The role of frequency resolution and temporal resolution in the detection of frequency modulation

John P. Madden
Speech and Hearing Department, Cleveland State University, 1983 East 24th Street, Cleveland, Ohio 44115

(Received 30 March 1993; accepted for publication 26 August 1993)

The experiment investigated subjects' ability to detect short-duration changes in frequency. In an adaptive, 2AFC task, three normal-hearing subjects were asked to distinguish a sinusoidal signal that increased in frequency in a series of discrete steps from a standard that was identical except that its frequency increased essentially continuously. The signals were 60 ms in duration with center frequencies of 0.25, 0.5, 1, 2, 3, 4, and 6 kHz. The smallest frequency increase between steps (FI) at which the stepped signal could be distinguished from the standard was determined as a function of the number of steps in the signal. As the number of steps increased and the step duration decreased, the FI at first decreased and then reached a roughly asymptotic level. Eventually, however, at a certain number of steps, the FI increased rapidly. The data were analyzed using a model of auditory temporal resolution that included a bank of bandpass filters, a nonlinearity, a temporal integrator, and a decision device. The analysis yielded ERDs that ranged from 3.8 to 5.0 ms and did not change systematically with frequency. Detector efficiency varied considerably, being greatest at 0.5 and 1 kHz, and declining at higher and lower center frequencies.

PACS numbers: 43.66.Mk, 43.66.Fe, 43.66.Ba [LLF]

INTRODUCTION

Naturally occurring acoustic signals such as speech contain rapid amplitude and frequency modulations. The ability to follow these modulations depends, in part, on the temporal resolution capacity of the auditory system, and thus there have been many studies of auditory temporal resolution. Most of these studies have involved either gap detection (e.g., Plomp, 1964; Fitzgibbons, 1983; Shailer and Moore, 1985; Formby and Forrest, 1991) or the temporal modulation transfer function (e.g., Viemeister, 1979; Bacon and Viemeister, 1985; Eddins, 1992). Also, Moore and his colleagues (Moore et al., 1988; Plack and Moore, 1990) measured the shape of the ear's temporal "window" by determining the threshold of a brief sinusoidal signal preceded and followed by noise bursts. The signals used in these experiments change rapidly in amplitude but are spectrally static. In contrast to the large and growing literature in this area, the number of studies of temporal resolution using frequency modulated (FM) signals is very small. Huffman sequences (Huffman, 1962) have been used in a few experiments (e.g., Patterson and Green, 1970; Jesteadt et al., 1976), and studies of the detection of frequency modulation in sinusoids have implications for temporal acuity (see Kay, 1982, for a review). However, these approaches have often been limited by the presence of spectral artifacts that can serve as confounding cues. For example, the detectability of FM sinusoids improves at rates higher than 100 Hz, apparently due to the resolution of spectral sidebands (Kay, 1982). Thus, it has been difficult to measure the auditory temporal resolution of FM signals, and it cannot be assumed that estimates of temporal resolution derived from experiments with signals that change in amplitude also hold for FM signals.

Feth and his co-workers (Feth et al., 1989; Madden and Feth, 1992) recently developed a new means of investigation the temporal processing of FM signals. Briefly, in these experiments subjects were asked to discriminate between two FM sinusoid signals. The standard signal was a glide that made a transition from a lower frequency to a higher frequency over a smooth, linear path. The target signal, called the step signal, was identical to the standard except that its trajectory followed a series of brief steps. That is, the step signal remained at one frequency for a brief time before jumping to the next frequency. The long-term spectra of the standard and the step signals are very similar at moderate levels (see Madden and Feth, 1992 for a spectral comparison), and spectral artifacts appear not to have been a confounding factor in these studies. A schematic representation of these signals is shown in Fig. 1.

