Turning on a tone

William Morris Hartmann and Dan C. Sartor
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 5 October 1990; revised 14 January 1991; accepted 1 April 1991)

It is possible to choose the starting phase of a pure tone in a way that minimizes the onset noise when the tone is turned on abruptly. A spectral model shows that when the tone has a low frequency, minimum onset noise is expected for a starting phase of zero (turning on a sine tone) but when the tone has a high frequency, minimum onset noise is expected for a starting phase of ± 90 deg (turning on a cosine tone). Listening experiments confirm the above expectations and show that the transition between low- and high-frequency domains is sharp and depends upon both the electroacoustical transducer and the individual listener.

PACS numbers: 43.66.Nm, 43.60.Qv, 43.66.Yw [WAY]

INTRODUCTION

When an acoustical signal is turned on abruptly, there is an onset transient that is heard as a click or pop or thud. Often this transient is regarded as an unwanted noise, and efforts are made to reduce its effect. Unfortunately, the noise is particularly prominent if the signal is a pure tone, which happens to be the most useful signal of all in the study of hearing.

It is possible to reduce the effect of the onset transient by choosing the starting phase of the tone. A reasonable first guess is that the transient is minimized if the starting phase is zero so that the waveform is a "sine tone," as shown in Fig. 1(A). In the literature, this is often described as turning on the tone in "sine phase." By way of justification, one might use a spectral argument which says that the sine phase leads to a waveform without a discontinuity and therefore produces the minimum spectral splatter. At least one popular introductory textbook in hearing science makes this claim, and there is at least one electronic gating device used in psychoacoustical experiments that is designed according to this principle, turning on signals in sine phase. However, as will be shown below, this qualitative spectral argument is incomplete and the conclusions to which it leads are sometimes wrong.

I. THE ABRUPT SIGNAL AND ITS SPECTRUM

The abrupt pure-tone signal is defined as

$$x(t) = 0 \quad \text{for} \quad t<0,$$
$$x(t) = A\sin(\omega_0 t + \phi) \quad \text{for} \quad t \geqslant 0, \tag{1}$$

where ω_0 is the angular frequency of the tone ($=2\pi f_0$, where f_0 is the frequency in Hz), and A is the amplitude. Phase angle ϕ is the starting phase; it is zero for sine phase and ± 90 deg for cosine phase.

With A set equal to unity, the Fourier transform of $x(t)$ is given by

$$X(\omega) = \int_{-\infty}^{\infty} dt\, e^{-i\omega t} x(t)$$
$$X(\omega) = \frac{\omega_0 \cos(\phi) + i\omega \sin(\phi)}{\omega_0^2 - \omega^2}, \tag{2}$$

where i is $\sqrt{-1}$. The power spectrum is the squared magnitude of X,

$$P(\omega) = [\omega_0^2 \cos^2(\phi) + \omega^2 \sin^2(\phi)]/(\omega_0^2 - \omega^2)^2 \tag{3}$$

Equation (3) has two limits of special interest: sine phase ($\phi = 0$), where

$$P(\omega) = [\omega_0/(\omega_0^2 - \omega^2)]^2, \tag{4a}$$

and cosine phase ($\phi = \pm 90$), where

$$P(\omega) = [\omega/(\omega_0^2 - \omega^2)]^2. \tag{4b}$$

The spectral functions in Eqs. (4a) and (4b) are plotted in Fig. 2. The implications of these functions are clear. The spectrum has a peak when ω is equal to ω_0, the frequency of the pure tone. The rest of the spectrum represents splatter from the abrupt onset. For frequencies well above ω_0, the power spectrum of the abrupt *sine* tone [Eq. (4a)] decreases as the inverse fourth power of the frequency, whereas the power spectrum for the abrupt *cosine* [Eq. (4b)] decreases only as the inverse square. As a result, the abrupt cosine leads to more high-frequency power than the abrupt sine. The high-frequency splatter is perceived as a click sound, and it is smallest for sine phase. If the frequency of the tone is low, such that splatter to the higher frequencies is perceptually important, then using sine phase minimizes the noise.

The price that one pays for using sine phases is that the spectrum does not go to zero at low frequency. By contrast, the spectrum for cosine phase vanishes as the square of the frequency as the frequency goes to zero. As a result, the abrupt sine leads to more low-frequency power than the abrupt cosine. The low-frequency splatter may be perceived

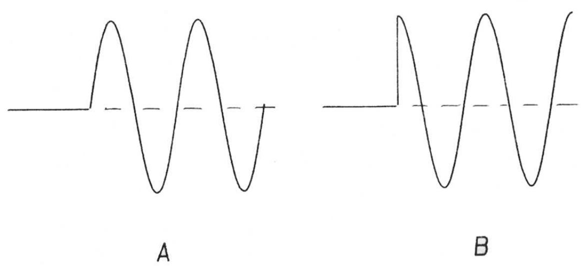

FIG. 1. (A) A waveform with zero phase (sine phase), and (B) a waveform with 90-deg phase (cosine phase).

