DISCRIMINATION IN DENSELY AND SPARSELY SPECTRALLED HARMONIC COMPLEXES

Astrid van Wieringen and Louis C.W. Pols

ABSTRACT

The ability to detect small changes in relatively high-pitched harmonic complexes is tested in three discrimination experiments.
In the first two tests discriminability in fundamental frequency and formant frequency was determined at four different fundamental frequencies, i.e. 120 Hz, 300 Hz, 400 Hz, and 500 Hz. In both tests the smallest detectable changes lie on average between 2.5% and 4.5% of the frequency.
The results indicate that discrimination depends strongly on the relation between the formant and the harmonics of the fundamental frequency (F0). In a third experiment it is confirmed that at a relatively high F0 one harmonic dominates perception if it coincides with a formant, whereas at a relatively low F0 more harmonics -possibly weighed- are taken into account. Discrimination is far more difficult in a sparsely spectralled harmonic complex, however, when the formant is located between two equally strong harmonics.

1. INTRODUCTION

The study of high-pitched voices is relevant to a number of fields within speech science. For one, it would help us in understanding the development of speech productions of infants better, while, for another such knowledge would be of benefit to the field of speech perception in general, as most experiments in this field are performed with an average male fundamental frequency. Furthermore, it is not without reason that techniques and knowledge involved in for instance speech synthesis focus on the male voice: the fairly low fundamental frequency is favoured most since it responds best to all the analysing techniques to date.

One drawback on the high fundamental frequency is the resolution of the formants, which becomes poorer as the interval between the harmonics becomes larger, this makes it extremely difficult to estimate the formant frequencies from the imaginary spectral envelope. Compare the FFT spectra of the two synthetic stimuli in Figure 1 on the next page. Both harmonic complexes have a formant frequency at 1250 Hz, but the fundamental frequency of the first is 120 Hz, while that of the second is 500 Hz. Clearly, the chance of estimating the formant frequency erroneously is greater when the glottal source lacks acoustic energy around the formant frequency itself.
Also, the shorter the pitch period the more difficult it becomes to estimate the fundamental frequency. From a perceptual point of view, however, we, as human beings, can cope just as well with high-pitched voices as with low-pitched ones. Understanding infant speech implies that we can detect quality differences at a relatively high fundamental frequency. To achieve this, different perceptual strategies may be used, according to the available spectral information. For instance, if the harmonics are widely spaced one harmonic may cue perception, whereas if they are not, the relative weight of more harmonics could be taken into account. The purpose of the study was to investigate how sensitive listeners are to changes in the fundamental and the formant

frequency and to what extent they are either aided or hampered by the scarce information in the relatively high-pitched harmonic complexes. To this end we have posed three hypotheses on the discriminability of relatively high-pitched harmonic complexes.

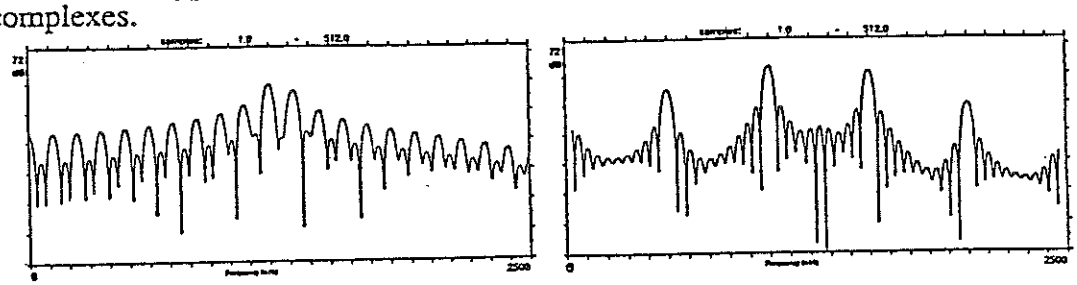

Figure 1. Spectra of two harmonic complexes differing in fundamental frequency

2. HYPOTHESES

1) Discriminability in fundamental frequency between two relatively low-pitched harmonic complexes is equally difficult as between two relatively high-pitched ones.

2) Discriminability in formant frequency between two relatively low-pitched harmonic complexes is equally difficult as between two relatively high-pitched ones.

3) Discrimination is dependent on the available spectral information. This means that the position of the formant frequency relative to that of the harmonics influences perception. This need not hold implicitly, however, that it is easier to discriminate between the lower-pitched harmonic complexes, where the formant frequency is well defined. If the formant frequency coincides with a harmonic of a relatively high-pitched harmonic complex a difference may be heard quickly as well, due to the prominence of the -widely spaced- harmonics. Should the formant be positioned between two harmonics or in any case too far away to enhance either of them, discriminability could be more difficult.
Before relating on our discrimination experiments we will discuss in short some literature concerning pitch and formant perception.

3. SHORT SURVEY ON LITERATURE

3.1 Literature on fundamental frequency discriminability

Just noticeable differences in fundamental frequency of **complex** stimuli were determined by Flanagan and Saslow (1958). The difference limens for pitch were obtained -after extensive training- on several combinations of four different vowels at three different sound pressure levels and two different fundamental frequencies. From 384 judgments (32 on every frequency changes * (AX&XA) * 3 SPL * 2 F0) the difference limens were found to be about 0.3% to 0.5% of the fundamental frequencies (80 Hz and 120 Hz). In comparison with **pure-tone** difference limens these are relatively small (Table 1). This could mean that the presence of a harmonic structure provides additional information to the auditory system. Moore (1984) investigated the discriminability of individual harmonics within a fixed complex. The difference limens of the first three harmonics -between 200 Hz and 600 Hz- were higher (approx. 0.25%) than those of the 9th and 11th harmonic (2.0-5.0%). However, difference limens for the whole complex exceed those of the single harmonics as they are 0.13-0.2%. This means that information is combined across harmonics and that the low frequency harmonics are important in a complex. The absence of the fundamental frequency component seems to improve discrimination as well. However, we should be careful in comparing values as they are dependent on the experimental designs. In

Table 1 the method, duration and level of a number of studies are listed with their frequency range to illustrate this.

Difference limens have not only been obtained for static, yet also for dynamic stimuli. In order to determine difference limens more characteristic of speech, Klatt (1973) measured the discrimination of changes in the slope of linear fundamental frequency contours of otherwise stationary 5-formant vowel /e/-like stimuli of which the average fundamental frequency remained constant. On the one hand he measured three parameters, i.e. linear ramp and rate-of-change of F0 as well as high pass filtering. On the other hand he included formant transitions appropriate for the syllable /ya/. Apparently these formant transitions did not interfere much in the detection process as they degrade the difference limens only from 0.3 to 0.5 Hz for steady F0 and from 2 to 2.5 Hz for ramp F0.

Table 1. F0 difference limens in Hertz for pure and complex tones.

