Analysis and prediction of difference limen data for formant frequencies

Nord, L. and Sventelius, E.

journal: STL-QPSR
volume: 20
number: 3-4
year: 1979
pages: 060-072

http://www.speech.kth.se/qpsr
IV. SPEECH PERCEPTION

A. ANALYSIS AND PREDICTION OF DIFFERENCE LIMEN DATA FOR FORMANT FREQUENCIES*

L. Nord and Eva Sventelius**

Abstract

Difference limen data for formant frequencies were obtained using discrimination tests with synthetic, vowel-like stimuli, as a replication of a similar study by Flanagan (1955). A spectral distance measure proposed by Plomp (1970) was compared with the results from the perceptual tests. The correlation proved to be high (r=0.9).

Introduction

We were concerned in this experiment with the problem of relating perceptual (subjective) distances between stationary vowel-like sounds with a physical (objective) distance measure, taking into account known facts about the hearing mechanism. Such a distance measure can be regarded as a tool testing the perception model behind it, via its ability to predict data from perception tests.

Background

Fig. IV-A-1 presents a simplified model of the processing involved in a discrimination task. At a peripheral level an auditory analysis is made of the two incoming sound stimuli and auditorily based spectral images are formed. From these a distance measure is calculated and a discrimination decision is made. If the sounds occur close together temporally and with little or no quality variation, i.e. without causing phonemic shifts, the listener's stored references of phoneme inventories would not have to be resorted to as a basis for judgement.

A discrimination test, in contrast to an identification test, would thus depend to a lesser degree upon information from higher linguistic levels.

* to appear in Experiments in Speech Perception, PERILUS 1, Department of Linguistics, Stockholm University.
** Stockholm University, Department of Linguistics.
We used sound stimuli of the type used by Flanagan (1955) in his classical study. Specifically, Flanagan used synthetic vowel-like stimuli with systematic shifts of F₁ or F₂ at discrete intervals from the reference sounds. In an AB test listeners judged whether they could discriminate between reference sound and test sound. The difference limen, taken as the step in formant frequency for a discrimination of 50%, was found to lie in the range of 3 to 5% of the formant frequency.

**Distance measure**

The distance measure tested in the present experiment was formulated by Plomp (1970) in a study of the timbre of complex tones. Timbre was there defined as "that attribute of sensation in terms of which a listener can judge that two steady complex tones having the same loudness, pitch and duration, are dissimilar". Distance was defined as a difference between amplitude patterns of the frequency spectra of the sounds. The frequency resolution was taken to be 1/3 octave bands which corresponds well with the critical band theory of human hearing (Zwicker 1961).

Thus the spectral distance between two sounds sᵢ and sⱼ was defined by the equation:

\[
D_{ij} = \sqrt{\sum_{n=1}^{m} \left( L_{in} - L_{jn} \right)^p}
\]

(Eq. 1)

where

- \( D_{ij} \) = difference in frequency spectrum between the sounds sᵢ and sⱼ,
- \( L_{in} \) = SPL of sᵢ in band n,
- \( m \) = total number of frequency bands,
- \( p = 2 \), which gives a Euclidean distance, that is, the distance \( D \) can be seen as a distance between two points in an m-dimensional space, where the coordinates correspond to the intensity levels of the filter bands.

Plomp found a good correlation between distance measure based on frequency spectrum differences and dissimilarity indices for musical tones.

In Lindblom (1978) the measure was adopted and tested against several bodies of published data, such as a) dissimilarity estimates of
Fig. IV-A-1. Schematic model of discrimination processing.

Fig. IV-A-2. Hypothetical discrimination test result. The dependent variable is expressed as a formant frequency deviation (left) and as a spectral distance from the reference sound (right).
Swedish vowels (Hanson 1967), b) American English vowels in \(/\hbox{-}d/\) frames (Singh and Woods 1971), and c) vowel identification in noise (Nooteboom 1968 and Pickett 1957). Good correlations were also obtained with these data.

The purpose of testing the distance measure with the present particular set of stimuli is: the discrimination curves for the Flanagan experiment show some asymmetries, specifically, a formant shift in one direction from the value of the reference vowel gives an earlier, i.e. a more sensitive discrimination response compared to a change in the other direction.

Thus the question can be posed: will the distance measure predict these asymmetries and make the transformation from frequency (in Hertz) to distance (in dB) resulting in a one-to-one relation between distance and percentage dissimilarity as indicated in the hypothetical results shown in Fig. IV-A-2? Moreover, is the measure sensitive enough to predict the perceived discriminations despite the fixed 1/3 octave bands (approximating the critical band) having a rather poor frequency resolution?

