Discrimination of simplified vowel spectra

Lijzenga, Johannes

Chapter 5

Frequency discrimination of stylized synthetic vowels with two formants

Johannes Lyzenga
J. Wiebe Horst

Department of Otorhinolaryngology/Audiology
University Hospital Groningen
P.O.Box 30.001, 9700 RB Groningen, The Netherlands

Abstract

Just-noticeable differences (jnds) in the formant frequencies of synthetic two-formant "vowels" were measured for normal hearing subjects. The jnds were examined for a change in only the first or the second formant, and for a combined change of both formants. For the combined change, we used two quantitative relations between the two formant frequencies; one in which the relative changes in both formants were equal, and one with a double relative change for the first formant. The formant frequencies ranged from 500 to 600 Hz for the first, and from 2000 to 2100 Hz for the second formant. For the slope of the formant peaks, 50 and 100 dB/oct were used for the first formant region, combined with 100 and 200 dB/oct for the second formant region, respectively. For the fundamental frequencies of the complexes 100 and 200 Hz were used. For the single formant changes, two phase relations were used between the individual components of the complexes; a "natural" phase relation, and a random-phase relation. These results were compared to the jnds for a Gaussian noise that was filtered with the same spectral envelopes as the harmonic complexes. For the combined formant changes the jnds were measured for only the natural phase relation. A three-interval, three-alternative forced-choice task was used. All measurements were performed with roving stimulus level. For the single formant changes, the phase relations had no effect on the results. These results for the harmonic stimuli as well as the noise bands can be described by a model using a spectral profile comparison, in which information is combined over a limited frequency range by comparing the spectral profiles of the two presented stimuli. For the combined formant changes, the jnd could be explained by combining the perceptual differences from the two separately changed formants. In this combination these perceptual differences were summed as independent variables.

Introduction

This is the last paper in a series of three on formant-frequency discrimination for synthetic vowels. The investigations were started using highly stylized vowels with just a single formant (Lyzenga and Horst, 1995). By stepwise increasing the complexity of the stimuli via more natural one formant vowels (Lyzenga and Horst, 1997), we have now arrived at synthetic vowels containing two formants. The vowels in natural speech contain a number of formants, the first two or three of which characterize the vowel. In the present study, the formants of our two-formant vowels are either changed separately or simultaneously in the same direction.

Lyzenga and Horst (1995) investigated jnds for synthetic vowels with a single formant near 2000 Hz. Their stimuli had either a triangular or a trapezoidal spectral envelope. They found that the position of the formant frequency relative to the harmonics was an important influence on the jnd. For the triangular envelope, jnds were smaller for formant frequencies halfway between two harmonics than for formant frequencies at or near a harmonic. For the trapezoidal envelope, jnds were smaller for formant frequencies coinciding with a harmonic than for formant frequencies between harmonics. The splitting of the results into more than one group induced them to use more than one model to describe their data. For the triangles, under the conditions where the formant frequencies were not halfway between two harmonics, they proposed a model based on the modulation depth of the stimuli. On the other hand, they proposed a place model for the trapezia with formant frequencies not coinciding with a harmonic. They suggested that a modified place model might describe their data under the conditions where the formant frequency was halfway between two harmonics for the triangles, and coinciding with a harmonic for the trapezia. So, to encompass all explanations, they used a hybrid model.

The need to use a hybrid model was corroborated in a subsequent paper (Lyzenga and Horst, 1997). In this paper they reported jnds for synthetic vowels with a single formant in the range of either the first or the second formant. Again their stimuli were characterized by two spectral envelope shapes, but now they used the triangular envelope and a more natural envelope according to the Klatt synthesizer (Klatt, 1980). They found phase effects in the jnds for the second formant region with a low fundamental frequency, indicating temporal discrimination processes. In accordance with Lyzenga and Horst (1995), they found good correspondence between the jnd and the modulation depth for the triangular stimuli with formant frequencies coinciding with a harmonic. For such formant frequencies, their data for the Klatt-shaped envelopes could be explained by a modified place model (with a detection threshold of 2 dB, whereas Lyzenga and Horst, 1995, used a threshold of 1 dB in all their place models), in which information was combined over a limited frequency range by comparing the spectral profiles of the two presented stimuli. (For the trapezoidal envelope that was used by Lyzenga and Horst, 1995, the predictions of the present modified place model are very close to those of the place model that they proposed for these stimuli.) For the stimuli of the second formant region with a formant frequency halfway between two harmonics, Lyzenga and Horst (1997) found evidence for temporal discrimination processes. For these stimuli, they proposed a discrimination model based on the detection of changes in the sharpness of the minima of the temporal envelope of the stimuli, which discords with the modified place model suggested by Lyzenga and Horst (1995). One of the questions we wish to address in the present investigation is whether it is necessary to use such combined spectral-temporal modelling to describe formant-frequency discrimination for stimuli containing two formants.

In the present investigation we combined two of the stimuli of Lyzenga and Horst (1997) with the Klatt-shaped envelope, to form a synthetic vowel with two formants. To study the influence of the presence of a second peak in the spectrum, we investigated the jnd while changing one formant and keeping the other stationary. So, either the first formant was changed and the second was stationary, or vice versa. Earlier investigations using vowels that contain more than one formant, while changing them separately, have been performed by e.g., Flanagan (1955), Mermelstein (1978), Kewley-Port and Watson (1994), and Hawks (1994). A large range of jnds was found in these studies. Flanagan, like Kewley-Port and Watson, investigated jnds for a change of either the first or the second formant for a range of widely spread formant frequencies. Kewley-Port and Watson found jnds decreasing rather abruptly from 7 to 2% for formant frequencies from 200 to 800 Hz and then gradually sloping to about 1% for a formant frequency of 2800 Hz. Over this range Flanagan found relative jnds decreasing from 5 to 3%. Mermelstein (1978) and Hawks (1994) investigated jnds for single and combined changes of the first two formants. The relative jnds Mermelstein found for a single change in the first formant (14% and 5.5%) were slightly larger than those for a single change in the second formant (4.2% and 7%). Hawks found jnds near 2% for a single change in both the first and the second formant. All in all, in most of these studies jnds were larger for the first than for the second formant region. However, because of the non-overlapping spread of the parameter values used over all these studies, models that might describe the data cannot be easily tested using these results.

