Center for Computer Research in Music and Acoustics February 1975

Department of Music
Report No. STAN-M-4

ON THE LOUDNESS OF COMPLEX, TIME-VARIANT TONES

by

James A. Moorer

DEPARTMENT OF MUSIC
Stanford University

Center for Computer Research in Music and Acoustics February 1975

Department of Music
Report No. STAN-M-4

ON THE LOUDNESS OF COMPLEX, TIME-VARIANT TONES

by

James A. Moorer

ABSTRACT

This is part of a proposal to the NSF Division of Psychobiology.

This study of loudness is motivated by the discovery that a set of complex, time-variant tones appear to behave differently with respect to loudness than would be predicted by the methods proposed in the literature. It is possible that the time-variant behavior of the sounds influences the loudness, so that a more complete theory of loudness must take this behavior into account. We thus propose to study these data and attempt to either verify the existing theories of loudness or formulate a more comprehensive hypothesis of loudness, building upon the currently existing theories, and to test this hypothesis by synthesizing new tones, doing equalization experiments, and comparing the results with the predictions of the model of loudness perception.

ON THE LOUDNESS OF COMPLEX, TIME-VARIANT TONES

A. INTRODUCTION

This study of loudness is motivated by the discovery that a set of complex, time-variant tones appear to behave differently with respect to loudness than would be predicted by the methods proposed in the literature [Zwicker and Scharf, 1965; Stevens, 1956]. It is of the utmost importance to timbre research to equalize tones for pitch, loudness, and duration, so as to avoid confounding dimensions in stimulii used to study the already farr too multidimensional attribute of timbre. Control for these factors is sought in the form of empirical equalization. In Grey [1975], a computer analysis based approach was used whereby 16 tones were analyzed. The analysis data was used to resynthesize the tones under computer control. The equalization experiment (See section III.B) resulted in a set of tones that were judged equal in pitch, loudness, and duration. When the data used to synthesize these 16 tones were examined, radical differences in the spectra were found. A preliminary investigation seemed to indicate that no current theory of loudness was adequate to explain these differences, although this preliminary study is based on such limited data that the differences are not statistically significant. It is possible that the time-variant behavior of the sounds influences the loudness, so that a more complete theory of loudness must take this behavior into account. We thus propose to study these data and attempt to either verify the existing theories of loudness or formulate a more comprehensive hypothesis of loudness, building upon the currently existing theories, and to test this hypothesis by synthesizing new tones, repeating the equalization experiment, and comparing the results with the predictions of the model of loudness perception.

B. HISTORICAL REVIEW

Steady-state Theories of Loudness

The steady-state theories of loudness attempt to answer the following questions:

1) For a pure tone, how does the perceived loudness relate to the frequency of the tone and the sound pressure level (SPL) of the tone?

2) How do we determine the loudness of a complex sound given its spectrum?

The second question is by far the more complex and many theories have been advanced over the years to deal with this question. The question is sometimes called the problem of *loudness summation*. Generally, the methods involve dividing the frequency spectrum into distinct regions and making rules to evaluate the contribution of each region to the loudness of the total sound. As we will see, it was discovered that the loudness is not just the simple sum of the loudnesses in each band.

The most common method of determining the perceived loudness of a tone of some kind is the method of adjustment. The tone in question (*probe* tone) and a *reference* tone (usually a 1000 Hz sinusoid at a calibrated amplitude) were presented to the subject. The subject then adjusted the volume of the *probe* tone until he was confident that the probe and the reference tone were the same loudness. This procedure has a number of inherent bias problems, but it seems to be suitable for many purposes. One of the problems is that subjects often do not know how to compare the loudness of two tones that differ widely in timbre, such as a low-frequency band-limited noise and a pure tone. When working with multi-tone complexes (sums of pure sinusoids), sometimes the subject will listen to the tones individually rather than as a gestalt. For these reasons, many of the experiments in loudness show a wide spread in the results.

Loudness of Sinusoids

The pioneering work in loudness was done by Fletcher and Munson [1933, 1934]. A set of curves was established by comparison of a sinusoid of variable frequency to a sinusoid at a fixed frequency and known energy. These curves, which relate the perceived loudness to the energy of the signal for a full range of frequencies and amplitudes, appear in numerous references on loudness and serve as the classical measure of the ear's sensitivity at different frequencies and amplitudes. The experimental technique was a version of the method of constant stimuli. The subjects were seated in a sound-proof booth and were required only to listen and then operate a simple switch to indicate whether they thought the probe tone was louder or softer than the reference tone. The tones were presented in sequence, of 1 second duration, with a .5 second pause between presentation of the probe tone and the reference tone. The probe tone was always first and the probe and the reference tones for a particular setting were presented twice. These data were converted to equal-loudness contours. This determination answers question 1) above, and the results have stood virtually unchallanged. Only their use in attempting to predict the loudness of a complex tone has been questioned. Much of this earlier work was summarized by Stevens and Davis [1938] in their book "Hearing: Its Psychology and Phisiology".

Stevens [1955] surveyed much of the literature to that point and put forth a rule-of-thumb for the simplified calculation of loudness which was essentially a *power law*. This was an attempt to produce a new scale to replace the Fletcher-Munson curves which related the intensity in dB to the loudness level in sones. He listed the values which had been previously determined by a number of different researchers. There were a total of 178 measurements recorded. All of the measurements were for the *half-loudness* judgment; that is, they listed SPL of two pure tones one of which was judged as being 'half as loud' as the other. He found extreme variability in the results. The range of SPL extended from 2.1 dB to 24 dB, roughly a 100-fold range. He took the somewhat bold step of reconciling the disparate reports by suggesting that the median of the 178 values, exactly 10.0 dB, be accepted a the basis of a new sone scale and that it be considered that loudness doubled for each 10 dB increase in stimulus intensity, a power law. He also expressed some doubt as to the validity of simple summation as proposed by Beranek *et al* [1951].