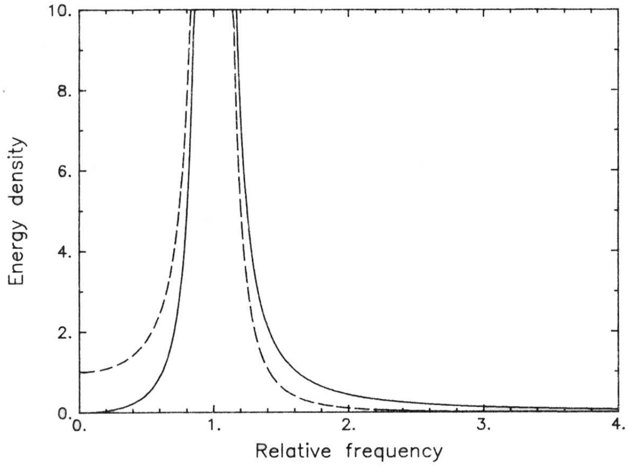

FIG. 2. Power spectra calculated for sine phase (dashed line) and cosine phase (solid line) from Eqs. (4a) and (4b), respectively. The vertical axis shows energy density, in units of $1/\omega_0^2$. The horizontal axis is the frequency parameter relative to the sine tone frequency, i.e., ω/ω_0.

as a thud, and it ought to be smallest for cosine phase. If the frequency of the tone is high, such that splatter to the lower frequencies is perceptually important, then using cosine phase should minimize the noise.[1]

II. FORCED-CHOICE EXPERIMENTS

The spectral calculation of Sec. I predicts that the starting phase that leads to the least onset noise, or to the most, depends upon the frequency of the tone. To test that prediction, we did psychoacoustical experiments, with listeners comparing abrupt tones having different starting phases. In separate runs a listener heard high- and low-frequency tones.

A. Methods

The listener was in a sound-isolated room, listening monaurally to tones through the right earphone of a Sennheiser HD480 headphone set.[2] The tones were turned on abruptly with a carefully controlled starting phase. Five different starting phases were used: -90, -45, 0, 45, and 90 deg. (Positive starting phase was defined as an initial outward motion of the diaphram.) These phases were paired with one another to make 20 possible pairs of tones for a two-interval forced-choice experiment (all permutations except that no starting phase was ever paired with itself in a trial).

During an experimental run, the frequency and level of the tones were kept fixed, and the listener heard five passes through the 20 possible pairs, making a total of 100 forced-choice judgments. On each pass the pairs were presented in a different random order. The listener's task was to press one of two buttons indicating which tone had the greater onset noise. There was no feedback.

Each tone had a full-on duration of 70 ms and was turned off smoothly with a raised-cosine envelope of 15-ms duration. Therefore, turn-off noise played no role. The gap between the two tones of a trial was 600 ms. The experiment was self paced. After a listener made a response, there was a gap of 1 s before the next pair of tones appeared.[3] A typical run lasted about 4 min. The level of the tones was set to be 80 dB SPL during full-on portions. The level was set at 1000 Hz using a sound level meter and a flat-plate coupler.

Tones were generated digitally using a PDP 11/73 computer and a DS16 buffer and 16-bit digital-to-analog converter from Microtechnology Unlimited. The sample rate was 48 kHz, and the output from the DAC was filtered by a Wavetek-Rockland 752A Brickwall filter, set to pass the band below 18 kHz. The signal rise time was therefore approximately 1000/18, or 56 μs. Waveforms, with the desired frequency and starting phases, were computed on-the-fly, using an efficient procedure, and loaded into 4096 words of the buffer between the experimental intervals. The signals were turned off by a computer-controlled amplifier before the end of the buffer readout.

Experimental runs were done with sine tone frequencies of 125, 250, 500, 1000, 2000, and 4000 Hz. Then, depending upon the results for each listener, runs at intermediate frequencies were sometimes done to sharpen the psychometric function. The intermediate frequencies were 300, 400, 707, and 1414 Hz.

B. Listeners

There were four listeners, one female (A, age 20) and three males (B, age 50; C, age 19; and D, age 21). Listeners B and D were the authors; listeners A and C did not know about onset phase or its role in the experiment. In preparation for the experiment, the listeners did a training session in which they heard a series of 1000-Hz tones with smooth envelopes. The envelopes had 15-ms raised-cosine attacks and decays, and full-on durations of 50 ms. Listeners were asked to note the difference between these smooth tones and abrupt tones from the experiment, also 1000 Hz. It was explained that the difference was due to onset noise and that the experimental task would be for the listener to choose the tone that had the greater onset noise.

C. Results

A listener's judgments were sorted according to the phases of the first and second tones of a trial to search for several effects as follows.