For the stimuli of the first experiment, we used an orderly set of parameters to enable thorough testing of possible models to describe formant-frequency discrimination. Under a number of stimulus conditions, Lyzenga and Horst (1997) found a phase effect, which indicates that temporal discrimination processes were involved. To check the occurrence of such temporal processes for the two-formant vowels, we used two different phase relations between the harmonics of the stimuli. The first is the phase relation as generated by the Klatt synthesizer. This condition resembles hearing a speaker from close by. The second is a random-phase relation, more akin to hearing a speaker in a room, where the original phase relation of the sound has been disturbed by the added reflections of the walls. Since the position of the formant frequency relative to the harmonics was found to be an important influence on the jnds, we employed the two positions also used by Lyzenga and Horst (1997): at a harmonic, and halfway between two harmonics. To enable easy comparison of the present jnds with their single-formant jnds, we used the same two fundamentals and slopes as Lyzenga and Horst.

In the second experiment we allowed both formants to change simultaneously. We used two different relations between the two formant frequencies. With the choice of these rules we tried to cover a range in which the changes in both formants were roughly of equal influence on the discrimination. According to the first rule, the relative change in the first formant was twice that of the second formant ($\Delta F_1/F_1 = 2\Delta F_2/F_2$). This rule was chosen to compensate for the fact that in most studies larger jnds were found for the first than for the second formant region. According to the second rule, the changes in the first and the second formant were chosen equal ($\Delta F_1/F_1 = \Delta F_2/F_2$). This rule was chosen because we expected that the jnds for the second formant would be disturbed more by the presence of a first formant than vice versa (in many stimuli the level of the first formant was much larger than that of the second formant). In the stimulus generation, we used the same two fundamentals and slopes as in the first experiment. Because of the large influence of the position of the formants relative to the harmonics, we used both the position at a harmonic and halfway between two harmonics for both formants.

Mermelstein (1978) measured a few jnds for single and combined changes of the first two formants. Hawks (1994) investigated discrimination for single changes of the second formant, and for a number of combinations of changes in the first three formants. In a "pilot fashion" he measured jnds for single and combined changes in the first two formants. Both Mermelstein and Hawks found smaller jnds for a parallel combined change of two formants than for single changes of these formants. Mermelstein proposed a model for the prediction of jnds for combined formant changes from the separate formant changes. In this model the frequency changes in the separate formants are added as "independent parameters" to form a measure of the combined frequency change. For parallel formant changes, Hawks proposed a "weighted Euclidian distance model" that predicts a jnd for combined formant changes directly from the jnds for single formant changes. In the present study we will check for our stimuli which rule is the more appropriate one.

I. General methods

A. Stimuli

The experiments were carried out using bandlimited harmonic complexes as stimuli. We used a smoothed spectral envelope, equivalent to the shape of two-formant vowels as generated by the digital Klatt synthesizer (Klatt, 1980). The fundamental frequency (F_o) of the harmonic complexes was either 100 or 200 Hz. We used two center frequencies (F_C) in the formant generation of the Klatt synthesizer: one in the region of the first and one in the region of the second formant. The center frequencies of the formants were chosen either to coincide with a harmonic, or in the middle between two harmonics. For the stationary formant these positions will be referred to as the peak-relations 1 and 2, respectively. In the first formant region the center frequencies were 500 and either 550 or 600 Hz, and in the second formant region they were 2000 and either 2050 or 2100 Hz. Two combinations of values for the slope (G) were employed for the formants: 50 dB/oct for the first and 100 dB/oct for the second formant, or 100 dB/oct for the first and 200 dB/oct for the second formant. Using the Klatt synthesizer, the relative formant peak levels in the spectral envelope depend on the distance between these two formants. Furthermore, the actual formant peak levels depend on the positions of the formants relative to the harmonics. Because of this we found attenuations of the second formant peak relative to the first between 0 and 27 dB (dependent of the stimulus parameters). This is illustrated in figure 1, in which some examples are shown of the harmonic stimulus spectra used in both experiments. For the stimuli shown in panels (c) and (g), the attenuations of the second formants are about 18 and 0 dB, respectively. In the first experiment we used two phase relations between the harmonics of the complexes: the phase relation as it is generated by the digital Klatt synthesizer (from here on called the Klatt-phase relation), and a random phase. For brevity, the stimuli with the Klatt phase and the peak-relations 1 and 2 (for the stationary formants) are indicated as Klatt-phase 1 and Klatt-phase 2, respectively. For the stimuli with the random phase, the stationary formant coincided with a harmonic (peak-relation 1). In the first experiment, we also investigated the jnd for a Gaussian noise that was filtered with the same spectral envelopes. The attenuation of the second formant peak relative to the first was between 15 and 18 dB, depending on the distance between the two formants. In the second experiment we only used the Klatt-phase relation in the stimulus generation. The formant positions at a harmonic, and between two harmonics were used for both the first and the second formant. The relative changes in the two formants were either equal, or twice as large in the first as in the second formant. In this experiment, the jnds will be considered in terms of either the first or the second formant, depending on which formant is dominant in the frequency discrimination. For the remaining formant, the positions at a harmonic and between two harmonics will be referred to as the peak-relations 1 and 2, respectively.

The stimulus generation and presentation procedures were equal to those described by Lyzenga and Horst (1995) for the Roving Level condition. The spectral envelope shape was calculated first, after which the harmonic components were added with the appropriate amplitude and phase. For the noise stimuli and the stimuli with random-phase relations, three sets of stimuli were made with different random relations. During the measurements each stimulus was picked at random from one of these three sets. In this way the correlation between the stimuli was reduced.

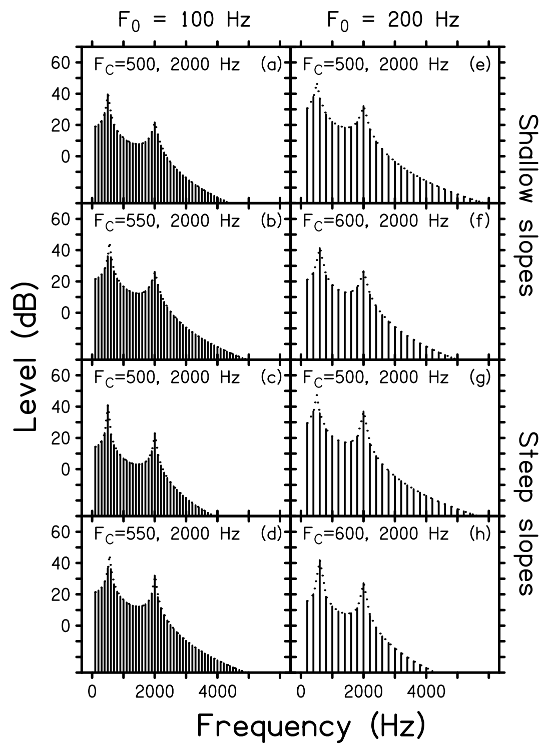

Figure 1: The spectra of some stimuli. The left column shows stimuli for the 100-Hz fundamental, and the right column for the 200-Hz fundamental. The two rows at the top display the stimuli with the shallow slopes, and the two rows at the bottom those with the steep slopes. The dotted lines indicate the spectral envelopes according to the Klatt synthesizer.

