PERCEPTUAL CONSEQUENCES OF EQUALIZING LOUDNESS DIFFERENCES OF VOWELS VARYING IN VOICE QUALITY

PERTTI HURME

Department of Communication, University of Jyväskylä
40100 Jyväskylä, Finland

ABSTRACT

Listening tests were organized in which vowels varying in sound pressure level (SPL) and voice quality were rated. The vowels were produced at 60, 70 and 80 dB; the SPL differences were equalized for the listening tests. The following voice qualities were simulated: normal, nasal, strained, breathy, rough. The results show that original SPL differences are reflected in the perceptual ratings in a voice quality dependent manner.

INTRODUCTION

The preliminary observations reported here are part of a research project on the perceptual, acoustic and clinical properties of normal and dysphonic voice [1]. One of our methods is to study the effect of controlled variations in the acoustic properties of voice on perceptual dimensions. Data from such studies may be useful in the construction of tools for examining normal and dysphonic voices.

The human voice shows much variation. Voices differ in fundamental frequency (related to pitch), sound pressure level (to loudness) and long-term spectral characteristics (to voice quality [2], resulting from long-term laryngeal and supralaryngeal settings) as well as other properties (e.g. temporal). The object of the present study is to evaluate voices varying in sound pressure level (SPL) and spectral characteristics on several perceptual dimensions.

More specifically, this paper investigates the perceptual impressions listeners extract from speech produced by simulating several voice qualities at three SPLs. Impressions of voices are certainly affected by voice quality. They are also influenced by SPL: for instance, voices with high SPL are generally judged as more "carrying" or powerful than those with low SPL. However, does this also hold when SPL differences originally present in the voices are technically eliminated by equalizing the SPLs to a common level? If impressions of voices manipulated this way are affected by the SPL, this must be due to reasons other than absolute SPL, probably the relative spectral properties of the voices. This is possible, as both momentary spectra of individual sounds and long-term spectra of speech are highly affected by sound pressure level [3, cf. also 4]. In general, the fundamental frequency dominates the spectra of sounds or speech with low SPL, whereas in sounds or speech with high SPL higher harmonics (especially in the first formant region) are much more prominent.

PROCEDURE

The vowel [ɑ] was produced by the present author in five simulated voice qualities: normal, nasal, strained (or tense, pressed), breathy (with a strong high-frequency noise component) and rough, cf. [5]. These qualities were clear and extreme. The five qualities were produced at three sound pressure levels: 60, 70 and 80 dB, by means of visual feedback from a decibel-meter when producing the 15 vowel tokens. Care was taken to produce the vowel in the same manner at all SPLs. The vowels were then digitally extracted from the master tape [6] and adjusted in SPL to the same level, i.e. the amplitude of vowels produced at 60 and 70 dB was raised to the level of those produced at 80 dB. Thus, the original loudness differences were leveled out. These vowels were then recorded in pseudorandom order on a test tape.

The vowels, presented at about 80 dB in the relatively small room where the listening tests were conducted, were rated by two groups of students of speech communication and logopedics (N=28) with some experience in assessing voices. The vowels were rated on the dimensions good/poor, "carrying" (powerful, stentorian, tragend)/weak, powerless and pleasant/unpleasant by all listeners. In addition, the vowels were rated by half of the group (N=14) on the dimensions nasal/normal, strained/normal, breathy/normal and rough/normal. Scales of seven points were used in the assessment in the first three dimensions, e.g. extremely pleasant (3), moderately pleasant (2), slightly pleasant (1), neither pleasant nor unpleasant (0), slightly unpleasant (-1), moderately unpleasant (-2), extremely unpleasant (-3) and of four points in the latter four dimensions, e.g. not at all nasal (0), slightly nasal (-1), moderately nasal (-2), extremely nasal (-3). The results of the listening tests were analysed statistically by means of Chi-square tests.

RESULTS AND DISCUSSION

Table 1 shows the results of the listening test for the five voice qualities on the seven perceptual dimensions (as median values, pooled for all subjects and the three SPLs). The median values are on the whole as expected: for instance, vowels produced with nasal voice quality are on the average moderately nasal and those produced with rough voice quality moderately rough. Vowels produced with normal voice quality are rated as moderately good and pleasant, neither breathy, rough nor strained, slightly nasal.