Sine-cosine effect: A large positive value means that when a sine (0 deg) is compared with a cosine (either -90 or $+90$ deg), the sine makes more noise. Trials using an onset phase of 45 deg did not contribute to this statistic.

Sine effect: A large positive value means that when a sine (0 deg) is compared with any other phase (-90, -45, 45, or 90 deg) the sine makes more noise. Half of the sine effect is attributable to the sine–cosine effect; the other half depends upon comparison with 45-deg starting phase.

Cosine effect: A large positive value means that when a cosine (either 90 or -90 deg) is compared with any other phase (-45,0, or 45) the cosine makes more noise. One-third of the cosine effect is attributable to the sine–cosine effect; the rest depends upon comparison with the 45-deg starting phase.

Normalization: The above effects are normalized by their maximum possible values so that results reported below are in the range from $+1$ (the effect is present 100% of the time) to -1 (the opposite effect is present 100% of the time.)

For each listener the three effects were plotted as psychometric functions against the frequency of the pure tone, with intermediate frequencies chosen to lead to the most informative psychometric function. Then the entire experiment was repeated to give a second set of psychometric functions. These were averaged with the first set to give final functions, shown in Fig. 3 A–D.

The psychometric functions for the sine–cosine effect, shown in the figures by closed circles, indicate little indecision on the part of listeners. Values tend to be close to $+1$ or -1. Values between -0.5 and 0.5 occur only rarely, indicating strong opinions one way or the other. As expected from the spectral argument in Sec. I, the sine–cosine functions show that, for low-frequency tones, the cosine phase made more noise (negative sine–cosine effect) and, for high-frequency tones, the sine phase made more noise. The crossover point between these two domains was different for each listener: 1000, 500, 300, and 1200 Hz for A, B, C, and D, respectively.

The sine effect and the cosine effect, given by symbols S and C, are strong when they are negative but less strong when they are positive. Our interpretation of this result is that a starting phase of 45 deg is generally rather noisy. The sine effect will serve as an example: For a low-frequency tone there is a strong negative sine effect because both the cosine phase and the 45-deg phase make more noise than the sine phase. The choice of a starting phase of zero (sine phase) is especially quiet for a low-frequency tone and a starting phase of 45 deg does not approach its quietness.

For a high-frequency tone there is a positive sine effect, but the effect is not particularly strong. Although the sine phase makes more noise than the cosine phase, the sine phase is not appreciably noisier than the 45-deg phase. Therefore, the sine-phase tone is not always judged to be louder than the 45-deg phase tone. The choice of a starting phase of 90 deg (cosine phase) is especially quiet for a high-frequency tone and a starting phase of 45 deg does not approach its quietness.

Because the sine effect and the cosine effect are never strongly positive, it is clear that the starting phase of 45 deg is not quiet, but neither is it especially noisy. It is not hard to prove that if a starting phase of 45-deg were particularly noisy, then the sum of the sine effect and the cosine effect would be strongly negative. Experimentally, however, sine and cosine effects usually have opposite signs so that they cancel, and the sum is never strongly negative or positive.

D. Transducer dependence

When a signal is turned on abruptly, the entire signal processing chain is shock excited. We therefore expect that the transient response of the transducer will play an important role in experiments in which the onset noise is evaluated. If the linear response adequately characterizes the transient behavior, then the frequency response, amplitude and phase, is an adequate measure of transient response. The experiments described above used headphones with a wide