Author	Type	Method	Length	Level	Difference limens (Hz)
Harris (1952)	pure	AX	1400 ms	30 dB	0.78 (250 Hz); 2.1 (2000 Hz)
Nordmark (1968)	pure	adjustment	800 ms	45 dB	0.34 (250 Hz); 1.5 (2000 Hz)
Moore (1973)	pure	2IFC	200 ms	60 dB	0.88 (250 Hz); 5.3 (2000 Hz)
Wier (1977)	pure	2IFC	500 ms	40 dB	1.9 (1000 Hz); 3.2 Hz (2000 Hz)
Flanagan (1958)	complex	AX and XA	500 ms		0.28-0.48 Hz
Klatt (1973)	complex	ABX	250 ms		0.3-0.5 Hz for steady tones
		AX and XA	250 ms		2.0-2.5 Hz for linear ramps

3.2 Literature on formant frequency discriminability

Flanagan (1955) obtained difference limens for formant frequency by changing either the first or the second formant of four-formant synthetic vowels in discrete steps. The frequency changes were determined for three different standard vowels at F0=120 Hz. Although there is some dependency on the proximity of the other formants, the difference limen is of the order of 3-5% of the formant frequency for both the first and second formants. The asymmetry observed for some vowels can be partly explained by the fact that the amplitude is not independent of the formant frequency, which leads to a comparatively large increase in the amplitude of the first and/or second formant, when these are close together.

3.3 Literature on quality differences

Vowel quality depends upon the presence of energy at certain frequencies, energy which is determined by the harmonics of the fundamental frequency and the resonance frequencies of the vocal tract. The influence of the fundamental frequency on the perception of vowel quality was analysed among others by Slawson (1968). He varied the fundamental frequency together with the formant frequencies and found that when the fundamental frequency of the second sound of each pair was raised an octave above the fundamental frequency of the first sound, identification was almost the same when the lower two formants of the sounds with the higher fundamental frequency were multiplied by a factor about of 1.1. Large differences in vowel quality occurred when the two lower formants of the second sounds were shifted only slightly in frequency. The impact on quality differences also depended on the type of vowel, the compact

vowels suffering more from the shifting of the spectrum envelope than the non-compact ones. Apparently, the auditory system uses different criteria for the various vowels. For the front vowels it tends to weigh the position of the second and third formant, while for back vowels a single resonance is conveyed.
Yet, viewing the abovementioned with respect to relatively high-pitched harmonic complexes it is not clear how the perceptual mechanism operates. Possibly one -very prominent- harmonic is extracted as a formant. As such is not inferred from identification tests we will test the ability to detect small quality differences in a discrimination test.
Ryalls and Lieberman (1982) performed a forced-choice identification test with vowels having a fundamental frequency of 100 Hz, 135 Hz, and 250 Hz. They found that the identification of a vowel is almost totally determined by the formant frequency values, changes in fundamental frequency being of minor effect. They did find that identification is aided by a lower fundamental frequency (133 (male) vs. 264 (female) errors). In other words, the mechanism which extracts formant frequencies is aided by a densely spectralled harmonic complex as less spectral information causes higher error rates.
Although we do not seem to have any trouble in identifying infant speech it is not unthinkable that the perceptual mechanism operates differently for densely or sparsely spectralled harmonic complexes. In a discrimination task less spectral information may not necessarily cause a higher error rate, though. It is possible that individual harmonics are weighed or selected according to the (prominence of the) available frequency components.

4. EXPERIMENTS 1 AND 2

4.1 General description

The purpose of the tests on the just noticeable differences in fundamental frequency (experiment 1) and formant frequency (experiment 2) was to learn to what extent the difference limens vary within a relatively large fundamental frequency range. In both experiments we made use of stationary synthetic stimuli. These were generated by a digital formant synthesiser, designed to imitate different parameters of a synthetic voiced speech-like sound on the microVAX II (Weenink, 1987). Relatively high-pitched sounds such as ours require the highest possible sample frequency to accurately manipulate the signals in steps of 1 Hz. Therefore, all signals were sampled internally at 1.2 megaHerz, after which they were downsampled to 20 kHz.
The other parameters of these stationary speech signals were as follows: a pulse was used as glottal function, the quality Q (the ratio of formant frequency to -3dB bandwidth) of the formant resonances was 10 and all signals were 200 ms in duration.
In order to avoid clicks a HAMMING window function was swept smoothly over the first and last 5 ms of the signals.

4.2 Experiment 1: on fundamental frequency discriminability

In the test on the just noticeable difference in fundamental frequency one two-formant /a/-like harmonic complex was offered at 4 different fundamental frequencies, i.e. 120 Hz, 300 Hz, 400 Hz and 500 Hz. The harmonic complexes ranged in discrete steps of 1 Hertz from -7.5% to +7.5% frequency deviation.
The two formant frequencies were 1200 Hz (F1) and 2000 Hz (F2). These values may seem rather high, especially for the 120 Hz condition. We know that the formant frequencies of the average female voice is approximately 10% higher than that of the average male voice (Weenink, 1985). Those of children or infants will be even higher. From a pilot test the abovementioned formant frequency pair seemed to be the most representative for the vowel phoneme /a/ at a high fundamental frequency of 450 Hz.

After randomizing and balancing them appropriately for an AX paradigm (AX and XA, X=ref.) the 38, 94, 122, and 154 pairs of stimuli for 120 Hz, 300 Hz, 400 Hz, and 500 Hz, respectively, were offered to twelve naive subjects. The four different subsets -kept separately per condition- were offered in different sequences for every two listeners (Latin square method). By using the Latin square method we were reasonably certain that the presented sequence was not of influence on the responses. The 120 Hz subset was offered before and after the sequence to examine any possible effect of training. Listeners were requested to score *same* or *different* for each pair presented on their scoring sheets.

4.2.1 Results

The discrimination curves of the pooled positive and negative response scores are plotted in Figure 2. Each frequency deviation point in the discrimination curve represents the mean of 48 judgments. As is seen in the plots all discrimination curves were also fitted by a 3rd-order polynomial function. This was done with regard to the variability in responses and it seems as though the function is a good approximation. From these polynomial functions we derived the average discriminable changes at the frequencies corresponding to the 50%, 75%, and 90% difference responses. The approximates are given below in Table 2 in Hertz as well as in percent of the fundamental frequency.

Table 2. Just discriminable differences in fundamental frequency in Hertz and percent. The difference limens are derived from the polynomial curve fit at the 50%, 75%, and 90% different responses.
The table corresponds to Figure 2.

