ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

DIFFERENCE LIMEN FOR FORMANT FREQUENCY DISCRIMINATION AT HIGH FUNDAMENTAL FREQUENCIES

Fredrik Karlsson and Anders Eriksson
Department of Philosophy and Linguistics, Umeå University
S-901 87 Umeå, Sweden
fredkarl@ling.umu.se anderse@ling.umu.se
http://www.ling.umu.se

10.21437/Eurospeech.1999-383

ABSTRACT

In previous studies of formant frequency discrimination, variation in the stimuli has mainly concerned the formant frequencies while other factors which may affect formant frequency discrimination have largely been ignored. In most studies, fundamental frequencies typical of adult male speakers have been used. In the study presented here, formant frequency discrimination as a function of fundamental frequency level was studied. Contrary to expectations, fundamental frequency level did not seem to affect formant frequency discrimination. This was true even when the fundamental frequency was higher than the tested formant frequency. The implication of these findings is that formant frequency discrimination must, at least partly, rely on some other mechanism than reconstruction of the vocal tract transfer function. It is suggested that listeners use differences in the amplitude of partials to discriminate between complex vowel like sounds. Asymmetries in the distribution of these amplitudes may also explain the corresponding asymmetries found in perception.

Keywords: perception, formant, DL, F0

1. INTRODUCTION

1.1 Difference Limen for formant frequency discrimination

Formant frequency discrimination has been studied in a series of studies since the mid-fifties (e.g. [1–8]). In an early study by Flanagan [1], the Difference Limen (DL) for formant frequency discrimination, expressed as a Weber fraction, was found to range between 0.03 and 0.05 (for 50% correct discrimination). Later studies (e.g. [5,6,7]) have reported similar values. Reported values for 50% correct discrimination of deviations in F_1 range between 0.03 and 0.14 in the above cited studies. Weber fractions for F_1 and F_2 do not differ in any systematic way in these studies. The results reported in [2,3] are somewhat lower, but were obtained with highly trained subjects and are thus not comparable in that respect.

1.2 Asymmetry

Discrimination has been observed to be asymmetric with respect to the reference frequency in some studies. In Flanagan's [1] study the effect was strongest in the cases where the frequency of the test formant was close to a neighbouring formant. In those cases discrimination was better on the side facing the adjacent formant. The explanation Flanagan offered was that when formants are close, small deviations in frequency in one of them may cause a great change in amplitude in the adjacent formant and in particular a greater difference in amplitude relations between the two. Flanagan's results were partially corroborated in the study by Nord and Sventelius [7], but the asymmetry was generally smaller and in one case in the opposite direction. In the context of the results presented here the most relevant comparison is the asymmetry found for F_0=120 Hz and F_1=300 Hz. Here Flanagan found a DL of 0.057 for frequencies below the reference and 0.040 for frequencies above the reference. Corresponding values reported in [6] were 0.057 and 0.037 respectively.

1.3 Discrimination as a function of F_0

The variation in the stimuli used in previous studies has almost exclusively concerned variation in formant frequencies. Other factors which may affect formant frequency discrimination, for example fundamental frequency, has largely been ignored. In most studies, fundamental frequencies typical of adult male speakers (100–120 Hz) have been used. In two studies [2,3] the effect of variation in fundamental frequency was studied, but only over a very limited range – 101 Hz and 126 Hz for stimuli simulating a male speaker in [3] and 160 Hz, 200 Hz and 240 Hz for the stimuli simulating a female speaker in [2].

In the study presented here, formant frequency discrimination as a function of fundamental frequency level was studied for a wide range of fundamental frequencies ranging from 120 Hz to 370 Hz for synthesized vowels with formant frequencies typical for an adult male speaker. For the stimuli with the highest fundamental frequencies used, the format frequency for the tested formant was actually lower than the fundamental.

Based on the results obtained in [3] one would expect discrimination to deteriorate as the fundamental frequency is raised. Within the limited range used in [3] (126 Hz to 101 Hz), DL was approximately 20% lower for the stimuli with the lower F_0. For the stimuli with formant frequencies comparable to the ones used in the present study there was, however, no definite trend in the DLs.