B. Procedure

The jnds were measured using an adaptive three-interval, three-alternative forced-choice method (3IFC). Before the actual jnd measurements, the absolute threshold of the reference tone was estimated, after which stimuli were presented at a level about 40 dB above this threshold. We added a pink background noise to the stimuli at a level of -40 dB relative to the stimuli; therefore, this procedure produced background noise levels close to the absolute thresholds. We used a "Roving Level" (RL) condition for the levels of all the tones: in this condition the levels of all stimuli were randomized, around one fixed level value, within trials over a 16-dB range in 0.5-dB steps (Henning, 1966). The background noise was roved in level along with the stimulus, keeping the signal-to-noise ratio in the stimulus constant.

The procedure by which the jnds were estimated was equal to the one used, and described in detail by Lyzenga and Horst (1995). In short, subjects were asked to identify the odd tone in a series containing two reference tones and one target tone. They were given immediate feedback. The frequency difference between the target and the reference tones was adapted according to decision rules that were chosen so that the procedure converges at 63% correct responses, which corresponds to a d' of 1 for the 3IFC paradigm. Data were collected until the direction of the center frequency adaption was reversed five times. On the average, one jnd measurement contained 71 trials. The whole set of jnds was measured three times (involving at least 200 trials per jnd), in one group containing the stimuli of both experiments in a pseudorandom order. The jnds were estimated from the averaged scores with the same algorithm as used by Lyzenga and Horst (1995). Since the subjects' bias toward one of the three signal intervals was found to be very small, it has been neglected in the calculations.

C. Subjects

Six normal-hearing subjects participated in the experiments. All were adults, four female and two male, with ages ranging from 25 to 48 years. For the harmonic stimuli, all subjects performed a series of jnd estimates for a different set of phase conditions, in such a way that three subjects participated for most conditions. In the first experiment, four subjects participated for a number of jnds under the Klatt-phase conditions. All six subjects contributed jnds for the filtered noise bands of the first experiment. All subjects had participated in frequency discrimination experiments before and were well trained. For all six subjects no improvements in the scores were observed during the course of the measurements.

II. Experiment 1. Single formant changes

The results of the first experiment are presented in figure 2. Each column contains the just-noticeable differences, measured for one region of the center frequency of the changed formant and one value of the slope, in six panels with individual results. The rows contain the jnds for the four combinations of formant frequency and slope. Each panel contains jnds for two values of the center frequency: one coinciding with and one halfway between the stimulus harmonics, denoted as /|\ and /||\, respectively. The smaller value of these pairs of center frequencies is plotted on the left and the larger on the right, except for the first formant region with the 200-Hz fundamental. Here the center frequencies are plotted in reversed order (i.e. 600 Hz, 500 Hz). For the 100-Hz fundamental, the jnds for the Klatt-phase 1, Klatt-phase 2 and random phase are shown as square, circular, and hour-glass symbols, respectively. For the 200-Hz fundamental, the jnds for these three conditions are shown as triangular-up, triangular-down, and picnic-table symbols. For clarity, the jnds for the 100-Hz and 200-Hz fundamentals are connected with a solid and a dashed line, respectively. The jnds for the filtered noise bands are shown as asterisks, connected by a dotted line. For the noise bands the value of the center frequency relative to the position of the harmonics is not relevant; we chose 500 and 600 Hz for the first formant and 2000 and 2100 Hz for the second formant, plotted at the positions /|\ and /||\, respectively.

The average jnds for the single formant changes are shown in figure 3. Each row contains average jnds for one of the four combinations of formant frequency and slope. The columns on the left and on the right contain averages for the fundamentals of 100 Hz and 200 Hz, respectively. The meaning of the symbols, and the annotations /|\ and /||\, is the same as in figure 2. For clarity the jnds for the Klatt-phase 1, Klatt-phase 2, and the random phase are connected with a solid, a long dashed, and a short dashed line, respectively. The averages for the filtered noise bands (the asterisks) are connected by a dotted line. For comparison, the predictions of the modified place model of Lyzenga and Horst (1997) are shown for each condition by means of the little