	Number of trials	50% (Hz)	(%)	75% (Hz)	(%)	90% (Hz)	(%)
F0							
120 Hz(1)	38	2.0	1.6	4.5	3.8	7.0	5.8
120 Hz(2)	38	2.0	1.6	4.0	3.3	6.0	5.0
300 Hz	94	4.0	1.3	6.5	2.2	10.0	3.3
400 Hz	122	6.0	1.5	10.0	2.5	14.0	3.5
500 Hz	154	7.0	1.4	11.0	2.2	15.0	3.0
X			1.5		2.8		4.1

Since the difference limen is generally defined as the frequency which is discriminated correctly 75% of the time we can see in the data of Table 2 that the difference limen is on average 2.8% of the fundamental frequency. Due to a number of reasons the difference limens are on average much larger than those obtained by Flanagan or Klatt (0.3%-0.5%).
First, our values are not true difference limens in the sense that they indicate the maximum perceptual sensitivity of the ear. Our subjects were naive, untrained and received every stimulus only twice. Second, the relatively high difference limens for the 120 Hz conditions may be the result of the relatively high F1 and F2 (for this rather low F0 subset) and/or of accustoming to the test (120 Hz (1) was always offered at the beginning of the test).

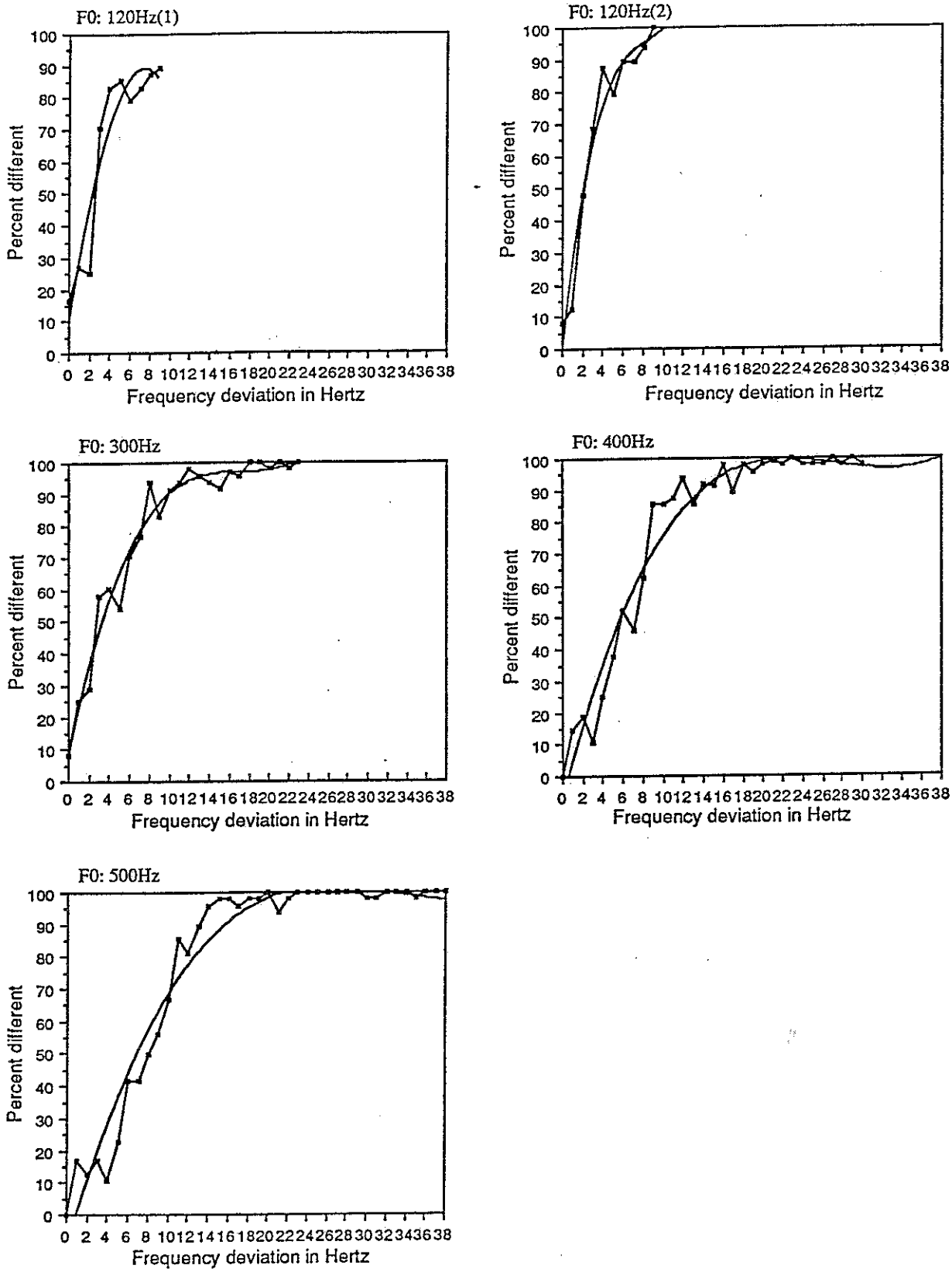

Figure 2. Sigmoids and 3rd-order polynomials of percent different responses as a function of fundamental frequency deviation for positive and negative responses pooled.

From our results we may conclude that equal differences are perceived - percentagewise- regardless of the fundamental frequency. This holds that it is as easy to determine small change in men's voices as it is in women's or children's ones.
The relatively small difference limens in the 400 Hz and 500 Hz conditions are possibly due to 1) the formant frequencies often being multiples of the fundamental (a favourable condition) or 2) the mere appropriateness of the two formant frequencies for /a/ at that fundamental frequency. Recall that this formant frequency pair was judged from a hundred other ones as being the most representative of the vowel /a/ at F0= 450 Hz. At 300 Hz another formant pair would possibly have been chosen.
The latter could hold that the relatively high values for the 120 Hz series are the result of the rather inappropriate relationship between the low fundamental frequency and the two formant frequencies. Instead of avoiding extra influences by using the same formant frequency pair for all conditions, the perceived timbre of this series may have caused the task to have been harder than that of the higher-pitched ones. For the lower series the formant frequency pair is more like F2/F3 for /a/ than F1/F2. Ryalls and Lieberman (1982) mention problems in this respect in an identification task. They found the error rate to be lower when their signals contained *appropriate formant frequencies* than when they contained a male fundamental frequency with female formant frequencies (133 errors for a low male F0 vs. 406 errors for a high male F0). Needless to say, only by rerunning the test a number of times will we gain better comprehension of these factors. We will proceed now with discussing the second discrimination experiment, which is meant to determine whether formant frequency discriminability is dependent on the fundamental frequency.

4.3 Experiment 2: on formant frequency discriminability

The formant frequency condition consisted of more subsets. There were three different one-formant harmonic complexes, i.e. 1000 Hz, 1250 Hz, and 1500 Hz at four different fundamental frequencies, i.e. 120 Hz, 300 Hz, 400 Hz, and 500 Hz. These formant frequency values were varied in discrete steps of 5 Hertz from -5% to +5% frequency deviation. The twelve subsets -together consisting of 704 pairs of stimuli- were randomized and balanced according to an AX paradigm (AX and XA, X=ref.) and all offered to twelve naive listeners.

