The effect of amplitude envelope on the pitch of sine wave tones

W. M. Hartmann

Physics Department, Michigan State University, East Lansing, Michigan 48824 [a)]
and Laboratory of Psychophysics, Harvard University, Cambridge, Massachusetts 02138
(Received 20 June 1977; revised 7 December 1977)

Psychophysical experiments show that the pitch of a short sine wave tone depends upon the amplitude envelope of the tone. Subjects find that the pitch of an exponentially decaying tone (1dB/ms) is higher than the pitch of a (20-ms) rectangularly gated tone of equal frequency. The percentage difference in frequency required to produce equal pitches with the two envelopes depends upon frequency f_0: 2.6% at $f_0 = 412$ Hz, 1.4% at $f_0 = 825$ Hz, 1% at $f_0 = 1650$ Hz, and 0.7% at $f_0 = 3300$ Hz. The pitch change is insensitive to the relative intensities of the two tones. The spectra of tones with the two different envelopes suggest no obvious explanation for the pitch change. However, the weighted time-varying spectra for tones with two different envelopes evolve differently with time. Alternatively the pitch change can be derived from a modified version of the auditory phase theory of Huggins.

PACS numbers: 43.66.Hg, 43.66.Fe

INTRODUCTION

The pitch of sine-wave tones is an interesting topic in psychophysics because one may plausibly regard experiments with sine waves as probing auditory mechanisms that are basic and elementary. Of particular interest in the development of theories of hearing are experimental studies of external factors which can change the perception of pitch of a sine wave. The pitch of a sine wave is often different in the two ears (van den Brink, 1970); it varies with intensity of the signal (Verschuure and van Meeteren, 1975). Pitch is altered by a long preceding satiating tone (Christman and Williams, 1963), by shorter leading tones (Hartmann and Blumenstock, 1976), and by bands of noise (Webster and Muerdter, 1965).

In this paper it is noted that the pitch of sine-wave tones also depends upon the shape of the amplitude envelope of the tone. The pitch of a sine-wave tone with an exponentially decaying envelope is higher than the pitch of a sine wave with the same frequency but with a rectangular envelope.

The experiments presented below are of several kinds. Section I discusses a survey study, with 15 subjects in a short experiment with rectangular and exponential envelopes. This survey establishes the existence of the effect of a pitch shift with envelope change. Section II describes a parametric study, with three subjects, in which both intensity and frequency range were varied. Appendix C eliminates from consideration several possible explanations of the pitch shift. In the discussion of Sec. III contact is made with previous work on pitch discrimination of short tones. The discussion includes a spectral study of the stimuli and explores the possible relationship between the present results and a modified version of the auditory-phase principle of Huggins (1952).

[a)]Permanent address.

I. SURVEY EXPERIMENT

A. Procedure

In the survey experiment subjects compared the pitches of two sine-wave tones, one with a 20-ms rectangular envelope, the other with an exponential envelope with a 120-dB decay in a time of 120 ms. This comparison experiment is called the R–E experiment in the rest of the paper. The rectangular (envelope) tone was heard at 89 dB SPL; the exponential (envelope) tone had an initial (maximum) amplitude corresponding to 95 dB SPL. These tones seemed to be equally loud and to have equal duration. Appendix A shows that these two tones have equal energy. The psychophysical procedure used a two-interval forced-choice up–down staircase pattern. (See Appendix B.) On each trial the subject heard a rectangular tone and an exponential tone (each first with probability $\frac{1}{2}$ on any trial). A 500-ms gap separated the tones. The subject indicated, with push buttons, which tone, the first or the second, had the higher pitch.

During the course of a block of trials, the frequency of the standard, the rectangular tone, varied according to a pseudorandom schedule with no successive repetitions among the values, 800, 810, 815, 825, 830, 835, 840, 845, and 850 Hz. This range will be referred to as the $f_0 = 825$-Hz range. The variable of interest in the staircase cycle was the difference in frequency between the exponential and rectangular tones. This difference took on values, −30, −20, −10, 0, +10, +20, and +30 Hz, so that the stimuli in any run were always distributed symmetrically about zero shift. Other precautions against biasing the results to favor one direction were taken. The subjects were not informed of the trends of their responses or of those of others until all data had been collected. This statement does not apply to subjects numbered 2, 4, 5, and 9, including the author and three colleagues who were aware, if only dimly, of the trend of responses. The data of these four subjects did not differ systematically from those of the others. The

 0001-4966/78/6304-1105$00.80

rate of the experiment was set by the subjects, but after several cycles subjects typically ran the experiment at its maximum rate, one run of 64 judgments in 150 s. After each run, subjects in the survey experiment rested for at least three minutes.

The stimuli were presented diotically to the subjects through TDH-39 headphones with 001 cushions, while the subjects were seated in individual quiet rooms, IAC 1200A. The stimuli were generated by a voltage-controlled function generator, Wavetek, VCG 116 controlled by a computer through a D/A converter. The oscillator frequency was monitored with a digital counter often during the course of a run and compared with the calculated frequency displayed on a video screen. The tones were shaped by a programmable attenuator, Charybdis model A, also controlled by the computer. The turn on and turn off of tones was uncorrelated with the phase of the sine-wave signal. The electrical signals were observed on a scope and found to have no obvious overshoot, ringing, or transient distortion components. The Wavetek function generator output was low-pass filtered to remove upper harmonics before amplitude shaping but no filtering followed amplitude shaping. The signals were presented in a constant noise background with spectrum level of 10 dB *re* 20 μPa/Hz in a band from 500 to 1500 Hz. The noise background was included principally to provide an unambiguous basis for determining an effective duration of the exponentially decaying tone. Other aspects of the noise background are noted in Sec. III.