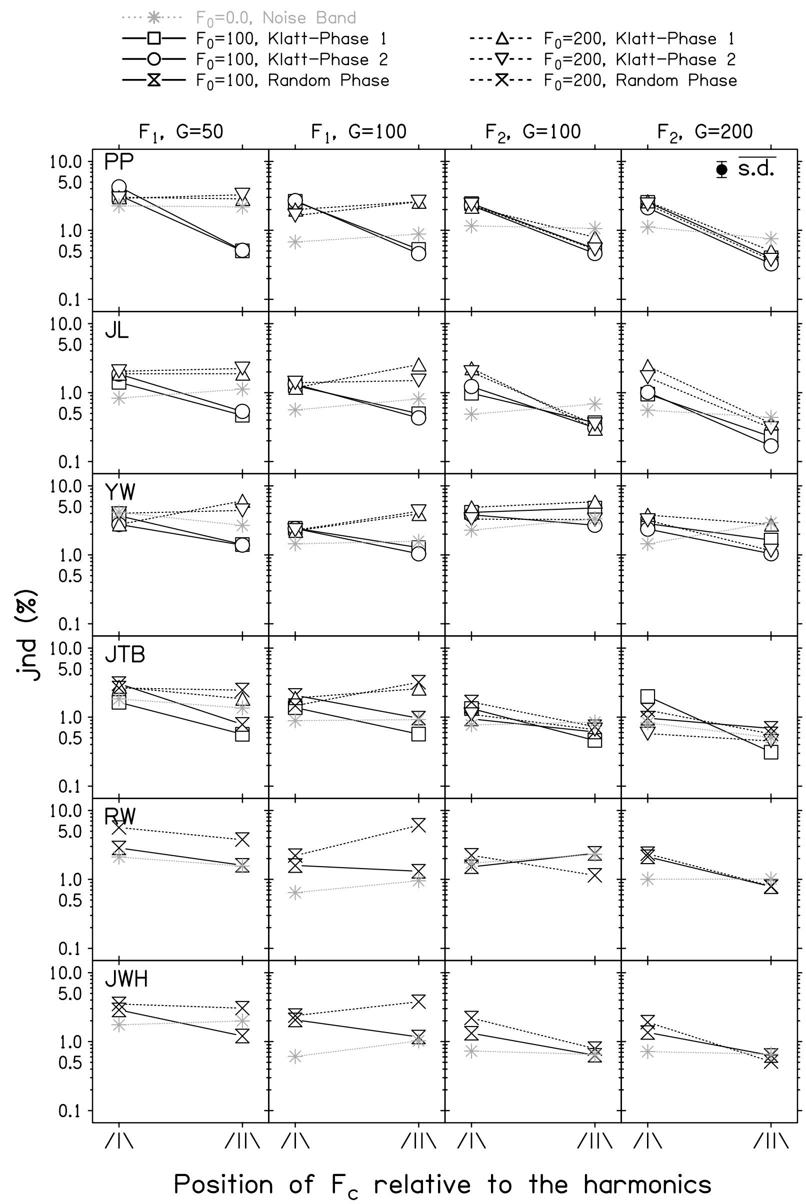

Figure 2: Individual jnds for single formant changes. Each column contains the jnds measured for one formant region and slope combination, showing individual results in six panels. The error bar in the right top corner indicates the mean standard deviation of the individual results. The jnds for the 100-Hz fundamental, combined with the Klatt-phase 1, the Klatt-phase 2, and the random phase are shown as square, circular and hour-glass symbols, respectively. For clarity, these symbols are connected with solid lines. For the 200-Hz fundamental, the jnds for the same three phase relations are shown as triangular-up, triangular-down and picnic-table symbols, respectively. These symbols are connected with dashed lines. The jnds for the noise bands are shown as asterisks, connected by a dotted line.

line segments at the sides of each panel. The settings of this model are identical to those used by Lyzenga and Horst (a Q_{10} of 5 for the Roex-filters, and a detection threshold of 2 dB). The solid and the dotted line segments represent the expected jnds for the harmonic stimuli and the noise bands, respectively.

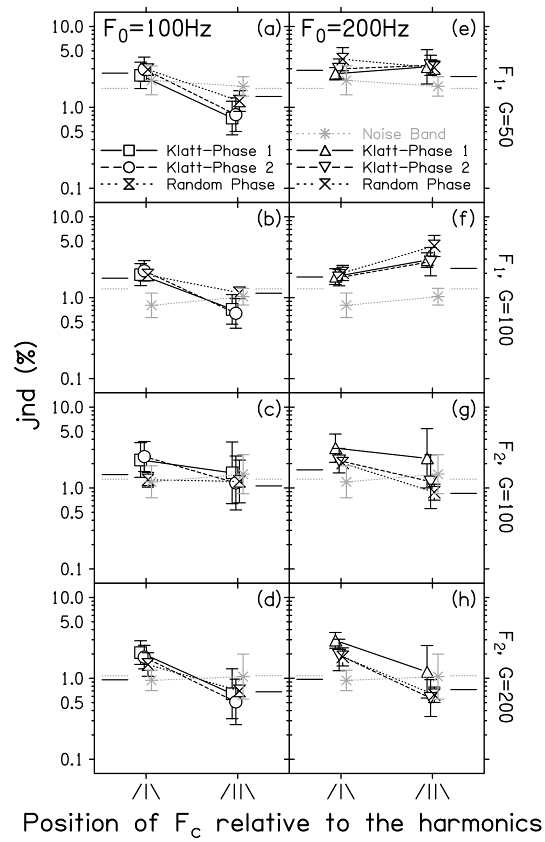

Figure 3: Average jnds for single formant changes. Each rows shows jnds for the four formant region and slope combinations. The left column contains the averages for the 100-Hz, and the right column for the 200-Hz fundamental. The meaning of the symbols is the same as in figure 3. For clarity, the jnds for the Klatt-phase 1 are connected by a solid line, those for the Klatt-phase 2 by a long-dashed line, and those for the random phase by a short-dashed line. The error bars indicate the standard deviations of the averages. The little solid and dotted line segments at the side of each panel indicate the expectations of the modified place model for the harmonic stimuli, and the noise bands, respectively.

In the region of the first formant ($F_C \approx 500$ Hz), the average jnds for the filtered noise bands are 2.0% for a slope of 50 dB/oct, and 0.9% for a 100-dB/oct slope. These values correspond to just audible level differences in the flanks of the stimuli of 1.4 dB and 1.3 dB, respectively. In the second formant region ($F_C \approx 2000$ Hz) the average jnds are 1.3% and 1.0% for the slopes of 100 and 200 dB/oct, respectively. These jnds correspond to level differences of 1.9 and 2.9 dB. So, for the second formant we find a larger level difference in the flanks of the stimuli, necessary for discrimination, than for the first formant. For both formant regions, Lyzenga and Horst (1997) found a level difference of just over 1 dB for noise bands with one formant. So, for the two-formant noise bands discrimination of a change in the first formant is not affected by the presence of the second formant, whereas for the second formant the jnds are increased by the presence of the first formant. Furthermore, for the second formant region, Lyzenga and Horst (1997) found a poor correspondence between the jnds and the predictions of the modified place model. For these stimuli they suggested a temporal discrimination process. For all of the present noise-band jnds in figure 3 we see good correspondence between the data and the modified place model. This implies that for the second formant the frequency discrimination mechanism is affected by the presence of the stationary first formant in such a way that it can no longer use temporal information and has to resort to a spectral process.

When considering all the average jnds for the harmonic stimuli, we find no significant differences between the two phase conditions and peak relations. On the other hand, for the second formant region and the 200-Hz fundamental, the averaged jnds for the Klatt-phase 1 are systematically somewhat larger than those for the Klatt-phase 2 and the random phase (significant with $p < 0.01$). However, such a difference is not observed in the individual results of figure 2. For each condition, the jnds in figure 3 are averaged over a different group of subjects; for the second formant region some spread in the average jnds is produced by the relatively large jnds of subject YW. Therefore, the increased jnds for the Klatt-phase 1 appear to be artifacts of the averaging, and it is reasonable to assume that the jnds are independent of the phase conditions and the peak relations.

On the average, the jnds of the first formant region are moderately larger than those of the second formant region (with a factor of 1.6, which is significantly larger than unity with $p < 0.01$). All jnds show modest, but significant, dependencies on the slope, on the fundamental, and on the position of the center frequency relative to the harmonics ($p < 0.01$). The predictions of the modified place model are practically equal to those for the single-formant stimuli of Lyzenga and Horst (1997). For the second formant region for the 200-dB/oct slope and a center frequency coinciding with a harmonic, the predictions of the model are somewhat lower than the jnds. However, these deviations are roughly equal to the standard deviation of the individual jnds. Under most conditions, the jnds are close to the predictions of the modified place model.

Figure 4 shows a direct comparison of the present data with the data found by Lyzenga and Horst (1997) for single-formant vowels. Each entry in this figure shows the quotient of the present two-formant jnd and the single-formant jnd. So, this figure shows the direct influence of adding a stationary formant to a single-formant vowel. A value larger than unity indicates a larger two-formant than single-formant jnd, a smaller value indicates the opposite. In the first formant region we find no consistent deviations from unity. Some deviations may be expected because different subjects participated for each phase relation in the two compared experiments. For the second formant region we find a consistent increase in the noise-band jnds, and in the harmonic jnds for the Klatt-phase relation. For the 100-Hz fundamental and a center frequency that lies halfway between harmonics, we find an increase for both phase relations. So, figure 4 shows that the jnds of the first formant region are hardly affected by adding a stationary second formant to the stimuli, and that for the second formant region many jnds are increased when a stationary first formant is added.