4.3.1 Results

Figure 3 illustrates the sigmoids and the 3rd-order polynomial functions of the pooled positive and negative response scores. The difference limens listed in Table 3 are derived from these polynomial functions.
Grossly speaking, the difference limens fall within Flanagan's range (3-5% of the formant frequency), the total average being 3.8% of the formant frequency. The derived thresholds at 75% averaged over the four F0 conditions are 4.5%, 3.3%, and 3.5% of the frequency for 1000 Hz, 1250 Hz, and 1500 Hz, respectively, whereas those averaged over the three formant frequency conditions are 4.1%, 4.0%, 3.4%, and 3.7% of 120 Hz, 300 Hz, 400 Hz and 500 Hz, respectively. The dotted lines in the table indicate that the intended difference limen could not be derived from Figure 3 (the averages are based on the actual values only).
Our results do not indicate that it is more difficult to detect changes in formant frequency at a relatively high-pitched vowel than it is at a relatively low-pitched one. The variation in fundamental and formant frequency conditions may rather be explained by the relation between the formant frequency and the harmonics of the fundamental frequency. If the relation is favourable, i.e. if the formant (nearly) coincides with a harmonic, the difference limen will be low. In the conditions consisting of F1=1000 Hz and F1=1500 Hz at F0=500 Hz both formants are located at multiples of the fundamental. Not only are their difference limens relatively low (3.5% and 3.8%), but

they also yield a small error rate at zero-frequency and a 100% different response score at the most extreme frequency deviation. Contrariwise, harmonic complexes with a less favourable relationship between the fundamental and the formant frequencies, as for instance F1=1250 Hz at F0=500 Hz, show a higher error rate at zero-frequency (30%) and fail to reach the 100% different response score.

Table 3. Just discriminable differences in formant frequency. The difference limens in Hertz and percent are derived from the polynomial curve fit at the 50%, 75%, and 90% different responses.
The table corresponds to Figure 3.
[1] Number of trials per formant frequency condition (to be multiplied by four fundamental frequency conditions)
[2] As read from the discrimination curve, not the polynomial function.

		1000 Hz (# 48) [1]		1250 Hz (# 60)		1500 Hz (# 68)		
F0		Hz	%	Hz	%	Hz	%	$\bar{X}$
120 Hz	50%	25	2.5	25	2.0	34	2.3	**2.3**
	75%	50	5.0	46	3.7	55	3.7	**4.1**
	90%	----	----	----	----	65	4.3 [2]	**4.3**
300 Hz	50%	22	2.2	25	2.0	40	2.6	**2.3**
	75%	45	4.5	44	3.5	60	4.0	**4.0**
	90%	50	5.0	----	----	75	5.0	**5.0**
400 Hz	50%	27	2.7	15	1.2	16	1.1	**1.7**
	75%	50	5.0	35	2.8	33	2.3	**3.4**
	90%	----	----	----	----	65	4.3	**4.3**
500 Hz	50%	25	2.5	40	3.2	42	2.8	**2.8**
	75%	35	3.5	----	----	57	3.8	**3.7**
	90%	42	4.2	----	----	67	4.5	**4.4**
			4.5		**3.3**		**3.5**	**3.8**

Comparison of the twelve conditions in Figure 3 leads to the following conclusions:

- F1=1000 Hz at F0=500 Hz and F1=1500 Hz at F0=400 Hz and F0=500 Hz yield both a low error rate at zero-frequency difference (9%, 5%, and 10%) as well as the highest (100%) difference response at the extreme frequency deviations .
- F1=1000 Hz at F0=400 Hz and F1=1250 Hz at F0=500 Hz yield a high error rate at the zero-frequency difference (18% and 32%, respectively) and fail to achieve 75% different responses.
- F1= 1250 Hz and F1=1500 Hz at F0=120 Hz and F0=300 Hz have a relatively low error rate at zero-frequency and a high percent (yet not 100%) of difference responses at the extreme frequency deviations.

