Tone-segregation by phase: On the phase sensitivity of the single ear

Michael Kubovy and Ray Jordan

Department of Psychology, Yale University, Yale Station Box 11A, New Haven, Connecticut 06520
(Received 9 February 1978; revised 27 November 1978)

A monaural complex tone is synthesized from 12 harmonically related pure tones, played in phase. In each of 12 segments, one of the tones (the target) is played out of phase so that the sequence of targets is increasing or decreasing in frequency. If the target is at least 30° out of phase, the targets are perceptually segregated. This tone-segregation by phase raises doubts concerning several current theories of pitch perception. The phenomenon is conjectured to be caused by the ear's nonlinear compressive transfer characteristic or by a temporal analysis of the stimulus.

PACS numbers: 43.66.Nm, 43.66.Hg, 43.66.Lj

INTRODUCTION

Recent theories of pitch perception (Goldstein, 1973; Terhardt, 1974; Wightman, 1973b) are predicated on the empirical law that "within limits, all stimuli made up of the same spectral components have the same pitch, regardless of the relative phases of the components" (Wightman, 1973b, p. 407). Thus, one of the goals of these theories is to explain the loss of information about relative phases of stimulus components. In this paper we describe a new paradigm useful in the investigation of phase effects in pitch perception, and we show, using this paradigm, that there is a sense in which pitch perception is affected by phase. Our data highlight the multidimensional nature of the pitch percept: certain complex tones, consisting only of harmonically related components, are shown to possess at least two pitches. The low pitch of these tones will be shown to be unaffected by phase, but the existence and prominence of a second, high pitch, of these tones will be shown to be quite sensitive to phase.

The paper is organized as follows: In the first part of the paper we describe four experiments that refute the hypothesis regarding the loss of information about relative phases of stimulus components. In the second part of the paper we show that these findings could either be accounted for by a temporal fine-structure theory of pitch or by postulating a nonlinear compressive stage in the auditory system prior to the pitch-perception mechanism.

I. EXPERIMENT I: DEMONSTRATION OF TONE-SEGREGATION BY PHASE

Our stimuli resembled those employed by Patterson (1973) and Wightman (1973a), and our paradigm those of Licklider (1956a) and Schroeder (1959). Pierce (1960) summarizes Schroeder's observations as follows:

"How sensitive is the ear to phase changes? This depends on the original phases of the components. For instance, when the harmonics are all in phase, giving a sequence of pulses, a reversal of phase of any harmonic produces a sensation of tone accompanying the sound, the pitch corresponding approximately to the frequency of the harmonic reversed. Indeed, one can play a tune by successively reversing the phases of various harmonics." (p. 42)

The stimuli we used to replicate and extend these observations were complex tones made up of 12 equal-amplitude, consecutive harmonics of 200 Hz ranging from 600 to 2800 Hz. In a canonical configuration (which was never actually played) all components would be in sine phase. The actual stimulus consisted of a sequence of 12 consecutive segments lasting about 1 sec each. At the beginning of each segment, the phases of all components were reset to sine phase, except for one component, called the target, which was reset to, say, 90° out of phase. The 12 segments were arranged so that the frequency of the component that was out of phase was a monotonic increasing (or decreasing) function of serial position, and the entire stimulus was played cyclically.

After listening for a while, observers described the stimulus as follows with features listed in the order of their distinctiveness: (a) A series of sharp clicks, delimiting, (b) segments all of which have the same pitch, equivalent to 200 Hz (the "missing" fundamental of the complex tone), (c) which change somewhat in timbre, and (d) a faint, but unmistakable, tone segregated in each segment, forming a cyclical ascending (or descending) series in pitch.

A. Method

1. Stimuli

All acoustic signals were 12-component, complex tones similar to the one described above. A stimulus was formed from a series of 12 acoustic segments, in each of which a different harmonic was the target. For each stimulus, the 12 segments were arrayed in monotonically ascending or descending order according to target frequencies. One of 14 values of target phase shift was chosen: 1°, 5°, 10°, 20°, 30°, 40°, 50°, 60°, 70°, 80°, 90°, 120°, 150°, or 180°. Thus 28 stimuli were created.