Because the rectangular and exponential tones sounded different, the subjects were instructed to try to ignore the tone quality differences and to concentrate on the pitch of the two tones. Some subjects volunteered that after the experiment was underway, this advice seemed easy and natural to follow.

A control experiment was run in conjunction with the R–E experiment described above. The control experiment was identical to the R–E experiment except that both tones of every pair had rectangular envelopes and were of equal amplitude. Half the subjects did the control runs first. The control runs served to rank the subjects because they involved only a simple discrimination task.

B. Results

The 15 subjects in the survey experiment were chosen haphazardly. They performed in two R–E experiment runs and two control runs. The subjects in the survey were numbered according to their performance on the control experiment. This performance is indicated in Fig. 1(a) by plotting twice the fractional JND, $K \equiv 2(\Delta f_{75} - \Delta f_{50})/f_0$ as the length of the line. This is the correct quantity to compare with the error bars in the R–E experiment shown below. Figure 1(b) shows the results of the R–E experiment plotted as follows.

The quantity $f_E - f_R$ is the difference in the frequencies of exponential and rectangular tones which have equal pitches according to the psychometric functions. The percentage change $(f_E - f_R)/f_0$, is obtained by dividing by the nominal frequency of the range, here 825 Hz.

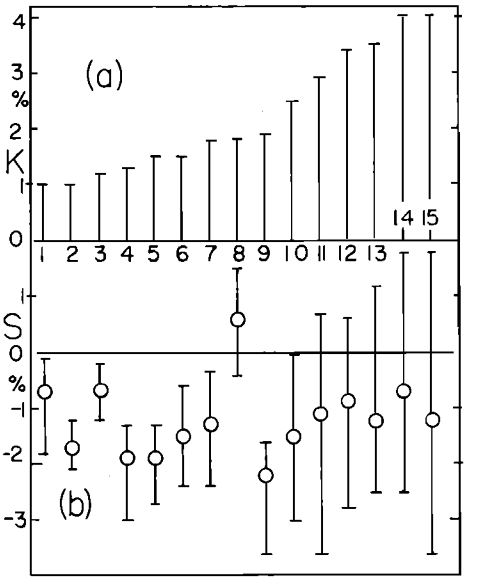

FIG. 1. The top panel (a) shows $k = 200\,(\Delta f_{75} - \Delta f_{50})/f_0$, i. e., twice the JND as a percent of the average frequency (825 Hz) for 15 subjects in the survey control experiment, comparing the pitches of two rectangularly gated sine-waves tones. The bottom panel (b) shows the corresponding results of the R–E survey experiments. Variable $S = 100\,(f_E - f_R)/f_0$, is the frequency of an exponential tone minus the frequency of the rectangular tone as a percent of the average frequency (825 Hz) for the case that the two tones have equal pitch. This point of subjective equality was determined by the 50% point on a psychometric function. The 75% and 25% points are indicated by the extremities of the error bars.

In other words, the figure shows the negative of the increase in pitch due to the exponential envelope. The circle represents the 50% point on the psychometric function. The top and bottom of the error bar represent, respectively, 75% and 25% points.

The data quite clearly reveal a shift in pitch due to the different envelopes. All but one subject concluded that the exponentially shaped tone had a higher pitch than the rectangular tone for equal frequency. For those 14 subjects showing a pitch shift of the same sign the average shift was − 11.3 Hz with a standard deviation (N weight) of 3.8 Hz. This corresponds to a shift of 1.5% of the absolute minimum frequency of 770 Hz and 1.3% of the maximum frequency of 880 Hz. The best estimate of the shift is 1.4% of the mean frequency 825 Hz. A total of five subjects perceived a shift that is greater than 1.45%, which is 25 cents (one-quarter of an equitempered semitone).

The standard deviation noted above is the standard error of the mean of the 50% point averaged across subjects. The mean of the interquartile spacing is 20.8 Hz.

As noted in Appendix B the modified staircase procedure was apparently an unnecessary precaution. Subjects did not exhibit significantly the response biases which were feared. Except for two subjects who produced unusable psychometric functions, Fig. 1 includes the results from all subjects ever tested.

II. PARAMETRIC STUDY

Three subjects, numbers 1, 2, and 12 were selected from the subjects in the survey experiment to run in a more extensive parametric study of the effect observed in Sec. I. The basic stimulus in the configuration was kept the same as that for the survey experiment, except that the noise background was extended in range to lie between 0 and 5000 Hz, while maintaining the same noise-power density 10 dB. Subjects made two runs in five minutes. Two kinds of parameter variations were carried out: an intensity variation and a frequency variation. In each of the conditions studied the subjects judged 72 staircase cycles, 4.5 times the number used in the survey. The six different conditions were done in haphazard order over five days of experimenting.

It was important first to investigate the possibility that the pitch effect observed in the R–E survey experiment is somehow exclusively a loudness effect. Initially there was reason to believe that loudness effects might be unimportant because the range of the tone frequencies 770–880 Hz is one where tone pitch is relatively insensitive to loudness variations.

In the experiment to search for loudness effects the exponential tone was the same as in the survey experiment. The rectangular tone remained 20 ms long but its amplitude was increased by 6 dB and decreased by 6 dB on alternate groups of runs. When the amplitude was increased, the rectangle amplitude was equal to the peak exponential amplitude 95 dB. In this condition the rectangular tone was unquestionably louder than the exponential tone. When the rectangular amplitude was decreased to 83 dB, it was unquestionably weaker than the exponentially decaying tone. Initially the considerable difference in amplitude made judgments difficult; with practice subjects learned to ignore loudness differences.

The three points shown in Figs. 2–4 in the region f_0 = 825 Hz in these figures allow one to compare the R–E experiments with nominal frequency of 825 Hz and rectangular amplitudes of 83, 89, and 95 dB.