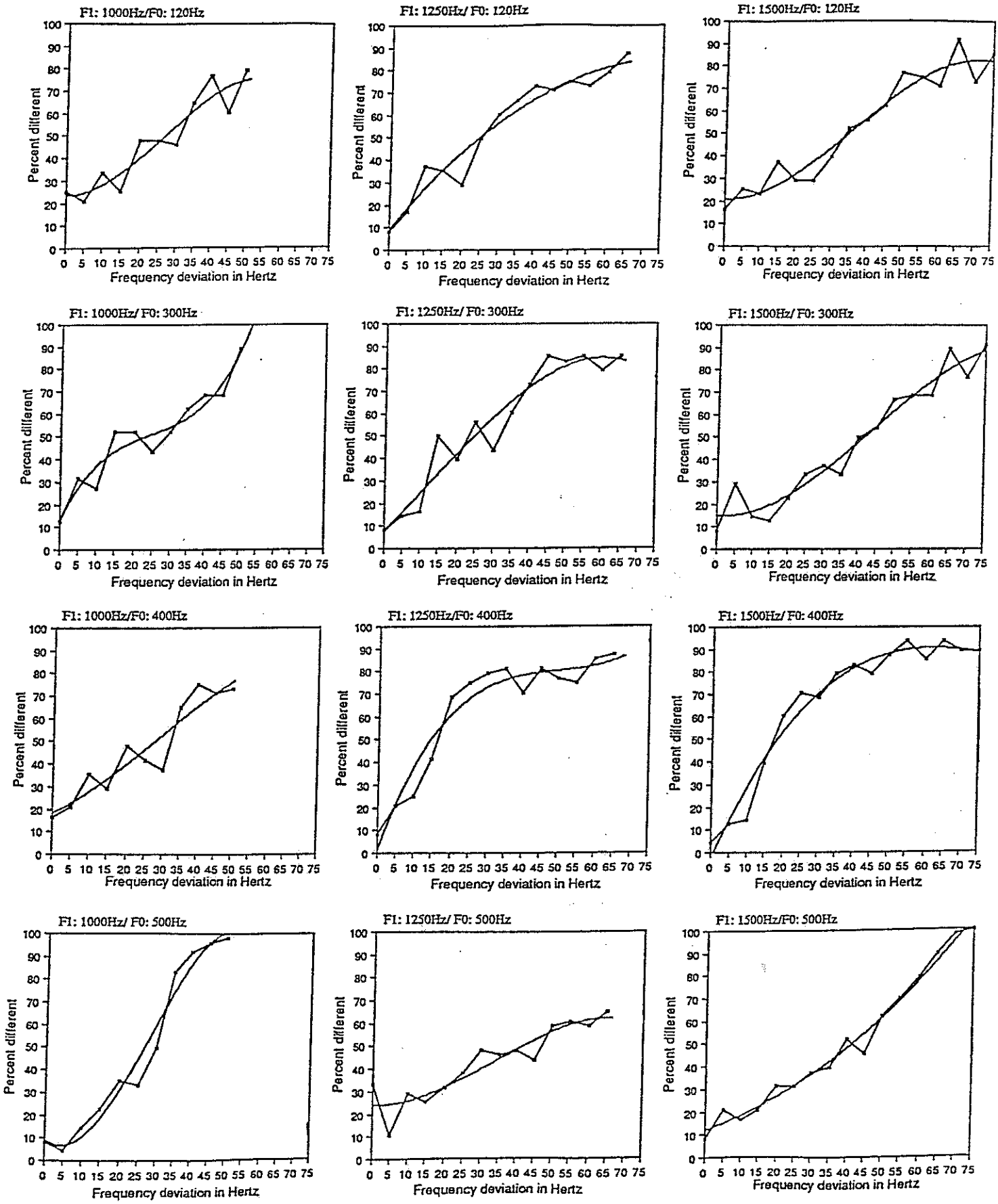

Figure 3. Sigmoids and 3rd-order polynomials of percent different responses as a function of formant frequency deviation for positive and negative responses pooled.

The variation within the conditions as reflected by the difference limen and the course of the sigmoid is explained by the small number of trials and the inexperience of the subjects. The difference across the conditions also prompts another explanation. It appears that discriminability is very much dependent on a favourable relation between the formant and the harmonics of the fundamental, this can occur just as well in sparsely as in densely spectralled complexes. If the formant is situated rather unfortunately in relation to the harmonics of the fundamental the impact of only a few harmonics in the higher pitched harmonic complexes will be large. In the following section we have classified the results of our formant frequency experiment according to three conditions, i.e. so-called 1) favourable, 2) unfavourable, and 3) intermediate. A third discrimination experiment, based on a working model will test the classifications.

4.3.2 Favourable conditions

In the results of our formant frequency experiment it is seen (Figure 3) that the sigmoids belonging to F1= 1500 Hz at F0= 400 Hz and at F0=500 Hz and F1=1000 Hz at F0=500 Hz display the lowest error rate at zero-frequency and attain a 100% different responses at the most extreme deviation points. In the last set the 1000-Hz harmonic is clearly the strongest. The first formant frequency deviates from 950 Hz to 1050 Hz, this being too small a range to bring about any other effect or to diminish the 'power' of the harmonic. Essentially the same applies to the same fundamental frequency condition with F1=1500 Hz. The slope of the discrimination curve is not steep, but the most extreme frequency changes are heard by all. A perhaps even better example is the same formant frequency condition at F0=400 Hz. The formant frequency lies very near to the 1600-Hz harmonic and approximates it even better as it increases in frequency to the most extreme frequency deviations offered, i.e. 1575 Hz.

4.3.3 Unfavourable conditions

In considering F1=1250 Hz at F0=400 Hz we assume that at the first few frequency deviations the influence of the 1200-Hz harmonic is strongest. However, this influence becomes noticeably less as the frequency increases to 1315 Hz. Despite the fact that the frequency deviation becomes larger there is *no harmonic nearby to aid perception*, resulting in a relatively low percentage of 'different' responses at the most extreme frequency deviations. At F1=1000 Hz with F0=400 Hz the formant frequency lies at first exactly halfway between the 800-Hz and 1200-Hz harmonics. Neither this frequency nor the deviation of up to 50 Hz is supported by either harmonic. The error rate at zero-frequency deviation is high and a 75% difference response is hardly reached.The same applies to F1=1250 Hz at F0=500 Hz. The formant frequency deviates between two harmonics, neither of which is very prominent. The distance between the two harmonics is too large to be of any aid in perceiving small changes.

4.3.4 Intermediate conditions

At the lowest pitch condition, 120 Hz, the abovementioned effect is absent and a weighing mechanism of the harmonics, caused by the changing position of the formant frequency, may explain the percentage of different responses in the discrimination curves.
Due to the fact that there are more harmonics in the 300 Hz subsets than in the higher ones, difficulty in detecting differences will be less severe. At F1=1000 Hz and F1=1250 Hz discrimination requires more effort as the frequency deviation becomes larger, due to the diminishing influence of the 1000-Hz and the 1200-Hz harmonics, respectively. At 1500 Hz the integration of the harmonics and the formant frequency is possibly causing discrimination to slow down, as the sigmoid is rather level.

4.4 Working model

In order to test whether and to what extent discrimination is dependent on the relation between the fundamental and formant frequency we performed a third test consisting of three different conditions (Figure 4):
1) Condition A contains harmonic complexes with a favourable relation between the formant frequency and the harmonics of the fundamental frequency. This would apply to all the higher F0 conditions with a formant falling (nearly) in line with one of the harmonics. No matter how the formant frequency varies, the harmonics are so widely spaced that the most prominent harmonic remains prominent. Apparently, the listener is aided by the sparse number of harmonics. The more prominent the harmonics are, the easier it will be to perceive a difference.
2) Condition B contains the most unfavourable conditions, i.e. those in which the formant occurs exactly in between two harmonics. Now the widely spaced harmonics influences discrimination unfavourably. The high fundamental frequency conditions with the formant not being a multiple of the fundamental belong to this condition.
3) Condition C contains no prominent harmonics as the spectrum is amply filled. Discrimination does not occur by selection of one prominent harmonic, but by weighing all the frequency components of the spectrum.

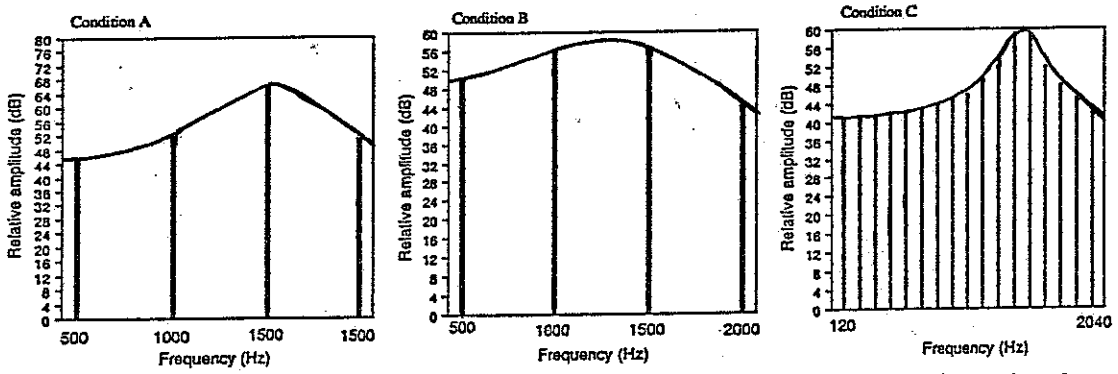

Figure 4. Sketches of conditions A, B, and C. The frequency of the harmonics (in Hertz) are plotted as a function of the amplitude (in dB).

