inter.noise 2000
The 29th International Congress and Exhibition on Noise Control Engineering
27-30 August 2000, Nice, FRANCE

I-INCE Classification: 7.9

AUDITORY REPRESENTATIONS OF TRANSIENTS: THE IMPORTANCE OF PHASE IN THE ANALYSIS OF RATTLE EVENTS

G.H. Wakefield*, L.M. Heller, L.H. Carney***, M. Mellody******

* The University of Michigan, EECS Department, The University of Michigan, 1301 Beal Ave., Ann Arbor, MI, 48109, Ann Arbor, United States Of America

** Naval Submarine Medical Research Laboratory, Naval Submarine Medical Research Laboratory, Groton, CN, 06349-5900, Groton, United States Of America

*** Boston University, Hearing Research Center and Dept. of Biomedical Engineering, Boston University, 44 Cummington St., Boston, MA 02215, 02215, Boston, United States Of America

**** The University of Michigan, Applied Physics Program, The University of Michigan, 2477 Randall Laboratory, Ann Arbor, MI 48109, 48109, Ann Arbor, United States Of America

Tel.: 734-763-1254 / Fax: 734-763-1260 / Email: ghw@eecs.umich.edu

Keywords:
COMPUTATIONAL AUDITORY MODEL, PHASE PERCEPTION, PERCEPTION OF TRANSIENTS, AUDITORY REPRESENTATIONS

ABSTRACT
Engineering the sound quality of an automobile has drawn upon well known perceptual studies of loudness to set design targets for various noise sources, such as wind and power train. Attempts to apply these studies to design target specifications for squeak and rattle noises have met with limited success, primarily because the perceptual response to transient signals differs in several important dimensions from the response to wideband, steady-state noises, where the phase spectrum, for example, contributes little. We present psychophysical results that detail how the perception of a transient's magnitude spectrum can be affected by its phase spectrum, introduce a computational auditory model, and demonstrate the model's use in visualizing the transient events.

1 - INTRODUCTION
Despite their clear need, objective metrics for quantifying the perceptual response to acoustic signals are often of limited utility when setting design targets for product sound quality. Even the ISO532B standard for loudness, which is the most often cited example of an objective metric in sound quality engineering, is limited by the stringent conditions under which the metric has been validated. Deviations in the signal from wide-sense stationary wideband noise, the signal class upon which the metric was developed, can result in poor correlation between the calculated sone level and the listener's perception of loudness [1]. When signal conditions fall outside the validated range of an objective metric, as they often do, there are two alternatives. Direct measurement of the listener's subjective response, though often costly from an experimental point of view, is, nevertheless, the most valid means for characterizing the perceptual attributes of a class of sounds. Besides the time required for such testing, direct measurement also suffers from measurement error and potential specificity. Like all measuring devices, there is measurement noise associated with all psychophysical threshold and suprathreshold techniques. Such noise can obscure the subtle, but costly, differences among alternative product designs. Specificity often limits the ability to generalize the results of listening tests from one class of signals to another.
Visualization techniques represent a second alternative to objective metrics. In their most basic form, objective, as well as subjective, measurement techniques characterize magnitude along a particular perceptual dimension, or attribute, such as loudness. This measurement, however, is an aggregate of the

actual perceptual response, which typically contains far more detail than the single measure can reflect. Visualization techniques are intended to better indicate the details of the perceptual response.

Sometimes visualization is part of the objective metric. For example, in the ISO532B standard, the specific loudness of each output from a bank of bandpass filters is calculated prior to computing the (aggregate) loudness [2]. The dependence of specific loudness as a function of frequency and time can be used to understand the dependence of loudness on the spectro-temporal properties of the signal. Such detail is important when attempting to reduce the loudness of a signal class by manipulating the acoustic attributes of the signal, but the success of this approach depends on the accuracy of the visualization. ISO532B is computed on a relatively simple and primitive representation of the auditory system. Substantially more powerful and accurate models are under development in hearing science, which may bring better visualization of perceptual response.

The purpose of the present paper is two-fold. We introduce experimental data pertaining to the auditory perception of transients, which bears on issues surrounding the characterization and control of squeak and rattle sources in automobiles. Secondly, we introduce a computational model of the earliest stages of acoustic processing by the human auditory system. This model refines the filterbank model used in the ISO532B standard and thereby provides a significantly more detailed visualization of the transient signals in the experiment.

