# Two phase effects in roughness perception[a)]

Daniel Pressnitzer
*Institut de Recherche et de Coordination Acoustique/Musique (IRCAM), 1 place Igor Stravinsky, F-75004 Paris, France*

Stephen McAdams
*Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, EPHE, 28 rue Serpente, F-75006 Paris, France and IRCAM, 1 place Igor Stravinsky, F-75004 Paris, France*

(Received 20 February 1998; accepted for publication 13 February 1999)

The respective influences of spectral and temporal aspects of sound in roughness perception are examined by way of phase manipulations. In a first experiment, the phase of the central component of three-component signals is shown to modify perceived roughness, for a given amplitude spectrum, regardless of whether it modifies the waveform envelope. A second experiment shows that the shape of the waveform envelope, for a given amplitude spectrum and a given modulation depth, also influences perceived roughness. We interpret both of these results by considering the envelope of an internal representation that is deduced from the physical signal by taking into account peripheral auditory processing. The results indicate that the modulation depth of such an internal representation is not the only determinant of roughness, but that an effect of temporal asymmetry is also to be taken into account. 
[S0001-4966(99)03805-9]

PACS numbers: 43.66.Jh, 43.66.Ki, 43.66.Nm [DWG]

## INTRODUCTION

Auditory roughness is a term that was introduced by von Helmholtz (1877) to describe the percept experienced when two sounds with proximal frequency components are heard simultaneously. He proposed that this attribute was linked to musical dissonance: for instance, an interval of a minor second played in the medium register of the piano produces roughness. The understanding of what kinds of sensory mechanisms are responsible for roughness perception is of interest when considered in the light of recent experimental results in the fields of developmental and comparative psychology. Babies (Schellenberg and Trainor, 1996) and members of other species (Fay, 1994; Hulse *et al.*, 1995) also seem to be sensitive to auditory roughness.

Experimental studies seeking to quantify roughness perception have often studied the effects of the frequency composition of stimuli (Plomp and Levelt, 1965; Plomp and Steeneken, 1968; Kameoka and Kuriyagawa, 1969). The presence of frequency components within the limits of a critical band is considered, in these studies, to be the source of the beats that produce the percept of roughness. Consequently, models of roughness perception have been proposed that are based on the spectral composition of energy falling within critical bands (Hutchinson and Knopoff, 1978).

A different approach to roughness is to study the influence of temporal parameters by means of amplitude-modulated stimuli (Mathes and Miller, 1947; Terhardt, 1974; Fastl, 1977). A dependence of roughness on the frequency and depth of the modulation was demonstrated. The interpretation proposed is that roughness is determined by the envelope fluctuations of the signal within an auditory filter. These results have inspired another kind of model in which roughness estimates are based on the rms value of the signal envelope after auditory filtering and after a modulation-frequency bandpass filter (Aures, 1985; Daniel and Weber, 1997).

These two approaches generally produce similar predictions as proximal frequency components give rise to amplitude beats and amplitude modulation gives rise to proximal frequency components. A method to pit them against each other is to use phase manipulations. This was done by Mathes and Miller (1947) who compared the roughness produced by Sinusoidally Amplitude-Modulated (SAM) tones and Quasi-Frequency-Modulated (QFM) tones. QFM tones have the same amplitude spectrum as SAM tones but display almost no amplitude modulation. The QFM tones were judged less rough than the SAM tones. This finding, confirmed by Terhardt (1974), has been interpreted in terms of differences in envelope rms after auditory filtering (Mathes and Miller, 1947; Terhardt, 1974).

Some authors claimed that this interpretation was flawed, as roughness differences could be heard for sounds having the same waveform envelope and frequency composition but different phase spectra (cf. discussion among Smoorenburg and Terhardt in Terhardt, 1970). In these cases, differences in temporal fine structure on the one hand or differences in the phases of the combination tones (Buunen *et al.*, 1974) on the other hand could influence the resulting percepts. However, to our knowledge, no experimental data have been produced to support this claim. The aim of the present study is to provide such quantitative experimental data and to discuss them in relation to the different models of roughness perception.

[a)]Portions of these results were first presented at the 4th French Acoustics Conference (Pressnitzer and McAdams, 1997) and at the 11th International Symposium on Hearing (Pressnitzer and McAdams, 1998).

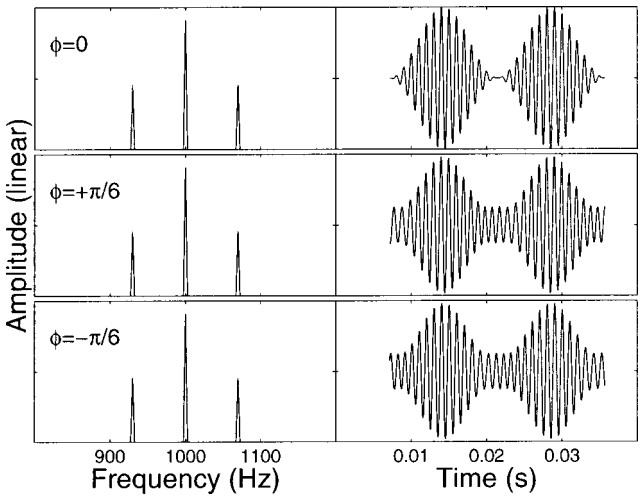


FIG. 1. Stimuli for Experiment 1. Amplitude spectra and waveforms of a pAM tone are plotted for three phase conditions at $(f_c, f_m)$=(1000 Hz, 70 Hz).

## I. EXPERIMENT 1

### A. Experimental hypotheses

Consider three spectral components with frequencies $(f_c),(f_c-f_m),(f_c+f_m)$ and with relative amplitudes of 1,1/2, and 1/2, respectively. In order to study the influence of phase relations among these components, it suffices to set the starting phase of the sidebands to zero and to vary the starting phase $\phi$ of the central component (Goldstein, 1967b). For $\phi=0$, a SAM tone with frequency $f_c$ modulated at a frequency $f_m$ with a modulation depth of 100% is obtained. By varying $\phi$, we can produce a family of sounds that have the same amplitude spectrum as SAM tones but different temporal waveforms. We will refer to these as ''pseudo-AM'' (pAM) tones:

$$\mathrm{pAM}(t)=\tfrac{1}{2}\cos[2\pi(f_c-f_m)t]+\cos(2\pi f_c t+\phi)+\tfrac{1}{2}\cos[2\pi(f_c+f_m)t]. \quad (1)$$

If the only determinant of roughness was the distribution of energy across frequency, all pAM tones with identical amplitude spectra should be equally rough.

