Discrimination of fundamental frequency of synthesized vowel sounds in a noise background

Discrimination of fundamental frequency of synthesized vowel sounds in a noise background

Michael T. M. Scheffers[a)]
Institute for Perception Research, P. O. Box 513, 5600 MB Eindhoven, The Netherlands

(Received 11 November 1982; accepted for publication 11 April 1984)

An experiment was carried out, investigating the relationship between the just noticeable difference of fundamental frequency (jnd_{f_0}) of three stationary synthesized vowel sounds in noise and the signal-to-noise ratio. To this end the S/N ratios were measured at which listeners could just discriminate a series of changes in f_0 in the range from 10% to 0.5%. Similar measurements were obtained for pulse trains and for pure tones as a reference for the results. A measure of S/N ratio based on an approximation of the critical bandwidth appeared to provide a fairly good predictor of the masked threshold of each signal, measured in a second experiment. Using this measure, it was found that a given change in the fundamental of a pulse train could be discriminated at a lower S/N ratio than in a pure tone with a frequency equal to that fundamental. The results for the vowel sounds were found to be in between those for a low-frequency pure tone and those for a pulse train. Owing to the signal-generation method (viz., changing f_0 by changing the sampling frequency), three cues could in principle be used to discriminate a change in the fundamental of a vowel: A change in the residue pitch, a change in the pitch of a single prominent harmonic, or a change in the spectral envelope of the signal. It can be inferred from the results that the subjects used that particular cue which yielded best performance. Which cue was optimal depended not only on the vowel but also on f_0 and on the presented change in f_0. It seems, however, that the pitch of a single harmonic was the cue most often used. Another interesting result is that changes in f_0 greater than about 5%, could for each signal be discriminated when the signal was just above masked threshold.

PACS numbers: 43.66.Fe, 43.66.Hg, 43.66.Dc, 43.70.Gr, 43.70.Ve [RDS]

INTRODUCTION

This paper describes an experiment investigating the relationship between the just noticeable difference of fundamental frequency (jnd_{f_0}) of vowel sounds in noise and the signal-to-noise ratio. The experiment was carried out as part of a research project investigating the role of pitch in the perceptual separation of speech sounds from an interfering background. It was carried out primarily to obtain data for the evaluation of a model that simulates human perception of *residue pitch* [1] of speech sounds (Scheffers, 1983). The experimental results, however, seem interesting enough to be described separately.

To the best of my knowledge there have been no reports of experiments on the jnd_{f_0} of vowels in noise. Data from experiments on the jnd_{f_0} of filtered pulse trains as a function of signal-to-noise ratio, such as reported by Hoekstra (1979) and Horst (1982), are not fully adequate for predicting the jnd_{f_0} of vowels. Hoekstra used signals with a trapezoidal spectral envelope, while the formants of vowels have roughly a triangular shape. In case the formant slopes are steep, the harmonic nearest to the center of the formant will mask some of the higher harmonics. At low S/N ratios in particular, the number of aurally detectable harmonics will thus be considerably smaller than for a signal with a trapezoidal, envelope. This will probably result in a larger jnd_{f_0}. Signals with a formantlike spectral envelope were used by Horst (1982). He reported that the jnd_{f_0} of these signals depended not only on the center frequency of the formant but also on its slopes and on the position of harmonics in the envelope. In vowels, different formant slopes will be found and the position of harmonics with respect to the center frequency of the formant will vary from formant to formant and from vowel to vowel. Moreover, if some noise is added to a vowel, harmonics in different formants may be above masked threshold. The combined influence of these harmonics on the jnd_{f_0} is difficult to predict. In an experiment by Flanagan and Saslow (1958) it appeared that even when no noise is present, an effect of the vowel spectrum can be found. They reported the jnd_{f_0} for vowel sounds with a sharp first formant at low frequencies, like the vowels /i/ and /u/, to be greater than for vowels that have a broader region of high-level harmonics, like /ɑ/ and /æ/. They assumed that a sharp low-frequency formant would mask more of the harmonics that convey information on the change in f_0 than would a shallower formant.

In one experiment reported here, the relationship between the signal-to-noise ratio and the jnd_{f_0} was investigated for three stationary synthesized vowel sounds, viz., the Dutch /i/, /ɑ/, and /u/ at three different f_0's. In the same experiment, the jnd_f for pure tones and pulse trains in noise was also investigated to provide a reference for the results. In a second experiment, masked thresholds were determined for the signals of the first experiment.

[a)] Present address: Nederlandse Philips Bedrijven BV, CAB Elcoma, Building BE-5, P.O. Box 218, 5600 MD Eindhoven, The Netherlands.

 0001-4966/84/080428-07$00.80

I. METHOD

Subjects were presented with two successive vowel sounds which differed in fundamental frequency. The vowels were presented in a background of pink noise that was turned on 300 ms before the onset of the first vowel and was turned off 300 ms after the offset of the second one. Vowel duration equaled 200 ms. The two vowels were separated by an interval of 600 ms. In a 2AFC task the subjects were asked to indicate which of the two sounds had a higher pitch. They received feedback after each response. Response time was not limited. The f_0 difference was fixed for each run. The noise level was kept constant while the level of the vowel sounds could be varied. The adaptive attenuation procedure PEST (Taylor and Creelman, 1967) was used to determine the signal-to-noise (S/N) ratio at which the subject gave 75% correct responses for a given difference in fundamental frequency.