As might be expected, as the number of steps was increased (while holding overall duration constant and therefore decreasing the duration of the individual steps) discrimination became more difficult, and listener performance decreased monotonically to chance. It was inferred that at the threshold of discrimination, the internal representation of the step signal is "blurred" by the temporal processing system to a point at which it is indistinguishable from the standard glide. The results indicated that for a signal with a center frequency of 1000 Hz, resolution thresholds occurred at step durations of about 5 to 10 ms, depending on the subject. The effects of several stimulus parameters on the resolution of the step signal were noted (Feth et al., 1989). No effect of overall stimulus length was found. Resolution thresholds did not change signifi-

sults. However, because the frequency increments decreased, it is also possible that the results reflect limitations placed on the auditory system. Although a range of transition sizes was used, the effect of frequency resolution was not addressed specifically in these studies, and the results are ambiguous in this respect. If step-glide discrimination was limited entirely by frequency resolution, one would expect a relatively linear decrease in discrimination performance with frequency, reflecting the almost linear increase in auditory filter bandwidth that occurs above 0.25 kHz (Patterson and Moore, 1986), and this was not the case for center frequencies between 0.25 and 2 kHz. However, it seems probable that performance at 4 kHz was affected by frequency resolution, since as overall transition size increased, discrimination performance increased as well. The major purpose of the study was to clarify this situation by investigating the respective roles of frequency resolution and temporal resolution in the discrimination of FM signals of this type.

The experiment is superficially analogous to the TMTF. In the case of the TMTF, temporal resolution is measured in terms of depth of amplitude modulation (e.g., Viemeister, 1979). At modulation rates above about 2 Hz, as the amplitude fluctuations of the signal become more rapid, modulation depth at threshold increases. In the present case, temporal resolution was measured in terms of the size of the frequency increment between steps (the FI). For example, the FI at the threshold of discrimination was determined for a two-step signal. That is, the smallest frequency transition at which a two-step signal could be distinguished from a glidelike standard (described below) of equal overall duration and transition size was determined. This measurement was repeated for signals with increasingly greater numbers of steps at various center frequencies.

It was predicted that as the number of steps increased (and step duration decreased, since the overall duration of the signal was held constant), at first the FI would remain relatively constant, because discrimination was being limited by frequency resolution. Eventually, however, as the number of steps continued to increase, the FI at threshold would begin to increase as the limitations imposed by temporal smoothing began to affect discrimination.

I. METHOD

A. Subjects

Three normal-hearing subjects participated. All had thresholds of 15 dB HL or less at the audiometric test frequencies from 250 to 8000 Hz.

B. Stimuli

The main features of the signals have been described in the introduction and in the previous studies mentioned (Feth et al., 1989; Madden and Feth, 1992). The step signals were like those used in Madden and Feth in that the "corners" of the steps were slightly rounded to minimize artifacts from spectral "splatter." Instead of the glide used in the two previous studies, a 17-step signal was used as the standard. A 17-step signal contains approximately twice the number of steps contained in a signal of 1-kHz center frequency at discrimination threshold (when the latter is specified). A 17-step signal is represented by the dashed line, the step signal by the solid line.

FIG. 1. A schematic representation of the step and glide signals. The glide signal is represented by the dashed line, the step signal by the solid line.
was considered to be 50 ms. Stimuli were generated at center frequencies of 0.25, 0.5, 1, 2, 4, and 6 kHz. The size of the stimulus transition varied adaptively. The maximum overall transition size was 500 Hz at the 0.25-kHz center frequency, 1000 Hz at 0.5 kHz, 2000 Hz at 1 kHz, and 4000 Hz at all other center frequencies. Transition sizes were dictated by the limits imposed by center frequencies and by the overall bandwidth available after filtering. The stimuli were presented at a sensation level of 35 dB.

C. Procedure

A two-interval, forced-choice (2AFC) task, in which the target was the step signal, was used to determine the FI threshold. A three-down, one-up adaptive procedure was used that estimated the 79% correct point on the psychometric function (Levitt, 1971). That is, after three consecutive correct responses the overall frequency transition of the signals was decreased, thus decreasing the FI. After one incorrect response the size of the transition was increased. Typically, the size of the increase or decrease was halved after the first four reversals, and these four reversals were discarded. The size of the change varied, depending on the center frequency. At low center frequencies, where FIs were relatively small, a 10-Hz change in transition size was used after the first four reversals. At high center frequencies, where FIs were larger, the change in size ranged up to 50 Hz. Further variation in these rules was necessary in some cases to accommodate the range of difficulty of the signals. For more difficult discriminations, involving signals with large numbers of steps, it was useful, to obtain the most consistent results, to discard as many as the first eight reversals. This gave the subject more opportunity to “home in on” the threshold area. The arithmetic mean of the FIs at all reversals after the discarded ones were used to calculate the FI for a run. A single stimulus run consisted of 100 trials, with a pause after the first 50 trials.