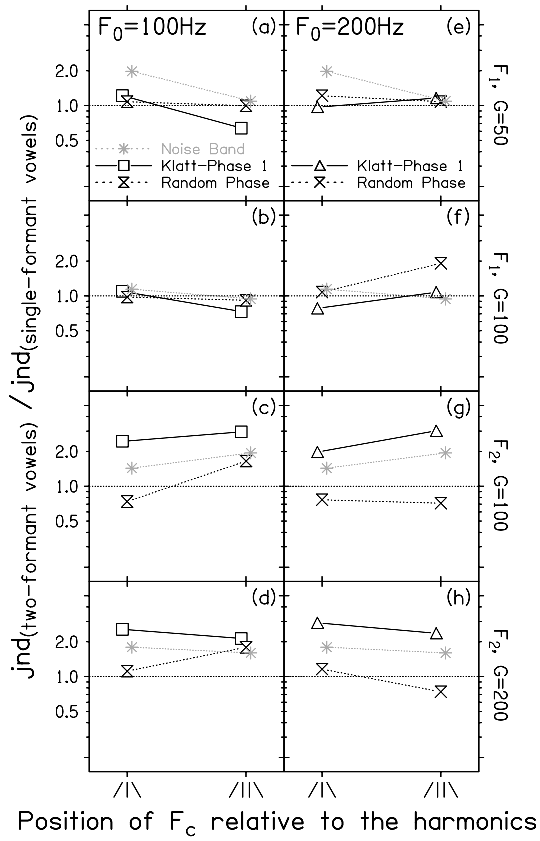

Figure 4: The average jnds for single formant changes in two-formant vowels related to single-formant jnds. The format of this figure is identical to that of figure 3. Each entry is the quotient of the jnd for a single formant change of figure 3 and that of the corresponding single-formant stimulus of Lyzenga and Horst (1997). No entries are given for the Klatt-phase 2.

III. Experiment 2. Combined formant changes

The individual jnds for the combined formant changes of the second experiment are presented in figure 5. The format of each panel is identical to those of figure 2. All stimuli have a Klatt-phase relation. Each column contains the just-noticeable differences, measured for one combination of the slopes of the two formants, in five panels with individual results. The leftmost two columns contain jnds for the formant relation: $\Delta F_1/F_1 = 2\Delta F_2/F_2$. These jnds are expressed as a percentage of the change in the first formant, because the first formant was found to be dominant in the discrimination process: the jnds resemble the corresponding jnds for the singly-changed first formant. For these jnds, the position of the first formant relative to the harmonics is indicated by the annotations /|\ and /||\. Whether the second formant coincides with a harmonic, or lies between two harmonics, is indicated by the peak-relations 1 and 2, respectively. The rightmost two columns contain jnds for the relation: $\Delta F_1/F_1 = \Delta F_2/F_2$. These jnds are expressed as a percentage of the change in the second formant, because we found that the changes in the second formant were the more important in the discrimination process here. For these jnds the position of the first formant relative to the harmonics is indicated by the peak relation. The jnds for the peak-relations 1 and 2 are shown as square and circular symbols for the 100-Hz fundamental, and as triangular-up and triangular-down symbols for the 200-Hz fundamental. For clarity, the jnds for the 100-Hz and 200-Hz fundamentals are connected with a solid and a dashed line, respectively. Due to an unfortunate choice in the distribution of the stimulus conditions over the subjects, RW and JWH contributed half the jnd series for both peak relations; under each condition their jnds for the center frequencies at and between harmonics belong to a different peak relation. Therefore, their jnds for each fundamental are not connected with lines.

The average jnds for combined formant changes are shown in figure 6. The upper-most two rows contain the jnds for the formant relation: $\Delta F_1/F_1 = 2\Delta F_2/F_2$, and the lower-most two rows for the relation: $\Delta F_1/F_1 = \Delta F_2/F_2$. The jnds for the 100-Hz fundamental are shown in the left column, and those for the 200-Hz fundamental in the right column. The symbols have the same meaning as in figure 5. The predictions of the modified place model are shown for all conditions by means of the little line segments at the sides of each panel. The model has been extended with a summation rule to predict the jnd for a combined formant change from the changes in the excitation due to both formants. The excitation difference of the first formant ΔE_1 (calculated below 1250 Hz) and that of the second formant ΔE_2 (calculated above 1250 Hz) are summed as independently perceived variables. The total perceived excitation difference ΔE_t is defined as:

$$\Delta E_t = \sqrt{\Delta E_1^2 + \Delta E_2^2} \tag{1}$$

The value of ΔE_t that ensues is compared with the detection threshold of 2 dB. The solid and the dotted line segments represent the expected jnds for the peak-relations 1 and 2, respectively.

Figure 6 shows that, when the relative change in the first formant is twice that of the second formant, the averaged jnds for peak-relation 1 are slightly larger than those for peak-relation 2 (on the average, they differ with a factor of 1.4, which just significantly differs from unity with $p < 0.01$). Furthermore, the jnds for peak relation 1 are very close to those for a singly-changed first formant shown in figure 3 (the differences between these two groups are not significant), and those for peak-relation 2 are slightly smaller than those for single formant changes (with a factor of 1.3, which is just significant with $p < 0.01$). This indicates that the changes in the second formants do not greatly influence the jnds (especially when the second formants coincide with a harmonic). So, we find that the first formant is dominant in the discrimination process here. The predictions of the modified place model extended with the summation rule are close to those of the modified place model for the first formant as they are shown in figure 3, and, in agreement with the jnds, we consistently find slightly smaller predictions for peak-relation 2 than for peak-relation 1. Most jnds are close to the predictions of this extended model. For the 200-Hz fundamental and a center frequency halfway between harmonics, the average jnds for peak-relation 2 are smaller than those for peak-relation 1. These relatively small jnds for the peak-relation 2 can also be seen in the individual results of subject JL, but not in those of subject YW (see figure 5). This inconsistency between subjects is displayed by the relatively large standard deviation of the average jnds for the peak-relation 2. So, here the changes in the second formant