The figure of 10 dB was disputed by Warren [1970] who claimed that due to inherent bias in the methodology, the figure is consistently overestimated. By using single-loudness judgments, he eliminated the known bias in other methods and converged on a figure of 6 dB for the half-loudness judgment. A *single-loudness judgment* is done by having each subject listen to a single pair of tones and judge their relative loudness. Needless to say, this procedure requires a large number of subjects, since each subject renders only one judgment.

It may be helpful to discuss the sources of the bias in loudness experiments. With the Method of Constant Stimuli, the experimenter chooses a range of stimuli centering about the expected midpoint of the responses. Since this procedure biases responses toward the midpoint of the range of stimuli which have been presented, convincing but false verification of the pre-experimental expectations are found. With the Method of Adjustment, the loudness judgments may be influenced by the starting position of the attenuator, the range of intensities available to the subject, and the function relating attenuation to the angle turned. In the Method of Limits, the experimenter presents a series of stimuli in regularly increasing or decreasing steps. Thus, starting with a faint stimulus, successively stronger stimuli are presented until the subject reports a level equal to half the loudness of the reference tone. The order of presentation is then reversed. Loudness judgments with this method are influenced by the direction in which intensity changes, and there is no assurance of the validity of the conventional assumption that errors may be balanced by combining increasing and decreasing presentation orders. Again, the range of stimuli presented may influence loudness estimates. One method which avoids these biases, employs *only the first relative loudness judgment of subjects presented with just two sounds at different intensities*. This is the so-called *single-loudness judgment* method. Under these conditions, all effects of prior judgments are eliminated. The major practical difficulty is that large numbers of subjects are required for a systematic study, since only one judgment is obtained for each subject.

Loudness of Complex Tones by Direct Summation

Many researchers have attempted to predict the loudness of complex tones. Gates [1937] suggested that the loudness of a complex noise could be found by summing the loudnesses in the several octave bands into which the noise could be analyzed. Using the method of adjustment, Howes [1950] tested the validity of loudness summation of multi-tone complexes of widely (>250 mel) separated sinusoids. The probe tone was the sum of 2 to 11 sinusoids which were not harmonic, but were separated by a constant frequency difference in mels. Again, the reference tone was a 1000 Hz sinusoid of known SPL. Howes found that at higher levels (>60 db per tone), the total perceived loudness was less than the sum of the loudness of the component sinusoids. He proposed a model of hearing based on masking which became the inspiration for much of what was to follow, although his results relating to the exact form of the deviation from additivity was questioned by Beranek *et al* [1951].

Beranek put forth a model for loudness based on summing the contributions from each octave-band. The method is as follows:

1) Divide the frequency spectrum of the stimulus into frequency bands, each at least greater in width than a critical band for hearing (which was given in a footnote as being about 40 Hz for frequencies between 100 and 1000 Hz, and increasing to about 250 Hz at 8000 Hz), but not greater in width than about 600 mels. The SPL of each band is determined.

2) By means of the equal-loudness contours for pure tones as measured by Fletcher and Munson [1933, 1934], find the loudness level in phons for each band corresponding to the SPL of each band.

3) Using the relation between loudness in phons and loudness level in sones, also as determined by Fletcher and Munson, determine the loudness in sones contributed by each band. This relation was essentially a power law for loudnesses greater than 40 phons, and deviated in a smooth manner below that point.

4) The individual values of loudness are then summed to obtain the total loudness in sones.

To review, the method breaks the sound into bands, weights each band according to the sensitivity of the ear at that point, converts to a subjective scale that accounts for the rapid fall-off at low levels, and sums the results for each band. This approach was condensed into a loudness chart for octave-band data by Mintz and Tyzzer [1952].

Non-additive models

Quietzsch [1955] made measurements and computed spectra for 37 noises, many of which were natural sounds, and found some deviation from simple additivity as proposed by Beranek *et al.* He invented several simple spectral-based rules to help adjust the sum to a more accurate value. Suggesting that simple models of loudness addition could not be entirely correct, Stevens [1956] proposed a model where adjacent octave bands were allowed to interfere with one another, so that the total loudness was less than the sum of the loudnesses from the octave bands. More specifically, the relation postulated was the following:

St = Sm + F(sum S - Sm)

Where St is the total loudness,
Sm is the loudness in the loudest band,
S are the loudnesses in the other bands, and
F is a constant representing the interference between adjacent bands.

The constant, F, was calculated for several different conditions. A value of .3 was advanced as suitable for any steady-state noise whose spectrum exhibits "reasonable" continuity. The constant was adjusted down if bands more narrow than octaves were used. Stevens also provided a graph of the dependence of F upon the width of the band.

Zwicker *et al* [1957] found that the concept of the critical band made sense when applied to loudness. Since the results of Zwicker and Feldtkeller [1955] as well as those of Bauch [1956] suggested that an *increase in the width of a band of noise has little or no effect on loudness until a critical band width is reached, after which the loudness increases*, a series of experiments were begun which resulted in a theory of loudness summation [Zwicker and Scharf, 1965]. The width of the critical band in loudness summation was found to be consistent with the critical band width derived from observations on other parameters such as thresholds, masking, and phase relations [Zwicker, Flottorp, and Stevens, 1957]. A discrete subdivision of the audible frequency range into discrete critical bands was given by Zwicker [1961].

Scharf [1961] also measured the loudness of a 4-tone complex as a function of over-all frequency spread both in the quiet and against various levels of a uniform masking noise. The loudness was found to decrease when the over-all frequency spread exceeded a critical bandwidth.

Zwicker [1958] made a calculation of the exponent for the power law which relates the stimulation of the organ of Corti to the perceived loudness, and suggested again that the sound pressure level in each "Frequenzgruppe" (commonly translated *critical bandwidth*) could be used to determine total loudness. Scharf [March, 1959] showed that for very low levels, less than 5dB sensation level (about 11dB SPL), the spreading of energy of a multitone complex beyond a critical bandwidth does not increase loudness of the complex and may even decrease it. For higher levels, however, the critical bandwidth was found to be dependent only on the center frequency of the complex, and not on the level.