The stimuli were generated numerically on a Digital Equipment Corporation PDP-11/40 computer which
calculated 10,000 waveform values per second of sound. They were played monaurally by a PDP-11/34 computer, through a D/A converter, a 24 dB/octave low-pass filter (Krohn-Hite model 3202) set at 4000 Hz, an amplifier, into Grason-Stadler TDH-49 earphones. The filtering process introduced frequency-dependent phase shifts which made phase alignments accurate to within 2°.

The segments were 307.2 msec long; thus a string of 12 segments lasted 3.6864 sec. (Exact frequencies and stimulus durations were dependent on the pulse rate of the computer clock which, in a postexperiment test, was seen to be pulsing approximately 5% slower than its nominal value.) Each segment began with all components in sine phase, except for the target. Thus, at the beginning of each segment the phases of all components, background and target, were abruptly changed marking the onset of each segment with a strong click. This was done to avoid associating a transient sound with the target, while leaving the background unaffected. Each component of the stimulus was played at approximately 50 dB SPL. The experiment was conducted in an acoustic chamber.

2. Procedure

Eight paid subjects (Yale University students and staff) who had little or no previous experience in auditory experiments and who had never previously listened to these sounds participated in the experiment.

For practice, several of the sequences with high phase-shift angles were linked so that a continuous rising-falling-rising pattern was produced. Subjects listened to these sounds for 10 min. They were told to attend to the higher pitched rising and falling scales, but were never told which scales ascended or which descended. Subjects then received 14 practice stimulus strings which they were to judge as ascending or descending. No feedback about performance was given. Subjects listened to the continuous sequence for five additional minutes before beginning the scored portion of the task. For each subject, the 28 different stimulus strings were randomly presented twice to each ear, for a total of 112 trials. Subjects were instructed to judge whether each sequence was ascending or descending, even if they had to guess for some trials. Listeners paced themselves by tapping a key on a computer terminal to begin a trial. The experiment lasted less than an hour.

B. Results and discussion

Results from this experiment are plotted in Fig. 1 as median percent correct direction identifications as a function of amount of phase shift. Performance rises from chance level (50% correct) at phase shifts below 10° to close to 100% correct at phase shifts above 40°. Thus we have demonstrated tone-segregation by phase.

In order to allay concern that tone-segregation by phase may fail with complex tones other than the ones studied in experiment I, a stimulus was generated with all components in cosine phase, except the target which was 90° out of phase. The authors verified that the tone-segregation was at least as pronounced as when all components except the target were in sine phase.

As a further test of the generality of the phenomenon, the following experiment was carried out.

II. EXPERIMENT II: REPLICACTION FOR HIGH AND LOW HARMONIC NUMBERS

The following experiment addressed the issue of whether phase sensitivity was preserved at high frequencies at which phase information is not preserved for localization.

A. Method

1. Stimuli

The stimuli were 12-component complex tones, similar to those used in experiment I. There were two sets of stimuli: low-frequency stimuli and high-frequency stimuli. All had a fundamental frequency of 100 Hz. The low-frequency stimuli were made up of harmonics of 100 Hz ranging from 300 to 1400 Hz; the high-frequency stimuli were made up of harmonics of 100 Hz ranging from 1500 to 2600 Hz. For each stimulus, one of three values of target phase shift was employed (22.5°, 45°, or 90°), and each formed an ascending or a descending series. Thus, 12 stimuli were employed. In all other respects these stimuli were like those of experiment I (except that the computer clock was now accurate).

2. Procedure

Ten unpaid volunteers participated in the experiment. The procedure was similar to that of experiment I except that: (a) preexperiment training was curtailed;
(b) stimuli were played starting at a random point in the sequence of segments; (c) stimuli were repeated cyclically without interruption until the subject responded; and (d) trials were initiated by the experimenter. Each subject went through four random permutations of the 12 stimuli, for a total of 48 trials. The experiment lasted less than half an hour.

B. Results and discussion

Results from this experiment are plotted in the two panels of Fig. 2: on the left the results for the low-frequency conditions, on the right the results for the high-frequency conditions. The results are plotted as median percent correct direction identification as a function of amount of phase shift. Performance is essentially identical in these two conditions and it replicates the results obtained in experiment I. These results demonstrate that the phenomenon of tone-segregation by phase is robust over a wide range of frequencies.