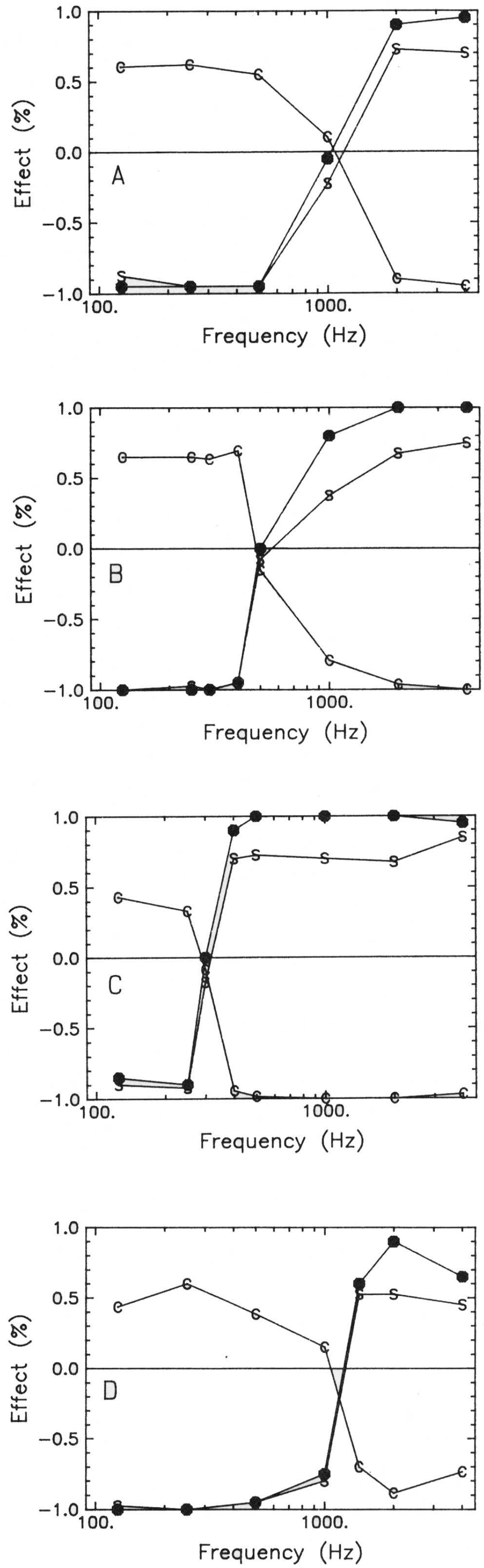

FIG. 3. Psychometric functions for four listeners, A, B, C, and D. The solid circles show the sine–cosine effect. When the function is 1.0, the sine phase is judged to be noisier than cosine phase 100% of the time. When the function is -1.0, the opposite is true. Symbols S and C show the sine effect and the cosine effect, respectively.

amplitude response. The experiments in the present section used two different transducers with narrower frequency ranges.

1. Transducers

A first set of runs used a TDH49 headphone with an MX-41/AR cushion. This headphone is a standard for American audiometers; it also has poor low-frequency response, in part because the cushion rests flat on the pinna and does not seal.

A second set of runs used a transducer from the opposite end of the electroacoustic range, a woofer. The woofer experiments employed a large Advent (circa 1975) loudspeaker with its crossover network and high-frequency driver internally disconnected. The speaker thereby became a single 10-in. acoustical-suspension driver in a sealed box. The speaker was placed in the sound-isolated enclosure 59 in. from the listener's head with the woofer at head level. The frequency response, measured with an omnidirectional microphone at three positions near the listener's head, was relatively flat at low frequencies up to 2 kHz, dropped by 25 dB from 2 to 4 kHz, and was then flat from 4 to 10 kHz. Positive starting phase corresponded to an initial outward motion for the woofer cone. Apart from the change in transducers, the TDH49 experiments and the woofer experiments were the same as the experiments described above.

2. Results

The form of the psychometric functions with the TDH49 headphones and with the woofer was the same as in the case of the HD480 headphones described in subsection C above. Again, the psychometric functions were sharp, indicating an abrupt crossover between the two domains. The crossover frequencies, however, were changed by the changes in transducer, as shown in Table I. Using the headphones with poor low-frequency response increased the crossover frequency for all four listeners, by as much as an octave for two of them. Using the woofer with poor high-frequency response decreased the crossover frequency for three of the four listeners.

These changes are in agreement with expectation based upon the spectral argument. When the low-frequency response is weak, the low-frequency splatter created in sine phase is reduced. Therefore, the frequency of the tone must be increased in order for the thud from the sine-phase onset to be as noisy as the click from the cosine-phase onset. Similarly, when the high-frequency response is weak, the click created in cosine phase is reduced. Therefore, the frequency of the tone must be decreased in order for the high-frequency splatter caused by cosine-phase onset to be as strong as the low-frequency splatter of the sine-phase onset.

TABLE I. Crossover frequencies in Hz, between the domain where cosine phase is noisier than sine phase and the domain where the reverse occurs. At the crossover, the sine–cosine effect changes sign. There are four listeners and three transducers.

Transducer \ Listener	A	B	C	D
HD480	1000	500	300	1200
TDH49	2000	1000	400	1700
Woofer	600	280	400	800

III. ADJUSTMENT EXPERIMENTS

The experiments of Sec. II clearly showed that the starting phase that must be used in order to minimize onset noise depends upon the frequency of the signal. The experimental evidence was quantitative and unambiguous. However, the experiments used only five different starting phases, and it was not possible to learn whether some intermediate starting phase might produce more noise, or less noise. The experiments of the present section were run in order to fill the gaps left by the forced-choice experiments. In contrast to the formal forced-choice experiments, the present experiments used an adjustment method; they were informal and introspective.

A. Methods

The listener was seated in a sound-isolated room, listening to signals through the right earphone of a Sennheiser HD480 headphone set. The signals were identical to those of Sec. II; tone pulses were repeated indefinitely at a rate of 72 pulses per minute.

By means of a ten-turn potentiometer on a response box the listener could adjust the starting phase of the tone. Control was done in the digital domain, the computer reading the setting of the potentiometer and then changing the computation of the next tone pulse accordingly. With a pushbutton the listener could change the frequency of the tone. The listener's task was to move around in the space of starting phases and tone frequency and to observe changes in the loudness of the clicks, the thuds and the noise in general. Frequencies of 125, 250, 500, 1000, 2000, and 4000 Hz were examined, as in the forced-choice experiments. The authors, listeners B and D, participated in these experiments.