The S/N ratios were determined for f_0 differences of 0.5%, 1%, 2%, 5%, and 10% at three f_0 frequencies viz., 75, 150, and 300 Hz. The values for the f_0 differences were chosen on the basis of data on the jnd_f of pure tones and on the jnd_{f_0} of complex harmonic sounds in isolation and in noise (Ritsma, 1963; Henning, 1967; Hoekstra, 1979). A pure tone and a pulse train were included in the signal set for general comparison. Three frequencies were used for the pure tone, viz., 150, 300, and 1000 Hz. The pulse trains had a flat spectrum up to about 3.5 kHz (owing to low-pass filtering at 4 kHz) and fundamental frequencies equal to the ones used for the vowel sounds.

Masked thresholds of all signals were measured using a similar experimental setup. A stimulus in this experiment consisted of two noise intervals of 800 ms each, separated by a silent interval of 200 ms. The signal was presented in random order in one of the two intervals, starting 300 ms after noise onset. Signal duration again equaled 200 ms. The subjects were asked to indicate in which interval the signal had occurred. They received feedback after each response.

A. Stimuli

Before describing the stimuli, it may be helpful to define the following two notions: The term *signal* is used for the target sound (the vowels, the pure tones, and the pulse trains). The term *stimulus* refers to the stimulus complex, i.e., the noise plus the two signals for the jnd experiment, and the two noise intervals plus the signal for the experiment in which masked thresholds were measured.

The signals were constructed in the following way. The sampled waveform of one period of each signal was calculated for each of the three fundamental frequencies, using a minicomputer. A software serial five-formant speech synthesizer (Vogten and Willems, 1977) was used for calculating the waveforms of the vowels. The pulse trains consisted of a repetition of unipolar pulses with a width of about 0.1 ms (one sample of the sampling frequency). The time samples had a resolution of 12 bits. Sampled waveforms were stored on a magnetic disk. The signals were generated during the experiment by D/A converting a continuous repetition of this single period. Signal level could be varied in steps of 1 dB by use of a digitally controlled attenuator. The signals were low-pass filtered at 4 kHz (off-band attenuation rate 24 dB/oct).

Two methods can be chosen to obtain variations in the fundamental frequency: (1) The period of the fundamental—and thereby the frequencies of the harmonics—can be varied while keeping the spectral envelope of the sound constant, or (2) the sampling frequency can be varied thereby varying not only the frequencies of the fundamental and of the harmonics, but also the spectral envelope. In the first method, not only the frequencies of the harmonics will vary with the change in f_0, but also their amplitudes. This can give extra indications for the presented f_0 difference. In the second one, the shift of the spectral envelope will give an extra cue. Variation of the sampling frequency was chosen in the present experiment for practical reasons. Either the first or the second signal was generated in random order at a sampling frequency of 9 kHz, while the other one was generated at a higher sampling frequency. The sampling frequency could be varied in steps of 1 Hz, thus providing a minimum variation of f_0 of 0.011%.

Three vowel sounds were used, viz., the Dutch /u/, /i/, and /ɑ/. They represent three typical formant configurations: low F_1 and F_2; low F_1, high F_2; and medium F_1 and F_2, respectively. Formant frequencies and bandwidths were adapted from a study of Dutch vowel sounds by Govaerts (1974). Figure 1 shows line spectra of the three vowel sounds generated at an f_0 of 75 Hz. The spectral envelope of each vowel sound was the same for the three reference fundamental frequencies used. Thus the spectra for the 150-Hz fundamental can be obtained by omitting the odd harmonics of the 75-Hz spectra. Those for the 300-Hz fundamental can be obtained by omitting the odd harmonics of the 150-Hz spectra. The latter spectra are indicated in Fig. 1 by dashed lines for the 300-Hz harmonics.

Pink noise (3 dB/oct attenuation) was used as a masker because it will have a rather flat response in the peripheral auditory system. It was bandpass filtered from 60 Hz to 6 kHz (off-band attenuation rate 40 dB/oct) and presented at a spectral level of 30 dB SPL/Hz at 1000 Hz. After gating, the signals and the masker were added in an analog signal mixer. Onset and offset ramps of signals and masker were Gaussian-shaped and had durations of 20 ms.