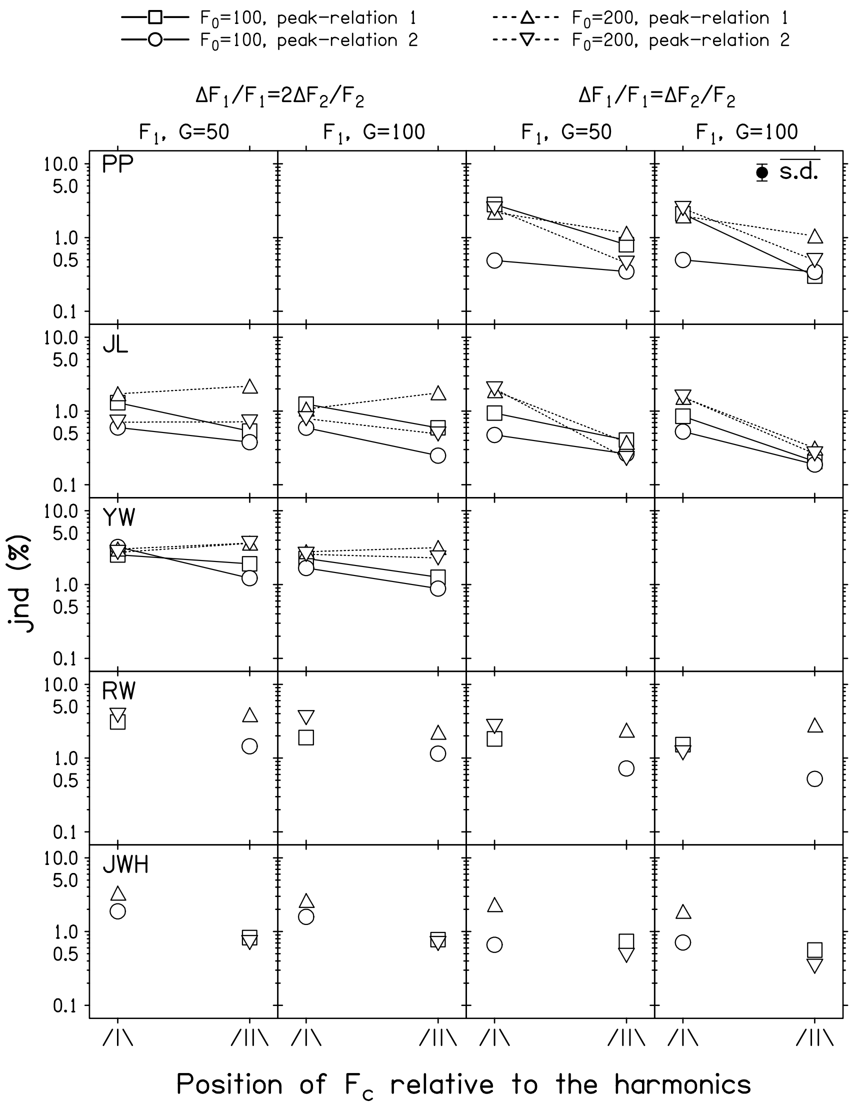

Figure 5: The individual jnds for combined formant changes. The format of this figure is the same as in figure 2. For the 100-Hz fundamental, the jnds for the peak-relations 1 and 2 are shown as square and circular symbols, respectively. For clarity, these symbols are connected with solid lines. For the 200-Hz fundamental, the jnds for these two peak relations are shown as triangular-up and triangular-down symbols, respectively. These symbols are connected with dashed lines. The error bar in the right top corner indicates the mean standard deviation of the individual results. The upper-most two rows display data for a relative change in the first formant that was twice that of the second formant. The lower-most two rows display data for equal relative changes in both formants.

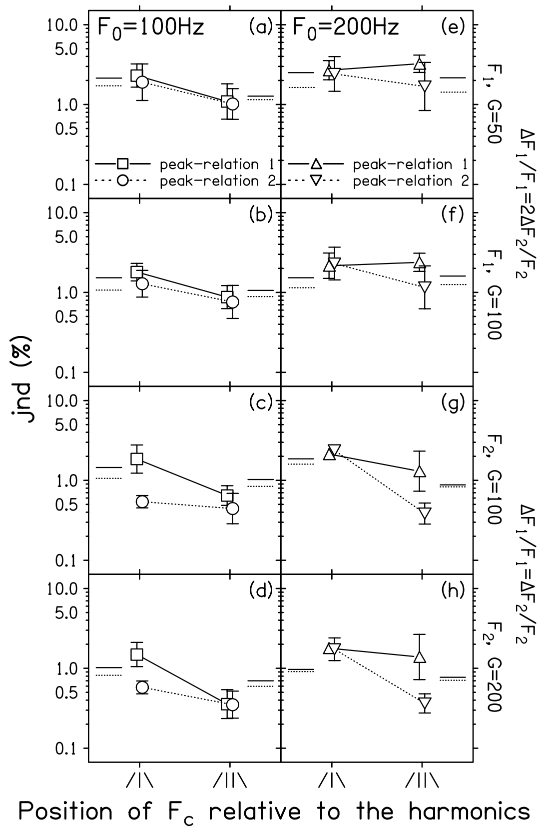

Figure 6: Average jnds for combined formant changes. The format of this figure is the same as that of figure 3, the meaning of the symbols is the same as in figure 5. The error bars indicate the standard deviations of the averages. The little solid and dotted line segments at the side of each panel indicate the expectations of the extended modified place model for the peak-relations 1 and 2, respectively.

were used in the discrimination process by subject JL, and possibly by subject JWH, but not by subject YW. In summary, when the first formant changes twice as fast as the second, the changes in the first formant dominate both the data and the model predictions, and we find good correspondence between the data and the predictions..

When the changes in both formants are equal, the jnds for the peak-relations 1 and 2 show a large difference under four conditions, indicating that the change in the first formant has influenced the discrimination. These differences are found for the 100-Hz fundamental when the center frequency coincides with a harmonic, and for the 200-Hz fundamental with a center frequency halfway between two harmonics. In the first case, the differences between the two peak relations are also significantly present in the individual results of figure 5 ($p < 0.01$). In the latter case, the relatively large jnds for peak-relation 1 bear inconsistencies between the subjects, and here the individual jnds for the two peak relations no not differ significantly. Under this condition, a difference between the individual jnds for the two peak relations is found in figure 5 for subject PP, but not for subject JL. Furthermore, due to differences in the individual jnds of the subjects RW and JWH, the difference between the average jnds of the peak-relations 1 and 2 appears to be exaggerated. These inconsistencies are reflected in the large standard deviations of the average jnds for peak-relation 1. Under the conditions where the jnds show differences between the two peak relations, the predictions of the extended modified place model show only small differences. For the remaining jnds, we find reasonably good correspondence between the jnds and the predictions of this model.