In [2] fundamental frequency was varied in three steps (160, 200 and 240 Hz respectively) for a simulated female voice. DL varied quite considerably but it is difficult to see any distinct pattern with respect to F_0 variation. The results are not unambiguous, but where there is a trend, it indicates that a lower fundamental frequency yields a lower DL.

Further reasons to expect discrimination to deteriorate for at least the highest fundamental frequencies used here (320 Hz and 370 Hz) is the fact that there is little or no energy in the region of the mid frequency of the reference formant ± one bandwidth. There is thus not enough information available to reconstruct the vocal tract resonances with any reasonable degree of accuracy.

2. METHOD

2.1 Stimuli

Stimuli consisted of synthesized four-formant vowels which were produced using a Klatt-type formant synthesizer. Mid-formant frequencies and F_0 values are shown in Table 1. Only the first formant was varied in the experiment, F_2–F_4 remained unchanged. Formant frequencies were constant throughout the stimuli. Formant bandwidths were 60 Hz for F_1, 90 Hz for F_2, 150 Hz for F_3 and 200 Hz for F_4. Six series of stimuli were produced with fundamental frequencies ranging from 120 Hz to 370 Hz. In each series F_1 was varied in 10 steps of 8 Hz on each side of the reference frequency, producing 21 test stimuli per series. Stimulus duration was 230 ms for all stimuli.

Table 1. Fundamental frequencies and formant frequencies for the test stimuli. n = 0, 1, ..., 10.

F_0	F_1	F_2	F_3	F_4
120	300±8∗n (220-380)	1500	2500	3550
170	300±8∗n (220-380)	1500	2500	3550
220	300±8∗n (220-380)	1500	2500	3550
270	300±8∗n (220-380)	1500	2500	3550
320	300±8∗n (220-380)	1500	2500	3550
370	300±8∗n (220-380)	1500	2500	3550

2.2 Subjects

16 subjects (6 male and 10 female), aged 19–46, participated in the experiment. One subject lacked formal linguistic education, 12 subjects had some phonetic knowledge and 3 subjects had studied phonetics more than 1 year.

2.3 Procedure

The study consisted of six tests, one for each fundamental frequency level. A test involved pair-wise comparisons between a reference and a test stimulus with an inter stimulus interval of 400 ms. The reference was always presented first. To avoid presentation order effects, five versions of each test were constructed which differed only with respect to the presentation order of the stimulus pairs. Subjects were told that the differences between stimuli could be quite small and they were encouraged to listen very carefully. Stimulus presentation and data gathering was handled entirely by Java™ Applets™ on the grounds of the portability to different computer environments. Information concerning the subjects age, gender, phonetic background and hearing status were collected with the help of an anonymous questionnaire before the testing started. After the test was completed, the subjects were asked to comment on the difficulty of the test and report if there were any external factors that may have influenced their ability to perform the test. Two subjects reported that they did not have sufficient computer knowledge to feel comfortable in the testing environment and with the testing procedure.

3. RESULTS

All data analysis was made using the SPSS™ statistical analysis package. Cumulative frequencies for ‘correct response’ were computed. These values were then converted into z-scores. Using linear regression analysis, the 50% point was computed. The 50% limit was chosen to facilitate comparison with previous studies were the same criterion had been used. This type of analysis was performed for all values pooled (referred to as mean values in the text), all values related to stimuli whose formant frequencies were lower than the reference (ΔF_1 below the reference) and higher than the reference (ΔF_1 above the reference).

3.1 Difference Limen

DLs based on pooled responses expressed as the Weber fraction at 50% discrimination varied little between the six series used in this study. Values ranged between 0.08 and 0.10 which falls in the middle of the range of values reported in the above cited studies. From this point of view earlier results thus receive support.

3.2 Asymmetry

The asymmetries found in some of the studies reported above were also found in our results. For F_0=120 Hz and F_1=300 Hz the results are directly comparable to the results in [1,7] although our DLs were somewhat higher. DL for stimuli with formant frequencies below the reference was found to be 0.11 and whereas stimuli with formant frequencies above the reference were discriminated slightly better, DL=0.07.

3.3 Discrimination as a function of F_0

As mentioned in 3.1, mean DL disregarding asymmetry, varied little in the six test series, with fundamental frequency values ranging between 120 Hz and 370 Hz. These results are summarized in Table 2, columns 6 and 7.

Table 2. DLs for the 6 test series. DL (**ΔF_1**) is reported in Hz.