If $\phi$ is nonzero, the envelope of the pAM tone is generally not sinusodal anymore. If $\phi$ takes for instance the value of $+\pi/6$, the envelope looks ''flatter'' (Fig. 1). The envelope actually gets flatter and flatter as $\phi$ is varied from 0 to $+\pi/2$, where a QFM tone is obtained (the analytical expression of the envelope is derived in Appendix A). By systematically varying $\phi$ we can therefore investigate the relation between envelope rms magnitude and roughness for a given amplitude spectrum.

Finally, pAM sounds with opposite phase values have the same amplitude spectrum and waveform envelope but different temporal fine structures (Fig. 1, bottom two panels). If only the characteristics of the envelope contribute to roughness, these two sounds should be identically rough.

### B. Method

#### 1. Stimuli and Apparatus

Seven series of pAM stimuli were used in which a given center frequency and a ''modulation'' frequency were paired: $(f_c, f_m)$=(125 Hz, 30 Hz), (250, 40), (500, 50), (1000, 70), (2000, 70), (4000, 70), and (8000, 70). In each case, $f_m$ was chosen to produce maximum roughness for a pure tone of frequency $f_c$ amplitude-modulated sinusoidally at $f_m$ (Zwicker and Fastl, 1990). The phase $\phi$ of the central component was varied from $-\pi/2$ to $+\pi/2$ in steps of $\pi/6$. Each sound was 1 s in duration and had raised-cosine onset and offset ramps of 50 ms. The stimuli were obtained by additive synthesis with a 16-bit resolution and a sampling rate of 44.1 kHz. The signals were played by a NeXT station through an ISPW DSP card, Pro IO digital-to-analog converters, and Canford power amplifier. They were presented over Sennheiser HD 520 II headphones at 60 dB (A-weighted), as measured by a Bruel & Kjaer 2209 sound-level meter with a flat-plate coupler. This level has generally been used in past studies of roughness. The distortion products of the whole sound reproduction chain for our stimuli were less than 60 dB below signal amplitude as measured by a Bruel & Kjaer spectral analyzer. Subjects were seated in a Soluna S1 double-walled sound-isolation booth and responded by clicking with a mouse on a graphic interface.

#### 2. Procedure

The notion of roughness was first introduced to listeners by means of a demonstration. A SAM tone with a carrier of 1000 Hz and a modulation frequency of 70 Hz was presented. Listeners could vary its modulation depth in a continuous fashion, thus changing concomitantly its amplitude spectrum and its envelope (the level of the carrier was kept constant). Listeners were instructed that the change in the quality of the tone they heard was called roughness. This example is considered to produce no roughness if the modulation depth is zero and the maximum roughness obtainable with a single SAM tone if the modulation depth is one (Zwicker and Fastl, 1990). Further demonstrations were given with SAM tones at the carrier and modulation frequencies used in the experiment. Following this familiarization phase, the experiment began. The stimuli were presented in separate blocks corresponding to the different $f_c$'s. In each block, all 21 pairs of nonidentical stimuli were presented in both orders in a randomized fashion, giving a total of 42 trials per block. For each trial, listeners were asked to decide which sound was rougher (2AFC). No feedback was given. After a few practice trials (5 to 10), all blocks were run just once.

#### 3. Subjects

A first group of 15 subjects ran the conditions $f_c$ =500,1000,2000 Hz. Subsequently, another group of 15 subjects ran the conditions $f_c$=125,250,4000,8000 Hz. The listeners consisted of 17 men and 13 women with ages varying from 24 to 32 (mean=27). They all had self-reported normal hearing and were not queried about musical training or previous experience in psychoacoustic experiments.

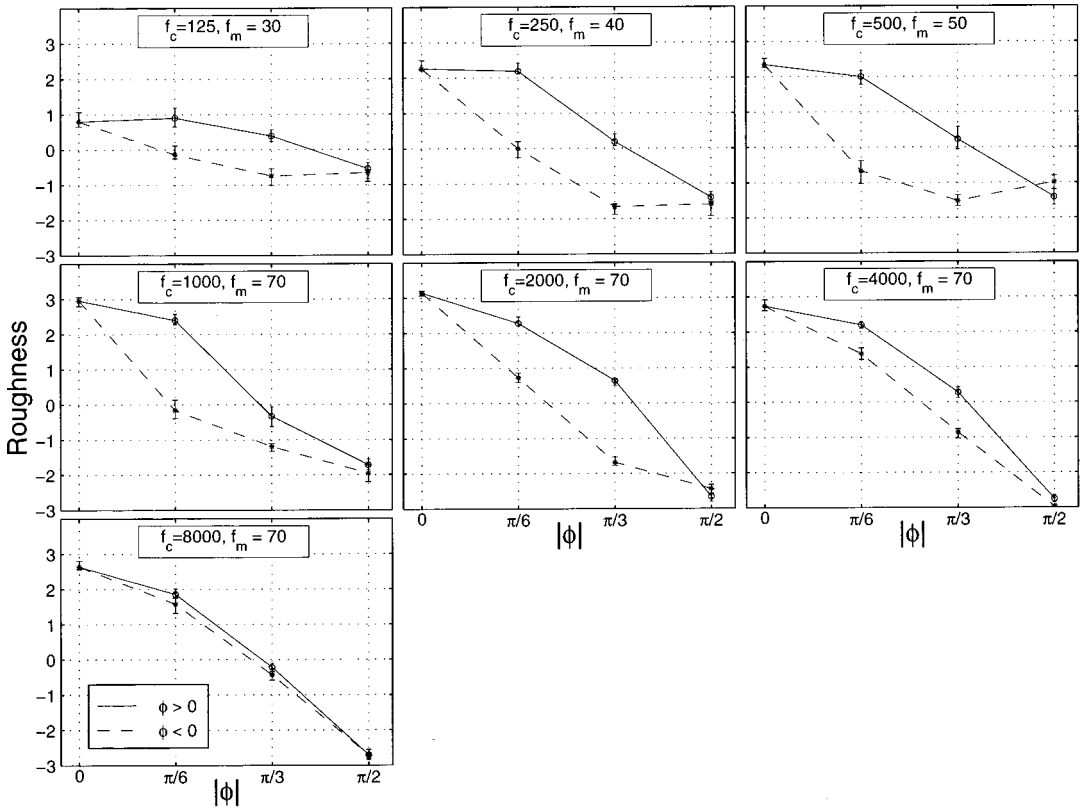


FIG. 2. Results of Experiment 1. The different panels correspond to the seven ($f_c$, $f_m$) conditions, as indicated (in Hz) on top of each panel. The roughness values estimated by the BTL method and the standard deviations estimated by bootstrap are plotted as a function of the phase $\phi$ of the central component. Solid lines represent $\phi>0$ and dashed lines represent $\phi<0$.