Contrary to the vowel dissimilarity and identification tests mentioned above, this type of test will be of a more psychoacoustic nature, hopefully not relying on stored linguistic information.

In brief summary, our task was to a) make discrimination tests on vowel-like stimuli, replicating Flanagan's experiment of determining the difference limen for formant frequencies. b) calculate the spectral distances formulated by Plomp according to Eq. (1), c) correlate the perceptual and the physical measures and evaluate the degree of correlation and especially determine whether any asymmetries in the discrimination curves will be predicted. In addition, is it possible to express DL values in terms of one single distance instead of different percentages of frequency deviation?

**Experiment**

**Material**

For the difference limen test six sets of synthetic four formant vowels were generated with the series synthesizer OVE III. Formant
values were chosen in accordance with the Flanagan study. The effects of voice source, radiation and correction for higher poles resulted in a spectrum slope of -6 dB/octave for a formant configuration corresponding to a neutral vowel. In each set of synthetic stimuli $F_1$ or $F_2$ were shifted systematically in discrete steps on both sides of the stimulus defined as the reference sound. Fifteen stimuli were generated for each set with $F_1=\pi n \cdot 10$ Hz $(n=0,1,...,7)$ or $F_2=\pi n \cdot 25$ Hz $(n=0,1,...,7)$. Maximal shift of $F_1$ and $F_2$ was \pm 70 Hz and \pm 125 Hz, respectively. The six vowel sets covered most of the $F_1-F_2$ vowel plane in the following manner: starting with a neutral formant pattern of 500, 1500, 2500 and 3550 Hz, $F_1$ was given a reference value of either 300, 500 or 700 Hz and in the same way $F_2$ was given a value of 1000, 1500 or 2000 Hz. $F_0$ was held constant at 120 Hz and the duration of the sounds were 750 msec, with bandwidths of approximately 50, 90 and 120 Hz for the first, second and third formant. These bandwidths were smaller than the values given by Flanagan ($B_1 = 130$ Hz, $B_2 = 150$ Hz, and $B_3 = 185$ Hz). The amplitude level of the voice source feeding the formant circuits was held constant for all sounds resulting in a level of the stimuli that changed with the formant configuration. This was different from Flanagan's study where all the stimuli had an intensity of 70 dB.

Random lists were generated for the AB discrimination test. The reference sound in each set was paired with a sound having a shifted first or second formant. The duration between the sounds was 500 msec and the duration between stimulus pairs was 3 sec. Each list contained 14 AB pairs for analysis, 6 identical pairs (AA) for control with 2 AB pairs placed initially and not analyzed.

Fig. IV-A-3 shows the spectrum envelope for the six reference sounds. The triangles refer to the formant peaks of the stimuli used by Flanagan (Flanagan 1955, figs. 1 and 2). As can be seen there are certain differences between the two sets. Flanagan's stimuli were synthesized using a source with a flat spectrum and only four formants. Although the spectrum slopes of the Flanagan stimuli are steeper, resulting in lower relative levels of $F_2$, $F_3$ and $F_4$, the similarities ought to be sufficient to allow for a comparison of the discrimination test results.
Fig. 15. V and F - Specra of the synthetic reference stimuli used in the distort.
Calculation of distances

The synthetic stimuli were analyzed with a Brüel & Kjaer Band Pass Filter (type 1617). Levels from thirteen 1/3 octave filters were measured and spectral distances were derived (see Eq. 1). To obtain a good fit with critical bands at low frequencies the lowest three and the next lowest two 1/3 octave bands respectively were integrated to form the two lowest critical bands. Fig. IV-A-4 shows an example of the derivation of the distance measure.

Listening test

The stimuli lists were presented over headphones at a comfortable listening level to subjects whose task was to decide whether they heard pairs of sounds that were the same or different.

Three test sessions with different groups of subjects were made. Two of the groups could be considered as unfamiliar with this type of testing situation. The number of stimuli sets varied between groups as did the number of repetitions for the group. A mean value of 27 listeners gave 39 responses per stimulus. Corresponding figures for the Flanagan test were 4 subjects and 20 responses. The listeners reported that the task was not difficult and an analysis of the responses shows consistent results with no opposite trends.

Results

Fig. IV-A-5 shows the result of the discrimination test for small shifts of formant 1 or formant 2. The lines of best fit were drawn by hand. The test results from the Flanagan study are also shown to provide a comparison. The correspondence between those two sets of data is good. Taking the 50% discrimination value as a DL value for the different test sounds, a mean value of 3.3% of the formant frequency is obtained, which can be compared with the corresponding value of 3.9% for the Flanagan study. The spread of DL values for the six stimuli sets is similar for the two studies.