B. Results

The observations in the adjustment experiments were consistent with the results in the formal forced-choice experiments. For a low-frequency tone, 125, 250, and 500 Hz, the click sound was dominant in the onset noise. As the starting phase was increased from zero, a click was heard after a threshold phase had been exceeded; the threshold was never more than 20 deg. The click increased as phase increased but the increase was negatively accelerated or else reached a plateau before a phase of 90 deg. This dependence of the click loudness on phase agrees with the observation in the forced-choice experiments that for a low-frequency tone the *cosine effect* is rarely larger than 1/3, the value that can be attributed to the *sine–cosine* effect. In other words, a 90-deg starting phase does not make appreciably more noise than a 45-deg onset phase. Both are a great deal noisier than a zero-degree starting phase.

For the higher frequencies, 1000, 2000, and 4000 Hz, the low-frequency thud was dominant in the onset noise, a sound that decreased as the starting phase increased. The loudness of low-frequency noise as a function of phase tended to be rather flat near 0 deg, decreasing to small values between 30 and 60 deg.

We conclude that the loudness of onset noise, either high- or low-frequency, tends to be insensitive to phase at values of the phase where the loudness is large, and only to depend strongly on phase where the loudness is becoming small. As a result, there is important overlap between the regions of thud and click so that it is not normally possible to find an optimum starting phase where both are small.

IV. LOUDNESS CALCULATIONS

This section takes the first logical step in an attempt to apply psychoacoustical theory to a calculation of the loudness of onset noise. The calculation is done by summing partial loudness from different critical bands according to a method of Stevens (1956), as described by Green (1976). This method is readily applicable to steady-state signals: it is not evident that it should apply to transient signals such as ours. However, we are encouraged to make this application because it is clear from the experiments above that listeners do perform a frequency analysis of the transient. Further, in order to calculate the loudness of the transient, one must eliminate the contribution of the sine tone itself, and this is most easily done in the frequency domain.

A. The schema

The calculation is straightforward. It begins with the spectrum of Eq. (3), which is an energy density, i.e., a power multiplied by a time per unit bandwidth. By integrating Eq. (3) over a critical band, we obtain the energy within the critical band. We make the simplifying assumption that level, measured in phons, can be obtained by subtracting the threshold of hearing for sine tones in the corresponding critical band. By using measured thresholds of hearing, we include in the calculation the effects of individual ears and of individual transducers. According to the formal definition of loudness, this assumption includes the approximation that equal-loudness contours are parallel.

The next step in Stevens' method is to convert the level for each critical band to units of sones by a formula that assumes that loudness grows as the 0.3 power of intensity. Then the loudness in individual critical bands are summed in a way that strongly weights the loudest band. In the sum over critical bands, one critical band is omitted, the band that contains the pure tone. Any onset noise within that band is presumably masked by the tone. The sum over critical bands leads to the total loudness for the onset noise. It is a function of sine tone frequency and of onset phase. The calculation thereby predicts which onset phase leads to the greatest noise for each sine tone frequency. We next describe the above steps in detail.

B. The model

The first step in the model is to compute the energy in the k th critical band, with frequency edges ω_1 and ω_2. The energy is given by

$$E_k = \int_{\omega_1}^{\omega_2} d\omega P(\omega). \tag{5}$$

Function $P(\omega)$, given in Eq. (3), has a singularity at $\omega = \omega_0$, but because the critical band containing the pure tone is omitted from the calculation, the limits ω_1 and ω_2 are either both less than ω_0 or else they are both greater than ω_0. Therefore, the singularity poses no problem. The indefinite integral of $P(\omega)$ can be done in closed form. If the trigonometric functions of ϕ are written in terms of $\cos(2\phi)$, the integral becomes the sum of two standard forms. The result is that

$$E_k = E(x_1) - E(x_2), \tag{6}$$

where

$$E(x) = \frac{1}{2\omega_0}\left[\frac{x}{1-x^2} + \frac{1}{2}\cos(2\phi)\ln\left(\frac{1+x}{1-x}\right)\right], \tag{7}$$

and $x_1 = \omega_1/\omega_0$, and $x_2 = \omega_2/\omega_0$.

Equation (7) is valid for ω_1 and ω_2 both less than ω_0. If ω_1 and ω_2 are both greater than ω_0, then the correct expression for $E(x)$ is just the same as above, but x_1 and x_2 are the reciprocals of the ratios given above. In all cases, therefore, the quantity x in Eq. (7) is less than unity.