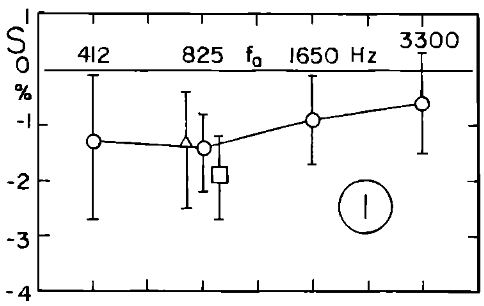

FIG. 2. R–E experiment results for subject 1. Variable S is $100\ (f_E-f_R)/f_0$, where f_0 is the nominal frequency of the range, when the two tones have equal pitches as determined by a psychometric function. The circles show the results for nominal frequencies of 412, 825, 1650, and 3300 Hz with a 89-dB rectangular tone. The points denoted by a triangle and a square show the results when the rectangular tone is presented at 83 and 95 dB, respectively, in the nominal range of 825 Hz. Each of the six points on the graph is based upon 72 staircase cycles, i.e., 576 judgments. The error bars shown extend between upper (75%) lower (25%) quartile points.

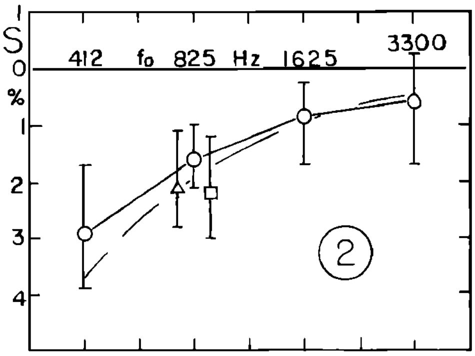

FIG. 3. R–E experiment results for subject 2. See caption for Fig. 2. The dashed curve is calculated for a constant shift of 16 Hz.

The results of this experiment make it seem highly unlikely that the pitch-shift effect is the result of overall loudness differences between rectangular and exponential tones. Despite the loudness variation, the pitch shift remains negative and, for the best ranked subjects, shows little change.

The major effect shown in Figs. 2–4 is the dependence of the pitch shift in the R–E experiment on the frequency range of the sine-wave signal. The experiments in the ranges 412, 1650, and 3300 Hz were simply done by scaling all frequencies of the stimuli in the standard 825-Hz range by a factor of the appropriate integral power of 2. All other experimental conditions were the same in all frequency ranges, e.g., the noise band was maintained at 10 dB, 0–5000 Hz.

The average of the results for subjects 1, 2, and 12 resembles closely the results for subject 2. Unfortunately the data do not permit one to eliminate conclusively either a shift which is a constant number of hertz or a shift which is a constant fraction of the frequency range f_0. Shifts of 16 Hz or 1.5% provide the best fits for those two rules. The best fit to the data, however, is a shift which increases with f_0 but does not increase as rapidly as f_0. The formula $f_R - f_E = 8 + 0.005 f_0$ provides a reasonable fit.

Auxiliary experiments with these three subjects tested for certain stimulus errors. The results are given in Appendix C.

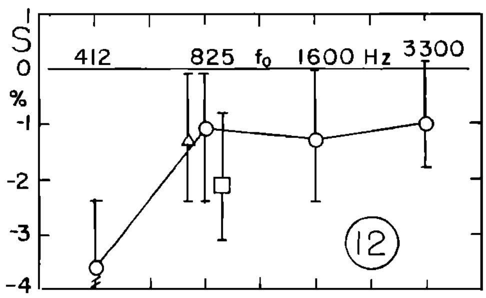

FIG. 4. R–E experiment results for subject 12. See caption for Fig. 2. The hatched error bar indicates that the lower-quartile point was not reached in the experiment.

III. DISCUSSION

A. Previous experiments

There does not seem to be previous work studying the effect of envelope on pitch perception. Studies have been made of the effect of different envelopes on pitch *discrimination*, when the envelopes for both pitches to be compared were the same. The study by Ronken (1970) includes a number of points of contact with the present work that deserve mention. In his first appendix Ronken noted that for rectangular-envelope durations as long as 20 ms the phase of the sine-wave signal at onset has no effect on discrimination. This observation supports the view that the choice of random phase in the present experiment is of little consequence.

Ronken also presented his signals in a noise background, one that was 4 dB more intense than ours. He concluded that for a decaying signal such as the exponentially enveloped tone there is an effective duration, based upon the ratio of signal power to the power in a 1 Hz band of the noise floor. According to Ronken's procedure the effective duration of the exponential tone used in the present study is 37 ms. From the work of Ronken, and others whom he quotes, it appears that the exponential and rectangular tones used in the present experiment lead to similar pitch discriminations. For rectangular tones of 20-ms duration Ronken found a JND, $(\Delta f_{75} - \Delta f_{50}) = 10$ Hz; Liang and Chistovich (1961) found JND = 6 Hz. These can be compared with the JND found in the survey control experiment (Sec. I) of 9 Hz. For an exponentially enveloped tone with 60-dB decay time of 60 ms, Ronken found JND = 6 Hz, and Stevens (1952) found JND = 8 Hz. The similarity of all these numbers suggests that the experiment is well controlled and that the exponential and rectangular tones selected are similarly discriminable.

B. Long-term spectra

The usual analysis of auditory experience classifies pressure amplitude variations according to their time scales. Pitch and tone color are associated with repetitive variations on a time scale from 10^{-2} to 10^{-4} s. For such variations a Fourier analysis is a natural representation. Amplitude variations on a time scale longer than that of the longest periodic variation or longer than 0.1 s are classified as part of an amplitude envelope. This analysis is not the only possible analysis (Gabor, 1950), but it is probably the best analysis for tones with a definite pitch because it corresponds closely with naive description of the perception of the tones. For the stimuli used in the R–E experiment there is no difficulty in identifying an envelope and a signal, with a simple Fourier transform, shaped by that envelope. Yet is is found that the pitches of the stimuli are not exclusively controlled by the underlying signal. Instead the pitch depends, though weakly, upon the amplitude envelope.