At each center frequency, FI thresholds were determined for step signals containing from two to as many as eleven steps. The number of steps in a signal was defined as equal to the number of steady-state portions of the signal. Both the step signal and the standard began and ended with a steady-state portion. In general, if a subject consistently reached the maximum possible transition size at a certain number of steps, signals with higher numbers of steps were not run at that center frequency.

Subjects were well practiced before data collection began, having served in several pilot studies using signals of the same type. Data collection was continued until discrimination performance was asymptotic. This meant that at least three runs per signal, each conducted on a different day. If performance did not improve on the third run, data collection was ended for that signal. If the third run showed improvement over the second, a fourth run was carried out, and so on, until performance failed to improve. The data from the last two runs were averaged, and therefore each data point is derived from the results of 200 trials.

Stimulus presentation and response collection were controlled by a PC. Stimulus and response intervals were signaled on a computer monitor. The subjects entered their responses on a computer keyboard and received visual feedback after every trial indicating the interval containing the target signal.

II. RESULTS

The individual and averaged results are shown for center frequencies from 0.25 to 1 kHz in Fig. 2(a) and center frequencies from 2 to 6 kHz in Fig. 2(b). FI at discrimination threshold is plotted as a function of the number of steps in the signal. The three lower center frequencies are plotted separately from the four higher center frequencies for clarity. The scale of the y axis used for the higher frequencies would result in a considerable loss of resolution if used for the lower frequencies. In the case of the individual subjects, the upper end point of each plot reflects the maximum possible FI at that center frequency, since data collection was terminated at that number of steps. For a particular center frequency, one subject often reached the maximum FI at a smaller number of steps than the other subjects. For example, S3 reached the 6-kHz maximum at five steps, whereas S2 reached the maximum at nine steps. In the case of such subjects, the FI at the final data point was used to calculate the average FI for larger numbers of steps.

Each of the plots in the averaged data can be divided into two portions. In the first portion, the FI first decreases and then reaches a level that is roughly asymptotic. This portion of the data can be interpreted as reflecting mainly the limits imposed on discrimination by the subjects’ frequency resolution capacities. The initial decrease in the FI might be accounted for by Green and Swets’ (1966) multiple look model, which predicts an increase in detectability proportional to the square root of the number of observations. On average, the FIs along this portion of the plots are similar at 0.25, 5, and 1 kHz, and increase at center frequencies above 1 kHz. The increase above 1 kHz is expected, because absolute frequency resolution decreases as frequency increases. However, the fact that the 0.25-, 5-, and 1-kHz plots are nearly superimposed on one another is somewhat surprising, in view of the fact that auditory filter bandwidth increases almost linearly in this frequency range (Patterson and Moore, 1986). The data are much “noisier” at 4 and 6 kHz, indicating that perhaps the task is somehow more difficult at these frequencies.

The second part of each plot is marked by an abrupt increase in the FI. This portion of the plot can be interpreted as reflecting mainly the subjects’ temporal resolution capacities. At the point of upturn, temporal resolution rapidly becomes the limiting factor in the subjects’ ability to distinguish the step signal from the standard. Temporal smoothing increasingly obscures the discontinuities of the step signal, and the subject eventually is unable to distinguish it from the standard, even at very high FIs. The upturn in the FI occurs sooner (at smaller step numbers) as center frequency increases, with the notable exception of the results at 0.25 kHz. In the averaged data, at 0.5 and 1 kHz the upturn occurs after eight steps, but at 4 and 6 kHz.
FIG. 2. Discrimination thresholds measured in terms of frequency increment (FI) between steps and plotted as a function of the number of steps in the signal. To permit better resolution of the results in the lower center frequencies, they are plotted separately from the higher center frequencies. The apparent leveling off after eight steps at 4 and 6 kHz reflects the fact that some subjects had reached the overall transition limits at these frequencies.
FIG. 3. The output of the auditory filter, the temporal integrator, and the decision device as a function of time in response to a five-step signal passed through a filter with a center frequency of 940 Hz. See the text for further details.