Energy E_k is placed on an absolute scale by a scale factor that we consider for the moment to be arbitrary. We find a level in phons by

$$\text{Ph}_k = 10\log_{10}(E_k) - \text{Th}_k + C, \tag{8}$$

where C is the arbitrary constant converted to a decibel scale. (It is at this step that the phons scale is taken to be equivalent to sensation level.) The threshold in the k th critical band, Th_k, is given by a measured value for a particular listener and transducer. The loudness, measured in sones, for the k th critical band is

$$L_k = 10^{0.03\text{Ph}_k - 1.2} \quad \text{for} \quad \text{Ph}_k > 0,$$
$$L_k = 0 \quad \text{for} \quad \text{Ph}_k \leqslant 0, \tag{9}$$

so that 40 phons corresponds to 1 sone.

Finally, the total loudness is given as

$$L_T = 0.13{\sum_k}' L_k + 0.87 L_{\max}, \tag{10}$$

where $L_{\max}$ is the largest value of L_k, and the prime on the sum over critical bands indicates that the band containing the pure tone is omitted.

C. Thresholds

Audiometric thresholds were measured for the four listeners, for the three transducers used in the forced-choice experiments, using a Békésy tracking program, implemented on the laboratory computer. The signal was pulsed, 250 ms on–250 ms off, using a raised cosine envelope with 10-ms edges. The frequency of the digital signal was controlled by the method of fractional addressing (Hartmann, 1987); it varied exponentially with time as the run progressed. Levels were varied by a computer-controlled amplifier at a rate of ± 2 dB/s. The audible range was covered by tracking runs in three ranges: 300–45 Hz, 200–3200 Hz, and 3000–16 000 Hz. Runs in the middle range swept the frequency over the course of 10 min; runs in the high and low ranges took 5 min each. In the low-frequency range the signal frequency decreased as time increased.

D. Results

Figure 4 shows the predictions of the loudness model for listener D and HD480 phone. On the horizontal axis is plotted the frequency of the pure tone that is abruptly turned on. Symbols appear at frequencies that are critical-band centers (Zwicker 1961), where the calculations were done. The vertical axis shows the loudness, as computed from Eq. (10). The raggedness in the graph arises because of structure in the measured threshold as a function of tone frequency.

The calculations shown in Fig. 4 exhibit the expected behavior. For low-frequency tones the onset noise is greater for cosine phase (square symbols). For high-frequency tones the onset noise is greater for sine phase (circle symbols). Onset noise for 45-deg phase (triangles) is intermediate. The numerical values of the loudness, between 0 and 1 sone, depend upon the additive constant C, here taken to be 20 dB. We favor the choice of 20 dB because, with the value, the calculated loudness rises from zero for those frequencies and onset phases where we observe the onset noise to become audible. More important, however, is the fact that the conclusions to be drawn from the calculations are rather insensitive to the value of C. The shape of the curves for different onset phases and the crossover behavior are essentially unchanged as C increases well beyond 20 dB.

The calculated crossover in Fig. 4 is at 1600 Hz, to be compared with the experimental value of 1200 Hz. Other calculated crossover frequencies are given in Table II, where they may be compared with experimental values in Table I.

The comparisons of the model calculations and the data show suggestive parallels in some cases, but not in others. In most cases the calculated crossovers are too high. Changing the constant C makes no difference to this result. We conclude that within the Stevens model of loudness summation, the calculated crossover cannot be brought into agreement with experiment. What appears to be missing from the model calculations is a realistic treatment of the masking of one part of the transient spectrum on another. The emphasis given to the loudest critical band in the model provides some representation of masking, but there is no asymmetry of masking in the model. By contrast, Zwicker's loudness summation (e.g., Zwicker and Scharf, 1965) includes upward spread of masking. The effect of including upward spread in our calculations would be to emphasize the loudness of the onset with sine phase, where the noise is prominent for low frequencies, and to de-emphasize the loudness of the onset with cosine phase, where noise is concentrated at higher frequency. As a result, the tone frequency would not have to be increased so much before the sine-phase signal was noiser than the cosine-phase signal. In other words, including the upward spread of masking would reduce the calculated crossover frequency, bringing calculations closer to experiment. Whether the improvement would actually lead to agreement between calculated and experimental values for most of the listeners is not known.

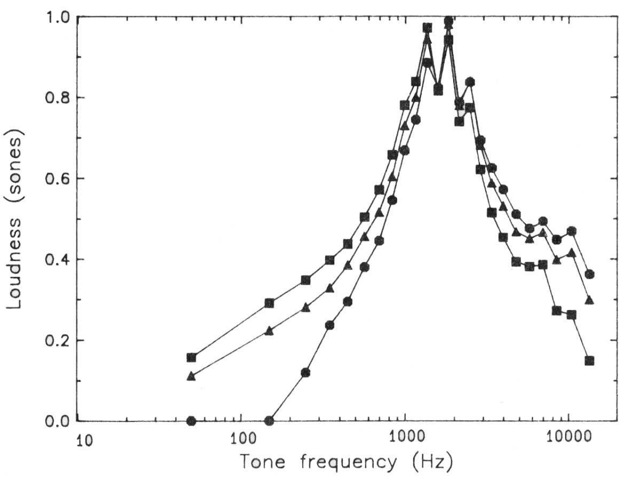

FIG. 4. Loudness of the onset transient calculated by Stevens' method as a function of the frequency of the pure tone that is abruptly turned on. Separate curves show different starting phases, circles for sine phase, squares for cosine phase, and triangles for 45-deg phase. Calculations are for listener D with the HD480 phone.