#### 4. Statistical analyses

The Bradley–Terry–Luce (BTL) method was used to construct a psychophysical scale from the binary paired-comparison judgments (David, 1988). The basic hypothesis for this method is that the stimuli can be arranged along a linear scale, roughness in our case. When presented with two sounds, listeners do not compare their ''true'' roughnesses but rather two random variables centered on the real values. Therefore, different outcomes are expected for the same comparison between two sounds, across listeners but also with repetitions for a same listener. It can then be shown that, with certain hypotheses on the random distributions, the proportion of times one sound is judged rougher than another is linked to the distance separating them on the perceptual scale (Bradley, 1953). The BTL method uses this principle to combine all comparison judgments across all listeners to produce a single *relative* scale that represents the contrast perceived between sounds in terms of roughness. The scale is normalized so that the sum of the values is equal to zero, hence the presence of negative values in the scale.

In order to test whether roughness differences were significant, they must be compared to the standard deviations of the results. These standard deviations are not readily available with the BTL method because judgments from all listeners are collapsed into proportions to build the scale. They were estimated by the bootstrap technique (Efron and Tibshirani, 1993). An empirical distribution was generated by resampling with replacement from the data set: a 100 sets of 15 samples were drawn. As a sample corresponds to the data matrix for a given subject, each set can be thought of as a simulated subject group. For each set, the BTL analysis was performed anew, giving in the end a distribution of estimated roughness values for each stimulus. These distributions were used to estimate the standard deviations displayed in the figures. The standard deviations were finally used to test for the significance of the difference between any two stimuli at a given risk, here $p<0.05$ (Efron and Tibshirani, 1993).

### C. Results

The results for the seven series of stimuli are presented in Fig. 2 in which the estimated roughness values are plotted as a function of $\phi$. The most obvious factor contributing to roughness is the absolute value of $\phi$, greater $|\phi|$ producing smaller roughness. This effect is similar in range across most $f_c$'s tested, except for $f_c=125$ Hz where a significant decrease in range is observed. However, for $f_c$'s at or below 4 kHz, there is also a strong effect of the sign of the phase $\phi$ for stimuli with a same amplitude spectrum and a same waveform envelope. This effect is consistent for the intermediate phase conditions for which the positive phase condition results in significantly higher roughness than the negative phase condition. This effect disappears for all $f_c$'s at $\phi=\pm\pi/2$ and for all phase differences at $f_c=8$ kHz. Another

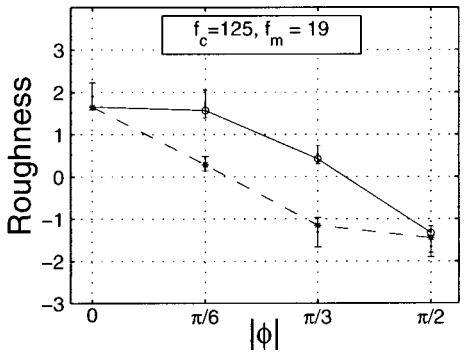


FIG. 3. Results of the control block of Experiment 1. This condition is to be compared with the top left panel of Fig. 2 that has the same $f_c$.

feature of the data to note is the relatively small size of the standard deviations, indicating a strong agreement across subjects.

### D. Discussion

#### 1. Dependency of roughness on the phase absolute value

The influence of $|\phi|$ reflects the effect of waveform envelope for a given amplitude spectrum. It can be interpreted by considering that the rms value of the envelope of the physical stimulus decreases as $|\phi|$ increases, and that this decrease is preserved after auditory filtering (Appendix C).

The overall reduction of variation range at $f_c = 125$ Hz could then be explained in terms of a critical band effect. At $f_c = 2$ kHz, the pAM stimuli cover 140 Hz, which is clearly less than the width of the auditory filter as estimated in terms of equivalent rectangular bandwidth (ERB=240 Hz; Glasberg and Moore, 1990). At $f_c = 125$ Hz, the stimuli cover 60 Hz which exceeds the ERB of 38 Hz at this center frequency. In this case, interactions occur mainly between adjacent components (separated by 30 Hz).

A control block was included in the experiment for 15 subjects. In this block, the modulation frequency was chosen so that the ratio $f_m$/ERB was identical to the one used for the condition $f_c = 1$ kHz. A series of pAM sounds with $(f_c, f_m) = (125,19)$ Hz was thus generated. We expected the effect of $|\phi|$ to be more pronounced in this case as all components could interact within an auditory filter. The results obtained are presented in Fig. 3. The global contrast increases, confirming our hypothesis. The greater variability in the results is probably due to the weak roughness produced by such a low modulation frequency; experimental judgments are harder to make in this case. The lessened contrast at this $f_c$ in the original data suggests that the effect of waveform envelope, when manipulated by phase relationships, is more pronounced if the three components can interact within a single auditory filter. This is in agreement with previous findings (Mathes and Miller, 1947; Goldstein, 1967b; Terhardt, 1974).

The effect of $|\phi|$ is therefore in good agreement with the hypothesis of the dependency of roughness on envelope fluctuation after auditory filtering as measured by envelope rms magnitude.[1] However, the significant influence of phase for a given waveform envelope and amplitude spectrum is not expected with this hypothesis. Let us examine different factors that may underlie the observed differences.

#### 2. Combination tones

Combination tones generated by the three components of the pAM stimuli may have played a role in the roughness judgments. Several studies have shown that among these tones, the most important is the first-order cubic difference tone (CDT) that is situated at $2(f_c)-(f_c+f_m)=f_c-f_m$ and that can interact with the acoustic component at the same frequency in our stimuli (Goldstein, 1967a; Plomp, 1976; Greenwood, 1991). Studying the influence of phase on the residue pitch of a three-component signal resembling the pAM tones, Buunen *et al.* (1974) proposed that the lower acoustic component was internally modified by a vector sum with the CDT. The phase of the CDT was found to vary as a function of the phase of the primaries, which may explain the differences observed in perceived roughness when $\phi$ varies.