Scharf [1959] also showed that the loudness of a complex did not vary substantially when the number of components was varied, keeping the center frequency, the frequency spread, and the SPL constant, even though inhibition (masking) is greater when the components are dense. The frequency spreads used were 175, 1600, and 3400 Hz for a center frequency of 1500 Hz. This ranges from entirely within one critical bandwidth, to a spread of more than 12 critical bandwidths. It was suggested that there are two opposing effects in operation, that as the number of components is increased, the loudness is enhanced by increasing the number of *active* critical bands, and diminished by increasing mutual inhibition.

Scharf [1962] measured the loudness of three-tone complexes centered at 2000 Hz as a function of the intensity relations among the three components. Generally, the loudness of a 3-tone complex whose over-all frequency spread exceeded a critical band was greatest when the intensities of the components were equal, but when the frequency spread was confined to a single critical bandwidth, the loudness changed little for a given SPL.

Zwicker [1963] produced a set of curves that show the partly masked specific loudness of tones, which he suggested could be used to determine the loudness of a complex tone, as will be described below. The specific loudness is essentially the perceived loudness density. When integrated over a modified frequency scale, it provides the total perceived loudness. The frequency scale is transformed to what Zwicker calls "tonalness", which represents length along the basilar membrane. This frequency scale is linear in critical bandwidths, but nonlinear with frequency. It is, however, monotonic with frequency. The partial masking referred to is the phenomenon whereby a the loudness of a tone is reduced by masking with a narrow band of noise, but the tone is still audible. Zwicker shows that the loudness of the tone corresponds to the strength of the tone above and beyond the masking threshold established by the narrow-band noise. The tone and the noise are not in the same critical band.

Scharf [1964] helped verify these results by showing that the masking effects were asymmetric in frequency. That the effects of a masking noise extended more on the high-frequency side than on the low frequency side. These results were consistent with Zwicker's specific loudness curves. One of the results of these studies was that if the frequency spread of the stimulus did not exceed a critical bandwidth, the loudness depended only on the total energy within the band, and not on the exact details of the spectrum.

The model put forth by Zwicker and Scharf [1965] was a model for the steady-state evaluation of loudness. The procedure is as follows:

1) Start with the physical spectrum of the sound. Divide it into components or into critical bands.

2) Derive the excitation function at the organ or Corti, as a function of "tonalness", a measure which is related to the distance along the basilar membrane as measured from the helicotrema. It is related in a monotonic fashion to the frequency of excitation. This is, as mentioned before, a measure of frequency which is linear in critical bands. The excitation function is derived by taking the loudness masking pattern for each component (or critical band) and superimposing (not adding) them. The excitation function is then the maximum of these superimposed patterns as a function of tonalness. Masking patterns for several different frequencies may be found in Zwicker [1963].

3) Compute the specific loudness from the excitation function. This is done by a modified power law.

4) The integral of the specific loudness as a function of tonalness is then the total loudness of the sound.

This method is quite complete, but has met with some opposition. Robinson and Whittle [1964], for instance, made determinations of the relation between the SPL of octave bands of noise to the SPL of equally loud pure tones at the center frequencies and found some discrepancies with the earlier data of Zwicker [1958]. Since the model did not deal with time or transient behavior, a great deal of data related to temporal effects of loudness has yet to be included in a comprehensive theory of loudness.

An excellent summary and discussion of the model, including discussion on the physical mechanisms that may be responsible for the behavior, may be found in Zwicker [1970].

Temporal Behavior of Loudness

The studies have generally examined the effect of duration or the effect of rise time. With respect to the temporal behavior of loudness, we seek to answer the following questions (and more):

1) Does the perceived loudness of a tone change with its duration?

2) Does the perceived loudness of a tone change with its rise time or its fall time?

3) How much detail in the amplitude fluctuations of a tone can be perceived by the ear?

4) Is the critical bandwidth independent of temporal effects?

As early as 1929, Békésy [1929] found that one of the factors in loudness was the duration of the sound. Munson [1947] followed that up with a set of curves showing how the loudness increased as the duration was increased from 5 milliseconds to 200 milliseconds. The experimental technique was to present a warning tone, the probe tone, then a 1 second long reference tone. The probe and reference tones were then repeated. The subject was asked to judge whether the probe was louder or softer than the reference tone. No *equal* judgment was allowed. These data were, however, different from those of Békésy and of Garner [1949]. Munson's data were more extreme, showing a 25 dB attenuation at 10 millisecond duration, whereas Garner showed only an 8.5 dB attenuation. Two possible reasons for the discrepancies were postulated. The first was methodological, the second was based on the fact that Munson used extremely short attacks whereas Garner used 3 millisecond rise times, thus reducing greatly the click at the beginning of the tone.

Munson [1950] used the temporal effect of backward masking to compute the masking pattern of two pure tones. This avoided the problem of beats in determining the loudness of the probe tone, but also points out very strongly the temporal dependence of loudness. The primary, or reference tone was 400 Ms duration, followed by 20 Ms of silence, followed by the probe tone, which was about 80 Ms. The masking pattern when derived in this manner is not as strong as the simultaneous masking pattern, but resembles the masking behavior of noise on pure tones in shape. A masking pattern plots the threshold of hearing of the tone being masked as a function of frequency. The tone being masked must exceed this threshold before it can be heard at all. When it does exceed the threshold, its loudness seems to be related more to the intensity above and beyond the threshold established by the masking tone, rather than the absolute intensity of the tone being masked.

Kryter and Pearsons [1963] found that over a range of durations from 1.5 to 12 seconds, a wide variety of sounds were judged equally loud when the SPL was reduced by 4.5 dB for each doubling of duration. Variations in rise and decay times from .5 to 4 seconds did not significantly influence the judgements. This measurement implies a *power law* relating duration to loudness. This would be the case, for instance, if the ear were a linear filter, and the exponent of the power law would reflect the "time constant" of the ear. It would be very convenient if the ear operated entirely by power laws of various kinds.