An interesting observation in the present data is that the asymmetries obtained for the stimuli shown in Figs. IV-A-5c and 5f do not appear in the Flanagan study where asymmetry is reported for Fig. IV-A-5d and an asymmetry in the opposite direction for stimulus IV-A-5f.
Fig. TV-A-4. An example of 1/3 octave band pass filter.
Fig. IV-A-5a-f. Results of the discrimination tests with varying $F_1$ (left column) and varying $F_2$ (right column). The dotted lines indicate results from Flanagan (1959).
We will now examine how well the spectral distance measure correlates with the perceptual data and return to the issue of asymmetry later. Fig. IV-A-6 shows the discrimination results as a function of spectral distance. The six curves are placed in the same way as in Fig. IV-A-5 with the left column containing stimuli where \( F_1 \) is shifted and the right column stimuli where \( F_2 \) is shifted. Curves have been fitted by hand. As in earlier figures the symbol (o) corresponds to data for a stimulus in which a formant has been lowered and the symbol (x) corresponds to a stimulus for which a formant value has been raised relative to the reference vowel.

On the whole, the correlation seems to be satisfactory between the perceptual and the physical measure. (The calculation of the correlation coefficients and the DL values do not include response values over 95%.)

One example of a less satisfactory outcome is the curve in Fig. IV-A-6c where (o) and (x) describe two different relations for the variables. For a given difference value, for example 60%, there exist two distances: 3 and 5 dB.

One example of a good correlation can be seen in curve IV-A-6f. Despite the strong asymmetry in IV-A-5f, with different positive and negative DL values, the transformation from frequency deviation to distance has resulted in a single-valued relation.

A 'true' distance measure would also be of use when comparing different types of spectra. A possible outcome would be one single distance for the DL values of the six sets of sounds. However, this did not happen and a fairly large spread of distance DL values can be seen in Fig. IV-A-6.

**Summary**

Discrimination tests for the determination of Difference Limen for formant frequencies gave a value of 3-4%, which is in close agreement with an earlier study by Flanagan (1955). A good correlation was found between perceptual discrimination values and a spectral distance measure suggested by Plomp (1970). The correlation was good also for perceptual results showing a strong asymmetry. The spectral distance measure is not powerful enough to transform the DL values for all the test sounds tried, into one single value.
Fig. IV-A-6a-f. Discrimination tests results as a function of spectral distance according to Plomp (see text). Left column: $F_1$ is shifted, right column: $F_2$ is shifted.

- (o): a formant is lowered relative to the reference sound.
- (x): a formant is raised relative to the reference sound.
Specifically the discrimination test for a stimulus with $F_2$ and $F_3$ lying close together resulted in an opposite asymmetry as compared to the results obtained by Flanagan (see below).

Discussion

What could be the explanation for the opposite asymmetries for stimulus specified in Fig. IV-A-3f with the formant frequencies 500, 2000, 2500 and 3550 Hz? Within the scope of this study we cannot give a precise explanation but a few factors that could be of importance will be mentioned. Recapitulating, in Flanagan's experiment listeners were more sensitive to an $F_2$ rise, while our listeners were more sensitive to an $F_2$ lowering. The explanation that Flanagan gives for the asymmetry seems intuitively correct: when two formants close to one another become even closer, the gradual build-up of that formant complex will increase rapidly, thus making subjects more sensitive to an $F_2$ raising than to a lowering. A closer look at the spectral change when $F_2$ is shifted only by one or two steps ($\pm 25, \pm 50$ Hz) does not show a drastic change in level of the $F_2-F_3$ complex, as the distance between these formants still is rather large. It seems therefore as if the marked asymmetry cannot be explained by the above rule-of-thumb describing formant intensity relations. One could perhaps argue that this increase in intensity build-up will not appear unless the proper auditory transform (analysis) is made. The 1/3 octave band filter analysis and derivation of spectral distance measure attempted in this study, however, will give a larger change in spectral distance when $F_2$ is lowered, thus giving a good correlation between our asymmetric discrimination curve and the distance measure (see Figs. IV-A-5f and 6f).

One factor not mentioned above is the influence of phoneme boundaries in the perceptual test. In Fig. IV-A-1 the box marked STORED REFERENCES is only indicated. Informal tests with Swedish listeners show that some place a phoneme boundary close to the reference vowel, which might explain that asymmetry. Further tests with American and Swedish listeners should elucidate this question.
It is still an open question as to what grounds a listener uses to make his judgement in this type of test. What is the relative importance of timbre compared with loudness of the stimuli? As our stimuli changed in intensity we have a larger loudness range within and between stimuli sets compared to Flanagan's stimuli which has a constant intensity.

References