2 - PERCEPTUAL RESPONSE TO TRANSIENTS: THE ROLE OF PHASE

The typical duration of signals found in automotive systems ranges over several orders of magnitude. Therefore, the fact that "behavioral" time constants in auditory perception vary over several orders of magnitude [3] must be taken into account when attempting to quantify the perceptual attributes of automotive sounds. Experiments on infra-pitch suggest that listeners can hear periodicities in wideband repeated noise on the order of 2 seconds or more. The ability to integrate information about steady-state deterministic signals, such as sinusoids, is on a time scale of 200 ms. Temporal ordering among acoustic events requires separation in onset around 25 ms. The sensitivity to fluctuations in a signal's temporal envelope structure, including onset asynchrony in wideband signals, is on the order of 2-4 ms. Even smaller time constants are observed in the response of auditory neurons, which can be modeled, to a first approximation, by a lowpass filter with a cutoff frequency of 2500 Hz.

In the present paper, we are particularly interested in transients with temporal support on the order of the auditory system's shortest time constant of 2-4 ms. We select this regime because it serves as the bound above which signals are more likely to be perceived as evolving in their attributes over time. In this sense, by understanding the perceptual response to such brief signals, we stand a chance to better understand what is perceived to vary as the signal becomes longer in duration.

3 - PSYCHOPHYSICAL EXPERIMENT

Listeners were asked in a same-different adaptive psychophysical procedure to discriminate between a 4-ms signal (standard) and a "smoothed" version of the signal (comparison) in which the magnitude spectrum was altered while leaving the phase spectrum the same. The degree of smoothing was adjusted from trial to trial according to a "two-down, one-up" Levitt tracker. Reversal points were averaged to determine the degree of smoothing that yields a psychophysical threshold of 70.7% correct for each run of the test. Thresholds are reported as the average of the levels obtained in five such runs.

The bandwidth of the standard and comparison was 100 Hz to 10 kHz. For the standard, the level (in decibels) of each spectral component was uniformly distributed i.i.d. over a range of 20 dB. The phase spectrum of the standard was uniformly distributed i.i.d. over 0 to 2π. The same phase spectrum was used for the comparison, while the magnitude spectrum was varied using a spectral smoothing function. Within a given block of trials, the phase and magnitude spectra remained constant.

Thresholds for a number of different magnitude and phase spectra were measured in three highly trained listeners. While we observed individual differences across the three listeners in overall sensitivity, everyone showed evidence that the *phase spectrum* of the transient can significantly affect the ability to detect smoothing in the standard's magnitude spectrum.

Typical results are shown in two panels of Figure 1 for Subject 1. In each panel, the heavy and thin lines denote the magnitude spectra of the standard and comparison at threshold, respectively. The top panel is an "easy" phase condition − the standard and comparison spectra differ by no more than 6.5 dB anywhere in the frequency with an average squared difference of 1.45 dB. The right panel is a "hard" phase condition. In this case, the comparison's magnitude spectrum at threshold is much flatter than the standard, particularly in the 3-8 kHz region. In comparison with the "easy" phase condition, the largest deviation between the two spectra is 23.7 dB and the average squared difference is 6.8 dB. We observed