The phenomenon of tone-segregation by phase is similar to the demonstrations by Dulfhuis (1970a, 1970b) that the audibility of a high harmonic (in periodic pulse trains) is sensitive to the phase of that harmonic. Specifically, Dulfhuis (1970a) found that low-order harmonics ($n<10$, for pulse rates of 25, 50, and 100 Hz) were equally audible if they were cophasic or antiphasic, but that the audibility of high-order harmonics ($n>25$ for pulse rates of 25 or 50 Hz, and $n>25$ for a pulse rate of 100 Hz) was quite sensitive to phase. In the present experiment, the low-frequency stimuli contained harmonics from 3 to 14, whereas the high-frequency stimuli contained harmonics from 15 to 26. From Dulfhuis's observations it would be expected that segregation be weaker in the low-frequency stimuli than in the high-frequency ones. Although our data do not bear out this implication, they do not refute it either, because four harmonics of the low-frequency stimuli had harmonic numbers above those hypothesized by Dulfhuis to be phase insensitive. Since it is possible that these four harmonics enable our subjects to identify the direction of the sequence, we conducted the following experiment.

III. EXPERIMENT III: EVIDENCE IN SUPPORT OF THE DULFHUIS HYPOTHESIS

A. Method

1. Stimuli

The stimuli were 12-component complex tones, similar to those used in experiment I. All had a fundamental frequency of 200 Hz (harmonics ranging from 500 to 2800 Hz). Each stimulus was formed from a series of four segments; in each segment a different harmonic (phase-shifted by 90°) was the target. For each stimulus, the four segments were arrayed in monotonically ascending or descending order according to target frequencies. An ascending and descending version of each of three types of stimuli was created: (a) low-harmonic targets (harmonics 3–6); (b) intermediate-harmonic targets (harmonics 7–10); and (c) high-harmonic targets (harmonics 11–14). Thus six stimuli were created. In all other respects the stimuli were like those employed in experiment I.

2. Procedure

Four paid subjects participated in the experiment, none of whom had participated in the earlier experiments. The procedure was similar to that of experiment I except that each subject went through four random permutations of the six stimuli (intermixed with other similar stimuli, which will not concern us here).

B. Results and discussion

Percent correct direction judgments were 56.3% for the low-harmonic targets, 59.4% for the intermediate-harmonic targets, and 78.1% for the high-harmonic targets. Only the latter proportion differs significantly from the 50% chance performance level ($z = 6.19$, $p < 0.001$).

The performance level of 78.1% is lower than the performance we obtained in experiment I for the same phase of the target (which we estimate to be close to perfect). Thus it would appear the lower harmonics do contribute to the detection of the sequence direction. These lower harmonics cannot however produce any appreciable segregation on their own, since the subjects were unable to detect the sequence of targets in them better than would be expected by chance. On the whole, these results corroborate Dulfhuis's observation on the absence of an effect of phase on the audibility of low harmonics of complex tones.

Up to this point, our observations support models of pitch perception which do not assume the loss of phase information; in particular they support temporal fine-structure theories of pitch (i.e., Licklider, 1956b; Schouten, 1962; see Plomp, 1976, for a review).
Insofar as these results support temporal fine-structure theories of pitch perception, they are incompatible with the theories of pitch perception we cited at the beginning of this paper (Goldstein, 1973; Terhardt, 1973; Wightman, 1973b). Before we consider the question of the support of our data for temporal-fine structure theories, we should compare the results of experiments I, II, and III to the earlier experiments by Patterson (1973) and Wightman (1973a). In the latter two experiments it was demonstrated that subjects match the same pitch to complex tones in cosine phase as to complex tones in random phase. As we will show presently, our results are not incompatible with the earlier ones. Presumably, in the earlier experiments attention was focused on the low pitch (the “missing” fundamental) of the complex tones, whereas in our experiments attention was focused on a second pitch property of the tones. The next experiment was designed to show that the low pitch of the stimuli we employed in the preceding experiments is not affected by manipulations of phase.

IV. EXPERIMENT IV: THE INSENSITIVITY OF LOW PITCH TO PHASE

A. Method

1. Stimuli

The test stimuli were the 12 individual segments from one of the stimuli used in experiment I; in each segment one of the 12 components was 90° out of phase. The probe tone was a variable-frequency square wave passed through a 24 dB/octave low-pass filter (Krohn-Hite model 3202) set at 4000 Hz, and equated in loudness to the test stimuli. The subject could choose at will to listen to the test or the probe tone by operating a switch.