	ΔF_1 below reference		ΔF_1 above reference		Mean DL	
F_0	**ΔF_1**	**ΔF_1/F_2**	**ΔF_1**	**ΔF_1/F_2**	**ΔF_1**	**ΔF_1/F_2**
120	33	0.108	21	0.070	29	0.097
170	**20**	**0.065**	**33**	**0.110**	25	0.083
220	**17**	**0.056**	**34**	**0.113**	25	0.082
270	51	0.171	14	0.046	27	0.089
320	**18**	**0.059**	**34**	**0.112**	26	0.088
370	35	0.115	21	0.070	28	0.093

The above mentioned asymmetry is also clearly present in the six series and the magnitude of the asymmetry is on the same order as that reported in other studies (Table 2, columns 2–5). What may also be seen, however, is that the direction of the asymmetry varies as a function of fundamental frequency. This is best seen in Figure 1, where absolute values for DL as well as DLs for stimuli with formant frequencies below and above the reference are plotted.

As may be seen in Figure 1, the deviations seem to be very systematic. With the exception of DLs for 270 Hz fundamental frequency, DL-values group into two distinct groups which form almost perfect mirror images. The test series for which asymmetry is in the opposite direction compared to previously reported asymmetries are marked in bold face in Table 2.

4. DISCUSSION

The mean DLs found in this study agree well with results from earlier studies. In this respect our results only lend further support to the results found in previous studies. The finding that formant frequency discrimination does not seem to depend on the fundamental frequency level, at least not within the

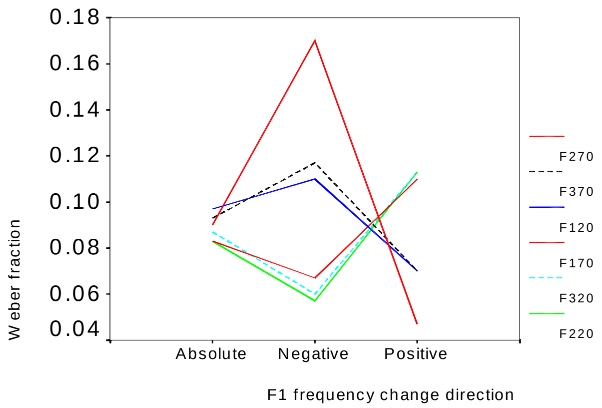

Figure 1. Weber fractions for the six test series plotted for test formant frequencies below the reference (middle), above the reference (right) and as a mean value for all stimuli (left).

range used here is, however, rather unexpected. Based on the results obtained in [2,3] and on considerations of the absence of energy content in the vicinity of the reference formant frequency for the highest fundamental frequencies used, we expected discrimination to deteriorate severely at least for the highest fundamental frequencies. As has been demonstrated above this prediction was not borne out. On the contrary, DLs were practically constant over the whole range of fundamental frequencies.

For fundamental frequencies in normal male speech, it is possible to reconstruct the vocal tract transfer function with at least reasonable accuracy. Identifying formants (i.e. underlying resonance peaks) should therefore not be an impossible task in perception. But for higher fundamental frequencies the underlying spectral envelope is severely undersampled if one wants to reconstruct the transfer function. It seems reasonable to suggest then, that discrimination of differences in vowel quality, and more generally timbre of complex sounds, must rely on some other mechanism than the reconstruction of the underlying transfer function.

If one looks at the results in the experiment presented here, there are then two main questions that beg an answer: What information in the stimuli do subjects use to discriminate between them if reconstruction of the transfer function is not possible? What could be the explanation for the observed asymmetries?

In [7] it was suggested that a spectral distance measure could be used to model perceptual distance of the stimulus sounds. The authors computed the area under the spectrum envelope for each stimulus sound and correlated the differences between the respective areas with auditory distance found in their perception tests. They found a reasonable degree of correlation between the two measures. We would like to suggest a slight variation of this theme. If one considers differences in the amplitudes of individual partials instead of the area under some, not reconstructible, spectral envelope, an interesting picture emerges. This may be clearly seen in Figure 2 were we have plotted the amplitude of individual partials for the reference stimulus and two stimulus sounds on each side of the reference. The lines connecting the points representing the amplitude values are only there to make it easier to separate the individual sounds, they have no meaning in themselves.