A rule was derived to predict the phase change of the CDT if the phases of the primaries change (Buunen *et al.*, 1974). In the case of the pAM tones, the phase change of the CDT reduces to two times the change in $\phi$. Therefore, between the conditions $\phi=\pi/2$ and $\phi=-\pi/2$ the phase of the CDT should shift by $2\pi$. This is consistent with the fact that both $\pm\pi/2$ conditions produced an equivalent roughness.

However, an interpretation based on CDT faces two problems. First, the phase effect disappears at $f_c = 8$ kHz, whereas combination tones are likely to be still present, even with a lessened amplitude (Greenwood, 1991). Second, even if the phase change of the CDT could be summed up in a consistent rule by Buunen *et al.* (1974), the relation between the absolute phase of the CDT and the absolute phases of the primaries varies widely between subjects (Buunen *et al.* 1974; Zurek and Leishowitz, 1976). This is true to the extent that individual data had to be presented to report the effect. Our results display on the contrary a remarkable agreement among listeners.

#### 3. Asymmetry of the internal envelope

Temporal factors can alternatively be considered to interpret the observed roughness differences between positive and negative phase conditions. Even if these conditions correspond to a same physical envelope, the temporal fine structure of the stimuli are not the same. These physical differences could modify an ''internal envelope'' that would take into account the effects of auditory peripheral filtering.

A computational model was used to investigate the effect of peripheral auditory filtering (Giguère and Woodland, 1994). It is an active (Davis, 1983) basilar membrane model that displays a realistic phase behavior. It has been successfully used to interpret previous experimental data demonstrating the influence of phase on internal envelope (Carlyon, 1996; Carlyon and Datta, 1997b). The details of the implementation we used are described in Appendix C.

Overall, the filters centered at frequencies higher than the stimulus $f_c$ are expected to produce most of the perceived roughness because of the spread of the excitation pat-

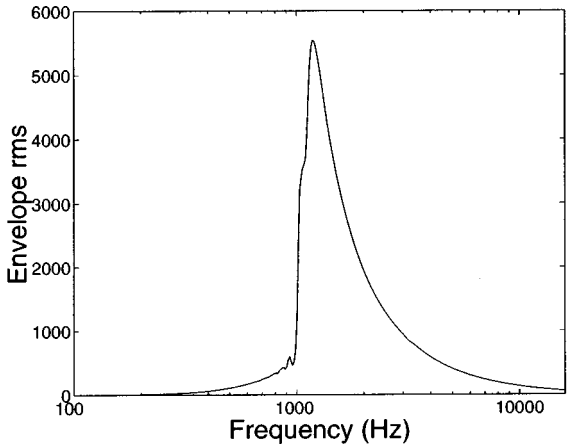


FIG. 4. Envelope modulation of the pAM tone with $(f_c, f_m) = (1000,70)$ Hz and $\phi = 0$ after auditory filtering. The filtering is simulated by the Giguère and Woodland (1994) model. The rms value of the Hilbert envelope of the signal present in each filter is displayed as a function of filter center frequency, in model's units.

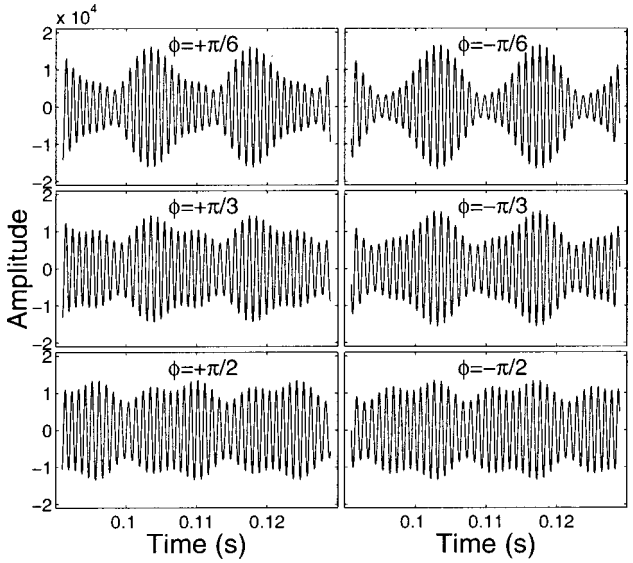


FIG. 5. Filtering of the pAM tones with $(f_c, f_m) = (1000,70)$ Hz by the Giguère and Woodland (1994) model. The output of the auditory filter that has the highest envelope rms (centered at 1180 Hz) is displayed for different values of $\phi$. Amplitude is in model's units.

terns toward high frequencies. This can be confirmed by computing the envelope rms value in individual channels of the auditory filtering simulation (Fig. 4).

The output of the filter that produces the highest envelope rms for the pAM tones with $(f_c, f_m) = (1000,70)$ Hz is displayed in Fig. 5, as an example of the output of the filters located above stimulus $f_c$. For all phases except $\phi = \pm\pi/2$, the modulations introduced by opposite phases reveal different shapes. In the positive phase conditions, the modulation shows an abrupt rise and a slow decay. This pattern is inverted for the negative conditions. This conversion of temporal fine-structure differences into envelope shape differences is for instance illustrated by the conditions $\phi = \pm\pi/6$, whose physical waveforms look very similar (Fig. 1, bottom panels), but whose filtered waveforms display asymmetrical envelopes (Fig. 5, top panels). These two sounds produced different roughnesses (Fig. 2).

The envelope asymmetry is actually introduced by the attenuation of the lower component of the pAM by auditory filtering, as we show in Appendix B. Therefore, it is not a critically dependent on the computational model used to simulate auditory filtering. In this Appendix, we show as well that the asymmetry has to disappear for the $\pm\pi/2$ conditions for all carrier frequencies. It also disappears progressively as the carrier frequency increases; the width of the critical band then becomes very large compared to the frequency span of the pAM tones and the auditory filters do not attenuate the lower component significantly more than the other components. The asymmetry is therefore present when significant differences in roughness were perceived and absent when the same roughnesses were judged.

