on the basis of (linguistically) categorized images. That is, subjects first identify each vowel in a pair as being one of a number of internal prototypes or categories (Oden & Massaro, 1978) and then make judgments of similarity not on the basis of the physical (acoustic) characteristics of the two vowels but on the basis of the internal perceptual attributes of features of their corresponding prototypes.

Clearly more research on this topic is needed, especially using an experimental design which examines the influence of quality of a speaker on the perception of phonetic quality both within and across speech-segments, as well as the influence of phonetic quality on the perception of quality of a speaker.

REFERENCES

AINSWORTH, W. A. The influence of precursive sequences on perception of synthesized vowels. *Language and Speech*, 1974, 17, 103-107.

CARTERETTE, E. C., & BARNEBEY, A. Recognition memory for voices. In A. Nooteboom & S. G. Nooteboom (Eds.), *Structure and process in speech perception*. New York: Springer Verlag, 1975. Pp. 246-266.

COLE, R., COLTHEART, M., & ALLARD, F. Memory of a speaker's voice: reaction time to same or different voiced letters. *Quarterly Journal of Experimental Psychology*, 1975, 26, 1-7.

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage (PAS). *Perception and Psychophysics*, 1969, 5, 365-373.

DECHOVITZ, D. R. Information conveyed by vowels, a negative finding. *Journal of the Acoustical Society of America*, 1977, 61, S39. (Abstract)

FUJISAKI, H., & KAWASHIMA, T. The roles of pitch and higher formants in the perception of vowels. *IEEE Transactions on Audio and Electroacoustics*, 1968, AU-16, 73-77.

GERSTMAN, L. J. Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics*, 1968, AU-16, 78-80.

HOLMGREN, G. L. Physical and psychological correlates of speaker recognition. *Journal of Speech and Hearing Research*, 1963, 10, 57-66.

JOOS, M. Acoustic phonetics. *Language Monographs*, 1948, 24(2), No. 23.

LADEFOGED, P. *Three areas of experimental phonetics*. London: Oxford Univer. Press, 1967.

LADEFOGED, P. *A course in phonetics*. New York: Harcourt Brace Jovanovich, 1975.

LADEFOGED, P., & BROADBENT, D. E. Information conveyed by vowels. *Journal of the Acoustical Society of America*, 1957, 29, 98-104.

MASSARO, D. *Understanding language*. New York: Academic Press, 1975.

ODEN, G. C., & MASSARO, D. Integration of featural information in speech perception. *Psychological Review*, 1978, 85, 172-191.

PETERSON, G. E. The phonetic value of vowels. *Language*, 1952, 27, 541-553.

PETERSON, G. E. Parameters of vowel quality. *Journal of Speech and Hearing Research*, 1961, 4, 10-29.

PETERSON, G. E., & BARNEY, H. L. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, 1952, 24, 175-184.

SLAWSON, A. W. Vowel quality and musical timbre as functions of spectrum envelope and fundamental frequency. *Journal of the Acoustical Society of America*, 1968, 43, 87-101.

SUMMERFIELD, A. Q., & HAGGARD, M. P. Vocal tract normalization as demonstrated by reaction times. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975. Pp. 115-142.

THOMPSON, C. L., & HOLLIEN, H. Some contextual effects on the perception of synthetic vowels. *Language and Speech*, 1970, 13, 1-13.

VOIERS, W. D. Perceptual bases of speaker identity. *Journal of the Acoustical Society of America*, 1964, 36, 1065-1073.

Accepted August 28, 1981.