A natural theoretical approach to this problem is to preserve the assumption that the pitch is somehow determined by the spectrum of the tone but to include the

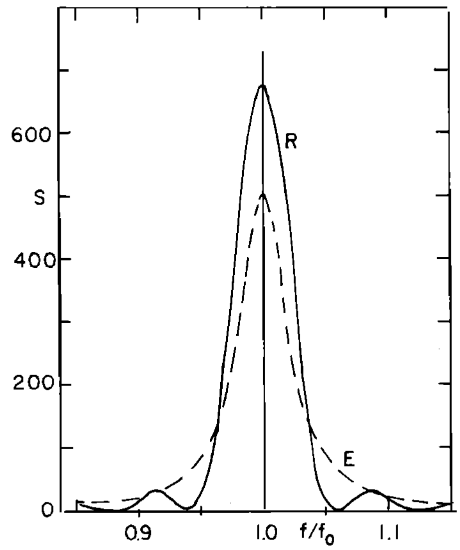

FIG. 5. The figure shows the power spectrum S_R for a rectangular tone (solid line) of 20-ms duration and the spectrum S_E for an exponential tone (dashed line) decaying at a rate of 1 dB/ms. The tones are sine waves with frequency f_0 = 825 Hz. The amplitude of the rectangular tone is half the initial amplitude of the exponential tone.

envelope in the Fourier analysis that determines the spectrum.

The rectangular (R) and exponential (E) tones have pressure functions of time,

$$p(t) = p_R \cos(\omega_0 + \phi) , \quad 0 < t < T , \qquad = 0 , \quad t > T , \tag{1}$$

and

$$= p_E \cos(\omega_0 t + \phi) e^{-Kt}, \quad t > 0 . \tag{2}$$

The power spectra of these tones are proportional to

$$|P(\omega)|^2 = \left| \int_{-\infty}^{\infty} dt\, p(t)\, e^{-i\omega t} \right|^2 . \tag{3}$$

The power spectrum, averaged over all phase angles, (see Appendix D) is

$$S = \langle |P(\omega)|^2 \rangle_\phi . \tag{4}$$

For rectangular tones

$$S_R = p_R^2 (Q_+ + Q_-) , \tag{5}$$

where

$$Q_\pm = \frac{\sin^2(\omega \pm \omega_0) T/2}{(\omega \pm \omega_0)^2} . \tag{6}$$

For exponential tones

$$S_E = p_E^2 (X_+ + X_-) , \tag{7}$$

where

$$X_\pm = \tfrac{1}{4}[(\omega \pm \omega_0)^2 + K^2]^{-1} . \tag{8}$$

The spectra for f_0 = 825-Hz tones, with rectangular envelope of duration T = 20-ms or − 1-dB/ms exponential envelope ($K = 115$ s^{-1}) are shown in Fig. 5. The spectra were computed in dimensionless units by setting $P_E = \omega_0$. To include the effect of the 6-dB reduction in rectangular tone for the equal loudness condition of the standard R–E experiment prefactor p_R was taken to be $\frac{1}{2}p_E$.

The two spectra are centered on the same frequency and have equivalent areas, as expected for equal energy tones. The detailed structure of the spectra depends upon the duration of the rectangular tone and on the exponential decay time. If these details are responsible for the shift of pitch with changing envelope then one would expect the shift to be a constant number of hertz for the present experiments where the temporal parameters of the tones were always the same. The observed frequency dependence of the shift, noted at the end of Sec. II, does not support such a conclusion, but the data do not conclusively rule it out.

C. Time-variant spectrum

More information is provided by a time-variant spectrum. The time-variant Fourier transform is

$$P(\omega, t) = \int_{-\infty}^{\infty} dt'\, e^{-i\omega t'}\, W(t, t')p(t') \ , \qquad (9)$$

where $P(\omega, t)$ is the Fourier coefficient of pressure viewed through a data window W at time t. A simple model for a data window is the exponential memory function (Flanagan, 1965)

$$W(t, t') = W(t - t') = \exp[\lambda(t' - t)]\theta(t - t') \ , \qquad (10)$$

so that events in the past contribute exponentially less to the Fourier representation and events in the future do not contribute at all. The time-variant power spectra, averaged over all initial phase angles, are given by the following expressions.

For a rectangular gated signal, Eq. (1),

$$\begin{aligned} S_R(\omega, t) &= \tfrac{1}{4}p_R^2\, e^{-2\lambda t}[z_+(\lambda, t) + z_-(\lambda, t)] \ , \quad t < T \\ &= \tfrac{1}{4}p_R^2\, e^{-2\lambda t}[z_+(\lambda, T) + z_-(\lambda, T)], \quad t \geq T \ , \end{aligned} \qquad (11)$$

where

$$z_\pm(\lambda, t) = \frac{1 + e^{2\lambda t} - 2e^{\lambda t}\cos(\omega \pm \omega_0)t}{\lambda^2 + (\omega \pm \omega_0)^2} \qquad (12)$$

For the exponentially decaying tone, Eq. (2),

$$S_E(\omega, t) = \tfrac{1}{4}p_E^2\, e^{-2\lambda t}[z_+(\lambda - K, t) + z_-(\lambda - K, t)] \ . \qquad (13)$$