It is important to note that the rms value of the modulation envelope after simulated auditory filtering cannot explain the observed results. It would lead to predict no effect of phase sign in most conditions, or even in some conditions a higher rms value for negative phase (see Appendix C). This precludes an interpretation in terms of the existing temporal models. The exploration of the effect of auditory filtering on positive and negative phase conditions therefore points to the hypothesis that the shape of the envelope could have an influence on roughness perception. Another experiment was designed to address directly this hypothesis.

## II. EXPERIMENT 2

### A. Experimental hypotheses

This experiment proposes to study the roughness produced by pure tones modulated with asymmetrical envelopes. A modulation resembling a sawtooth waveform, with a slow rise and an abrupt decay, was chosen. We chose not to use a real sawtooth to modulate our stimuli as the spectrum of such a waveform consists theoretically of an infinite number of harmonically related spectral components. In addition to the aliasing problems that the sampling of such a wave may cause (Stilson and Smith, 1996), resolved spectral components could appear in the lower auditory channels that might confuse listeners in their judgments. We rather constructed envelopes by truncating the spectrum of a sawtooth of frequency $f_m$. For a given center frequency $f_c$, only the spectral components that lay within a frequency distance of 1/2 ERB (Glasberg and Moore, 1990) of $f_c$ were retained:

$$E_{\text{sawtooth}}(t) = \sum_{n=1}^{n=N} \frac{1}{n} \cos(2\pi n f_m t - \pi/2) \quad \text{with } N \cdot f_m \leqslant \tfrac{1}{2}\,\text{ERB}(f_c). \tag{2}$$

The actual stimuli were then obtained by amplitude modulating a cosine wave at frequency $f_c$ with the corresponding envelope:

$$x_{\text{sawtooth}}(t) = \left(1 + m \cdot \frac{E_{\text{sawtooth}}(t)}{\max[E_{\text{sawtooth}}(t)]}\right) \cdot \cos(2\pi f_c t - \pi/2). \tag{3}$$

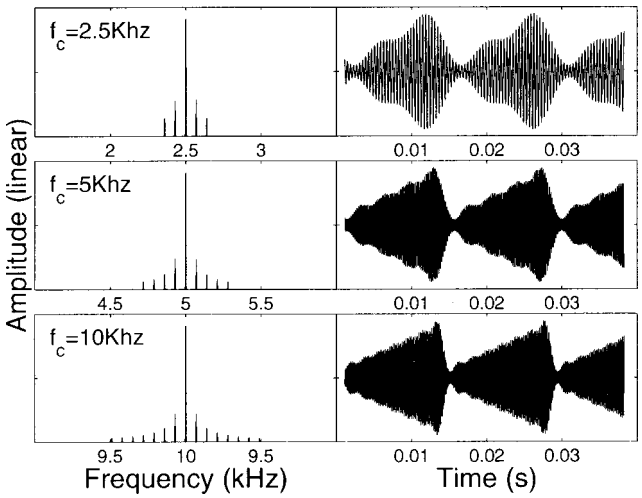


FIG. 6. Stimuli for Experiment 2. Amplitude spectra and waveforms are plotted for the ''sawtooth'' conditions with a modulation depth of $m=0.8$.

Such sounds look like sawtooth amplitude-modulated tones (see Fig. 6) and will therefore be referred to as ''sawtooth'' stimuli. Another set of stimuli was generated by inverting the starting phases of all components (envelope and carrier) to $+\pi/2$. These sounds are in fact a time reversal of the ''sawtooth'' ones and will be referred to as ''reversed'' stimuli. This phase manipulation allows us to compare stimuli with a given amplitude spectrum and a given modulation depth, but different physical envelope shapes. These differences in shape resemble those hypothesized to have been introduced by auditory filtering in Experiment 1. We checked that they should also be preserved after auditory filtering with a model simulation shown in Fig. 7 (Giguère and Woodland, 1994). If envelope asymmetry has an effect on roughness perception, a change in roughness should be observed between the ''sawtooth'' and ''reversed'' conditions.

## B. Method

### 1. Stimuli

The modulation frequency $f_m$ was chosen to be 70 Hz. The first center frequency for which $E_{\text{sawtooth}}(t)$ could have more than one spectral component according to the band-limiting criterion was 2500 Hz. In order to make the stimuli as similar as possible to those of the first experiment, various carrier frequencies were tested and for a given carrier frequency different values of modulation depth $m$ were used. Three series of amplitude-modulated tones were generated: $(f_c, f_m)=$(2500 Hz; 70 Hz), (5000; 70), (10 000; 70). Three modulation depths were used: $m=0.4$, 0.6, 0.8. An experimental block therefore includes sounds with varying modulation depth, and for a given amplitude spectrum and modulation depth ''sawtooth'' and ''reversed'' modulations are present.

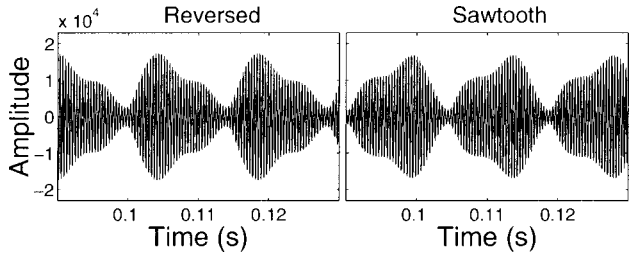


FIG. 7. Filtering of the ''reversed'' and ''sawtooth'' tones, with $(f_c, f_m)=(2500,70)$ Hz and $m=0.8$, by the Giguère and Woodland (1994) model. The output of the auditory filter with the highest envelope rms (centered at 2818 Hz) is displayed. Amplitude is in model's units.

### 2. Apparatus and procedure

The apparatus and procedure remained identical to those in Experiment 1, except that the headphones were changed to Sennheiser 420 (the headphones used in experiment 1 being no longer available). For a given $f_c$ block, all 15 possible pairs of nonidentical stimuli were presented in both orders in a randomized fashion, giving a total of 30 trials. All blocks were run once by each subject.

### 3. Subjects

Ten subjects participated in the experiment. The listeners consisted of 6 men and 4 women with ages varying from 24 to 45 (mean=28). All listeners had self-reported normal hearing and were not queried about musical training or previous experience in psychoacoustic experiments.

