Perceptual Description of Long-term Speaker-Characterizing Voice Features by means of Semantic Differential Ratings of Running Speech

W.P.F. Fagel and L.W.A. van Herpt
Amsterdam, the Netherlands

1. Introduction

For several reasons it is interesting to know which acoustic characteristics define our subjective impressions of running speech, so far as these impressions concern long-term speaker-characterizing voice features, or 'extra-linguistic voice features' according to the definition given by Laver and Trudgill (1979). These reasons vary from the need of more objective methods for the diagnosis of speech disturbances to the desirability of a reliable procedure for a phonetic description of dialects and sociolects.

If we want to analyse the relation between acoustic and perceptual features in an adequate way, we need an efficient and reliable instrument for collecting subjective ratings on speech in the first place. Our research is concerned with the development of such an instrument by means of the semantic differential technique (Osgood, Suci and Tannenbaum, 1957).

2. Rating Experiments

In Dutch, as in other languages, there are hundreds of adjectives that can be used to describe long-term speaker-characterizing voice features. These adjectives range from very general, like 'pleasant', 'slow' and 'powerful', to highly specialized like 'monophthongized', 'hyperfunctional' and 'breathy'. All these adjectives can be considered to represent potential criteria for judging a person's voice ('voice' being used here as including phonation *and* articulation).

However, such criteria differ strongly in reliability (consistency among judges when used for rating a speaker) and relevance (discriminability among speakers, proportion of variance in the speaker population that is accounted for). Apart from this, the criteria cannot a priori be considered to be independent of each other. On the contrary, it is clear that many criteria are strongly related to each other. This means that any arbitrary sample of descriptive adjectives selected to describe different speakers will yield more or less redundant information.

Therefore we conducted a number of experiments, the purpose of which was to construct a standardized procedure for perceptually describing a speaker's voice. We wanted this procedure to be based on the potential

judgement criteria mentioned above, covering all relevant perceptual parameters in an optimally reliable and economical way.

For this reason, some 800 terms referring to special attributes of speech were collected first by our colleagues of the Institute. Various bipolar 7-point rating scales were obtained by pairing contrasting items from this collection. After several preliminary experiments a rating form consisting of 35 bipolar rating scales was put into use.

Ten speakers were subsequently judged on these rating scales by 235 listeners. The speakers were all native, normal speakers of Dutch. (One speaker had a rather husky voice.) We had our speakers read uniform material from typewritten texts, thus eliminating differences in vocabulary and grammatical accuracy.

The subjects who served as judges did not know or see the speakers but merely listened to their tape-recorded voices. Most listeners (211) were students from training courses of Speech Therapists (one group from Antwerp, Belgium and 6 groups from various regions in the Netherlands). The listeners varied as to degree of training (first to third year of training). The remaining 24 listeners were students of Dutch from the University of Amsterdam.

A more detailed discussion of this experiment and its results will be available elsewhere (Fagel, van Herpt and Boves, 1983).

3. Results

First, the rating scores were subjected to a scaling analysis based on Thurstone's Law of Categorical Judgement (Torgerson, 1958). This analysis was used to obtain information about the linearity of the scales, their discriminative power, the reliability of the scores and the extent to which the psychological continuum underlying the scales can be considered as unidimensional.

Subsequent factor analysis of the correlations between the 35 rating scales yielded 5 orthogonal factors. The first factor was interpreted as 'melodiousness' or 'variety' the second factor as 'articulation quality' and the third as 'static voice quality', strongly associated with perceived clarity or brightness as well as with subjective strength of voice. The last two factors were clearly associated with pitch and tempo.

At the University of Nymegen the same 35-scale rating instrument was used in an experiment where 6 male speakers were judged twice by 117 listeners. Factor analysis of the scores yielded virtually the same factor structure as we found in our perceptual data.

A subset of 12 carefully selected scales proved to be sufficient to generate this factor structure for both groups of speakers. The selection was based on the results of the foregoing scaling analysis and on such criteria as factorial purity and interrater reliability.

The factor structure we found appeared to be very stable over different groups of raters, suggesting strong validity of the perceptual dimensions

involved. There is some evidence, though, that the third factor actually incorporates two different criteria, one 'clarity' criterion represented by items like 'husky--not husky' and 'dull--clear', and one 'potency' criterion, represented by 'powerful--weak' and 'loud--soft' for example. (The scale labels mentioned are, of course, only translations of the Dutch adjectives actually used in the experiments).

Since we want to check the stability of the factorial structure over different groups of speakers, a short rating form was designed on the basis of the results of our perceptual analysis. This rating form contains 14 bipolar rating scales, the 12 scales selected before plus two. Each of the 5 factors found in our study is represented by 2 rating scales, the third factor by four rating scales (two items for each criterion presumably confined in this factor). Although global evaluative ratings on the scales 'pleasant--unpleasant' and 'beautiful--ugly' are associated most strongly with the first factor, these ratings clearly cannot be predicted from speakers' scores on the first factor exclusively.

SPEAKER 1	-3	-2	-1	0	+1	+2	+3	
MONOTONOUS	★★★★★★★★★★★★★★★							MELODIOUS
EXPRESSIONLESS	★★★★★★★★★★★★★★★							EXPRESSIVE
SLOVENLY	★★★★★★★★★★★★★★★★★★★★★★★★							POLISHED
BROAD	★★★★★★★★★★★★★★★★★★★★★★★							CULTURED
DULL	★★★★★★★★★★★★★★★★★★★★★							CLEAR
HUSKY	★★★★★★★★★★★★★★★★★★★★★★★							NOT HUSKY
WEAK	★★★★★★★★★★★★★★★★★★★							POWERFUL
SOFT	★★★★★★★★★★★★★★★★★★★★★★							LOUD
SHRILL	★★★★★★★★★★★★★★							DEEP
HIGH	★★★★★★★★★★★★							LOW
DRAGGING	★★★★★★★★★★★★★★★★★							BRISK
SLOW	★★★★★★★★★★★★★							QUICK
UGLY	★★★★★★★★★★★★★★★							BEAUTIFUL
UNPLEASANT	★★★★★★★★★★★★★★							PLEASANT

Figure 1. Speech profile for speaker 1 (male).

Therefore these two scales were adopted in our rating form as a separate component to keep the possibility of checking on the relations between global evaluative reactions and the judgements on the perceptual criteria we isolated. This short-hand rating form will soon be used in new rating experiments with larger groups of speakers.

Speech profiles can be composed from ratings on the 14 scales involved. Figure 1 shows such a profile for one of our speakers.

At least for the 10 speakers used in our experiments these speech profiles proved to be very reliable if based on the scores of 25 listener-judges or more. Most reliable were scores on the scales 'monotonous--melodious' and 'expressionless--expressive'. Ratings on the tempo scales 'dragging--brisk' and 'slow--quick' turned out to be least reliable, that is to say, listeners disagreed most on these scales. The new rating experiments mentioned above will have to confirm this reliability information over a larger set of speakers.

Acknowledgements

References

Fagel W.P.F., van Herpt, L.W.A. and Boves, L. (1983). Analysis of the perceptual qualities of Dutch speakers' voice and pronunciation. Accepted for publication in *Speech Communication*.

Laver, J. and Trudgill P. (1979). Phonetic and linguistic markers in speech. In :K.R. Scherer and H. Giles (eds.), *Social Markers in speech*. University Press: Cambridge, 1-32.

Osgood, C.E., Suci, G.T. and Tannenbaum, P.H. (1957). *The Measurement of Meaning*. University of Illinois Press: Urbana.

Torgerson, W.S. (1958). *Theory and Methods of Scaling*. Wiley: New York.