Zwicker [1964] discovered an aural temporal effect analogous to the visual afterimage. White noise with a narrow band of frequencies suppressed was used as the stimulus. After exposure for one minute or more, the band-suppressed noise was turned off. The subjects reported hearing a pitch at the frequency of the suppressed band which decayed with time. It is not entirely clear, however, what implications for auditory theory this demonstrates.

Zwicker [1965] investigated the threshold of hearing of a 1000 Hz tone when masked by band-limited noise whose bandwidth and duration were varied. He found that the threshold was raised as the tone was made shorter in duration and as the bandwidth of the masking noise was increased. The threshold increase only occurs for tone durations less than 10 Ms. The amount of increase of the threshold depended very little on the level of the masking noise, but falls to zero near absolute threshold of hearing. This might be taken to imply some kind of temporal averaging process which required some amount of time to integrate the stimuli. The dependence on the noise bandwidth could be taken as indicative of the spread in frequency of a short tone burst as predicted by Fourier analysis of the tone burst. One might hypothesize that the short tones spread over a wider frequency, and thus can be suppressed by wider bandwidths of noise. This model, however, would predict an effective *spreading* of the critical bandwidth at short durations, due to the tone burst *leaking* into adjacent bands. There is, however, no clear evidence for this in the literature.

Scharf [1970] attempted to answer the question as to whether the critical bandwidth changes with decreasing duration. Even down to 5 Ms, he found no *statistically significant* changes in the critical bandwidth. This result is very important to the viability of the model of Zwicker and Scharf [1965].

Evans [1971] investigated the change in loudness level of a 1000 Hz tone burst as the duration of the burst was varied from 125 Ms to 4 seconds. A 2dB decrease in loudness was found at the shortest duration. A slight maximum was found at a duration of about 2 seconds.

Boone [1973] measured the loudness of a tone burst with a 5 Ms rise time as a function not only of the duration of the tone but also of the silent zone between the probe tone and the reference tone. His results implied that the loudness of a pure tone burst is significantly related to its energy. For a frequency of 1000 Hz, the loudness of the tone increased by about 10 dB per decade of duration, up to 30 Ms. This is what might be expected if we modeled the ear as a linear system with a time constant of 120 Ms. The duration of the silent time had a frequency-dependent effect in the rating of loudness.

Zwicker and Schütte [1973] tested the degree to which the ear can track random temporal changes of SPL by a masking scheme. Short test impulses were masked by a sequence of reproducible narrow-band noise with different bandwidths at a center frequency of 4 kHz at a defined time delay with respect to the beginning of the noise. The masked level-time pattern was expected to have a shape similar to the level-time pattern of the noise. This was found to be the case for bandwidths smaller than about 100 Hz. For larger bandwidths, the patterns were uncorrelated. This would imply that the ear can track temporal variations on the order of 10 Ms time constant, and that when more rapidly varying stimuli are presented, some sort of averaging of the temporal variations is done.

Robinson and Pollack studied the interaction between forward and backward masking to help determine the integrating period of the auditory system. A single click was used as the probe tone, masked by white noise which was not concurrent. It was found that the masking induced by both preceeding and following the click with noise (forward and backward masking) was greater than the sum of the separate masking effects. Up to 9 dB additional masking was observed. This was qualitatively explained by a running average of the stimulus, weighted by a function of unknown shape.

Studies of the variation of loudness with rise time were made by several groups. Vigran, Gjaevenes, and Arnesen [1964] measured the influence of rise time on loudness of a 750 Hz sinusoid and a broad-band noise. By changing the rise time from 25 ms to 1.6 seconds, an increase in loudness by about 3dB for the noise at the shortest rise time was found. The sinusoid increased by a lesser amount. This is a very small amount when compared to the change in loudness with the duration of the sound.

Dallos and Johnson [1966] measured the loudness of tone bursts which had known and identical rise and fall times. This made the amplitude of the tone burst vary with time in a trapezoidal manner. First, a linear attack portion of adjustable duration, then a steady state portion which was extended to make the *equivalent* durations of all the tones in the experiment the same, and then a linear decay portion which had a fall time the same as the attack time. They showed that as long as the equivalent duration remained constant, the loudness of the tone bursts remained constant as the rise/fall times were varied between 0 and 40 Ms.

Gjaevenes and Rimstad [1972] examined pure tones and wide-band noise with rise times from 30 Ms to 1 second. The duration of the tone, after the rise time, was 1 second. They found that the signals with the fastest onset had the highest loudness. The influence on loudness by the rise time was stronger for signals with spectral components at 250 Hz than for signals with spectral components at 1 KHz, and generally stronger for pure tones and wide-band noise than for narrow-band noise. The total change in loudness, however, was less than 2 dB in all cases as the rise time was swept between 30 Ms and 500 Ms.

All of these studies show that there is a variation in perceived loudness of a tone with the duration of the tone, to some extent with its rise time, and with what preceeds or follows it.

C. CURRENT CONCERNS IN LOUDNESS

The data we have already obtained gives us an ideal opportunity to test theories of loudness. The model of Zwicker and Scharf [1965] is sufficient to compute the loudness at a point in time. If our sound is constantly changing with time, but nonetheless produces a single impression of loudness, then we need a way to get a single number from this time-variant signal. The problem, then, is how to extend the current theories of loudness to deal with time-variant data.

In a pilot study, we have entered a simple version of the Zwicker and Scharf model into the computer and have computed the effective loudness using this model. Figure 1 shows the model of a single masking pattern, centered around 9 Bark. This was simulated by an exponential rise of .5 dB per Bark, an exponential fall of 3 dB per Bark, and a 3rd order polynomial joining the two, so that the curve and its first derivative are continuous. Figure 2 shows the computer simulation of the total masking pattern of the clarinet tone. Figure 3 and 4 show the pattern for the oboe and the flute respectively. These figures were obtained by superimposing the masking patterns for each critical band, and tracing the maximum, or the *envelope* of the superimposed patterns. The loudness, then, is the area between the threshold (not shown) and the envelope. This corresponds to the model of Zwicker and Scharf, except that not all the subtle details have been programmed and the masking pattern is only an approximation to the actual pattern.