## C. Results

The results are displayed in Fig. 8. The influence of modulation depth is visible for all center frequencies; a higher modulation depth introduces more roughness. A significant effect of the shape of the waveform envelope is also observed. For a given amplitude spectrum and a given modulation depth, the ''reversed'' stimulus is systematically

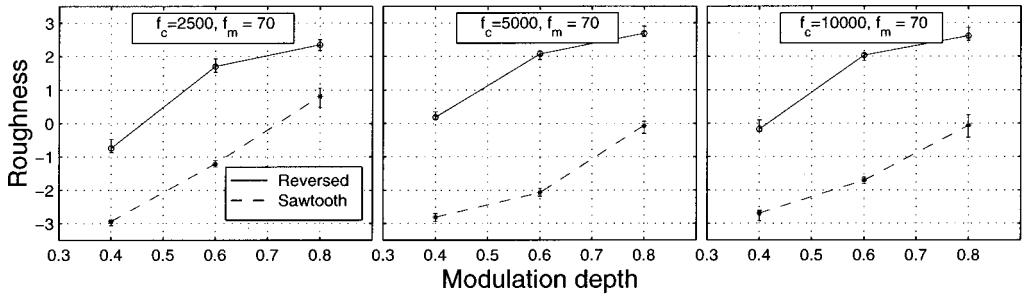


FIG. 8. Results of Experiment 2. The different panels correspond to the three $(f_c, f_m)$ conditions, as indicated (in Hz) on top of each panel. The roughness values are displayed as a function of modulation depth $m$. Solid lines represent the ''reversed'' modulation, dashed lines the ''sawtooth'' modulation.

judged rougher than the ''sawtooth'' stimulus. This effect is highly significant at all $f_c$'s for all modulation depths. It is, however, significantly smaller for the stimuli centered on $f_c = 2.5$ kHz.

### D. Discussion

#### 1. Effect of modulation depth

The increase in roughness with modulation depth can be explained by both spectral and temporal arguments. Indeed, in contrast to Experiment 1, the stimuli used here differ according to their amplitude spectra when the modulation depth increases; a greater modulation depth is accompanied by greater envelope rms, but also by greater sideband energy. This all leads to predict higher roughness. However, for a given modulation depth the signals have the same amplitude spectrum and waveform envelope rms. In this situation, the ''reversed'' and ''sawtooth'' conditions display equivalent envelope rms values after auditory filtering, as estimated by the model used for experiment 1 (Appendix C). An effect of the shape of the envelope is nevertheless observed.

#### 2. Effect of envelope shape

The significant effect of envelope shape is consistent with our previous hypotheses. ''Reversed'' sawtooth modulations that display an abrupt rise and a slow decay resemble the signals obtained after auditory filtering of the positive phase conditions of Experiment 1. They are rougher than ''sawtooth'' modulations that resemble the negative phase conditions.

This effect is more pronounced for high $f_c$'s. This is also consistent with our hypotheses. The asymmetry of the waveform envelope is less sharp for the lowest $f_c$ because of the small number of spectral components in the modulation. This could provide the basis for the reduction in the perceptual effect observed. In Experiment 1, the phase effect disappeared for high $f_c$'s as the asymmetry in the internal envelope disappeared. Here, the asymmetry is imposed on the acoustical stimulus and the effect remains. All this suggests that the phase effects are linked to characteristics of the envelope and not to details in the temporal fine structure of the stimuli, since the representation of such details would become progressively less reliable at high $f_c$'s due to the loss of phase locking.

## III. GENERAL DISCUSSION

### A. Spectral and temporal models

The effect of phase on roughness perception of sounds that have the same amplitude spectrum has been used as an argument against spectral models (Plomp and Steeneken, 1968; Terhardt, 1974). The only possibility to spectrally account for our results would be to consider the influence of combination tones on the ''internal spectrum,'' as the phase of the first-order cubic distortion product changes with the phase of its primaries (Buunen *et al.*, 1974). Such an effect cannot totally be ruled out as a contribution to the results of Experiment 1, even if we argued that these effects averaged out across listeners because of their variability. The effect of combination tones on the results of Experiment 2 is harder to imagine, however.

In contrast, the temporal models' hypothesis of the link between roughness and characteristics of the envelope can readily explain some of the features of our experimental data, namely the effect of $|\phi|$ in Experiment 1 and of $m$ in Experiment 2. However, these models fall short of explaining another major feature of the results, in that sounds with the same amplitude spectrum and rms envelope magnitude can produce different roughnesses.

### B. Phase dispersion

Phase effects have previously been reported for sounds with the same amplitude spectrum and waveform envelope. Thresholds of pure tones masked by harmonic sounds in positive Schroeder phase or negative Schroeder phase exhibited an influence of phase (Smith *et al.*, 1986; Kohlrausch and Sander, 1995; Carlyon and Datta, 1997a,b). The proposed explanation was that the phase-dispersive properties of the basilar membrane introduced differences in the modulation depth of the internal envelope near the resonance frequency of the masked tone. The model we used to simulate auditory filtering in the discussion of the results of Experiment 1 (Giguère and Woodland, 1994) has a realistic phase behavior and was able to predict the results of these former studies. However, in our case the envelope rms differences that appeared between some positive and negative phase conditions were very small and sometimes even went against the experimental results. The lack of large differences in envelope rms probably comes from the fact that the stimuli we used were narrowly band limited. The model, however, introduced an asymmetry in the filtered waveforms, that could be predicted by considering very general characteristics of auditory filtering. The phase effects we observed in both experiments for sounds with the same amplitude spectrum and envelope rms magnitude were therefore not attributed to a decrease in internal modulation depth but rather to a temporal asymmetry of the internal representation following auditory filtering.

### C. Ramped and damped sinusoids

Temporal asymmetry effects have been reported by Patterson (1994a,b) with sounds resembling the stimuli of Experiment 2. In these studies, exponentially modulated sinusoids and their time reversals were compared. Various combinations of carrier and modulation frequencies were tested. ''Ramped'' conditions sounded generally more pure-tone-like than ''damped'' conditions. The interpretation proposed by Patterson relies on the distribution of time intervals across auditory channels, more intervals corresponding to the period of the carrier being present in the ramped conditions. Therefore, the effect is less pronounced at high frequencies because of loss of phase locking. However, such an interpretation in terms of time intervals could not explain all the features of the results of Experiment 2 because the envelope shape effect persists (and increases) at high carrier frequen-

cies. We hypothesize that in our case subjects might have ''listened to the envelope'' rather than the carrier to make the roughness judgments.