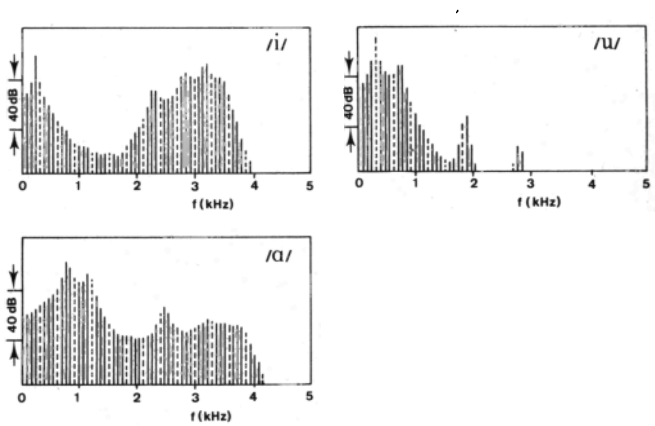

FIG. 1. Power spectra of the synthesized vowel sounds /i/, /u/, and /ɑ/ with a fundamental frequency of 75 Hz. The 300-Hz harmonics are dashed to indicate the spectral shape for that fundamental.

B. Subjects

Four male subjects with normal hearing took part in both experiments. Their ages ranged from 24 to 30 years. They were experienced in psychoacoustic experiments, but were not trained in pitch discrimination. One of the subjects attended the experiments three times in order to obtain an indication whether learning effects could occur during the experiment. The subjects were tested individually. They were seated in an Amplifon sound-attenuating booth. The stimuli were presented diotically through TDH 49-P headphones.

C. Specification of signal-to-noise ratio

Signal level, the dependent variable in the experiments, has not yet been defined because a nonstandard method for calculating signal-to-noise ratios was chosen. It was mentioned before that a typical characteristic of vowel spectra is that some parts contain a lot of energy (the formants) and others only a little. It may be assumed that, generally speaking, the strongest signal component will determine the masked threshold of the sound and that the harmonics that are above masked threshold, contribute to residue pitch perception. I therefore decided not to calculate the signal-to-noise ratio on the basis of the overall dB SPL level or the energy of the signal, for example, but on the basis of the "best" S/N ratio in a critical band (CB, cf. Zwicker and Feldtkeller, 1967). A 10% band was chosen as a practical approximation of the CB. The following method was used. The contiguous 10% band spectrum was measured for each signal, and the highest level found in this spectrum was noted. Using pink noise as a masker, the 10% band noise level is independent of frequency. The S/N ratio of a signal will be expressed as an S/N_{10} ratio, viz., the S/N ratio measured in a 10% band, centered at the frequency at which the highest 10% band signal level was found. These frequencies are given in Table I for the vowels and the pulse trains. Note that they can change considerably for a vowel when it is generated at a different fundamental frequency.

Table II gives the overall (20 Hz to 20 kHz) dB SPL levels of the vowels and the pulse train for the stimuli with an S/N_{10} of 0 dB. The pure tones had a level of 50 dB SPL for an S/N_{10} of 0 dB. Note that these are the levels of the signals before the noise was added. The overall noise level was 67 dB SPL.

Let us assume that a signal can just be detected when its onset increases the output of at least one of the auditory filters by 1 dB (about the DL of intensity) with respect to the steady-state response to the noise alone. Masked threshold of each of the signals should then be found at about an S/N_{10} of -6 dB [$10 \log(10^{1\,\mathrm{dB}/10} - 10^{0}) = -6$ dB].

TABLE I. Frequencies for the highest level in a 10% band spectrum of the stimulus signals.

f_0(Hz)	/u/	/i/	/ɑ/	Pulse train
75	300 Hz	225 Hz	750 Hz	3350 Hz
150	300 Hz	3225 Hz	750 Hz	3375 Hz
300	300 Hz	3300 Hz	900 Hz	3450 Hz

TABLE II. Overall dB SPL levels of the signals in the stimuli with an S/N_{10} of 0 dB.

f_0 (Hz)	/u/	/i/	/ɑ/	Pulse train
75	51 dB	54 dB	54 dB	60 dB
150	51 dB	54 dB	52 dB	60 dB
300	50 dB	53 dB	54 dB	60 dB

D. Terminating criteria for PEST

The implementation of the PEST procedure used differed on some points from the one given by Taylor and Creelman (1967). In our version (Zelle, 1978) a run is terminated either when the level adjustment step would have been reduced to 0.5 dB, or when the signal level has not been changed during 20 consecutive trials (percentage correct is then between 70% and 80%). In about 8% of runs, however, more than 100 trials would be needed to reach one of these criteria. A reasonable estimate of the 75% correct point can often be made in these instances on the basis of the responses to these 100 trials. A run was therefore terminated after 100 trials because longer runs tend to have a negative effect on the subject's performance. For estimating the 75% correct point the results at levels where less than ten presentations had been made were skipped. The remaining set was inspected to check that the percentage correct responses decreased with decreasing S/N ratio. If this was the case, a 75% point was interpolated. If not, then the subject was asked to do that particular run again, but *only once*. Because fixed f_0 differences were used, it was possible that a subject could not discriminate some given difference. This happened for 17 of the 90 stimuli with the smallest f_0 difference of 0.5%.