The finite constant λ eliminates the rather artificial zeros in the spectrum of the rectangular tone. The spectra evolve in time in different ways for rectangular and exponential tones. For $t \ll T$ both spectra are single broad peaks centered on $f = f_0$. As time increases the peaks become sharper in each case; some wiggles may appear next to the central peak for the rectangular spectrum, ghosts of the previous undamped spectral zeros. The total energy (area) in the rectangular spectrum grows until time T and then decays. The energy in the spectrum of the exponential tone reaches its maximum at some time considerably less than T (see appendix A) and then begins its final decay. More interesting, however, is the way in which the decay proceeds for $t > T$. For the rectangular tone the shape of the spectrum is constant for $t > T$. The only effect of increasing time is the uniform decay of all parts of the spectrum. For the exponential tone, however, the spectrum continues

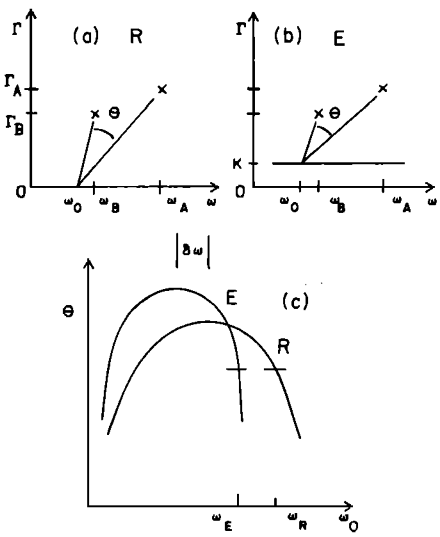

FIG. 6. The figures show schematically how two auditory filters in the Huggins model are used to determine pitch. (a) plots the complex poles for two filters A and B. For a constant sine-wave signal with frequency ω_0 the difference in phase shifts by the filters is θ. In (b) the same analysis takes place for a signal which decays with time constant K. (c) shows how θ, determined by the process in (a) and (b), varies with ω_0 for the cases of rectangular and exponential tones. It is supposed here that θ determines pitch, though some function or derivative of θ could be used instead. For given θ (pitch) the frequency of the exponential tone is less that the frequency of the rectangular tone.

to sharpen as time increases. If λ and K have similar values (regardless of which of the two is slightly larger) then the spectrum of the exponential tone can become very sharp as it decays and can develop a number of pronounced maxima and minima next to the central peak. It seems possible that this continuous dynamic change in spectral shape near f_0 changes the pitch of tones. Some subjects have, in fact, remarked that the pitch of the exponential tone rises as the tone decays. Other subjects, however, have not perceived such an effect.

D. Phase principle for complex frequency analysis

An entirely different point of view is the phase principle of Huggins (1952). Huggins postulated auditory filters in which phase shifts (not amplitude differences) are responsible for pitch perception. In his first model the difference in phase shift between two similar filters is the function of stimulus frequency that determines pitch. A major feature of Huggins' model is that it is supposed to apply to complex stimulus frequency, i.e., to damped sinusoid tones. Therefore, the model can be applied to the R–E experiment.

Suppose that the two filters have poles at $-\Gamma_A + i\omega_A$ and $-\Gamma_B + i\omega_B$, as shown in Fig. 6. The signal is represented by a pole at $-K + i\omega_0$. The differential phase shift θ from the two filters can easily be calculated from the geometry of the complex s plane. The value of the function $\theta_K(\omega_0)$ determines the pitch of the signal. A change in θ due to a change in ω_0 or due to a change in K then changes the pitch. If $\Gamma_A = \Gamma_B$, the situation discussed by Huggins, then the addition of damping to a signal does not change the center frequency of θ, only the sharpness of the peak is affected. Suppose, however, that for $\omega_A \gtrsim \omega_B$, $\Gamma_A \gtrsim \Gamma_B$. Then increasing the damping changes the location of the peak of function θ.

One plausible form for the variations of Γ and $\omega_A - \omega_B$ is to assume that Γ_A is proportional to ω_A, and $\omega_A - \omega_B$ scales with frequency range, i.e., both the damping and the "phase bandwidth" increase proportional to frequency. Then the frequency shift can be related to the properties of a single filter. The filter is not a sharp one, not necessarily a bandpass filter, and the definition of a filter Q in terms of half-power points does not apply. The definition of a generalized Q is just the ratio of the frequency of a pole to twice the damping constant of the pole. Let b be the ratio of filter frequencies, $b = \omega_B/\omega_A$. Then the frequency shift in the peak of $\theta(\omega_0)$ due to finite damping of the signal K is

$$\delta\omega = 2QK + [1+(2Q)^{-2}]^{1/2} \times [(\omega_A b - 2QK)^{1/2}(\omega_A - 2QK)^{1/2} - \omega_A b^{1/2}] \ . \quad (14)$$

The sign of $\delta\omega$ is negative for $K>0$; therefore the function θ shifts to lower frequencies. This is the right direction needed to produce a higher pitch with constant signal frequency and increasing signal damping. It is a good approximation to replace ω_A in Eq. (14) by ω_0. Then, to fit the R–E experiment at 825 Hz with $\delta\omega = 2\pi$ (12 Hz) and $K = 115\ \mathrm{s}^{-1}$ with b between 0.9 and 1.0 requires that Q be about equal to $\frac{1}{3}$. As noted by Huggins the phase theory is compatible with a strongly damped system. One can expand the square roots in Eq. (14) in powers of QK/ω_0. To first order, a good approximation in the regime of interesting parameters, $\delta\omega$ is independent of frequency ω_0. Therefore, this scaled model predicts that the pitch shift should be a constant number of hertz, the same result suggested by spectral theories.