TABLE II. Crossover frequencies in Hz calculated from the method of Stevens, as described, in Sec. IV. The calculated values can be compared with the experimental values shown in Table I.

Transducer \ Listener	A	B	C	D
HD480	1600	1000	1800	1600
TDH49	2600	1000	1400	1700
Woofer	1000	1000	1100	1000

The variation among listeners reflects an essential difficulty with calculations, such as the above, that treat the listener as a loudness meter that integrates accurately over different frequency bands. In fact, there is onset transient energy below the frequency of the tone and there is onset transient energy above. The former sounds like a thud, the latter sounds like a click, and listeners can, and do, direct their attention preferentially to one frequency region or another. We believe that such a selective attention effect is the explanation for the anomalous crossover frequencies found for listener C. Although he had normal high-frequency thresholds, listener C had very low crossover frequencies, which changed very little, even when the transducers were changed dramatically. Apparently, listener C was so attuned to the thud, that this sound dominated the click until the tone frequency became quite low. Because of the selective attention effect, one cannot expect a loudness calculation to account for the data except for the ideal listener who somehow manages to maintain a spectrally flat attention characteristic.

V. CONCLUSION

The noise caused by starting a tone with an abrupt onset can be minimized by appropriately choosing the starting phase. For a tone with low frequency, the starting phase for minimum noise is zero, turning on a sine function. For a tone with high frequency, minimum noise occurs for a starting phase of ± 90 degrees, turning on a cosine function. This frequency dependence is expected on the basis of a simple physical argument, and it is observed in psychoacoustical experiments. The experiments show that the crossover frequency, between low- and high-frequency domains, depends upon the transducer, again as expected from the physical argument.

The crossover between high- and low-frequency tones also depends upon the listener. The simple physical argument would suggest that those listeners with the best high-frequency hearing would be the most affected by the click caused by cosine-phase onset. Therefore, the crossover frequency should be highest in these individuals. This correlation between crossover frequency and audiometric thresholds was actually observed for the four listeners in our experiments. Listeners A and D had the best high-frequency hearing and also had the highest crossover frequencies. Further, listener A had unusually high thresholds at low frequency, a second factor that should promote high crossover frequency. These ideas were quantified by adopting Stevens' method to a calculation of the loudness of the onset transient. Calculated crossover frequencies were systematically too high, a problem that might be resolved, at least in part, by revising the model to include the asymmetry of masking.

Although the model calculations had some qualitative success, they did not begin to account for all the data. We suspect that the reason is that selective attention can play an important role in a listener's judgments of clicks and thuds. These two kinds of onset noise have their origins in different parts of the frequency spectrum. It is possible for a listener to direct his attention to the noise above the tone or the noise below. A naive listener may fix upon one of these spectral regions and ignore the other, such that extreme high or low frequencies are needed to change the response. This was, in fact, our initial experience with listener A. In her first round of experiments, Listener A was so aware of the high-frequency noise that a crossover did not occur until the tone frequency had been increased to 3000 Hz. Eventually listener A learned to hear the low-frequency part of the spectrum in runs with low-pass-filtered tones (data not reported here). After this experience, the data from listener A more closely resembled the data of other listeners.

To the extent that selective attention directs the evaluation of onset noise, simple loudness models cannot cope with the general listener. One expects that crossover frequencies will show rather wide individual variations. From our experience, however, we believe that, for any listeners, even for one whose attention is focused on the high-frequency range, a crossover frequency does exist. For a tone with a frequency above the crossover, one has the (apparently) surprising result that a cosine tone makes less noise than a sine tone.

ACKNOWLEDGMENTS

We are grateful to Dr. Soren Buus and Dr. B. C. J. Moore for useful conversations. Dr. Brad Rakerd wrote much of the Békésy tracking program. Associate Editor W. A. Yost brought Green's (1968) article to our attention. This work was supported in part by the National Institute on Deafness and other Communication Disorders of the National Institutes of Health under Grant No. DC00181.

APPENDIX: SMOOTHING ENVELOPES

The tone with an abrupt onset has an envelope that is a step function:

$$w(t) = 0, \quad \text{for} \quad t < 0,$$
$$w(t) = 1, \quad \text{for} \quad 0 \leqslant t. \tag{A1}$$

To reduce the onset noise, it is usual to multiply a signal by a smooth envelope. Here, we specifically consider two standard envelopes, the linear ramp and the raised cosine.