II. RESULTS

The results are presented in Fig. 2. The masked thresholds of the sounds, measured in the second experiment, are indicated near the right-hand side of each graph. No significant difference was found in the series of results for the subject who did the experiment three times. Also, his results did not differ significantly from those for the other three subjects, nor were significant interindividual differences found. The results were therefore averaged over all subjects. For a given f_0 difference, the standard deviation of each measuring point thus obtained appeared to vary only slightly with the reference f_0 (viz., 75, 150, and 300 Hz). In general, however, the standard deviations tended to increase with decreasing f_0 differences. An indication of the standard deviation, averaged over those for the three fundamentals, is given by the vertical bars on the lower horizontal axis of each graph. I have mentioned that in some cases a subject could not accomplish the task. When a point in a graph is the average of less than six measurements, a number near that point will indicate on how many measurements it is based.

It can be observed from Fig. 2 that the results for each signal follow a hyperboliclike curve. The horizontal asymptote of this curve is formed by the masked threshold of the signal, and the vertical asymptote—presumably—by the jnd_{f_0} of the signal in quiet. It can furthermore be observed that the curves for different fundamentals seem to be dis-

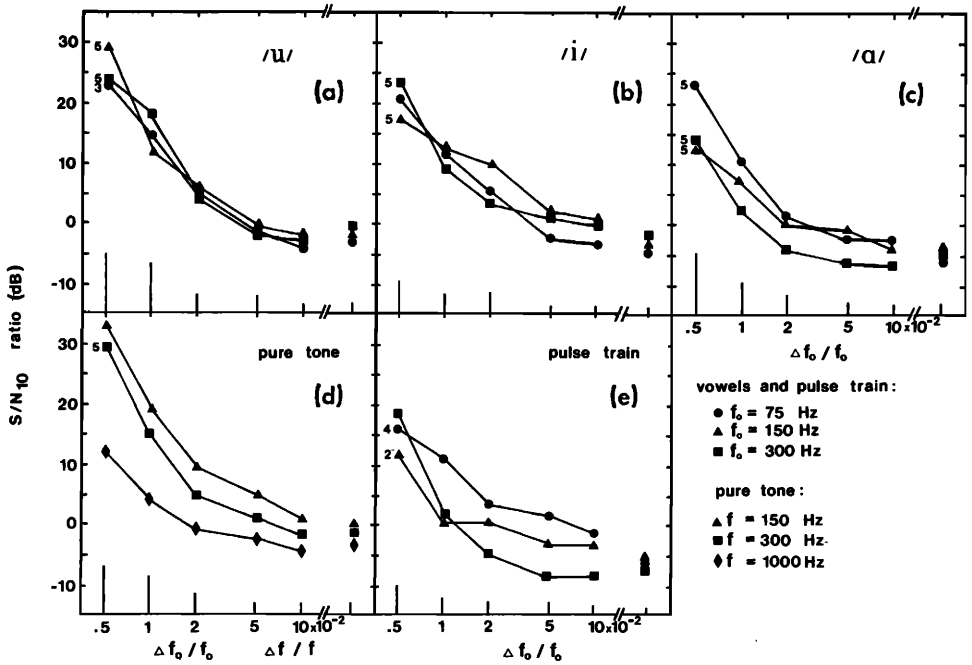

FIG. 2. Signal-to-noise ratios for 75% correct discrimination[2] of differences in the fundamental frequency of synthesized vowel sounds and pulse trains, and in the frequency of pure tones. The S/N_{10} ratio is the signal-to-noise ratio in a 10% band, centered at the frequency where the highest 10% band level of the signal is measured (see text). Vertical bars on the horizontal axis indicate one standard deviation (averaged over f_0). A number near a data point indicates on how many measurements the point is based if this number does not equal six. Masked thresholds of the signals are indicated by the isolated symbols near the right-hand vertical axis of each graph.

junct for the pure tone and the pulse train, but not for the vowels /i/ and /u/. The p values found in a sign test on the significance of the differences between the results for each signal at different fundamentals are given in Table III. It was investigated in this test whether the discrimination thresholds decreased for increasing fundamental frequency. Since the reverse was found for vowel /i/ (indicated by an asterisk in Table III), p values for a two-sided test are given.

Table III shows that the curves for the pure tone and for the pulse train are each significantly different. The results for vowel /u/ appear to be independent of the fundamentals that were used in the experiment. Furthermore, it appears that the discrimination thresholds of vowel /i/ are lower for the 75-Hz fundamental than for the 150- and 300-Hz fundamental. The curves for the latter two fundamentals are not significantly different. The curves for vowel /ɑ/ with a fundamental of 75 and 150 Hz are not significantly different, while the curve for the 300-Hz fundamental lies at lower S/N_{10} ratios than those for the other two fundamentals.

The masked thresholds of each signal are found to be independent of the fundamental of the signal in almost all cases. They differ little for different signals but tend to be lower for signals the energy of which is concentrated in a small band (the pure tones and vowel /u/) than for broadband signals (vowel /ɑ/ and, especially, the pulse train).

TABLE III. Results of a two-sided sign test on the difference between the curves for each signal in Fig. 2. The hypothesis was tested that the discrimination thresholds for a given f_0 difference decreased for increasing fundamental frequency. If the reverse was found, this is indicated by an asterisk.