In the matching experiment in Grey [1975], the loudness of 15 tones was matched with the loudness of a reference tone, which was the clarinet. These tones have thus been judged to be similar to the clarinet in loudness. We present the loudness below as computed by our simplified computer model. The loudness figures are in a scale that is similar to the sone scale, differing only by a scale factor.

TONE	LOUDNESS	COMMENT
OBOE 2	70.12	Tones with relatively fast attacks.
TENOR SAX	85.32	
OBOE 1	91.38	
CLARINET	92.41	Reference tone.
TROMBONE	92.87	With mute. Quite shrill.
ENGLISH HORN	98.38	
TRUMPET	101.53	
BASS CLARINET	104.95	
BASSOON	107.69	
SOPRANO SAX	108.18	
FRENCH HORN	113.22	
VIOLIN	115.28	
TENOR SAX	122.78	
VIOLIN	125.43	*sul tasto*
VIOLIN	126.36	*sul ponticello*
FLUTE	139.96	Most of these last tones have slow attacks.
AVERAGE	105.99	
MEDIAN	107.69	
STD. DEVIATION	17.20	
VARIABILITY	-2.4dB, +3.6dB	

There is an amount of discrepancy here, but it is no more than the experimental variation, thus no definitive statement may be made about the utility of the loudness model from these data. It is consistent with the literature, however, that the fastest rise times would sound somewhat louder. Likewise, that the slower rise times would sound somewhat softer.

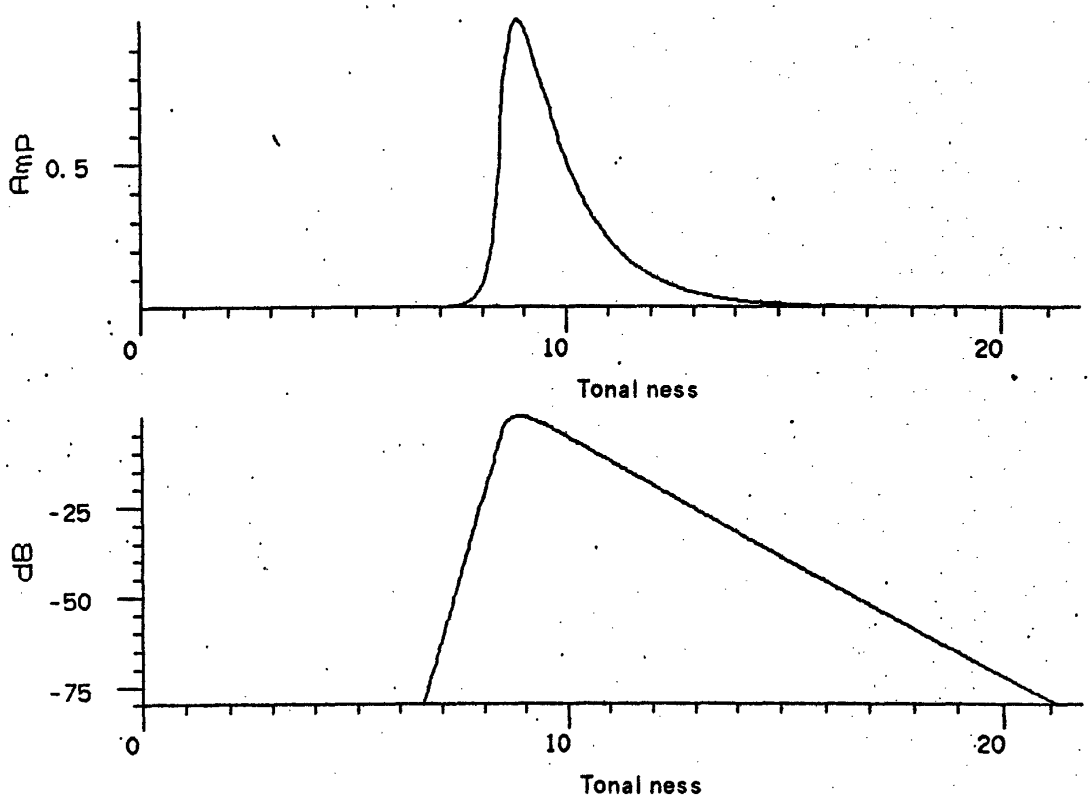
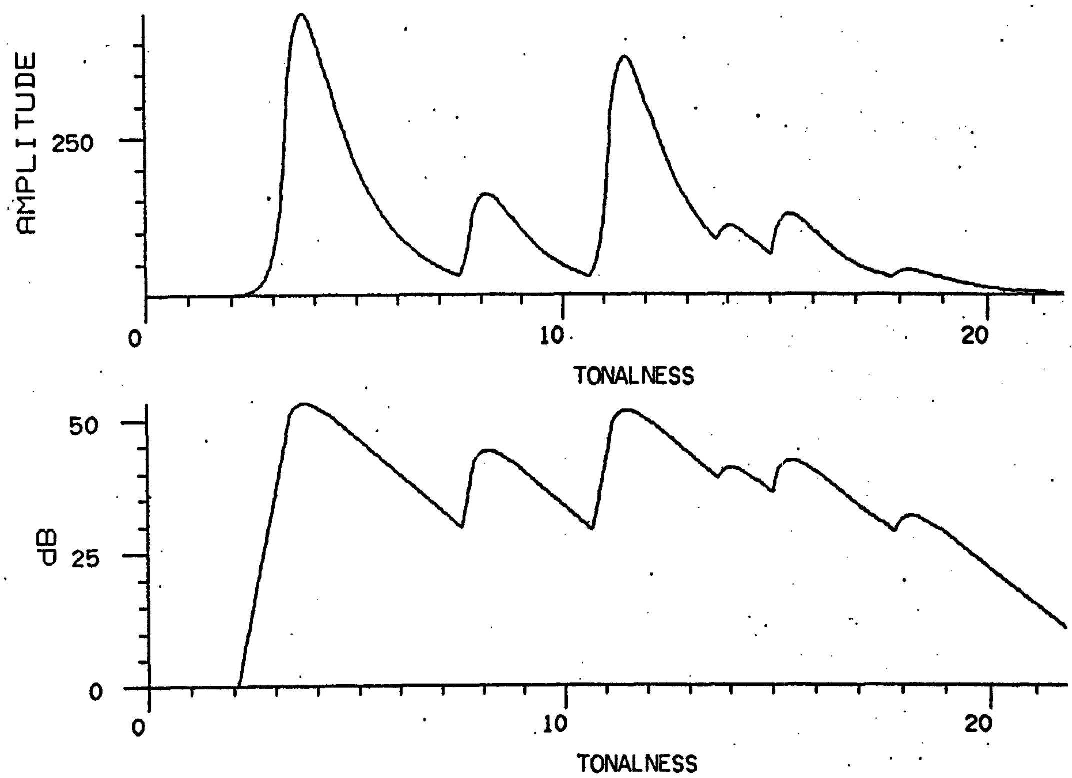
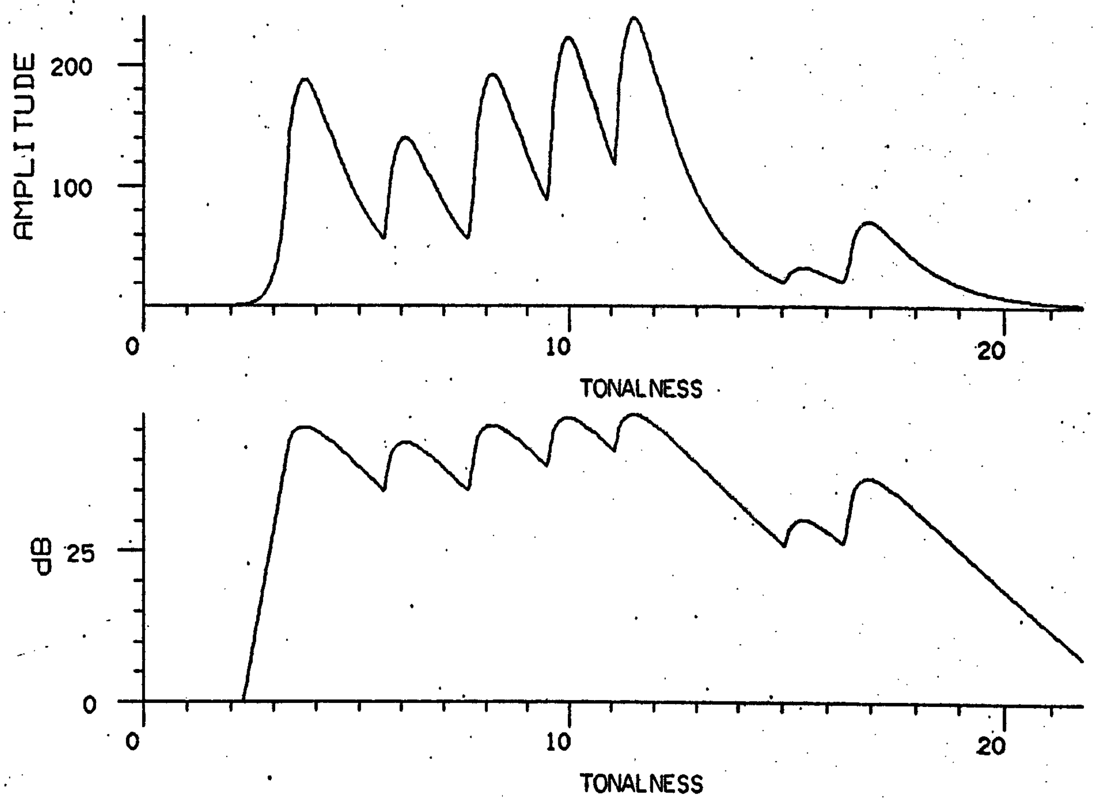