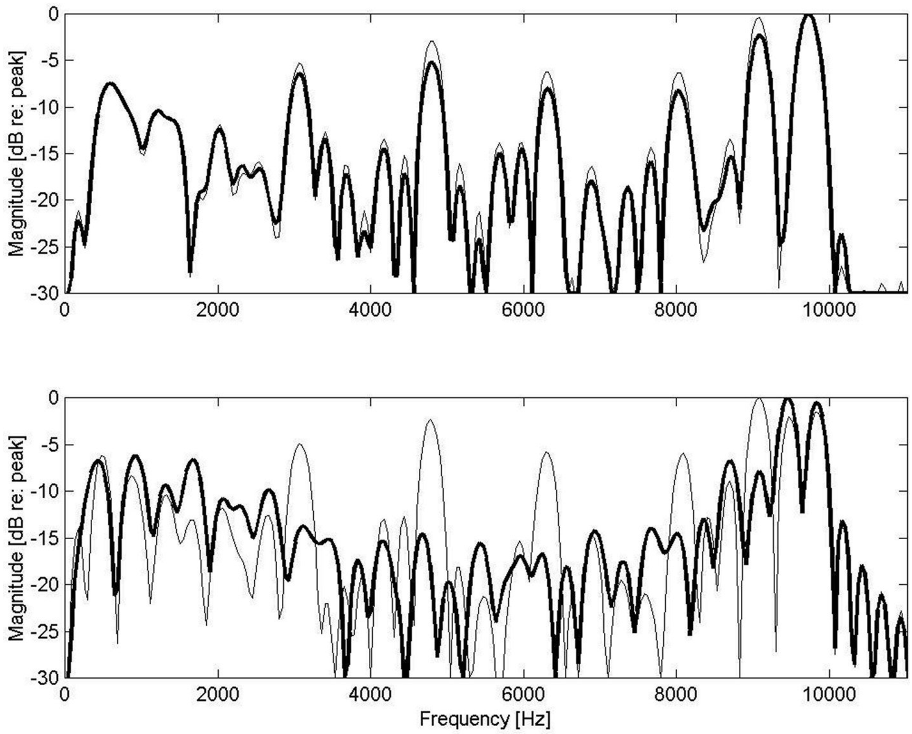

Figure 1: The magnitude spectrum of the standard is shown in the light solid lines of the two panels; the solid lines show the magnitude spectrum of the comparison stimulus at discrimination threshold for two different phase spectra; the top panel illustrates an "easy" discrimination condition, where the spectral differences between the standard and comparison are very small at threshold; the bottom panel illustrates a "hard" discrimination condition; in this case, the large physical differences between the spectra between 3.0 and 8 kHz are just discriminable.

similar phase effects for a variety of different random magnitude spectra. We also observed that phase conditions that were difficult for one subject were difficult for all subjects.

The phase effects are generally larger than what the (sparse) literature on the auditory perception of transients would suggest. Examination of the temporal and spectral properties of the signals, however, does not reveal the underlying source of such an effect. Peak factor, for example, does not appear to be important, nor does the concentration of energy in different bands of the spectrum. To obtain a better understanding of the possible causes for the results observed, we examined the response of a computational model of the human auditory system.

4 - COMPUTATIONAL MODEL OF THE AUDITORY PERIPHERY

The auditory nerve (AN) model used here is a simplified version of a model that was developed to study nonlinear properties of physiological AN responses [4]. The model used in this study consists of three main stages: 1) a linear filter bank for the tuning in the auditory periphery, 2) a model for the transduction and low-pass filtering of signals by sensory cells in the inner ear, and 3) a neural adaptation model for the synapse between the sensory cells and AN fibers [5]. The linear filterbank consists of 4-th order gammatone filters with bandwidths matched to psychophysical tuning curves for human listeners [6]. The sensory cells are modeled by an asymmetric saturating nonlinearity implemented using an arctangent, with an asymmetry of 3:1, followed by a 7-th order low-pass filter with a 3-dB cutoff at 2500 Hz. The low-pass filter limits the synchrony of the AN model responses; the rolloff in synchrony with increasing frequency matches that reported for physiological AN responses [7, 8]. The final stage of the model is a neural adaptation model, which is a time-varying implementation [4] of the three-stage diffusion model proposed by Westerman and Smith [9].

Preliminary analysis of the output of the AN model indicates that differences in the phase spectra of two signals are preserved. The two panels of Figure 2 show the output from the neural adaptation model in response to one of the 4-ms signals studied psychophysically. The top panel presents a gray-scaled image of the output amplitude of the neural adaptation model with time along the abscissa and center frequency of the filterbank along the ordinate. In general, the lagging of the response to the transient at the lower