2. Procedure

Two experienced listeners were asked to adjust the probe frequency so as to match its pitch with the pitch of the test tone. The test tones were presented in random order five times for a total of 60 pitch matches.

B. Results and discussion

Results from this experiment are plotted in Fig. 3 as mean matching probe-tone frequency as a function of target-tone frequency. By analysis of variance, target tone frequency had no effect on pitch matches \[F(11, 11) = 0.94\].

Our results show that the ear retains information about both the prominent low pitch of a complex tone and the phase-shifted tone. In this respect the pitch-extraction mechanism is functioning as an analytic sense (Helmholtz, 1862/1954, p. 62), recognizing more than one pitch in a complex tone.

Current psychoacoustic practice is not conducive to the discovery of tone-segregation phenomena such as tone-segregation by phase because of its emphasis on auditory stimuli that are static with respect to their power and phase spectra. In the present experiments, stimuli were not static with respect to phase spectrum, making possible the detection of a second-order invariant. The phenomenon is, as Bregman (in press) has argued about an earlier demonstration by Kubovy and Howard (1976), akin to the segregation by common fate in vision (Hochberg, 1971, p. 434): a group of dots can become perceptually segregated from another otherwise indistinguishable group of dots by moving coherently. In the Kubovy and Howard demonstration a set of six nonharmonically related dichotic tones were randomly distributed in a canonical array over localization space by assigning to each a different canonical interaural time disparity. In each segment of the stimulus a different tone was made into a target by displacing it from its position in the canonical localization array. The stimuli were constructed from segments in which targets formed an ascending or descending scale. Although the localization array was different from segment to segment, none of the individual segments contained sufficient information to determine which tone was the target. Only a device capable of comparing several segments could extract this information. Subjects were able to do so with ease when the segments were abutted and played in succession. If a cue causes tone-segregation and is essentially intersegmental, we call it a successive-difference cue.

Is tone-segregation by phase based on such a successive-difference cue? Although it would appear from experiment IV that it is, because segregation does not occur in individual segments, conceptual clarity requires us to distinguish between the type of dependence on change involved in tone-segregation by phase and in a successive-difference cue. In the latter, the segregation is based on information that is wholly intersegmental. In the present demonstration, in tone-segregation by phase, the information is entirely intrasegmental, since it is apparently caused by the concurrent difference in phase between the target and the remaining tones. Thus, it is appropriate to call it a concurrent-difference cue (for further discussion of the distinction between concurrent and successive difference cues, see Kubovy, in press).

There remains our puzzling observation that tone-segregation by phase cannot be perceived in the individual segment. We can only speculate that segre-
gation by phase is not an instantaneous process, but rather one that demands the accumulation of information regarding the nature of the object to be segregated, much as Bregman (1978) has shown that stream segregation requires the accumulation of perceptual evidence before it can take place.

We turn now to an explanation of tone-segregation by phase which does not require the auditory system to process phase information.

V. THEORY

A. The compressive transfer characteristic

Our theory follows Pfeiffer (1970) and Smoorenburg (1972, 1974) and assumes that the transfer characteristic of the auditory system is compressive and nonlinear, viz., that the relationship between pressure $p$ and the resulting displacement $d$ is

$$
d = \begin{cases} 
  p^\beta & \text{for } p > 0, \\
  -|p|^\beta & \text{for } p < 0.
\end{cases}
$$

Such a nonlinear transfer characteristic can transform a phase disparity into a power disparity. As an illustration, we applied this transfer characteristic (with $\beta = \frac{3}{4}$, as proposed by Pfeiffer, 1970) to one cycle from each segment of a stimulus similar to the one employed in experiment II. Specifically the stimulus was the low-frequency stimulus (300–1400 Hz) in which the target was phase shifted by 90° (the only difference being that in experiment II stimuli were aligned in sine phase, whereas in this simulation they were aligned in cosine phase). After applying this cubic-root transformation to the waveform, we calculated its spectrum. Figure 4 shows the results. Each of the 12 curves represents the spectrum of one of the segments over the range 100–1500 Hz. In each of these spectra we observe a maximum at the target frequency. These spectra should not be taken to imply that the "targets" dominate the percept. On the contrary, the 100-Hz "missing" fundamental is the most prominent pitch perceived, and the theory we are discussing is not designed to cope with that phenomenon. These spectra do account, however, for the relative prominence of one of the harmonics, namely the target, in the secondary pitch perceived.