IV. CONCLUSION

A survey experiment with 15 subjects showed that a tone with an exponentially decaying amplitude envelope has the same pitch as a tone with a rectangular envelope if the frequency of the exponential tone is lower than that of the rectangular tone. This effect was interpreted as a pitch shift with changing envelope. Comparison with an independent pitch-discrimination experiment showed that the pitch shift is not correlated with a subject's ability to discriminate pitches. Parametric studies with three subjects showed that the shift is essentially independent of overall relative loudness of the rectangular and exponential tones. Experiments in four different frequency ranges suggested that the shift is neither a constant number of hertz nor a constant fraction of the sine-wave frequency, but exhibits an intermediate dependence on frequency.

Further experiments showed that the pitch shift effect was unaffected by low-pass filtering of the rectangular tone or by truncating the exponential tone above the background noise level.

Three calculations were done that seem relevant to models of the auditory system. (1) The long-term spectra for rectangular and exponential tones were compared. (2) Time-variant spectra with an exponential causal window for the two tones were compared. (3) The phase principle of Huggins was modified to make the auditory filters constant Q, and the principle was applied to the rectangular and exponential tones.

The spectral calculations (1) and (2) may be relevant to a pitch theory in which the pitch sensation depends upon the details of a neural excitation pattern along some tonotopic coordinate. It is possible that the different shapes of the long-term spectra (Fig. 5) correlate with different patterns of excitation (or inhibition) that produce the pitch differences observed in the R–E experiment. However, the two long-term spectra are not dramatically different and there is no compelling reason to expect that they produce different pitch sensations. The time-variant spectra of the two tones, however, may differ dramatically, especially when the decay time of the temporal window is similar to the decay time of the exponential tone.

The Huggins model was invented to show how several broadly tuned filters could be used to achieve sharp frequency discrimination. The model is easily and naturally applied to the R–E experiment. In its original form the model predicts no pitch difference. If the model is modified so that the filters are constant Q then the model predicts a pitch difference in the right direction to agree with the results of the R–E experiment.

In the simplest forms the above three calculations predict that the R–E experiment should find a pitch difference which is a constant number of hertz for all frequencies. Obviously these theoretical ideas are speculative. Specific predictions of models based upon these ideas can be checked by further experiments using different temporal parameters for the envelopes and different envelope shapes.

ACKNOWLEDGMENTS

The hospitality of Professor David M. Green at the Laboratory of Psychophysics is gratefully acknowledged. This paper has benefited from helpful comments by David Green, Daniel Weber, Stephen Burbeck, and Edward Burns. The work was supported by NSF grant number BNS 76-20225.

APPENDIX A: ENERGY IN A PULSE

This appendix evaluates the acoustical energy in a burst of a sine wave

$$p(t) = p_0 \cos(\omega_0 t + \phi) \quad (A1)$$

turned on at time $t=0$. The constant pressure p_0 will be p_R for a rectangular envelope and p_E for an exponential envelope. For a rectangular envelope of unit amplitude and duration T the energy depends upon the phase angle at turn on and turn off. For unit acoustical impedance the average over all phase angles is

$$\langle E_R \rangle_\phi = \tfrac{1}{2} p_R^2 T \ . \quad (A2)$$

For specific phase relationships the variation may be at most

$$\Delta E_R = \pm p_R^2/\omega_0 \ . \quad (A3)$$

The largest variation in the experiments of this paper, for $T = 20$ ms and $f_0 = 400$ Hz, is only $\pm 2\%$ of the average, completely negligible. If the amplitude envelope is $\exp(-t/\tau)$ then the energy in the tone burst depends upon the phase angle ϕ. The average energy over all phase angles is

$$\langle E_E \rangle_\phi = \tfrac{1}{4} p_E^2 \tau . \quad \text{(A4)}$$

Maximum and minimum energy phase angles satisfy the equation

$$\omega_0 \tau = \tan 2\phi \ . \quad \text{(A5)}$$

They differ by $\frac{1}{2}\pi$, but they are not 0 and $\frac{1}{2}\pi$. The maximum variation as a fraction of the average energy is

$$\Delta E_E / \langle E_E \rangle_\phi = \pm [1 - (\omega_0 \tau)^2][1 + (\omega_0 \tau)^2]^{-3/2} \ , \quad \text{(A6)}$$

which is $(\omega_0 \tau)^{-1}$ for $\omega_0 \tau \gg 1$. For a decay rate of 1 dB/ms, $\tau = 8.686$ ms and the maximum variation, for $f_0 = 400$ Hz, is only 5% of the average.

In the standard condition of the experiments in this paper the constant rectangle sound level is 6 dB less than the maximum level of the exponential tone, i.e., $p_R = \frac{1}{2} p_E$. The energy ratio then becomes $E_R / E_E = 1.15$, negligibly different from unity. These two tones are judged equally loud; the judgements of the ear, in this case, agree with the measure of total signal energy.

The energy at time t in the Fourier transformer introduced in Sec. III is

$$\epsilon(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega, t)\, d\omega \quad \text{(A7)}$$

$$= \int_{-\infty}^{\infty} dt' \, W^2(t, t')\, w(t') \ , \quad \text{(A8)}$$

where w is the instantaneous signal power $= p^2$, averaged over all initial phase angles.

For the rectangular tone

$$\begin{aligned} \epsilon_R(t) &= (p_R^2 / 4\lambda)(1 - e^{-2\lambda t}) \ , && t \le T \\ &= (p_R^2 / 4\lambda)\, e^{-\lambda t}(e^{2\lambda t} - 1) \ , && t \ge T \end{aligned} \quad \text{(A9)}$$

and the maximum power is $\epsilon_R(T)$. For the exponential tone

$$\epsilon_E(t) = [p_E^2 / 4(\lambda - K)](e^{-2Kt} - e^{-2\lambda t}) \ . \quad \text{(A10)}$$

The maximum energy is defined in terms of the ratio

$$r \equiv \lambda / K \ . \quad \text{(A11)}$$

$$\epsilon_{E,\max} = \frac{p_E^2}{4K} \, r^{r/(1-r)} \quad \text{(A12)}$$

and occurs at time

$$t = \ln(r) / [2(r - 1)K]. \quad \text{(A13)}$$

The interesting exponential function in (A12) has value 1 at $r = 0$, $1/e$ at $r = 1$, and tends to the function $1/r$ for large r.