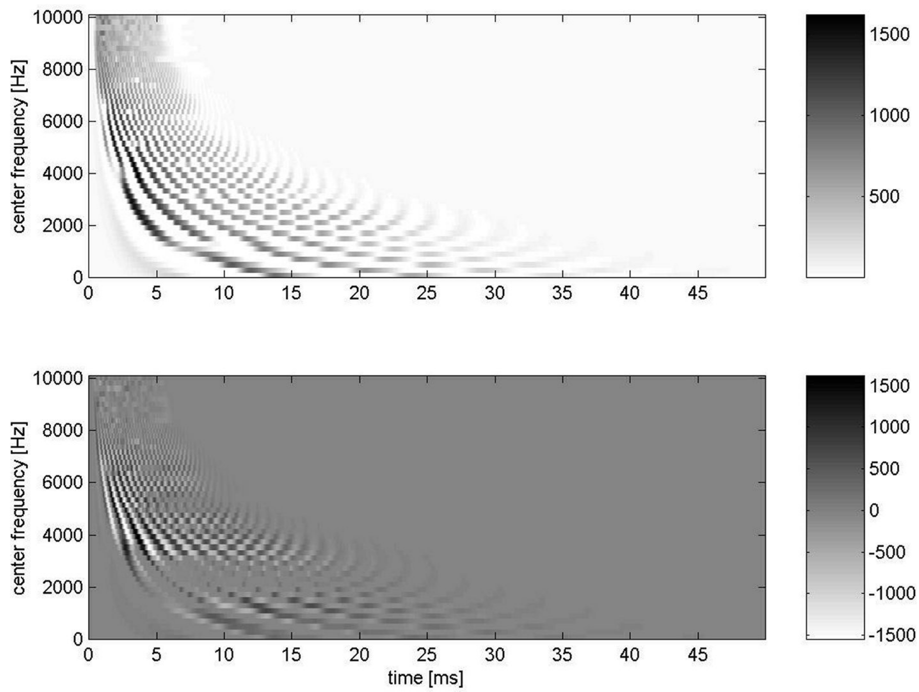

Figure 2: The output of the neural-adaptation stage of the AN model is shown in the top panel for the standard stimulus in "hard" phase condition; the image encodes the amplitude of the output as a function of the center frequency of the auditory filters and time; the bottom panel shows the differential between the outputs for the standards used in the "hard" and "easy" phase conditions.

frequencies is an attribute of the AN filterbank, rather than a property of the particular phase spectrum used in this example. Different average intensities across time correspond to the variations in spectral magnitude of the random transient. The bottom panel presents a gray-scaled image of the differential between the output amplitudes of two signals with the same magnitude spectrum but different phase spectra. The rippling observed in regions of this differential image is due to the differences in phase spectra across the two transients.

5 - DISCUSSION AND CONCLUSION

The sensitivity of the AN model to the phase spectrum of 4-ms wideband transients suggests that further characterization of the output, or differential, surfaces, will help us better understand the source of the phase effects observed psychophysically. Similar sensitivity is not as readily observed for the spectrogram. Our research is currently applying surface-representation techniques used in wavelets and time-frequency distributions to summarize the dependence of surface features on the signal's phase spectrum.

In the limit as the duration of the transient increases, we observe that phase plays a much weaker role in the discrimination of spectral magnitude [10]. We observed similar results in the present case when the 4-ms transients were concatenated to form 40-ms signals. Phase effects were not seen in the discrimination of spectral smoothing among the concatenated signals; thresholds were similar to the best thresholds observed for single transients. A much more surprising observation was that a delay in the concatenation of two transients by as much as 250 ms also eliminated much of the phase effect. Perceptual studies are underway to better understand this result.

While the source of the phase effects remains an open question, the fact that phase can have such an impact on discriminability has consequences on the perceptual characterization of transients in automotive systems. The perceptual response to transient signals has traditionally been summarized by loudness, which has often proved insufficient in accurately predicting the annoyance of such signals. When the magnitude spectrum of the transient has been used to characterize its timbre, relatively little gain has been achieved in predicting annoyance or other subjective attributes. The present data suggest that there may be a reason why poor correlation has been observed between objective metrics based on the magnitude spectrum alone and subjective response. It is our hope that the AN model will prove useful in teasing out the effects observed and in developing better objective metrics.

ACKNOWLEDGEMENTS
This research was supported by grants from the Office of Naval Research (MURI Z883402 and N0001499WR30037) and by a grant from the Department of Defense (US DOD N66604-96-C-H366).

REFERENCES

1. **M. Blommer, N. Otto, G. Wakefield, B.J. Feng, et al.**, Calculating the loudness of impulsive sounds, In *Society of Automotive Engineering: Noise and Vibration Conference*, 1995

2. **E. Zwicker and H. Fastl**, *Psychoacoustics - Facts and Models*, 1990

3. **D. A. Eddins and D.M. Green**, *Temporal integration and temporal resolution*, Hearing, B.C.J. Moore, Editor. Academic Press, 1995

4. **L.H. Carney**, A model for the responses of low-frequency auditory nerve fibers in cat, *Journal of the Acoustical Society of America*, Vol. 93, pp. 401-417, 1993

5. **M. G. Heinz, L.H. Carney, and H.S. Colburn** , Performance limits for frequency and intensity discrimination based on a computational auditory-nerve model, In *Association for Research in Otolaryngology*, 1999

6. **B. R. Glasberg and B.C.J. Moore**, Derivation of auditory filter shapes from notched-noise data, *Hearing Research* , Vol. 47, pp. 103-138, 1990

7. **D. H. Johnson**, The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones, *Journal of the Acoustical Society of America*, Vol. 68, pp. 1115-1122, 1980

8. **T. F. Weiss and C. Rose**, A comparison of synchronization filters in different auditory receptor organs, *Hearing Research*, Vol. 33, pp. 175-180, 1988

9. **L. A. Westerman and R. L. Smith**, A diffusion model of the transient response of the cochlear inner hair cell synapse, *Journal of the Acoustical Society of America*, Vol. 83, pp. 2266-2276, 1988

10. **B. C. J. Moore**, *Frequency Analysis and Masking*, Hearing, B.C.J. Moore, Editor. Academic Press, 1995