FIGURE 1: Computer simulations of masking patterns, modeled after Ziwcker [1955]. The model consists of exponential skirts, with a 3rd-order polynomial joining the two exponentials. These curves are centered around 9 Bark. The high-frequency skirt has a fall-off rate 7 times that of the low-frequency side. The polynomial matches the slopes as well as the values at the joints.



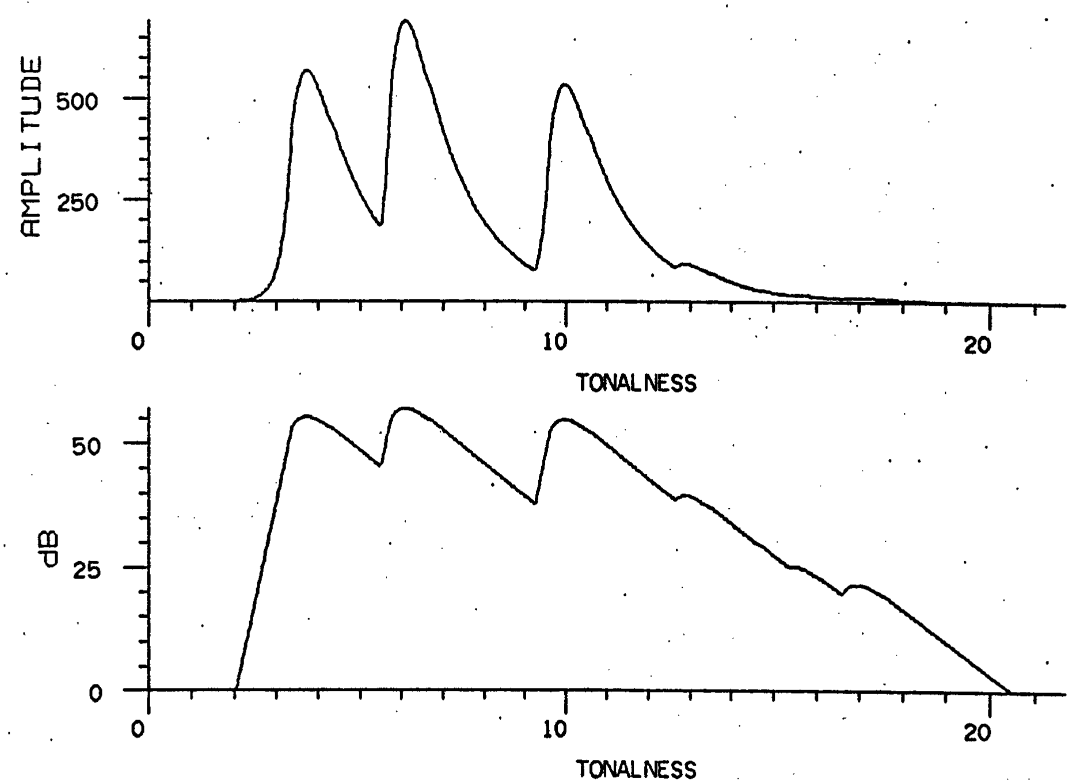
INTEGRAL = 92.41432

FIGURE 2: Simulated masking pattern based on the analysis of a clarinet tone. The integral under the linear pattern, minus the threshold, subjected to a simple transformation represents the loudness estimate, based on the model of Zwicker and Scharf [1965]. This represents the superimposition of the individual masking patterns. Where the harmonics are more than one critical bandwidth apart in frequency, they each get a separate masking pattern. When several harmonics fall within a single critical bandwidth, the energy is summed and used as the strength of the masking pattern.



INTEGRAL = 70.11962

FIGURE 3: Simulated masking pattern based on the analysis of an oboe tone. The oboe tone has many harmonics of almost equal amplitude, unlike the clarinet which represents largely odd harmonics.



INTEGRAL = 139.9591

FIGURE 4: Simulated masking pattern based on the analysis of a flute tone. This tone has a predominance of low harmonics, and thus presents quite a contrast to the previous oboe tone. Notice the differences among the integrals of the linear curves. Recall that these three tones (Figures 21, 22, and 23) have been equalized for loudness, so they are perceived as equally loud.

D. PERCEPTUAL RESEARCH ON LOUDNESS

The model of Zwicker and Scharf [1965] gives us a way to calculate the loudness of a sound at a point in time. The isolated notes from traditional music instruments we obtained change with time. Is it reasonable to update the model to extend it to time-variant tones, and if so, how would we go about doing this?

We propose to take the problem one step at a time. Since the greatest effect upon loudness is in the long-term changes, such as duration (in contrast to rise time), we may start by attempting to work long-term variations into the model. When the Zwicker model is applied to each point in time, the result is a plot of loudness versus time. At each point in time, the result is what the loudness would be if the spectral content at that point were present for a long time. Since we can get a judgment from a subject of the loudness of the tone *as a whole,* we need some way to derive a single number from this plot of loudness with time. The search for this method is one way of approaching the problem. In this case, we may choose to ignore the short-term effects, like rise time, until a reasonable theory for longer term effects can be verified.

Research Proposal:

We propose to do the following:

1) Enter the model of Zwicker and Scharf [1965, see Zwicker [1958,1963] also] into the computer. This involves digitizing the graphical information contained in the figures of the papers. We will reduce the complexity of the data and provide reasonable interpolations by fitting the data with high-order polynomials. Thus, we will have the masking patterns and the frequency-to-tonalness chart stored as continuous functions. This has already been done to a limited extent. The masking pattern and the frequency-to-tonalness chart have been digitized in a crude form. We propose to complete this task.

2) We propose to repeat the matching experiment using different tones as the reference tone. The tones we would like to pick are those which deviate the most from the predicted values. From the observed variance in the results of the matching experiment, we would expect that there will be no statistically significant correlation with the results of these experiments. If some correlation does appear, it would be a new phenomenon deserving further investigation. In any case, these new experiments will give us much more data to work with.

3) For the 16 tones that were equalized for loudness, compute the loudness from the model as a function of time. Once the model has been entered in the computer, this is a straightforward task.

At this point, the research forks into several paths. We will explore several models of what to do with the information at this point. These avenues will only be followed if significant deviations from the predicted values are obtained.

4) One possibility is to try to construct a graph of sone-seconds, the integral of the loudness chart over time, with loudness, such that the loudness function could merely be integrated and the total loudness figure read off the graph. Since our data only deals with one duration, this would give us only one point on the graph, but calculating that point and computing the spread of the data will surely tell us whether this method is worth pursuing. If the data clusters tightly, we will repeat the matching experiment with many more durations and determine the entire graph. The graph will probably have to be computed for many different loudness levels also.

This model assumes that the loudness of a tone depends on the energy accumulated with time, reaching an asymptotic value as the note persists, and decreasing slightly at longer durations.

4') Another possibility is to model the loudness mechanism as an averaging process, or filter, such that the loudness rises slowly as the tone proceeds, and approaches an asymptote which depends on the power in the signal. The value attained by this loudness function is then taken to be the loudness of the tone. This model can handle correctly the effects of rise time, in that a peaking effect can be introduced into the filter at short rise times. Again, the characteristics of the filter will probably have to depend on the total loudness level.

Literature Relating to Loudness

Bauch, H. Die Bedeutung der Frequenzgruppe für die Lautheit von Klangen. Acustica, 6, 40-45 (1956).

Békésy, G. V. Zur Theorie des Hörens. Physik. Zeits. 30,115 (1929).

Beranek, L. L., Marshall, J. L., and Cudworth, A. L. Calculation and Measurement of the Loudness of Sounds. JASA 23(3), 261-269 (1951).

Boone, M. M. Loudness Measurements on Pure Tone and Broad Band Impulsive Sounds. Acustica 29, 198-202 (1973).

Churcher, B. G. Calculation of Loudness Levels for Musical Sounds. JASA 34(10), 1634-1642 (1962).