IV. General discussion

An important goal of these experiments was the comparison of jnds for single formant changes of two-formant vowels with those of single-formant stimuli (of Lyzenga and Horst, 1997). Figure 4 shows the quotients of the average jnds for the former and the latter stimuli. For the noise bands, these quotients show that adding a stationary second formant to a stimulus with a changing formant in the region of the first formant hardly influences jnds. Adding a stationary first formant to noise bands in the second formant region increases the jnds. For these stimuli, Lyzenga and Horst (1997) could not describe their jnds accurately with the modified place model, and assumed that a temporal discrimination mechanism was involved. For the two-formant stimuli, all the noise-band jnds show good correspondence with the predictions of the modified place model. This indicates that for these stimuli only spectral discrimination mechanisms are involved.

For the harmonic stimuli, adding a stationary second formant to a stimulus with a changing formant in the region of the first formant hardly influences the jnd. Adding a stationary first formant to a stimulus with a changing second formant increases the jnd under a number of conditions: for the Klatt phase, and for the 100-Hz fundamental combined with a center frequency that lies halfway between two harmonics. Under many of these conditions Lyzenga and Horst (1997) proposed a temporal mechanism to describe their jnds. For the first formant region they found no temporal mechanisms, and these jnds are hardly influenced by the addition of a stationary second formant. Furthermore, all jnds for the two-formant stimuli show good correspondence with the modified place model. So, in analogy with our findings for the noise bands, we conclude that the addition of a stationary first formant disturbs the temporal discrimination mechanisms in the region of the second formant, and thereby reduces the discrimination to spectral mechanisms only.

The relative levels of both formants (i.e. the attenuation of the second formants relative to the first) are equal for the stimuli with the Klatt-phase 1 and those with the random phase. However, in figure 4 we can see that the jnds for these two phase relations are influenced differently by the addition of a stationary first formant. Furthermore, the relative levels of both formants are different for the Klatt-phases 1 and 2 (i.e. for the peak-relations 1 and 2), and for these two conditions we find practically the same jnds in figure 3. This implies that, for sufficiently distant formants, the influence of the first formant on the second is independent of the relative levels of the formants. This is an indication that the observed transition from temporal to spectral formant-frequency discrimination occurs in central auditory processes.

A second important goal of this study was the comparison of jnds for combined formant changes with those for single formant changes. When considering the jnds for combined formant changes with a relative change in the first formant that is twice that of the second formant, we find that all jnds for peak-relation 1 are very close to, and those for peak-relation 2 are only slightly smaller than those for a singly-changed first formant (this can be seen by comparing the top halves of the figures 6 and 3). So, here the discrimination mechanism is dominated by the changes in the first formant (especially when the second formant coincides with a harmonic). These jnds are in good agreement with the extended modified place model. However, since the discrimination is dominated by the change in one formant, this condition does not produce a very thorough testing of the extended version of the model.

For equal relative formant changes, the changes in the second formants exert the largest influence upon the jnds, but there are several jnds with a clear influence of the changes in the first formants (as can be seen by comparing the bottom halves of the figures 6 and 3). The occurrences of these latter jnds show no correspondence with the relative levels of the two formants. (Under four conditions the jnds for the peak-relations 1 and 2 show a large difference; in contradiction to our expectations, the influence of the first formant is larger when this formant is smaller.) Most jnds for combined formant changes are in good agreement with the extended modified place model. Only the jnds for the 200-Hz fundamental with a center frequency halfway between two harmonics show a difference for the two peak relations, that is not displayed in the predictions of the model. Such a difference is only found in the individual results of some subjects, so, under this condition the extended modified place model does not apply to the results of all the subjects.

Using the summation rule for independently perceived variables, predictions for the jnds for combined formant changes can also be computed directly from the jnds for single formant changes. Under the assumption that the perceived difference is proportional to the frequency difference, an expectation for the jnd for combined formant changes can be derived from equation (1) as:

$$jnd_t = \sqrt{\frac{1}{\left(\omega_1/jnd_1\right)^2 + \left(\omega_2/jnd_2\right)^2}} \tag{2}$$

where ω_1 and ω_2 are weights that reflect the rates of change of the first and the second formants, respectively. With the proper weights, this relation can be used to calculate predictions in terms of both formants, and as well in the unit Hertz as in percent. When calculating predictions in terms of the first formant, ω_1 must be equal to unity, and for predictions in terms of the second formant ω_2 must equal unity. The weight of the remaining formant must be chosen equal to the rate of change of this formant divided by that of the formant in terms of which the predictions are expressed. For this the rates of change must be expressed in the same units as the jnds (in either Hertz or percent). So, for the formant-frequency relation $\Delta F_1/F_1 = 2\Delta F_2/F_2$ the weights are $\omega_1 = 1$ and $\omega_2 = 1/2$, because the predictions are calculated in terms of the first formant, and the relative rate of change of the second formant is half that of the first formant. For the relation $\Delta F_1/F_1 = \Delta F_2/F_2$ both weights are equal to unity.

The described weighting rules correspond with the behavior of equation (1). When the rates of change of both formants are equal, the excitation differences show a certain growth when the formant frequencies are increased. When the rate of change of the second formant is halved, the corresponding excitation difference will grow at half the original rate. Because of the square in equation (1), its effect in this equation will be reduced with a factor of four. This reduction is equal to the influence of a weight $\omega_2 = 1/2$ in equation (2).

The predictions calculated with equation (2) are shown in figure 7, along with the replotted jnds of figure 6. In general these predictions are close to those of the extended modified place model in figure 6. However, they approximate the data somewhat better, notably for the aforementioned deviations of the extended modified place model for the 200-Hz fundamental combined with a center frequency halfway between two harmonics. This indicates a cumulative increase of errors in the modified model when it is extended for the description of combined formant changes.

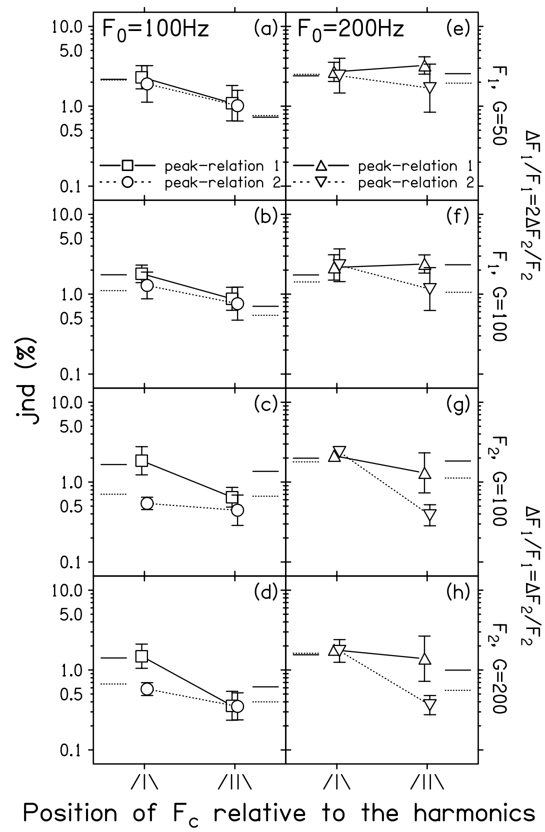

Figure 7: Average jnds for combined formant changes, replotted from figure 6. The format of this figure is the same as that of figure 6. The little solid and dotted line segments at the side of each panel indicate the results of combining the jnds for single formant changes with equation (2).