### D. Roughness and envelope perception

Our results lead us to propose that roughness is linked to envelope fluctuations after auditory filtering. We also argue that the extraction of features relevant to roughness perception cannot be reduced to a simple bandpass filtering of the envelope followed by rms evaluation, but that an effect of envelope asymmetry has to be taken into account. Why such an asymmetry might translate into more or less roughness is still unclear. The recently reported influence of temporal asymmetry on latter stages of auditory processing (Winter and Patterson 1998) may ultimately provide a way to understand this effect.

## IV. CONCLUSION

This study aimed to estimate the influence of some phase manipulations on roughness perception. We confirmed that phase changes that affect a signal's envelope rms magnitude while leaving its amplitude spectrum constant can have a clear effect on roughness perception, all the more so as the signals have a limited bandwidth compared to the auditory critical band. Differences in roughness between signals having the same amplitude spectrum and envelope rms magnitude across auditory filters were also found. We interpreted these differences on the basis of the shape of the modulation after auditory filtering: Modulations with an abrupt rise and a slow decay produce more roughness than modulations with a slow rise and an abrupt decay.


## ACKNOWLEDGMENTS

The authors would like to thank Bennett K. Smith for his insights on the effects of phase dispersion. We also thank Ian M. Winter and Roy D. Patterson for discussing with us some of their recent data. Finally, we thank Wes Grantham and two anonymous reviewers who provided helpful comments on a previous version of this manuscript. Part of this work was completed thanks to a Fyssen Foundation grant to the first author.


## APPENDIX A

Let us consider a signal composed of the sum of N cosine waves with amplitudes $A_i$, angular frequencies $w_i$, and phases $\phi_i$, with $i \in [1, N]$. The envelope $E(t)$ of this signal may be obtained by taking the absolute value of its Hilbert transform (Hartmann, 1997):

$$E(t)^2 = \sum_{i=1}^{N} A_i^2 + 2 \sum_{\substack{j>i \\ i=1}}^{N} A_i A_j \cos[(\omega_i - \omega_j)t + (\phi_i - \phi_j)]. \tag{A1}$$

In the case of the pAM stimuli given in Eq. (1), this envelope can be simplified as in Eq. (A2):

$$E(t)^2 = \underbrace{3/2}_{dc} + \underbrace{1/2 \cos(2\omega_m t)}_{a(t)} + \underbrace{2|\cos(\phi)|\cos(\omega_m t)}_{b(t)}. \tag{A2}$$

Three terms participate in the square of the envelope: a dc offset, a component $a(t)$ at angular frequency $2w_m$ that does not depend on $\phi$, and a component $b(t)$ at angular frequency $w_m$. If $\phi = 0$, then $b(t)$ has maximum amplitude and Eq. (A2) simplifies to a cosinusoidal modulation. The rms value of the envelope is then $\sqrt{2}/2$. If $\phi = \pm \pi/2$, then only $a(t)$ remains and the envelope has an rms value of $(1 - \sqrt{2}/2)/2$. As $b(t)$ depends only on the absolute value of $\phi$, two opposite phase conditions will produce the same envelope.

## APPENDIX B

Let us examine the influence on the pAM tones of an ideal bandpass filter that is phase linear and that halves the amplitude of the lower component without affecting the amplitude of the other components. In this case, simplifications of Eq. (A1) lead to the following expression of the envelope:

$$E(t)^2 = \mathrm{dc} + a(t) + b(t)$$

with

$$\begin{aligned} &\mathrm{dc} = 11/8 \\ &a(t) = 1/4 \cos(2\omega_m t) \\ &b(t) = \sqrt{1/4 + 2\cos^2(\phi)}\cos(\omega_m t + \theta) \\ &\theta = \begin{cases} \arctan[-1/3 \tan(\phi)] & \text{if } (\phi \neq \pm \pi/2) \\ -\phi & \text{if } (\phi = \pm \pi/2) \end{cases}. \end{aligned} \tag{B1}$$

The important thing to notice here, disregarding the numerical values specific to the example given, is that the phase of $b(t)$ depends on $\phi$. If $\phi > 0$, then $\theta < 0$ and $b(t)$ will be slightly delayed compared to $a(t)$. Since $b(t)$ has a larger amplitude than $a(t)$, this will result in an asymmetry in the shape of the envelope (a rapid rise and a slow decay). If $\phi < 0$, the sign of $\theta$ and therefore the shape of the envelope is reversed, producing a slow rise and an abrupt decay. If $\phi = \pm \pi/2$, the delay $\theta$ is equal to $\pm \pi/2$. The absolute phase difference between these two conditions corresponds to $\pi$, half a period of $b(t)$, therefore a complete period of $a(t)$; the envelope remains the same between positive and negative phase conditions if $\phi = \pm \pi/2$. This same pattern of dependency of the envelope on the phase $\phi$ will always be observed if the lower component of the pAM has a smaller amplitude than the higher one.

This example aims to illustrate what happens in the auditory filters of center frequency above $f_c$, since for such filters the lower component is more attenuated than the higher one (additional attenuation of the central component does not change the reasoning). The actual filtering at the level of the basilar membrane is not phase linear, but the small frequency span of the pAM allows one to make a linearization of phase around $f_c$ without introducing much error—such a linearization was also proposed by Goldstein (1967b) for comparing AM and QFM envelopes within a

single auditory filter. Therefore, any model of auditory filtering should introduce an asymmetry similar to the one we just described.

The asymmetry is reversed for filters centered below $f_c$; however, these filters will contain much less activity because of the spread of the excitation patterns toward high frequencies. At the level of gross activity at the input to the auditory nerve, we thus hypothesize that the critical-band filtering has the global effect of sharpening the leading slope of the modulation envelope for positive $\phi$ and flattening it for negative $\phi$.

## APPENDIX C

To simulate the effect of auditory filtering, we used the model of Giguère and Woodland (1994) as implemented in the AIM Release 8 software platform (Patterson *et al.*, 1995). All parameters were set to default values, except for the quality factor $Q_n$, which was increased from 2 to 8. In doing this we followed the suggestion made by Carlyon (1996, 1997b) that the original value leads to too broad tuning curves when compared to psychophysical data. The wave input level was set to 60 dB SPL. All simulations were made with 320 segments between 100 Hz and 16 kHz at a sampling rate of 71 680 Hz (Giguère *et al.*, 1997).