Dallos, P. J., and Johnson, K. R. Influence of Rise-Fall Time Upon Short-Tone Threshold. JASA 40(5), 1160-1163 (1966).

Durlach, N. I., and Braida, L. D. Intensity Perception. I. Preliminary Theory of Intensity Resolution. JASA 46(2), 372-383 (1969).

Evans, D. H. An Experimental Determination of the Growth of Auditory Sensation for Time Periods Greater Than 125 Ms. J. Auditory Research 11, 374-384 (1971).

Fletcher, H., and Munson, W. A. Loudness, Its Definition, Measurement and Calculation. JASA 5, 82-108 (1933).

Fletcher, H. Loudness, Pitch and the Timbre of Musical Tones and Their Relation to the Intensity, the Frequency and the Overtone Structure. JASA 6(2), 59-69 (1934).

Garner, W. R. The Loudness and Loudness Matching of Short Tones. JASA 21(4), 398-403 (1949).

Gjaevenes, K., and Rimstad, E. R. The Influence of Rise Time on Loudness. JASA 51(4), 1233-1239 (1972).

Grey, John M. An Exploration of Musical Timbre PhD Dissertation, Department of Psychology, Stanford University, published by Department of Music as Technical Report STAN-M-2, February, 1975.

Hellman, R. P., and Zwislocki, J. Some Factors Affecting the Estimation of Loudness. JASA 33(5), 687-694 (1961).

Howes (1950), D. H. The Loudness of Multicomponant Tones. American Journal of Psychology 63(1), 1-30 (1950).

Kryter, K. D., and Pearsons, K. S. Some Effects of Spectral Content and Duration on Perceived Noise Level. JASA 35(6), 866-883 (1963).

Mintz, F., and Tyzzer, F. G. A Loudness Chart for Octave-Band Data on Complex Sounds. JASA 24(1), 80-82 (1952).

Munson, W. A. The Growth of Auditory Sensation. JASA 19(4), 584-591 (1947).

Munson, W. A., and Gardner, M. B. Loudness Patterns - A New Approach. JASA 22(2), 177-190 (1950).

Quietzsch, von G. Objektive und Subjektive Lautstärkemessungen. Acustica 5(1), 49-66 (1955).

Rawdon-Smith, A. F., and Grindley, G. C. An Illusion in the Perception of Loudness. British Journal of Psychology 26, 191-195 (1935).

Robinson, D. W., and Whittle, L. S. The Loudness of Octave-Bands of Noise. Acustica 14, 24-35 (1964).

Robinson, C. E., and Pollack, I. Forward and Backward Masking: Testing a Discrete Perceptual-Moment Hypothesis in Audition. JASA 50, 1512-1519 (1971)

Robinson, C. E. Interaction Between Forward and Backward Masking: A Measure of the Integrating Period of the Auditory System. JASA 53(5), 1313-1316 (1973).

Scharf, B. Critical Bands and the Loudness of Complex Sounds Near Threshold. JASA 31(3), 365-370 (1959a).

Scharf, B. Loudness of Complex Sounds as a Function of the Number of Componants. JASA 31(6), 783-785 (1959b).

Scharf, B. Loudness Summation Under Masking. JASA 33(4), 503-511 (1961).

Scharf, B. Loudness Summation and Spectrum Shape. JASA 34(2), 228-233 (1962).

Scharf, B. Partial Masking. Acustica 14, 16-23 (1964).

Scharf, B. Loudness and Frequency Selectivity at Short Durations. in Frequency Analysis and Periodicity Detection in Hearing, R. Plomp & G. F. Smoorenburg, ed. A. W. Sijthoff, Leiden, 455-462 (1970).

Scharf, B. Loudness Summation Between Tones from Two Loudspeakers. JASA 56(2), 589-593 (1974).

Stevens, S. S., and Davis (1938), H. Hearing: Its Psychology and Physiology. John Wiley and Sons, Inc., New York (1938).

Stevens, S. S. The Measurement of Loudness. JASA 27(5), 815-829 (1955).

Stevens, S. S. Calculation of the Loudness of Complex Noise. JASA 28(5), 807-832 (1956).

Stevens, S. S. and Guirao, M. Loudness, Reciprocality, and Partition Scales. JASA 34(8), 1466-1471 (1962).

Swets, John A, and Green, David M. On the Width of Critical Bands. JASA 34(1), 108-113 (1962).

Vigran, E., Gjaevenes, K., and Arnesen, G. Two Experiments Concerning Rise Time and Loudness. JASA 36(8), 1468-1470 (1964).

Warren, R. M. Elimination of Biases in Loudness Judgments for Tones. JASA 48(6), pp1397-1403 (1970).

Zwicker, E., and Feldtkeller, R. Ueber die Lautstärke von Gleichformigen Geräuschen. Acustica 5, 303-316 (1955).

Zwicker, E., Flottorp, G., and Stevens, S. S. Critical Band Width in Loudness Summation. JASA 29(5), 548-557 (1957).

Zwicker, E. Subdivision of the Audible Frequency Range into Critical Bands (frequenzgruppen). JASA 33(2), 248 (1961).

Zwicker, E. Ueber die Lautheit von Ungedrosselten und Gedrosselten Schallen. Acustica 13, 194-211 (1963).

Zwicker, E., and Scharf, B. A Model of Loudness Summation. Psychological Review 72(1), 3-26 (1965).

Zwicker, E. Temporal Effects in Simultaneous Masking and Loudness. JASA 38, 132-141 (1965).

Zwicker, E. Masking and Psychological Excitation as Consequences of the Ear's Frequency Analysis. in Frequency Analysis and Periodicity Detection in Hearing, R. Plomp & G. F. Smoorenburg, ed. A. W. Sijthoff, Leiden, 455-462 (1970).

Zwicker, E., Schütte, H. On the Time Pattern of the Threshold of Tone Impulses Masked by Narrow Band Noise. Acustica 29, 343-347 (1973).

Zwicker, E. "Negative Afterimage" in Hearing, JASA 36, 2413-2415 (1964).