Presumably listeners average -and switch- between the two strategies. It is not clear to which conditions the F1=1000 Hz at F0=300 Hz and the F1=1250 Hz at F0=300 Hz conditions belong. These may switch between conditions C and B, depending on the position of the formant. As the harmonics are not spaced as far apart as in the higher conditions, differences will be quicker perceived.

5.0 EXPERIMENT 3: ON TIMBRE DISCRIMINABILITY

In order to determine what the influence of increasing or decreasing the amplitude of the harmonics -the indirect effect of shifting a formant frequency- will be on discriminability, we generated synthetic harmonic complexes according to the abovementioned conditions. The signals were generated digitally by adding a variable number of harmonics of which the frequency and amplitude per harmonic were changed. Such a systematic variation of the frequency and the amplitudes of harmonics is not possible with a source-filter model as used in the earlier experiments.
Appropriate amplitudes were derived by measuring the FFT spectra of the original data. If the formant frequency is positioned exactly on top of a harmonic the amplitude of the harmonic can be as high as 64 dB (this is a relative measure). The more the formant frequency changes -and the more this frequency departs from the harmonic- the more

the amplitude of that harmonic will decrease. The neighbouring one will increase, but the extent to which this happens depends on the distance between the two harmonics. If the formant frequency is situated between two harmonics the amplitudes of the two nearest harmonics are more or less the same (they may deviate by 1 dB or 2 dB).

For the harmonic complexes in experiment 3 the number of harmonics (situated in the frequency range of up to about 2 kHz) per fundamental frequency condition were 17, 6, 5 and 4 harmonics for the 120 Hz, 300 Hz, 400 Hz and 500 Hz conditions, respectively. Starting with the value as measured at the reference stimulus the amplitude of the harmonics were decreased in steps of 1 dB.

5.1 Data

Seven subsets -all held separately- were generated according to conditions A, B and C. As it was assumed that the harmonic complexes belonging to condition B would be more difficult to differentiate than those belonging to A, more harmonic complexes were made for the first.

For condition A the most prominent harmonic of F1=1000 Hz at F0=500Hz (the 2nd harmonic) was decreased in 5 steps from 64 dB to 59 dB. Together with 1 reference signal this results in 6 stimuli.
As F1=1250 Hz at F0=400 Hz does not coincide exactly with the harmonic, two subsets were generated, one in which only one harmonic was decreased (from 62 dB to 54 dB, 9 stimuli), the other in which two harmonics were varied (10 stimuli).

For condition B the 1000-Hz and 1500-Hz harmonics (originally both equally strong) of F1=1250 Hz at F0=500 Hz were varied in different ways with 1-dB steps, resulting in a total of 15 different harmonic complexes. The same applies to the two harmonics of F1=1000 Hz at F0=400 Hz, another difficult condition.

For condition C only one subset, i.e. F1=1500 Hz at F0=120 Hz was created. The 1440-Hz and the 1560-Hz harmonics were varied in 8 different ways, together with the reference stimulus this results in 9 stimuli.

In the earlier discussion of the formant frequency data with respect to a working model, it was unclear as to which condition the 300 Hz stimuli belonged. One intermediate condition was therefore generated, i.e. F1=1000 Hz at F0=300 Hz. Two of the six harmonics -the 900-Hz and the 1200-Hz ones were varied such that 9 different harmonic complexes (reference included) were obtained.
Altogether the test contained 73 different stimuli.

As in the previous tests all stimuli had a 200 ms duration. They were sampled at 20 kHz and the first and last 5 ms of the signal were smoothed by a HAMMING window to avoid clicks.
The signals were once again balanced for an AX paradigm (AX and XA), randomized and recorded on a high quality tape. All together there were 146 stimuli (73*2). The interstimulus time was 0.6 seconds and the interpair time was 2.0 seconds.
The stimuli were offered to five naive listeners who heard the signals binaurally over headphones. They were instructed to indicate on their scoring sheet whether they heard a difference or not. It was stressed that each pair should be judged independently and that they should not try to remember a reference sound. No mention was made as to the type of difference. A test tone preceded every ten stimuli as a means of reference.

5.2 Results and appropriateness of working model

Even though the results of the tests are based on only a small number of trials, there is such unanimity among the subjects that we consider it justified to draw some conclusions:

1) discrimination is easier if the formant falls in line with a harmonic. The harmonic is then so prominent that a decrease of 1 dB is clearly perceived (condition A).
2) a decrease of at least 3 dB is needed to hear a difference between a standard signal and a deviating one if the formant is not near a harmonic. This may be a decrease in one or both of the harmonics. If the formant is not suitably situated discrimination is aided by more harmonics. This seems plausible as the distance between the formant and the harmonics can be very large in the sparsely spectralled, i.e. high-pitched sounds (condition B).
3) in the densely spectralled samples the position of the formant is not very important as more than one harmonic is enhanced. Discrimination occurs by weighing many harmonics and not by selecting one (condition C).
4) Discrimination is cued by the amplitude changes of the most prominent harmonic (1 dB) in the favourable conditions.

5.2.1 Results condition A

The histograms in Figure 5 show the percentage different responses as a function of the harmonic variation. Either the amplitude of the prominent harmonic or the ratio of two changing harmonics are plotted on the abscissa. In all histograms the left bar belongs to the standard harmonic complex.
Only one of the three subsets belonging to condition A are illustrated as they are all the same. Apart from all the standard signals being, correctly, scored as *same,* every deviating harmonic complex was correctly scored as being *different.* It is therefore clear that if a relatively sparsely spectralled harmonic complex contains one very strong harmonic a difference as small as 1 dB is heard. This also holds for the 400 Hz condition in which two harmonics are varied. The second harmonic appears to be of no influence in a discrimination task. The amount with which the second harmonic is increased with decrease of the first is very small as there remains a large amplitude difference between the prominent harmonic and the neighbouring ones. The first harmonic remains approx. 14 dB or 20 dB stronger than the neighbouring ones in the 400 Hz and 500 Hz subsets, respectively.