It may be seen that the distribution of partial amplitudes differ quite considerably between the stimuli. The precise perceptual consequences of these differences remain to be studied in future perception experiments but it seems reasonable to suggest that differences on this order should be clearly audible and thus may explain why our subjects performed so well in the perception tests.

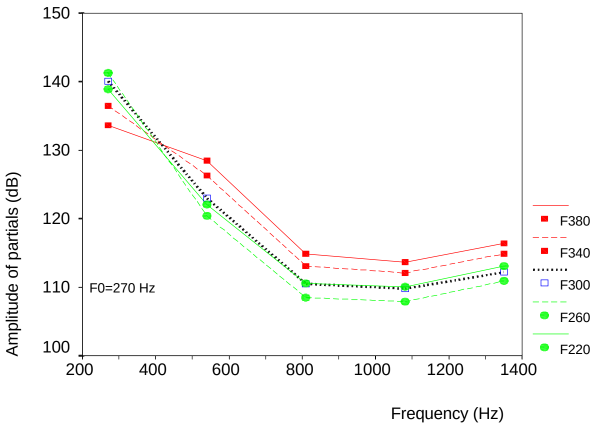

Figure 2. The amplitudes of individual partials for the reference stimulus and two stimulus sounds on each side of the reference. The F-values in the legend represent the centre frequencies of the first formants.

The second question concerning asymmetry may also be addressed with reference to these results. For the particular series of stimuli from which the data in Figure 2 are drawn one may see that deviations are greater for the stimuli representing first formant frequencies above the reference than for those below. Based on this observation one would like to suggest that if the change in amplitude of the partials as a function of a change in formant frequency is greater on one side of the reference compared to the other, discrimination against the reference should be easier for stimuli on that side. This prediction is borne out for all six series of test stimuli used in this experiment.

5. CONCLUSIONS

We have demonstrated that subjects are able to discriminate between vowel like sounds over a wide range of fundamental frequencies, including stimuli where the fundamental is above the tested formant. In addition we did not find Difference Limen to be affected to any appreciable degree as a function of fundamental frequency level. We suggest that discrimination is based on the perception of differences in distribution amplitudes of the partials of the perceived sounds. Preliminary investigations reported above indicate that the amplitude differences should be great enough to be clearly audible. Differences in the amplitude of partials for stimulus sounds with formant frequencies above or below the reference frequency could be used to account for the observed asymmetry in discrimination between test stimuli with centre frequencies above or below the reference. The precise consequences in perception of variations in the amplitude of partials, however, remain to be explored and described in future studies.

6. ACKNOWLEDGEMENT

Ola Andersson, research engineer at the Department of Philosophy and Linguistics at Umeå University, made all the programming necessary to produce and present the stimuli and collect the data.

7. REFERENCES

[1] Flanagan, J. (1955). A difference limen for vowel formant frequency. *J. Acoust. Soc. Am.*, **27**, 613–617.

[2] Kewley-Port, D. and C. S. Watson. (1994). Formant-frequency discrimination for isolated English vowels. *J. Acoust. Soc. Am.*, **95**, 485–496.

[3] Kewley-Port, D., X. Li, Y. Zheng and A. Neel. (1996). Fundamental frequency effects on thresholds for vowel formant discrimination. *J. Acoust. Soc. Am.*, **100**, 2462–2470.

[4] Kewley-Port, D. and Y. Zheng. (1998). Auditory models of formant frequency discrimination for isolated vowels. *J. Acoust. Soc. Am.*, **103**, 1654–1666.

[5] Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state consonant-bound vowels. *J. Acoust. Soc. Am.*, **63**, 572–580.

[6] Nakagawa, T., S. Saito and T. Yoshino. (1977). Tonal difference limen for second formant frequencies of synthesized Japanese vowels. *Annual Bulletin RILP*, **16**, 81–88.

[7] Nord, L. and E. Sventelius. (1979). Analysis and prediction of difference limen data for formant frequencies, *PERILUS*, (Vol. 1) 25–37. Stockholm: University of Stockholm, Department of Linguistics.

[8] Sommers, M. S. and D. Kewley-Port. (1996). Modelling formant frequency discrimination of female vowels. *J. Acoust. Soc. Am.*, **99**, 3770–3781.