A. Components of the model

(1) The auditory filters are from Slaney’s (1993) implementation of the Patterson–Holdsworth gammatone filter bank (Patterson et al., 1992), which is based on a fourth-order gammatone filter. The Patterson–Holdsworth filter has an impulse response of the form

$$g(t) = a^n \cos(\pi f_c t + \phi)/e^{2\pi b t}, \quad (1)$$

where $t$ is time, $g(t)$ is the impulse response, $a$ is amplitude, $n$ is the order of the filter (4), $f_c$ is the center frequency of the filter, and $b$ is a parameter that determines the bandwidth of the filter. $b$ is equal to 1.019 times the equivalent rectangular bandwidth (ERB) of the filter. The filter’s transfer function is designed to closely match the “roex” filter often used to account for auditory frequency resolution (e.g., Patterson and Moore, 1986). ERBs were determined using the formula

$$\text{ERB} = 24.7(4.37f/1000 + 1),$$

where $f$ is the center frequency of the filter in Hz (Glasberg and Moore, 1990). The top panel of Fig. 3 shows the output of a filter centered at 940 Hz for a 5-step signal with a center frequency of 1 kHz.

A second order nonlinearity converts the filtered waveform to a powerlike quantity by squaring the instantaneous amplitude of the filtered waveforms (as, for example, in Moore, 1989).
formed into peaks in the output of the decision device that are shown in dB as a function of time. The differences between the signals are shown in the bottom panel of Fig. 3. The ERDs and detection criteria derived from the fitting procedure are listed in Table I, and the graphs in Fig. 4 show the predictions of the model in comparison with the actual FI thresholds at the various center frequencies. Overall, the predictions match the general trend of the data reasonably well. With the exception of 2 kHz, the model at least approximates the point at which the sharpest increase in FI occurs. It is not clear why a better fit could not be achieved at 2 kHz. The tendency to underestimate the FIs at the smaller numbers of steps in the lower center frequencies is the only aspect of the predictions where the model systematically departs from the observed FI values. The FI predicted by the model does not change at the smaller numbers of steps, or changes much less than the observed FI. This may mean that the auditory system needs to "look" at more than one or two peaks in the detector output to achieve maximum efficiency, as was suggested in the results section.

IV. DISCUSSION

The observed results shown in Figs. 2 and 3 indicate an apparent decrease in temporal resolution at 0.25 kHz and above 1 kHz. Numerous studies also have noted decreased temporal resolution at low frequencies (e.g., Fitzgibbons, 1983; Plack and Moore, 1990), and several explanations have been proposed for this phenomenon. One is that the relatively slow response of the auditory filters at low frequencies may limit temporal resolution (e.g., Shailer and Moore, 1983). The "tinging" of the auditory filters may tend to smooth the temporal details of the step signal at the 0.25-kHz center frequency, but Plack and Moore (1990) and Moore et al. (1993) found that the auditory filter had little effect on temporal resolution at frequencies of 100 and 300 Hz, thus casting doubt on this assumption concerning several other parameters of the model. The size of the predicted FI varied considerably as a function of the relationship between the center frequency of the signal and the center frequency of the auditory filter through which the signal was passed. Therefore, for each combination of ERD and ΔL, the procedure cycled through a range of filter center frequencies to determine the smallest FI at which the threshold criterion was met. The smallest FI sometimes was obtained with a filter at a center frequency below the starting frequency of the signals. Because the magnitude of the filtered signal decreased as the center frequency of the filter decreased below the signal starting frequency, the question of the perceptual salience of the output of such filters on the periphery of the signal had to be addressed. Somewhat arbitrarily, it was decided that only filters with outputs greater than 0.5 times that of the filters centered at the signal center frequency (and therefore with the greatest possible output level) would be considered. Finally, and also somewhat arbitrarily, it was assumed that the peaks calculated by the decision device had to occur above the -20-dB level to be evaluated. For most signals, peaks occurred only above this level.