	/u/	/i/	/ɑ/	Pulse train
75–150 Hz	n.s.	$p<0.0002$*	n.s.	$p<0.0001$
150–300 Hz	n.s.	n.s.	$p<0.001$	$p<0.0001$
75–300 Hz	n.s.	$p<0.01$*	$p<0.005$	$p<0.0001$

	Pure tone
150–300 Hz	$p<0.006$
300–1000 Hz	$p<0.0001$
150–1000 Hz	$p<0.0001$

III. DISCUSSION

The signal-to-noise ratios were measured for 75% correct discrimination of a series of differences in fundamental frequency of synthesized vowels in noise. As a reference for the results, the same was done for pure tones and pulse trains. The results for the latter two signals will be discussed first.

A. Pure tones

The curves for the pure tones tend to be steeper for tones with lower frequencies, suggesting that the vertical as-

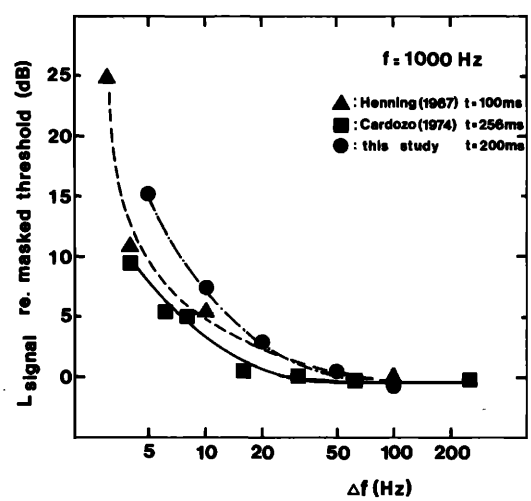

FIG. 3. Just noticeable difference of frequency of a pure tone of 1000 Hz as a function of signal-to-noise ratio. Results are plotted from Henning (1967), Cardozo (1974), and from the present study. S/N ratios are relative to masked threshold. The curves are fitted through the data points by eye.

ymptote is reached at higher df_0/f_0 values for lower frequencies. On the basis of data on the jnd_f of pure tones in quiet (e.g., Harris, 1952) my view is that the noise has hardly any effect on the jnd_f for S/N_{10} ratios greater than about 35 dB. The overall shift of the curves towards higher S/N_{10} ratios for lower frequencies will be discussed in Sec. III D.

The results for the pure tones of 1000 Hz can be compared with data from Henning (1967) and Cardozo (1974). The results of their experiments are plotted in Fig. 3 together with the present data. The jnd_f values of Cardozo's study are averaged over the three subjects in that experiment.

We can see from this figure that the present data correspond quite well to those of Henning and Cardozo although they seem to follow a slightly steeper curve. This difference (if significant) is probably related to the degree of training of the subjects. Both Henning and Cardozo used highly trained subjects while untrained listeners participated in the present experiment.

B. Pulse trains

A comparison between the results for the pulse train with a fundamental frequency of 150 and 300 Hz and those for a pure tone with corresponding frequencies, is given in Fig. 4.

Figure 4 shows that the discrimination threshold of an f_0 difference was lower for a pulse train than for a pure tone with a frequency corresponding to the fundamental of that pulse train. The differences between the curves for the pulse train and the pure tone appeared to be highly significant ($p < 0.0001$ for both the 150- and the 300-Hz curves). Similar results have been reported by, e.g., Henning and Grosberg (1968) and Walliser (1969) for filtered pulse trains in quiet. It was argued in both studies that the jnd_{f_0} depends on the spectral composition of the stimulus and that it corresponds to the smallest change in f_0 that results in a perceptual change of any partial. Another explanation is given in Goldstein's theory of pitch perception (Goldstein, 1973). According to his theory, the jnd_{f_0} depends on the accuracy with which information about the frequencies of harmonics that

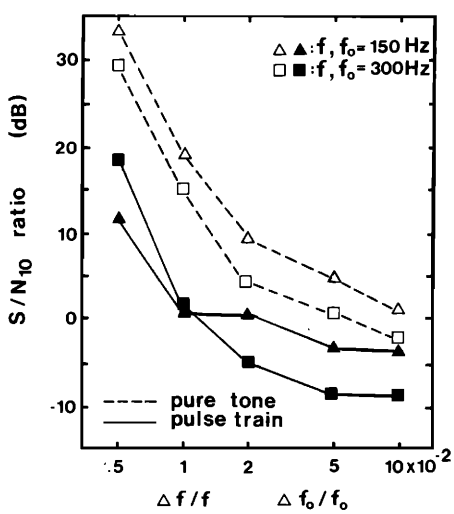

FIG. 4. Comparison between the jnds of a pure tone with frequencies of 150 and 300 Hz (dashed lines) and the jnds of pulse trains with corresponding fundamental frequencies (solid lines).

are resolved in the auditory frequency analysis is conveyed to a central processor that extracts the pitch [Goldstein, 1973, Eq. (15)]. Depending on the number of resolvable harmonics and their frequencies, the estimated jnd_{f_0} of a harmonic complex can be considerably smaller than the jnd_f of the fundamental.