APPENDIX B: SHIFTED STAIRCASE METHOD

In the R–E experiment subjects compared the pitch of a rectangular tone (R) with that of an exponential tone (E). The difference in frequencies of these tones $f_R - f_E$ was the experimental variable of principal interest. It took on the values −30, −20, −10, 0, 10, 20, and 30 Hz in a staircase pattern. Because subjects could easily distinguish between rectangular and exponential tones it was possible that response bias might be present in the judgments. Subjects might have attempted to use each of the two possible responses equally often or to base their judgments on some feature of the tones other than pitch.

To check for response bias of this type the staircase method was modified as follows. The range of $f_R - f_E$ was divided into two asymmetrical blocks that overlapped. On one block of trials the staircase values were shifted down; they were −30, −20, −10, 0, 10, 0, −10, −20, ... Hz. On the other block of trials the staircase values were shifted up; they were −10, 0, 10, 20, 30, 20, 10, 0, ... Hz. Separate psychometric functions were drawn for the up-shifted and down-shifted blocks. The following reasoning applied. Suppose that the judgments were free of response bias. Then the individual psychometric functions for up and down-shifted blocks would both be part of a common psychometric function for the entire range of values of $f_R - f_E$. The difference between the two psychometric functions at the three overlapping points would be zero. Suppose on the other hand, that subjects tended to use the two responses equally often. Then, for psychometric functions with positive slope, at points of overlap the function for the down-shifted block minus that for the up-shifted block would be a positive number.

Application of this test to the R–E experiment suggested that no significant response bias was present. The difference between the shifted psychometric functions was positive for nine subjects and negative for five subjects out of 15. The difference divided by the sum of the psychometric functions was less than 0.05 for subjects numbered 1–12.

Note that besides testing for response bias the shifted staircase method has a second advantage over the standard staircase technique. It tends to concentrate data points in the middle of the range of parameters where the psychometric function is close to 50%. Therefore, the method leads to a more constant variance.

APPENDIX C: A SEARCH FOR ARTIFACTS

Initially it seemed possible that the pitch shift reported in Secs. I and II might be caused by one of several possible stimulus artifacts, ringing of the stimulus system or some effect associated with the background noise. This section serves to eliminate these explanations for the pitch difference.

A. Ringing

Because the tones in this experiment are relatively short and involve sharp onset and offset transients, it is important that the physical system creating the stimulus not ring. It will be noted later in this section that ringing can cause considerable changes in the results of the R–E psychophysical experiment. Two tests were made to verify that ringing was not present in the stimulus system. First, physical measurements were made at one of the matched TDH 39 earphones in the circuit used in the R–E experiments. Measurements were made with B & K type 4145 condenser microphone and a 6-cm^3 coupler, ASA type 1. A B & K sound level meter, type 2203 on the fast linear scale, was used as a preamplifier. The earphone output was observed on an oscilloscope. The impulse response of the system showed no oscillatory structure. With the rectangular pulse, as used in the R–E experiment, the trailing edge similarly showed less than $\frac{1}{2}$ cycle of oscillation at 412 and 825 Hz.

A second test replaced the TDH 39 headphones with Beyer DT-48 headphones with B2-03-00 foam cushions. It seemed likely that if the pitch shift were caused by a transient distortion then the most likely source of the distortion was in the electromechanical conversion at the headphones, or in the circumaural cavity. One subject, number 2, ran 48 cycles of the survey R–E experiment with the DT-48 headphones. The psychometric function obtained had a 50% point and 25% and 75% points within 2 Hz of the corresponding values obtained with the TDH 39 headphones. In sum, it seems evident that physical ringing was not a factor in the experiments of Sec. I and II.

If physical ringing does enter the stimulus system, however, the effect can be dramatic. The effect of ringing was noted by repeating the R–E experiment with two new conditions and four subjects, numbers 1, 2, 12, and 15. In the first condition the signal was high-pass filtered at 600 Hz after amplitude shaping. In the second condition the signal was low-pass filtered at 1000 Hz after amplitude shaping. The filter was a Krohn-Hite model 3343 with 24-dB/oct asymptotic slope. Otherwise the stimuli were identical to the 825-Hz range signals used in the survey experiment. The high-pass filter introduced considerable ringing into the transient response of the entire system. At least five complete cycles with period about equal to $\frac{1}{600}$ s appeared at the earphone in the impulse response of the system with this filter. The impulse response of the maximally flat low-pass-filtered system, by contrast, included only two small bumps at about 1 and 2 ms on the side of the overall decay. Neither filter introduced amplitude changes greater than $\frac{1}{2}$ dB for all the frequencies presented.

The four subjects made judgments on 48 cycles of the standard survey experiment with each of these filters. The effects of the high-pass-filtered system are dramatic and quite similar for all four subjects. The high-pass filtering doubles the pitch difference, compared with the unfiltered condition of Sec. II. A plausible explanation for the results is as follows. According to Nabelek *et al.* (1973) the final part of a moving tone is most important for determining pitch. The high-pass filter has no effect on the end of the exponential tone, but it adds a low-frequency tail to the waveform of rectangular tone as the filter rings, near 600 Hz, after signal offset. The low-frequency tail lowers the pitch of the rectangular tone. Therefore, the exponential tone is perceived to be higher than the rectangular tone on an increasing fraction of the trials. The effect on pitch of the high-pass filtering is subtle, in that subjects are unaware of the effect.

