ANALYSIS AND PREDICTION OF DIFFERENCE LIMEN DATA FOR FORMANT FREQUENCIES. THE TESTING OF TWO DISTANCE MEASURES

Lennart Nord
Dept. of Speech Communication, KTH, Stockholm

Introduction [1]

The aim of this study is to relate perceptual distances drawn from difference limen data to a physical distance measure, taking into account some facts about the hearing mechanism. The distance measure and its correlation to the perceptual distance can be regarded as a way of testing the underlying perception model.

We used sound stimuli of the type Flanagan (1955) used in his study of difference limen for formant frequencies. Six groups of four-formant synthetic vowels were produced with the series synthesizer OVE III. In each group, F1 or F2 was systematically shifted in seven steps of 10 or 25 Hz above and below the reference vowels (see Table 1).

Table 1. Reference sounds for the stimulus sets

no	1	2	3	4	5	6	
F_4	3550	./.	./.	./.	./.	./.	(Hz)
F_3	2500	./.	./.	./.	./.	./.	
F_2	1500	./.	./.	1000	1500	2000	
F_1	300	500	700	500	./.	./.	

In an AB test, listeners judged whether they could discriminate between the reference sound and a test sound.

A number of distance measures based on a more or less refined processing of static speech sound spectra have been presented during the last years. The correlation between perceptual data and distance measures that are solely based on peripheral auditory processing may give poor results in identification tests which involve higher levels of processing. Hopefully, a better correlation will be reached if the perceptual data are taken from discrimination tests where less phonetic processing will play a role for the listener's judgement. This was the reason for choosing this particular set of stimuli.

Distance Measures

In the present study, distance measures based on two types of auditory representations of the sound stimuli were tested.

Plomp (1970) formulated a distance measure in a study of the timbre of complex tones. Basically, the distance between two steady state complex tones, having the same loudness, pitch and duration was defined as the distance between two points in an m-dimensional space where the m coordinates correspond to the intensity differences in m 1/3-octave bands (cf. the critical band theory of hearing, Zwicker, 1961). Plomp found a good correlation between the distance measure and dissimilarity indices for musical instrument tones, vowels, and pipe organ stops. In Lindblom (1978), the measure is tested on speech-like material and good results were likewise obtained.

Schroeder et al (1980) proposed a spectral distance measure based on the masking properties of auditory perception. The auditory spectrum is transformed into loudness per critical band (sone/Bark) as compared to Plomp's model which does not include masking effects, but represents the spectrum as intensity per critical band (dB/Bark). For further details, see Plomp (1970) and Schroeder et al (1980).

Lindblom and Bladon (1980) tested the "Schroeder et el" model in a discrimination test with synthetic vowels. Carlson and Granström (1979) examined a number of auditory models and tested them with a variety of different speech materials. As one of their auditory representations of the signal spectra, they adopted the "Schroeder et al" model but introduced the equal loudness curves of hearing in the transformation from intensity to loudness.

In the present study, the Carlson-Granström version of the sone/Bark representation of spectra was used.[2]

At least two ways of calculating the difference between two transformed spectra seem reasonable. Either a simple summation of the differences between spectra at every sample point (in every critical band) or a geometric summation, that is, the square root of the sum of the squared differences. The first distance is usually referred to as the city block distance, the second gives a Euclidean distance.

The correlation between the discrimination test data and the distance measures was evaluated for the six groups of vowel-like stimuli. A linear regression was assumed, but due to the usual s-shape of the discrimination curve sample points with discrimination values below 5 % and above 95 % were excluded from the regression analysis.

Results

The results from the DL test confirmed as a whole the results obtained by Flanagan (1955) with a DL value of 3-5 % of the first and second formant frequency.

However, when looking at details of the discrimination curves for the six sets of vowel stimuli, some deviations appeared. Asymmetries of the curves reported by Flanagan, that is, cases in which the listeners were more sensitive to a shift of a formant in one direction than in the other, did not show up in the expected way in the present study. As one example, the stimulus set with the reference formant values 500, 2000, 2500, 3550 Hz, gave asymmetric discrimination data results in the opposite direction compared to the Flanagan study. Our listeners were more sensitive to an F2-lowering than to a raising.

The explanation for the asymmetry obtained in the Flanagan study seems intuitively correct. When two formants approach each other, an increase of the F2F3 complex would effect the listener's response towards being more sensitive to an F2 raising than to an F2 lowering. A closer look at the local intensity change when F2 is raised 25 and 50 Hz, however, will not show any drastic increase. Informal tests with native English and American listeners (though with a knowledge of Swedish) showed the same asymmetries as for our Swedish listeners, which seems to rule out an explanation in terms of different phonemic border within the stimulus set. Presently, we have no explanation for the different asymmetries.

The correlation between the perceptual data and the distance measures proved to be high (r=0.8-0.9) for both models. The intensity per critical band representation gave a somewhat but not signficiantly better fit than the loudness per critical band representation. A distance metric using the Euclidean distance did not differ appreciably from the city block distance.

A notable result was that also the discrimination test data for the stimulus sets with the marked asymmetry discussed above, correlated well with the distance measure, thus giving some support to the auditory models that were tested.

However, any distance measure based on models which try to capture the peripheral auditory processing will not be able to predict listeners' responses as soon as the decision involves a phonetically based judgement, or, in other words, a more centrally located processing.

In this study it was thus not possible to achieve a single physical distance value as a transformed version of the Difference Limen for formant frequencies, independent of formant number, vowel spectrum etc.

For a discussion of the relative merits of different auditory models, see Carlson & Granström (1979).

Footnotes

1 Part of this work has been reported elsewhere. For a more detailed presentation, see Nord & Sventelius (1979).

2 The use of the computer programs developed by Carlson & Granström for the auditory models, distance measure calculations, and the repression analysis is thankfully acknowledged.

References

Bladon R.A.W. and B.E.F. Lindblom, 1980, Modelling the judgement of vowel quality differences. Submitted for publication.

Carlson R. and B. Granström. 1979. Model predictions of vowel dissimilarity. STL-QPSR 3-4/1979, 84-104.

Flanagan J.L.. 1955. A difference limen for vowel formant frequency J.Acoust.Soc.Am., 613-617.

Lindblom, B.E.F.. 1978. Phonetic aspects of linguistic explanation. Studia Linguistica XXXII, I-II, 137-153.

Nord L. and E. Sventelius. 1979. Analysis and prediction of difference limen data for formant frequencies. STL-QPSR 3-4/1979; also in PERILUS, Report I: Experiments in Speech Perception (eds. B. Lindblom & J. Lubker), 25-37.

Plomp R.. 1979. Timbre as a multidimensional attribute of complex tones. In Frequency Analysis and Periodicity Detection in Hearing (eds. R. Plomp & G.F. Smoorenburg). Sijthoff, Lieden, 397-414.

Schroeder M.R., B.S. Atal and J.L. Hall. 1979. Objective measure of certain speech signal degradations based on masking. In Frontiers of Speech Communication Research (eds. B. Lindblom & S. Öhman). Academic Press, London, 217-229.

Zwicker E.. 1961. Subdivision of the audible frequency range into critical bands (Frequenzgruppen). J.Acoust.Soc. Am. 284.