The importance of aurally resolvable components can also be seen in the present results for a pulse train. Whereas the S/N_{10} ratios for masked threshold are found to be independent of f_0, the S/N_{10} ratios that were required by the listeners to perceive f_0 differences greater than 2% increased with decreasing f_0. The highest 10% band level for the pulse trains was measured at about 3400 Hz. For the fundamental frequencies of 75 and 150 Hz, however, the harmonics in that region cannot be resolved by the auditory system (cf. Plomp, 1964). The curves could be plotted on the basis of the S/N_{10} ratios of aurally resolvable harmonics for these signals (harmonic number less than eight). The curve for the 75-Hz fundamental in Fig. 2(e) would then be shifted downwards by 6 dB, that for the 150-Hz fundamental by 3 dB. In that case the curves become essentially equal for f_0 differences greater than 2%. However, no clear effect of the fundamental frequency on the jnd_{f_0} can then be observed for the smaller f_0 differences either. The Goldstein theory predicts a dependency of the jnd_{f_0} on f_0 with a factor $\sqrt{2}$/oct for pulse trains in quiet [Goldstein, 1973, Eq. (15)]. Such an effect would be obscured by the variance in the measurements.

C. Vowel sounds

Three cues could in principle be used by the subjects to discriminate a difference in fundamental frequency of the vowel sounds: (1) a difference in the residue pitch of the sound; (2) a difference in the pitch of a single strong harmonic; and (3) a change in the spectral envelope of the sound. Assuming that residue pitch had been the discrimination cue for the pulse trains, a comparison between the results for the vowels and those for the pure tones and pulse trains can reveal which of the first two cues was probably used.

A general comparison is given in Fig. 5 of the curve for the 300-Hz pure tone, the results for the vowel sounds (each averaged over f_0), and the curves of the pulse train (averaged over f_0 after correcting the data for equal S/N_{10} ratio of aurally resolvable harmonics (see Sec. III B). Figure 5 shows that the results for the vowel sounds are roughly in between those for a low-frequency pure tone and those for a pulse train.

The curves for the vowel /u/ in Fig. 2(a) were found to be independent of f_0. None of these curves was found to differ significantly from the curve of the 300-Hz pure tone. From the spectrum of this vowel (see Fig. 1) we see that the 300-Hz component is by far the strongest harmonic. It is also present for all three f_0 values used. At low S/N ratios all other components will be masked by the noise, and only the pitch of this component will be perceived. The results indicate that at higher S/N ratios as well, discrimination was performed on the basis of the 300-Hz component. This would mean that the pitch of this component was a better cue for discrimination than the residue pitch.

A significant difference was found for vowel /i/

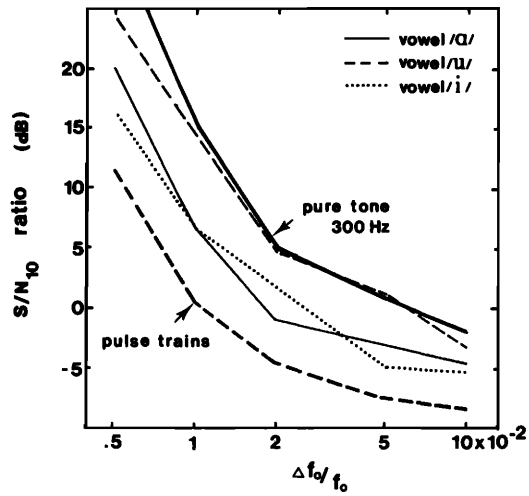

FIG. 5. Comparison between the results for a pure tone of 300 Hz (bold line), the results for the three vowels averaged over f_0. (/ɑ/: thin solid line; /u/: thin dashed line; /i/: thin dotted line), and the results for the pulse train (bold dashed line). The results for the pulse train are corrected for equal S/N_{10} ratio of aurally resolvable harmonics before averaging over f_0 (see text).

between the results for the 75-Hz fundamental and those for the 150- and 300-Hz fundamentals. Because no differences were found between the results for the 150- and 300-Hz fundamental, it may be assumed that another discrimination cue was used for those two fundamentals than for the 75-Hz f_0. The 225-Hz harmonic is the strongest component in the spectrum of vowel /i/ generated at a fundamental of 75 Hz. A similar curve as for vowel /u/ can therefore be expected. Figure 2(b) indeed shows that the curve for vowel /i/ at 75 Hz comes close to that for a low-frequency pure tone. No significant difference was in fact found between that curve for vowel /i/ and the average of the curves for the pure tones of 150 and 300 Hz. For vowel /i/ generated at fundamentals of 150 and 300 Hz, the highest 10% band level is found near the fourth formant (3200 Hz). For the 150-Hz fundamental, however, harmonics in that region are outside the *existence region* for residue pitch (Ritsma, 1962, 1963). For the 300-Hz fundamental, only a few harmonics around 3 kHz will be unmasked at low S/N ratios (see Fig. 1). Experiments by Houtsma and Goldstein (1972) have shown that such few harmonics with relatively high harmonic numbers do not evoke a clear residue pitch percept either. It is therefore improbable that the subjects used residue pitch as a cue for discrimination in these cases. On the basis of the results reported by Horst (1982) I assume that they used the shift in the spectral envelope of these high harmonics instead. This might, however, only be possible for f_0 differences greater than 2%. Flanagan (1955) also found values of 3% to 5% for the jnd of formant frequency. Whether or not the subjects used residue pitch as a cue for discriminating smaller f_0 differences could not be determined because of the relatively small number of measurements in the present experiments and the great variance in those results.