The ERDs and detection criteria derived from the fitting procedure are listed in Table I, and the graphs in Fig. 4 show the predictions of the model in comparison with the actual FI thresholds at the various center frequencies. Overall, the predictions match the general trend of the data reasonably well. With the exception of 2 kHz, the model at least approximates the point at which the sharpest increase in FI occurs. It is not clear why a better fit could not be achieved at 2 kHz. The tendency to underestimate the FIs at the smaller numbers of steps in the lower center frequencies is the only aspect of the predictions where the model systematically departs from the observed FI values. The FI predicted by the model does not change at the smaller numbers of steps, or changes much less than the observed FI. This may mean that the auditory system needs to "look" at more than one or two peaks in the detector output to achieve maximum efficiency, as was suggested in the results section.

IV. DISCUSSION

The observed results shown in Figs. 2 and 3 indicate an apparent decrease in temporal resolution at 0.25 kHz and above 1 kHz. Numerous studies also have noted decreased temporal resolution at low frequencies (e.g., Fitzgibbons, 1983; Plack and Moore, 1990), and several explanations have been proposed for this phenomenon. One is that the relatively slow response of the auditory filters at low frequencies may limit temporal resolution (e.g., Shailer and Moore, 1983). The "tinging" of the auditory filters may tend to smooth the temporal details of the step signal at the 0.25-kHz center frequency, but Plack and Moore (1990) and Moore et al. (1993) found that the auditory filter had little effect on temporal resolution at frequencies of 100 and 300 Hz, thus casting doubt on this assumption.
hypothesis. Both studies indicate instead that the ERD of the temporal window increases markedly at frequencies in this range.

A trend toward poorer temporal resolution in the higher frequencies is not evident in any previous studies of which the author is aware. Resolution thresholds for gapped sinusoids have been shown to be independent of frequency between 0.5 and 4 kHz (Formby and Forrest, 1991), and Eddins et al. (1992) provide evidence that gap detection thresholds for narrow-band noise in this frequency range are independent of frequency as well, if signal bandwidth is held constant. Data from another study by Eddins (1993) indicates that the time constant of the TMTF does not vary across frequency if signal bandwidth is constant. Plack and Moore (1990) found that the equivalent rectangular duration of the temporal window actually decreased as a function of frequency from 900 to 8100 Hz.

As is evident from Table I, however, a different interpretation of the data emerges from the modeling analysis. The ERDs given in Table I are clustered in a fairly narrow range (3.8 to 5.0 ms) and exhibit no clear tendency to change systematically with frequency. This suggests that...
the change in performance evident in Fig. 2 is not due to temporal resolution per se, as reflected in the width of the temporal window. It should also be noted that the ERDs are in the same range as those obtained by Plack and Moore (1991) for decrement detection in wideband noise, although they are somewhat smaller than ERDs derived from experiments that estimate the shape of the temporal window (Moore et al., 1988; Plack and Moore, 1990).

Unlike the ERD values, the ΔL values in Table I exhibit considerable variation, being roughly one-half as great at 0.5, 1, and 2 kHz as they are at the other center frequencies. These are the frequencies, of course, at which the best performance is seen in Figs. 2 and 3 in the sense that the smallest step durations are reached before greater FIs are needed for discrimination. The model analysis therefore indicates that the greater ability to resolve short-duration frequency changes at these center frequencies is due to greater detector efficiency. Plack and Moore (1990) reported an effect of frequency on detector efficiency in a task involving the detection of a tone burst between two noise bursts. In this case, however, the pattern was quite different, as efficiency increased almost linearly between 300 and 8100 Hz.