5.2.2 Results condition B

As was expected discriminability in the harmonic complexes belonging to condition B (the two lower histograms in Figure 5) is more difficult. Owing to the rather unfortunate relationship between the harmonics and the formant frequency, the listener is not aided by the harmonic structure of the harmonic complex in detecting small differences between two signals. In a sparsely spectralled harmonic complex, discrimination is even more difficult as the harmonics may lie too far apart from each other to cue detection. Once the formant frequency has increased to such an extent that a neighbouring harmonic has increased substantially in comparison to the other ones, a difference in two signals will become detectable again.
In comparison with the standard signal (cond. 57/56 dB) the harmonic to the left of the formant frequency in the 400 Hz subset will have to decrease by at least 3 dB, while the one to the right will have to increase by at least 4 dB (conditions 54/60, 53/60, 52/61 and 50/61) to bring about a 50% difference response. Smaller differences are only distinguished by few listeners. However, the differences between the harmonics are of a particular kind. Compare the percentage different responses of the 54/60 dB harmonics and the 53/59 dB ones in the same plot in Figure 5. The results indicate that

even though the amplitude difference between two harmonics of the same harmonic complex is the same and that of the harmonic complexes in comparison with the standard signal is reversed for both harmonic complexes, the influence of the harmonics is different. Apparently, the harmonic to the right of the formant frequency - which increases to 60 dB- is the discriminating cue, whereas the first harmonic decreases to the same level as the surrounding harmonics.
This increase of the harmonic at the right of the formant frequency is also more important than the ratio between the two harmonics in the 500 Hz subset. In this subset the harmonics lie even further apart than in the previous one. In comparison with the amplitudes of the harmonics of the standard signal -56/55 dB- an increase of 3 to 4 dB of the right 1500-Hz harmonic is enough to warrant some differentiability. However, a 4-5 dB decrease of the (left) 1000-Hz harmonic is of less influence, since no difference is heard between the standard signal and 51/58 dB, whereas a 30% different response is obtained for 52/59. Once again, when the relation between the formant frequency and the harmonic is favourable, i.e. the amplitude of the nearest harmonic is prominent in comparison to the other, discrimination noticeably increases.

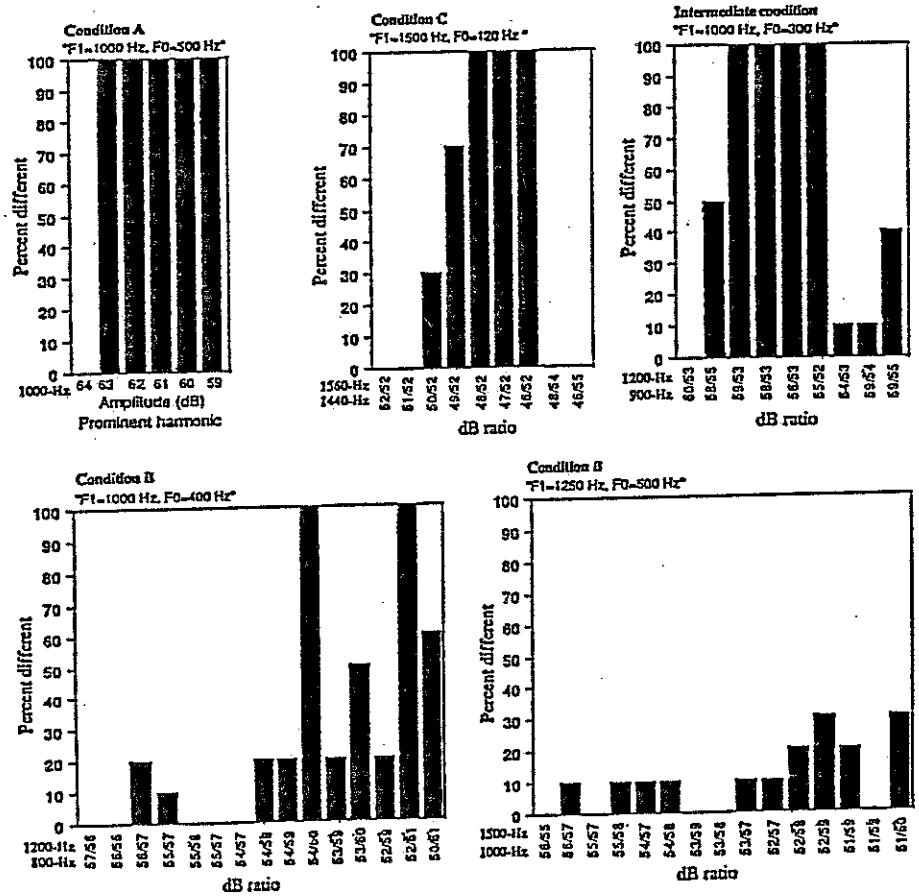

Figure 5 Histograms of percent different responses as a function of harmonic variation

5.2.3 Results condition C

In the most densely spectralled harmonic complexes containing 17 harmonics discrimination is fairly easily achieved. As the harmonic complex resembles F1=1500 at F0=120 Hz the strongest harmonics -both 52 dB- were the 1440-Hz and 1560-Hz ones. Differences between the standard signal and the deviating ones were heard once the 1440-Hz harmonic had been decreased by 3-5 dB. The cue is the difference in amplitude between the 1440-Hz harmonic of the standard and the deviation signal, as the other harmonics remain the same. Naturally, the ratio of the two harmonics in one harmonic complex becomes larger, but this does not appear to be the cue. For if the

1560-Hz harmonic is increased by 2 dB -while the 1440-Hz is decreased (48/54)- no difference between the signals is detected. Not only is the increase of the 1560-Hz harmonic or the larger ratio between the harmonics of the same signal of no influence, but the decrease of the 1440-Hz harmonic, which at first seemed to be the cue for discrimination is of no influence either.

To follow discrimination in a more sparsely spectralled harmonic complex we evaluated the 300 Hz series (intermediate condition). The relationship between the harmonics and the formant frequency are not very favourable here either, but better than with condition B, where the harmonics are even wider spaced. Of the two changing harmonics the first was 60 dB (standard signal), while the second was 53 dB. Unanimous discrimination is obtained if only the first is decreased, even if this is only by 1 dB. Remarkably though, decreasing the left harmonic is only functional up to 54 dB. Then the two harmonics within one harmonic complex are equal in amplitude, causing the standard signal and the deviating one to sound the same. It is difficult to explain this. For if the two equal amplitude harmonics are integrated in the right harmonic of the standard signal then the 60 dB harmonic should be of some influence, as it is the most prominent one. However, perhaps not prominent enough to cause discrimination. Also, discrimination is more difficult if the right harmonic is gaining in prominence while the left one is decreasing by a small amount.
The process of detecting small differences is more complicated than described in the working model. From our results on the 300 Hz data it appears that listeners switch between an overall strategy, i.e. one in which more than one harmonic is involved, and a (prominent) harmonic strategy. If there is no very prominent harmonic the relation between two harmonics of a harmonic complex become increasingly more important. If there is no prominent harmonic and the ratio between two harmonics is too small the harmonic complexes sound the same. Compare the standard signal 60/53 dB with 59/53 and 59/55. These results may imply a loudness detection task (of the most prominent harmonic) for condition A and a quality difference one for condition B.