No significant difference was found between the results for vowel /ɑ/, generated at f_0's of 75 and 150 Hz. Figure 2(c) gives the impression that the curves are mainly the same below 5-dB S/N_{10}. For both fundamentals, the 750-Hz harmonic in the first formant is the strongest spectral component. The results in question correspond well with those for the 1000-Hz pure tone. The slight difference between the results for the 75- and 150-Hz fundamental above a S/N_{10} ratio of 5 dB in Fig. 2(c) might indicate a shift to the use of residue pitch as a cue. The spectrum of vowel /ɑ/ contains a series of strong harmonics with about equal level in the range from 600 to 1200 Hz for the f_0 of 150 Hz (see Fig. 1). The residue pitch evoked by this set of harmonics might be a good cue for discrimination. The results for vowel /ɑ/ with a fundamental frequency of 300 Hz, correspond well to those for the pulse train generated at an f_0 of 300 Hz. The spectral composition for this fundamental also leads me to assume that the subjects used residue pitch as a cue for discriminating f_0 differences.

The interpretation of the results is supported by reports of the subjects. An independent listener, highly trained in analytical listening and in naming musical intervals, was asked to describe the sound characteristics that he perceived in the vowels presented at various S/N_{10} ratios. He confirmed the listeners reports, but also reported being able to hear residue pitch down to lower S/N_{10} ratios than those at which the data suggest that residue pitch was used as a cue for discrimination. This leads me to assume that even when the subjects could perceive residue pitch, they used another cue if that resulted in better performance.

D. Masked threshold

In general, masked threshold for a signal appears to be almost independent of f_0 when expressed in S/N_{10} ratio. For the pure tones, however, masked thresholds were found to systematically decrease with increasing frequency. A similar effect can be found in the masked thresholds reported by Fletcher (1953). It can also be observed from Fig. 2 that the masked thresholds of signals with predominant energy at low frequencies such as vowel /u/—and to some extent vowel /i/—are higher than those of signals with energy concentrated in a region at higher frequencies. This is probably caused by the fact that the approximation of the CB by a 10% band filter does not hold at low frequencies. Although there is recent evidence that the bandwidth of the auditory filter still decreases below 500 Hz, it has also been found that the efficiency of the detection mechanism following the filter, also decreases below that frequency (Patterson *et al.*, 1982; Fidell *et al.*, 1983; see also Moore and Glasberg, 1983). A constant bandwidth for frequencies below 500 Hz would therefore be a better approximation for predicting masked thresholds. The present data furthermore suggest that masked threshold decreases when the region of near-equal-level harmonics of a signal is broader. Taking these considerations into account, measured masked thresholds corresponded fairly well with the theoretically predicted value of − 6 dB (see Sec. I C).

IV. CONCLUSIONS

In the present experiments, subjects could in principle use three cues to discriminate differences in the fundamental frequency of vowel sounds, viz., a difference in the residue pitch of the sound, a difference in the pitch of a single har-

monic, or a shift in the spectral envelope of the signal. The conclusion may be drawn that they used that cue which yielded the lowest discrimination threshold for a given f_0 difference.

The pitch of a single strong harmonic was probably most often used for the vowels. If the level of this harmonic is much higher (more than, e.g., 10 dB) than that of all other signal components, it will be the only unmasked component at low S/N ratios. This cue can still yield lowest discrimination thresholds at higher S/N ratios. The use of this cue was most apparent for the vowel /u/, but could also be inferred from some of the results for the vowels /i/ and /ɑ/.

Owing to the experimental setup (changing the sampling frequency to obtain differences in f_0), a difference in fundamental frequency could result in a detectable change in the spectral envelope of a vowel, especially when the vowel spectrum contained a strong high-frequency formant. This cue was apparently used by the subjects for the vowel /i/ generated at f_0's of 150 and 300 Hz. It can probably only be used to discriminate f_0 differences greater than 2%.

Residue pitch can be used to discriminate f_0 differences when a number of harmonics within the existence region are above masked threshold, such as for the pulse train. This cue was probably also used for the vowel /ɑ/, generated at a fundamental of 150 and 300 Hz. For those two fundamentals the vowel spectrum contained a series of harmonics with about equal level within the *existence region*. The results for those signals indeed correspond well with the results for the pulse train.

In general it can be concluded that the jnd_{f_0} of periodic sounds in noise depends on the number of aurally resolvable harmonics that are above masked threshold. It can be greater or smaller than the jnd_f of a pure tone with a frequency corresponding to f_0 at the same S/N_{10} ratio. It furthermore appears that differences of more than 5% in the fundamental frequency of a periodic sound can, in general, be discriminated when the sound is just above masked threshold.