Summarizing the discussion to this point, it may be said that the model can account for the data, but the ERD and ΔL values that best predict the results are at odds with the findings of previous studies, mainly in that they indicate decreased detector efficiency at high center frequencies. This suggests that the same temporal integration mechanism is involved in the processing of both FM and non-FM signals but that the detection process differs. The obvious candidate for the detector difference is temporal coding of the waveform through neural synchrony. In most mammalian ears, synchrony fades above 2 kHz and is completely absent above 5 kHz (Rose et al., 1967; Anderson et al., 1971). This pattern parallels that found in the data of this experiment. Both level of excitation and temporal information would be available in the lower frequencies, but only level information would be present in the higher frequencies. It has been argued for some time that the auditory system can make use of temporal coding in frequency discrimination tasks involving pulsed tones (e.g., Moore, 1973; Moore and Glasberg, 1986). Moore and Glasberg (1989) contend that phase-locking information is used at least to some extent in the detection of frequency modulation as well, although their results indicate that temporal information is still present at 4 kHz. The present analysis also suggests that the auditory system makes use of both temporal and level information in tracking these FM signals. This analysis still does not account for the poorer detector efficiency at 0.25 kHz, however, since temporal information should be available there as well.

Some potential confounding factors must be addressed. The subjects almost certainly experienced greater changes in SL when listening to signals at the higher center frequencies than at the lower center frequencies. Sensation levels were based on thresholds determined using signals with an overall transition size of 400 Hz. At the higher center frequencies, transition sizes were considerably larger than this, extending into regions of reduced sensitivity, and were larger than the transitions at the lower frequencies. The results probably were not greatly affected by these changes, however, since pilot testing indicated that performance was relatively stable from about 30 to 50 dB SL. Nevertheless, it is possible that the upper ends of the transitions were not as audible at the higher center frequencies as at the lower ones.

There is some evidence that frequency resolution worsens for stimuli of short duration (e.g., Bacon and Viemeister, 1985), and this alternatively could explain the upturn in FI at short step durations. If this effect is frequency dependent, it also could account for the pattern of the results. This process fails to account for certain aspects of the data, however. Bacon and Viemeister demonstrated reduced frequency resolution in the detection of a signal centered within a 1-kHz masker of 50-ms duration. This would seem to predict progressively poorer frequency resolution at durations shorter than about 25 ms. Figure 2 indicates that the FI at 1000 Hz does not increase until step duration decreases to 7.1 ms. In addition, another experiment, not reported, measured the greatest number of steps, or shortest step duration, that could be detected in a signal for a given overall transition size. Using the same step-standard discrimination task, the number of steps was varied adaptively while the frequency transition was held constant. This procedure was repeated over a range of transition sizes. To the extent that a worsening of frequency resolution with decreased step duration limits performance, step detection should have continued to improve (larger numbers of steps should have been detected) as the transition size was increased, since the larger FIs would compensate for reduced frequency resolution. In fact, as transition size was increased, performance increased, but only up to a point, and then leveled off. For example, for signals at a center frequency of 1500 Hz, the detection threshold was about eight steps for transitions ranging from 400 to 2000 Hz (FIs of 57 to 286 Hz). It should be noted that eight steps also is about the average point of upturn for the 1000-Hz signal. The same pattern was found at frequencies above 1 kHz.

V. SUMMARY

For a signal that increased in frequency in a stepwise fashion, the experiment measured the smallest increment in frequency between steps at which the signal could be distinguished from a standard signal that approximated a glide. The results indicated that the subjects' performance was best for signals at 0.5 and 1 kHz and declined for signals above and below those frequencies. Analysis using a model incorporating an auditory filter bank, a nonlinearity, a temporal integrator and a detection device indicated that these differences in performance resulted from increased detector efficiency in the midfrequencies. Detection criterion values approximated those derived in previous temporal resolution experiments. The ERDs derived from the model did not vary systematically with frequency, and values were similar to those obtained in previous experiments with non-FM signals.
ACKNOWLEDGMENTS

This research was supported by a grant from the NIDCD. I would like to thank Bill James for writing the programs used in the study. I would also like to thank Peter Fitzgibbons and Chris Plack for their insightful and helpful comments.