Table 2 shows the median values of the ratings separately in the three SPLs in cases where there was a statistically significant difference in the perceptual ratings. The perceptual ratings are affected by the original SPL level in a different manner in the five voice qualities here investigated. Normal vowels appear to be most susceptible to the effect of SPL. High SPL (in the original signal) of normal vowels is associated with less good, less pleasant, more rough and more strained impressions. Normal vowels with low SPL show no difference to those with medium SPL. Nasal vowels with low SPL are associated with less carrying (weak) impressions, whereas high SPL goes with more carrying (powerful) impressions. Strained vowels originally produced with low SPL were rated as more pleasant than those with higher SPL. Rough and breathy vowels generally give a negative impression on all dimensions; low SPL in rough vowels is associated with less strain.

The results imply that SPL information is important in perceptual studies of voice quality. SPL should either be standardized (for instance by means of visual feedback) when producing the sample to be

Table 1: Median values of the estimations of the stimulus vowels produced with five voice qualities (normal, nasal, strained, breathy, rough) on seven perceptual dimensions (good, carrying, pleasant, nasal, strained, breathy, rough) by 28 Ss; column and row medians are also given.

PERCEPTUAL DIMENSIONS:	STIMULUS VOWELS:					
	NORMAL	NASAL	STRAINED	BREATHY	ROUGH	ALL
GOOD (3/-3)	2	0	-1	-3	-3	-1
CARRYING (3/-3)	1	1	1.5	-2	-2	0
PLEASANT (3/-3)	2	1	-1	-3	-3	-1
NASAL (0/-3)	-1	-2	-1	0	0	-1
STRAINED (0/-3)	0	0	-1	-2	-2	-1
BREATHY (0/-3)	0	0	0	-1	-3	-1
ROUGH (0/-3)	0	0	0	-2	-2	-1
ALL	2	1	1	-3	-3	

Table 2: Median values of the estimations on the perceptual dimensions with a statistically significant difference between stimulus vowels originally produced at 60, 70 and 80 dB.

STIMULUS VOWELS:	PERCEPTUAL DIMENSIONS:	p<	SPL 60	SPL 70	SPL 80
NORMAL	GOOD/POOR (3/-3)	0.01	2	2	1
NORMAL	PLEASANT/UNPLEASANT (3/-3)	0.01	2	2	1
NORMAL	ROUGH (0/-3)	0.01	0	0	-0.5
NORMAL	STRAINED (0/-3)	0.001	0	0	-1
NASAL	CARRYING/WEAK (3/-3)	0.01	-1	1	2
STRAINED	PLEASANT/UNPLEASANT (3/-3)	0.01	1	-1	-1
ROUGH	STRAINED (0/-3)	0.05	-1.5	-2.5	-2.5

evaluated or measured as an independent variable when recording speech for perceptual ratings. This is especially important when assessing dysphonic voices as they show much variation in SPL.

On the whole, the results indicate that extremes in the voice (here represented by SPLs of 60 and 80 dB) tend to be associated with less favourable ratings. However, the results presented here apply only to one speaker. Generalizations are perhaps to some extent valid for the Finnish speech culture as well. The cross-cultural comparison of voice quality (and presumably also the use of SPL) in different speech situations is a challenging but difficult task in view of the immense variation between languages and between cultures [7].

REFERENCES

[1] Hurme, P. & A. Sonninen 1986. Acoustic, perceptual and clinical studies of normal and dysphonic voice. Journal of Phonetics, 14, 489-492.

[2] Laver, J. 1980. The phonetic description of voice quality. Cambridge: University Press.

[3] Fant, G. 1973. Speech sounds and features. Cambridge, Mass.: MIT Press.

[4] Hurme, P. 1980. Auto-monitored speech level and average speech spectrum. Papers in Speech Research (Jyväskylä), 2, 119-127.

[5] Isshiki, N. & Y. Takeuchi 1970. Factor analysis of hoarseness. Studia Phonologica, 5, 37-44.

[6] Karjalainen, M. 1980. A computer-based system for speech and audio signal research. NAS-80 (Turku), 291-294.

[7] Ladefoged, P. & N. Antoñanzas-Barroso 1986. Computer measures of breathy voice quality. Working Papers in Phonetics (UCLA), 61, 79-86.