Masked thresholds of quite different sounds can, to a first approximation, be fairly well predicted on the basis of a 10% band spectrum.

ACKNOWLEDGMENTS

The author wishes to thank H. Duifhuis, S. G. Nooteboom, and A. J. M. Houtsma for the stimulating discussions and for their assistance in the preparation of this paper. He also thanks H. F. Arnoldus, A. W. Bezemer, H. W. Zelle, and J. 't Hart for their careful listening. This research was supported by the Netherlands Organization for the Advancement of Pure Research (Z.W.O.), through the Netherlands Foundation for Psychonomics, Grant 15-31-011.

[1]The term "residue pitch" will be used in this paper for the pitch corresponding to the fundamental of harmonic sounds in contrast to the pitch evoked by a single signal component. This first pitch is known by a number of names such as residue pitch, periodicity pitch, virtual pitch, low pitch, etc. Underlying most of these terms is a certain theory of pitch perception. It is not my intention to discuss here the validity of these theories for the present results. For this reason the rather historical term residue pitch (Schouten, 1940) has been chosen.

[2]Reliable percentages correct of about 75% were sometimes found around two S/N ratios, which could be as much as 20 dB apart, while between these values the percentages varied wildly. This occurred especially for the smallest f_0 difference of 0.5%. Subjects often reported for these stimuli that after PEST had lowered the S/N ratio below discrimination threshold, and was increasing the signal level again, they had lost the cue on which they had been discriminating. They claimed to hear a difference between the two signals but could not classify it as a higher or lower pitch. Only at a far higher S/N ratio could they get hold of the cue again. The lowest S/N ratio was taken in such cases.

Cardozo, B. L. (1974). "Frequency discrimination at the threshold," in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer, New York), pp. 164–177.

Fidell, S., Horonjeff, R., Teffeteller, S., and Green, D. M. (1983). "Effective masking bandwidths at low frequencies," J. Acoust. Soc. Am. **73**, 628–638.

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

Flanagan, J. L., and Saslow, M. G. (1958). "Pitch discrimination for synthetic vowels," J. Acoust. Soc. Am. **30**, 435–442.

Fletcher, H. (1953). *Speech and Hearing in Communication* (Van Nostrand, New York).

Goldstein, J. L. (1973). "An optimal processor theory for the central formation of pitch of complex tones," J. Acoust. Soc. Am. **54**, 1496–1516.

Govaerts, G. (1974). "Psychologische en fysische structuren van perceptueel geselecteerde klinkers, een onderzoek aan de hand van Zuidnederlandse klinkers," Doctoral thesis, University of Louvain (in Dutch).

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

Henning, G. B. (1967). "Frequency discrimination in noise," J. Acoust. Soc. Am. **41**, 774–777.

Henning, G. B., and Grosberg, S. L. (1968). "Effect of harmonic components on frequency discrimination," J. Acoust. Soc. Am. **44**, 1386–1389.

Hoekstra, A. (1979). "Frequency discrimination and frequency analysis in hearing, a psychophysical study of some aspects of the normal and abnormal auditory system," Doctoral thesis, University of Groningen.

Horst, J. W. (1982). "Discrimination of complex signals in hearing," Doctoral thesis, University of Groningen.

Houtsma, A. J. M., and Goldstein, J. L. (1972). "The central origin of the pitch of complex tones: Evidence from musical interval recognition," J. Acoust. Soc. Am. **51**, 520–529.

Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.

Patterson, R. D., Nimmo-Smith, I., Weber, D. L., and Milroy, R. (1982). "The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold," J. Acoust. Soc. Am. **72**, 1788–1803.

Plomp, R. (1964). "The ear as a frequency analyzer," J. Acoust. Soc. Am. **36**, 1628–1636.

Ritsma, R. J. (1962). "Existence region of the tonal residue I," J. Acoust. Soc. Am. **34**, 1224–1229.

Ritsma, R. J. (1963). "Existence region of the tonal residue II," J. Acoust. Soc. Am. **35**, 1241–1245.

Scheffers, M. T. M. (1983). "Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter," J. Acoust. Soc. Am. **74**, 1716–1725.

Schouten, J. F. (1940). "The residue, a new component in subjective sound analysis," Proc. K. Ned. Akad. Wet. **43**, 356–365.

Taylor, M. M., and Creelman, C. D. (1967). "PEST: Efficient estimates on probability functions," J. Acoust. Soc. Am. **41**, 782–787.

Vogten, L. L. M., and Willems, L. F. (1977). "The Formator: a speech analysis–synthesis system based on formant extraction from linear prediction coefficients," IPO Ann. Prog. Rep. **12**, 47–54.

Walliser, K. (1969). "Zur Unterschiedsschwelle der Periodentonhöhe," Acustica **21**, 329–336.

Zelle, H. W. (1978). "A computer implementation of PEST (Parameter Estimation by Sequential Testing) for auditory masking experiments," IPO Ann. Prog. Rep. **13**, 42–48.

Zwicker, E. C., and Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger* (Hirzel, Stuttgart), 2nd ed.