Furthermore, we compared the effect of the cubic-root transfer characteristic on the power spectra of the low-frequency complex tones studied in experiment II. As we see in Fig. 5, the spectra for the corresponding segments in the two types of tones are virtually identical.

We also compared the effect of the cubic-root transfer characteristic on the power spectra of the low-frequency complex tones studied in experiment II for four different phase shifts: 0°, 22.5°, 45°, and 90°.

As we see in Fig. 6, the prominence of the spectral peak at the target frequency is a monotonically increasing function of the phase shift.

These results support the conjecture that the transfer characteristic of the auditory system is compressive. This is in keeping with data in several sources (see

![Figure 4. Line spectra of twelve waveforms similar to segments of the low frequency stimuli of experiment II, after cubic-root transformation. In each waveform a different target was phase-shifted by 90° relative to the remaining components. The ordinate of each spectrum is the amplitude of the harmonic at each frequency (arbitrary units). The lines connecting points of the spectra are included for visual convenience only.](image-url)
FIG. 5. Line spectra of two waveforms similar to segments of experiment II, after cubic-root transformation. The bottom spectrum represents the low-frequency complex tone, with a target of 800 Hz; the top spectrum represents the high-frequency complex tone, with a target of 2000 Hz. In both segments the targets were shifted 90°. The ordinate of each spectrum is the amplitude of the harmonic at each frequency (arbitrary units). The lines connecting points of the spectra are included for visual convenience only.

Penner, 1978): (a) loudness is compressive, i.e., it grows as a power function of sound pressure, with a positive fractional exponent (Marks, 1974; Stevens, 1956); (b) a compressive transfer function can account for the form of the function relating intensity discrimination to absolute intensity (McGill and Goldberg, 1968a, 1968b; Penner, 1972); (c) neural response rate is a compressive function of stimulus energies (see Katsuki, 1961); and (d) a compressive transfer function makes it possible to reconcile widely divergent results in the literature on temporal integration (Penner, 1978).

C. Implications for theories of pitch

The demonstration of tone-segregation by phase disparity undermines current evidence in favor of theories of pitch which assume insensitivity to phase disparity. It does not, however, refute the theories themselves. For suppose that (a) tone-segregation by phase disparity were caused by a nonlinear transfer characteristic, and (b) the auditory system were not able to process monaural phase disparities as such. Suppose, furthermore, that we knew the nonlinear transfer characteristic of the ear. Then it should be possible to create stimuli which, after passing through the transfer characteristic of the ear, would have flat power spectra, and phase spectra in which all components would be in phase except for the target. To create such stimuli we would only have to apply the inverse of the known transfer characteristic to stimuli similar to the ones used in experiments I and II, and play the output of this transformation to subjects just as we did in experiments I and II.

Unfortunately, we do not know the transfer characteristic of the ear. Hence, in order to determine whether the single ear is or is not sensitive to phase disparity, it is necessary to search through the space of transfer characteristics for one whose inverse produces a stimulus which eliminated tone segregation by phase disparity. If such a transfer characteristic were found, we would have simultaneously produced a contender for a transfer characteristic and evidence in favor of the phase-insensitivity law. Until such a result is obtained, however, we must view with caution theories of pitch perception that imply insensitivity to phase disparity.

Given current evidence it is impossible to decide between temporal fine-structure theories and spectrum-
based theories of pitch perception. We have, however, provided some support for Duifhuis's temporal analysis approach in experiment III by showing that the audibility of high harmonics is more affected by phase differences than is the audibility of low harmonics. The temporal effects predicted by Duifhuis's theory would appear to cohere well with temporal fine-structure theories of pitch perception.

ACKNOWLEDGMENT

This research was supported by NSF Grant BNS 76-21018. Figures were drawn by Virginia L. Walters.

We are grateful to Robert N. Sorkin and to a reviewer for this Journal for drawing our attention to this note by Pierce.