When the relative change in the first formant (at approximately 500 Hz) is twice that of a second formant (at approximately 2000 Hz), the absolute change of the second formant is twice that of the first formant. This relation is identical to the one used by Mermelstein (1978) for his measurements with combined formant changes. Because the center frequency of his second formant was six times that of the first formant (F_2 = 2100 Hz and F_1 = 350 Hz), the relative change in the first formant was three times as large as for the second formant. So, when comparing this condition with the relative rates of change in the present study, we expect that the changes in the first formant are dominant in his data. This effect can be observed in the averages of his jnds; for a single change in the first formant the average jnd is 49 Hz, and for a combined change it is 38 Hz (in terms of the first formant). For the second formant this decrease is much more dramatic; for a change in the second formant the average jnd is 172 Hz, compared with 76 Hz for a combined formant change (in terms of the second formant). The average of his predictions for combined formant changes in terms of the first formant is 42 Hz, compared to the average jnd of 49 Hz. So, both in his data and predictions the changes in the first formant dominate. Therefore his data do not produce conclusive testing of models for combined formant changes.

Mermelstein (1978) predicted his jnds for combined formant changes from those for single formant changes (see his Table II) by considering "F_1 and F_2 to be independent parameters that contribute information to discriminability given by $\sqrt{\omega_1(\Delta F_1)^2 + \omega_2(\Delta F_2)^2}$, where ω_1 and ω_2 are appropriate weighting factors". Using this relation, Hawks (1994) could not replicate Mermelsteins predictions exactly. We were able to reproduce Mermelsteins predictions (which are in Hertz and in terms of F_2) by using our equation (2) with the weights $\omega_1 = 1/2$ and $\omega_2 = 1$. In Mermelsteins experiments, the rate of change of the first formant was half that of the second formant. So, in disagreement with his formula, it appears that Mermelstein actually used a relation, equivalent to equation (2), in which the weights were squared as well as the independent parameters.

Hawks (1994) used identical relative changes for the first and the second formant, but he performed only a few measurements with a single change in the first formant. For his predictions, he used the formula of Mermelstein (1978) to predict jnds directly. In a first option he used the weights $\omega_1 = \Delta F_1/(\Delta F_1+\Delta F_2)$, and $\omega_2 = \Delta F_2/(\Delta F_1+\Delta F_2)$. In a second option he used the relative changes of the formants instead of the absolute changes in these relations: since these relative changes were equal, both ω_1 and ω_2 were equal to 1/2. In these calculations the jnds were expressed in Hertz, and the predictions were given in terms of the second formant. For his vowels AH-UH and EH-AE he measured jnds for parallel formant changes of 1.87% and 1.10%, respectively (in terms of F_2). With the first weighting option, his predictions for these two jnds were 2.24% and 1.02%, respectively. With the second option, these prediction were 1.94% and 0.89%, respectively. Using our equation (2) we find similar predictions: 1.75% and 0.94%, respectively. So, all predictions are close to each other, and from these data we cannot ascertain which model functions better.

When applying the model of Hawks (1994) to our data, we find better correspondence between the predictions and our data when using the weights of the second option. The sum of the squared relative errors between the predictions and the data is 129 for the first option and 68 for the second option (summed over 32 data points). However, the predictions calculated using our equation (2) approximate the data better than those of the model of Hawks for both weighting options. For this model the sum (over 32 data points) of the squared relative errors between the predictions and the data is 7.1. A comparison of the squared absolute errors between the predictions and the data provides the same conclusion. The most important abberations of Hawks' model are encountered when the jnds for the single formant changes are about equal. Under such conditions the predictions are equal to, instead of smaller than these jnds. This problem can be avoided when using equation (2).

The extensive data of Hawks' actual experiments show clearly that jnds for parallel combined changes are smaller than those for single formant changes. His jnds for opposite changes in the first two formants are practically equal to those for a single change of the second formant. So, for opposite combined formant changes the equations (1) and (2) do not apply, they are only valid for parallel formant changes.

V. Conclusions

For frequency discrimination of single formant changes we find no influence of adding a stationary second formant to a stimulus containing a changing first formant. For changing second formants, the addition of a stationary first formant increases the jnd under many conditions. The resulting jnds are in good agreement with the predictions of the modified place model. It can, therefore, be concluded that adding a stationary formant disturbs temporal discrimination processes, and limits the discrimination mainly to spectral processes. This transition from temporal to spectral formant-frequency discrimination appears to occur in central auditory processes.

Under many conditions, we found smaller jnds for combined formant changes than for the corresponding single formant changes. A good description of most of these data could be achieved with the modified place model, by extending this model with a summation rule for the individual excitation differences of the two changed formants.

Using the summation rule directly in equation (2), we could accurately predict the jnds for combined formant changes from those for single formant changes. With this equation we could also describe the data for parallel combined formant changes of Mermelstein (1978) and Hawks (1994). For his predictions, Mermelstein apparently used a relation equivalent to equation (2). To predict his data for combined formant changes, Hawks used a "weighted Euclidian distance model". The predictions calculated with equation (2) agree better with to our data than those from the model of Hawks.

Increasing the complexity of the stimuli appears to decreases the number of feasible mechanisms used for discrimination. For the more complex stimuli of our investigations, temporal discrimination mechanisms have disappeared and only spectral mechanisms, analogous to the modified place model, remain.

Acknowledgments

This research was supported by the Netherlands Organization for Scientific Research (NWO) through the Foundation for Behavioral Sciences (SGW), and the Heinsius Houbolt Fund.

References

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

Hawks, J.W. (1994). "Difference limens for formant patterns of vowel sounds," J. Acoust. Soc. Am. 95, 1074–1084.

Henning, G.B. (1966). "Frequency discrimination of random-amplitude tones," J. Acoust. Soc. Am. 39, 336–339.

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

Klatt, D. H. (1980). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. 67, 971–995.

Lyzenga, J., and Horst, J.W. (1995). "Frequency discrimination of bandlimited harmonic complexes related to vowel formants," J. Acoust. Soc. Am. 98, 1943–1955.

Lyzenga, J., and Horst, J.W. (1997). "Frequency discrimination of stylized synthetic vowels with a single formant," J. Acoust. Soc. Am. (accepted).

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