6.0 CONCLUSIONS

1) Listeners are equally sensitive to detecting relative changes in fundamental frequency in stationary speech-like signals similar to those produced by infants and adults.
It would be interesting to know the true difference limens (in the sense of the maximum sensitivity) after training subjects extensively and after rerunning the test a number of times. It is expected that the values of the pitch experiment would decrease substantially, from over 2% down to about 0.5% (similar to those of Flanagan or Klatt).
2) Discriminability between two harmonic complexes differing in formant frequencies is similar for low or high fundamental frequencies. In determining true difference limens one would need a multi-formant harmonic complex. This would be a more natural situation, but possibly also more difficult as in our one-formant harmonic complexes there can be no interference from other formant frequencies, nor masking of lower formants by higher ones.
3) The mechanism by which discrimination is achieved differs according to the available spectral information. The differences in the results do not lie in the relatively high fundamental frequency, but rather in the relation between the position of the formant frequency and the harmonics. However, in the case of a relatively high fundamental frequency this relation may be very unfavourable -for instance if the formant frequency is positioned exactly in between two harmonics, which lie 500 Hz apart. However, at the high F0 there can also be favourable conditions. If the formant frequency falls exactly in line with a harmonic, discrimination is easier at a relatively high fundamental frequency than at a relatively low one. This is because of the prominence of the single harmonic.

It is difficult to conclude from a discrimination task whether listeners base their decisions on changes in amplitude or on timbre. As a difference of 1 dB is perceived when the harmonic is very prominent, whereas it is at least 3 dB in other conditions, listeners are possibly cued by amplitude changes (of the harmonic) in the harmonic is prominent and by quality changes when the harmonics are densely spaced or when one harmonic is not very prominent. These quality differences -judged from the (difference)ratio of the harmonics within and across harmonic complexes- are dependent on the relation between the harmonics of the fundamental frequency and the position of the formant.

6.1 DISCUSSION

The working model has been used to explain the results of a relatively small set of conditions so far. If it proves to be an adequate model after a more thorough validation, existing ideas from identification and matching experiments may be reflected on the discrimination process. Bladon (1982), for instance, finds the formant frequencies to depend on the fundamental frequency (unlike Fujisaki and Kawashima), at least for the higher-pitched vowels. If the harmonics are not closely situated to each other the formant will shift in the direction of the nearest harmonic. This means that that harmonic has become so prominent that identification will then occur by picking the harmonic peak instead of the formant one (which lacks energy at that frequency).
As our data indicates that both fundamental frequency and formant frequency differences are perceived equally well at different fundamental frequencies, but that the discrimination process is dependent on the available spectral information -to which listeners adjust their perceptual mechanism accordingly- we can discriminate and identify infant, women and male speech well. Presumably, the same strategies are used in a discrimination, identification or matching tasks. However, within these tasks there appear to be different strategies. For instance the amplitudes of the most prominent harmonics in the densely-spectralled harmonic complexes are more or less the same as in those harmonic complexes where a formant falls halfway between two harmonics. Still, the subsets in the discrimination test yield different results. As we can evaluate performance only in terms of the way the task has been specified, and in terms of the time necessary to arrive at an achievement, the procedures by which the values were obtained should be considered seriously. The procedure holds the type of stimulus, the manner in which it is generated, and the actual testing methods.
The matching criterion of a two-formant stimulus by a one-formant one appears to depend on the distance between the two formants (Chistovich et al., 1979). If the two formants were placed closer than the critical value of 3.5 Bark, the match was dependent upon the amplitudes of the formants, while if the formants were spaced more than 3.5 Bark matching seemed insensitive over a large range of amplitude variations. In our widely spaced harmonic complexes with a formant far from the skirts of a harmonic, discrimination also seemed insensitive over a large range of frequency deviations.
By performing three discrimination experiments we have excluded from explicit labeling of speech stimuli. It is important to start with a discrimination task as the identification of stimuli is dependent on discriminable differences. This holds that, contrary to Ryalls and Lieberman (1982), we do not find the auditory system to be necessarily aided by a denser spectralled harmonic complex. In our discrimination tests one prominent harmonic in a sparsely spectralled harmonic complex could be a very distinctive cue. At this stage it is not clear if there is a relationship between the critical bandwidth and the spacing of the harmonics. It is likely that the auditory system switches between a strategy in which the decision is based on a prominent harmonic in the sparsely spectralled harmonic complexes and on a weighed average of more harmonics when the spectrum is filled.

In this paper the perception of complex sounds has been described rather schematically and simply. To test more multidimensional parameters like timbre, one would have to use other testing paradigms. It is expected that perception would then depend on cues similar to those we find in our discrimination tests, but that a different perceptual strategy is used. More research is needed to examine these perceptual strategies.

REFERENCES

Bladon, A. (1982). "Arguments against formants in the auditory representation of speech", in: Auditory analysis and perception of speech (G. Fant and M.A.A. Tathem, eds.), 95-102.

Chistovich, L.A., Sheikin, R.L. and Lubinskaja, V.V. (1979). "Centres of Gravity and Spectral peaks as the determinants of vowel quality" in: Frontiers of speech communication research (B. Lindblom and S. Ohman, eds.), 143-157.

Flanagan, J.L. (1955a). "A difference limen for vowel formant frequency.", J. Acoust. Soc. Am. 27(3), 613-617.

Flanagan, J.L. and Saslow, M.G. (1958). "Pitch disrimination for synthetic vowels.", J. Acoust. Soc. Am. 30 (5), 435-442.

Harris, J.D. (1952). "Pitch discrimination", J. Acoust. Soc. Am. 24 (6), 750-755.

Klatt, D.H. (1973). "Discrimination of fundamental frequency contours in synthetic speech: implication for models of pitch perception". J. Acoust. Soc. Am. 53(1), 8-16.

Moore, B.C.J. (1973)."Frequency difference limens for short-duration tones". J. Acoust. Soc. Am. 54(3), 610-619.

Moore, B.C.J., Glasberg, B.R., and Shailer, M.J. (1984). "Frequency and intensity difference limens for harmonics within complex tones.", J. Acoust. Soc. Am. 75(2), 550-561.

Nordmark, J.O. (1968). "Mechanisms of frequency discrimination", J. Acoust. Soc. Am. 44 (6), 1533-1540.

Ryalls, J.H. and Lieberman, Ph. (1982). "Fundamental frequency and vowel perception", J. Acoust. Soc. Am. 72(5), 1631-1634.

Slawson, A.W.(1968). "Vowel quality and musical timbre as function of spectrum envelope and fundamental frequency", J. Acoust. Soc. Am. 43(1), 87-101.

Weenink, D.J.M. (1985). "Formant analysis of Dutch vowels from 10 children", IFA proceedings (9), 45-52.

Weenink, D.J.M. (1988). "Klinkers: een computerprogramma voor het genereren van klinkerachtige stimuli", IFA-report nr.100.

Wier, C.C., Jesteadt, W. and Green, D.M. (1977). "Frequency discrimination as a function of frequency and sensation level", J. Acoust. Soc. Am. 61 (1), 178-184.