The results of the R–E experiment with the low-pass-filtered system, with negligible ringing, are almost identical to those in the original unfiltered condition. Rectangular gated tones that have been low-pass filtered do sound different from those which have not been filtered. The former sound like a chirp, the latter sound like a cluck. Nevertheless, the pitch comparison experiment suggests that low-pass filtering produces negligible change in the pitch of the rectangular tone.

B. Background noise

A second kind of artifact that must be considered involves possible effects of the background noise on the pitches of the tones in the R–E experiment. The pitch of a sine wave tone is increased by adding broadband noise (Webster and Muerdter, 1965). One must consider the possibility that such an effect is operating in the present experiment. The following discussion argues that the noise background does *not* play a significant role in the pitch shift reported in this paper.

Firstly the pitch shift attributed here to amplitude envelope changes is more than three times larger than the 7-cent shift attributed to noise by Webster and Muerdter. Secondly, the signal to noise ratio is much greater in the present experiment (62 dB for rectangular tones) than in the experiment by Webster and Muerdter (apparently 26 dB).

Finally an auxiliary experiment suggested that the pitch of the exponential tone is not significantly affected by masking by background noise. In the auxiliary experiment the exponential tone was truncated after 43 ms. The truncation ensured that the instantaneous signal power was always at least 20 dB greater than the noise power in a critical band at the signal frequency. The author ran the R–E experiment with 36 cycles with a truncated exponential interleaved with 36 cycles using the standard exponential decay. The two decaying tones were found to be almost indistinguishable. The pitch shift found was 1.2%, close enough to the standard − 1.4% to conclude that noise is not a significant factor in the R–E experiment.

APPENDIX D: RANDOM PHASE ASSUMPTION

Consider a signal represented by Fourier components and an envelope V.

$$p(t) = V(t)\sum_i p_i \cos(\omega_i t + \phi_i) \quad . \tag{D1}$$

The Fourier transform,

$$P(\omega) = \int_{-\infty}^{\infty} dt\, e^{i\omega t}\, p(t) \quad , \tag{D2}$$

includes contributions associated with $\omega = +\omega_i$ and $\omega = -\omega_i$. The power spectrum $|P(\omega)|^2$ generally includes cross terms. This appendix notes that averaging the power spectrum over equally probable phase angles causes such cross terms to vanish.

There are two cases of interest. In case 1 the phase relationships among the components i, which may be harmonics, are fixed. Only one phase angle ϕ is a free parameter that relates the Fourier components to the envelope. This case obtains when, for example, a triangle wave is gated on, at random phase ϕ_1, by an electronic switch. Because the average over phase angles

$$\langle e^{i\phi_1}\rangle_1 = 0\,, \tag{D3}$$

the power spectrum does not include cross terms from positive and negative frequency half-planes, i.e.,

$$S(\omega)=\sum_i \tfrac{1}{4}P_i^2\{|V(\omega+\omega_i)|^2+|V(\omega-\omega_i)|^2\}$$
$$+\sum_{\substack{i,j\\(i\neq j)}} \tfrac{1}{4}P_iP_j\{V(\omega+\omega_i)\,V^*(\omega+\omega_j)\exp[i(\phi_j-\phi_i)]$$
$$+V(\omega-\omega_i)\,V^*(\omega-\omega_j)\exp[i(\phi_i-\phi_j)]\}\,. \tag{D4}$$

In case 2 the Fourier components are not harmonically related or otherwise correlated in phase; then an average over independent phase angles causes the second sum in Eq. (D4) to vanish. The results of this appendix are clearly still applicable when $V(t')$ in Eq. (D2) is replaced by the expression used in a time-variant power spectrum $V'(t')=V(t')\,W(t,t')$ so long as a Fourier transform $V'(\omega)$ exists.

Christman, R. J., and Williams, W. E. (1963). "Influence of the time interval on experimentally induced shifts of pitch," J. Acoust. Soc. Am. 35, 1030–1033.

Flanagan, J. L. (1965). *Speech Analysis, Synthesis and Perception* (Academic, New York).

Gabor, D. (1950). "Communications Theory and Physics," Philos. Mag. 41, 1161–1187.

Hartmann, W. M., and Blumenstock, B. J. (1976). "Time dependence of pitch perception-pitch step experiment," J. Acoust. Soc. Am. 60, S40(A).

Huggins, W. H. (1952), "A phase principle for complex-frequency analysis and its implications in auditory theory," J. Acoust. Soc. Am. 24, 582–589.

Liang, C., and Chistovich, L. A. (1961), "Frequency difference limens as a function of tonal duration," Sov. Phys. Acoust. 6, 75–80. Nabelek, I. V., Nabelek, A. K., and Hirsh, I. J. (1973). "Pitch of sound bursts with continuous or discontinuous or discontinuous change of frequency," J. Acoust. Soc. Am. 53, 1305–1312.

Ronken D. A. (1970). "Some effects of bandwidth duration constraints on frequency discrimination," J. Acoust. Soc. Am. 49, 1232–1240.

Stevens, K. N. (1952). "Frequency discrimination for damped waves," J. Acoust. Soc. Am. 24, 76–79.

van den Brink, G. (1970). "Experiments on binaural displacusis and tone perception," in *Frequency Analysis and Periodicity Detection*, edited by R. Plomp and G. F. Smoorenburg (A. W. Sijthoff, Leiden).

Verschuure, J., and van Meeteran, A. A. (1975). "The effect of intensity on pitch," Acustica 32, 33–44.

Webster, J. C., and Muerdter, D. R., (1965). "Pitch shifts due to low pass and high pass noise bands," J. Acoust. Soc. Am. 37, 382–383.