The linear ramp (LR) is defined by

$$\begin{aligned} w(t) &= 0, \quad \text{for} \quad t < 0, \\ w(t) &= t/T, \quad \text{for} \quad 0 \leqslant t \leqslant T, \\ w(t) &= 1, \quad \text{for} \quad T < t. \end{aligned} \tag{A2}$$

The raised cosine (RC) is defined by

$$\begin{aligned} w(t) &= 0, \quad \text{for} \quad t < 0, \\ w(t) &= \frac{1}{2} - \frac{1}{2}\cos\left(\frac{\pi t}{T}\right), \quad \text{for} \quad 0 \leqslant t \leqslant T, \\ w(t) &= 1, \quad \text{for} \quad T < t. \end{aligned} \tag{A3}$$

We are interested in the power spectrum of a signal made by shaping a pure tone by these envelopes. Especially, we want to know if the dependence on the phase of the pure tone appears for LR and RC envelopes as it did for the tone with the abrupt onset. The easiest way to proceed is to calculate the Fourier transform of the signal by convolving the Fourier transforms of the envelope and the pure tone. Because the general results are algebraically complicated, we examine high- and low-frequency limits, as a guide to the strength of "click" and "thud" onset noises, respectively.

1. High-frequency limit

The high-frequency limit can be described roughly by the spectral envelope. The general signal with an abrupt amplitude envelope has a discontinuity, and therefore the high-frequency spectral envelope decreases as $1/\omega^2$, or -6 dB/oct. The linear ramp has a discontinuity in the first derivative, and therefore the spectral envelope decreases as $1/\omega^4$ or -12 dB/oct. The raised cosine has a discontinuity in the second derivative, and therefore the spectral envelope decreases as $1/\omega^6$ or -18 dB/oct. These rules apply unless something special about the pure tone cause the signal to be smoother than its envelope. For the abrupt onset, the special pure tone is one with sine phase; here, the discontinuity is moved to the first derivative, so the spectral envelope decreases as $1/\omega^4$. For the other envelopes, the situation is not so simple. The reason is that they actually have two discontinuities, one at $t = 0$ and the other at $t = T$. Sometimes the spectral effects of the two tend to cancel; sometimes they tend to add constructively. The result depends upon onset phase ϕ and upon the pure tone frequency through the parameter $\omega_0 T$. In any case, the spectrum includes oscillatory terms in the parameter ωT. In sum, the actual spectrum is both complicated and rather specific to each pure-tone frequency.

2. Low-frequency limit

In general, the low-frequency limit of the spectrum is a constant unless some special signal feature reduces the constant to zero. For the abrupt envelope the special signal is the pure tone in cosine phase.

For the linear ramp, the low-frequency limit of the power spectrum is reduced by a factor of $1/(\omega_0 T)^2$ compared to the abrupt onset. Further, the limit is strictly zero, independent of phase ϕ, whenever an integral number of cycles fits exactly into the time interval T. For the raised cosine envelope, the low-frequency limit is reduced by a factor of $1/(\omega_0 T)^4$ compared to the abrupt onset, so long as $\omega_0 T \gg \pi$. The limit is zero, independent of phase, if an odd number of half cycles (but more than one) fits exactly into the duration T.

[1] There is evidence that the spectral representation of these onset phase effects may be perceptually significant. In a 1968 masking study, Green used maskers that were 10-ms tone bursts having an integral number of cycles. Some bursts began and ended in sine phase, while other bursts began and ended in cosine phase. Although the spectra of these tone bursts are not the same as the spectra for semi-infinite tones, given in Eq. (4), the ratio of the sine-phase spectrum to the cosine-phase spectrum is the same as in Eq. (4), namely $(\omega_0/\omega)^2$. Green found that a sine-phase burst was more effective than a cosine-phase burst in masking a target tone having a frequency below the frequency of the burst. The reverse was found for a target tone with a frequency above the frequency of the burst.

[2] The HD480 phones are of the open-air design. We suspect that their frequency response is less dependent upon the fit at the ears than is the frequency response of headphones with cushions designed to make a seal against the ear or head.

[3] Two listeners, A and D, preferred a slower timing. For them, the intertone interval and intertrial interval were both longer by 300 ms.

Green, D. M. (**1968**). "Sine and cosine masking," J. Acoust. Soc. Am. **44**, 168–175.

Green, D. M. (**1976**). *An Introduction to Hearing* (Erlbaum, Hillsdale, NJ), p. 288ff.

Hartmann, W. M. (**1987**). "Digital waveform generation by fractional addressing," J. Acoust. Soc. Am. **82**, 1883–1891.

Stevens, S. S. (**1956**). "Calculation of the loudness of complex noise," J. Acoust. Soc. Am. **28**, 807–832.

Zwicker, E. (**1961**). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)," J. Acoust. Soc. Am. **33**, 248.

Zwicker, E., and Scharf, B. (**1965**). "A model of loudness summation," Psych. Rev. **72**, 3–26.