In order to test if the transformation made by the filtering could account for the effect of phase observed in Experiment 1, we extracted the Hilbert envelope of the signals present in each channel, computed their rms value, and then summed up these values across channels. This does not give a perfect roughness estimate as a power function probably relates rms and roughness (Terhardt, 1974). In addition, the problem of roughness addition across channels is not taken into account here. However, all existing temporal roughness models would predict a monotonic link between overall rms value and perceived roughness.

The results of the computation for some representative $f_c$ and $f_m$ conditions are presented in Fig. C1. They can be compared to the experimental results of Fig. 2. The envelope rms value decreases with increasing $|\phi|$, which corresponds to the experimental results. However, the simulations are at clear variance concerning the effect of the sign of $\phi$.

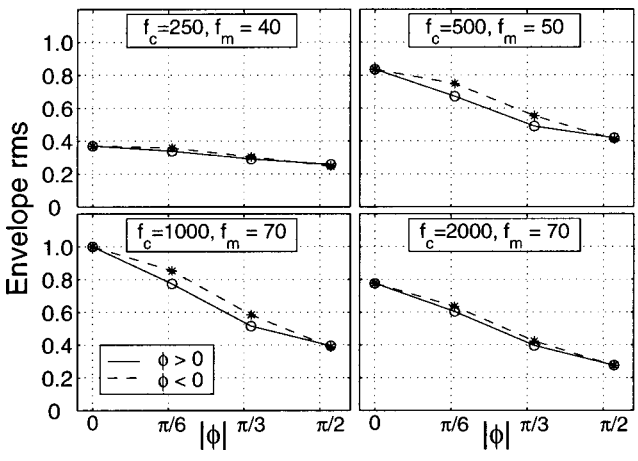


FIG. C1. Model simulation for the stimuli of Experiment 1. The sum of the rms value of the envelope of the signals present in each auditory filter is displayed as a function of $\phi$. The rms values have been normalized to their maximum across stimuli.

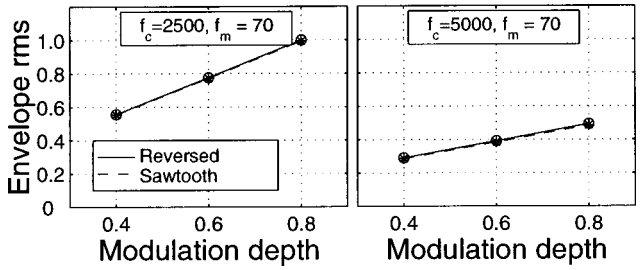


FIG. C2. Model simulation for the stimuli of Experiment 2. The sum of the rms value of the envelope of the signals present in each auditory filter is displayed as a function of *m*. The rms values have been normalized to their maximum across stimuli.

The same model was used with the stimuli of Experiment 2. The results of the computation for some representative $f_c$ and $f_m$ conditions are shown in Fig. C2. The increase of the rms value of the envelope with modulation depth generally goes with the perceived increase in roughness. However, the large effect of envelope asymmetry between the ''reversed'' and ''sawtooth'' conditions observed experimentally is not reproduced by the model.

[1]An aspect of the stimuli other than the envelope rms could be thought to have influenced the perceived roughness when $|\phi|$ is varied. In the conditions $|\phi|=\pi/2$ the envelope period is halved (see Appendix A). This change makes the apparent modulation frequency higher than the frequency that produces maximum roughness and could contribute to the small roughness values observed in these cases. However, the fact that the main contribution to roughness judgments came from variations of envelope rms and not of envelope period is suggested by two distinct points. For all $f_c$ conditions, the halving of the envelope period is only fully present for $|\phi|=\pi/2$. For the other phase values, the period remains the same but the observed roughness value decreases along with the envelope rms. For the control block at $f_c=125$ Hz and $f_m=19$ Hz, the modulation frequency is below that which produces maximum roughness. The halving of the envelope period, leading to an apparent modulation frequency of 38 Hz, should actually tend to produce more roughness in this particular case. However, the observed roughness values when the halving takes place are still small, in agreement with the low envelope rms value.


Aures, W. (**1985**). ''Ein Berechnungsverfahren der Rauhigkeit [A roughness calculation method],'' Acustica **58**, 268–281.

Bradley, R. A. (**1953**). ''Some statistical methods in taste testing and quality evaluation,'' Biometrics **9**, 22–39.

Buunen, T. J. F., Festen, J. M., Bilsen, F. A., and van den Brink, G. (**1974**). ''Phase effects in a three-component signal,'' J. Acoust. Soc. Am. **55**, 297–303.

Carlyon, R. P. (**1996**). ''Spread of excitation produced by maskers with damped and ramped envelopes,'' J. Acoust. Soc. Am. **99**, 3647–3655.

Carlyon, R. P., and Datta, A. J. (**1997a**). ''Excitation produced by Schroeder-phase complexes: Evidence for fast-acting compression in the auditory system,'' J. Acoust. Soc. Am. **101**, 3636–3647.

Carlyon, R. P., and Datta, A. J. (**1997b**). ''Masking period patterns of Schroeder-phase complexes: Effects of level, number of components, and phase of flanking components,'' J. Acoust. Soc. Am. **101**, 3648–2657.

Daniel, P., and Weber, R. (**1997**). ''Psychoacoustical roughness: Implementation of an optimized model,'' Acustica **83**, 113–123.

David, H. A. **(1988)**. *The Method of Paired Comparisons*, 2nd ed. (Oxford University Press, New York).

Davis, H. (**1983**). ''An active process in cochlear mechanics,'' Hearing Res. **9**, 79–90.

Efron, B., and Tibshirani, R. J. **(1993)**. *An Introduction to the Bootstrap* (Chapman & Hall, London).

Fastl, H. **(1977)**. ''Roughness and temporal masking patterns of sinusoidally amplitude modulated broadband noise,'' in Evans, E. F. and Wilson, J. P. (Eds.), *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 403–415.

Fay, R. R. (**1994**). ''Perception of temporal acoustic patterns by the goldfish (Carassius auratus),'' Hearing Res. **76**, 158–172.

Giguère, C., Smoorenburg, G. F., and Kunov, H. (**1997**). ''The generation of psychoacoustic combination tones in relation to two-tone suppression effects in a computational model,'' J. Acoust. Soc. Am. **102**, 2821–2830.

Giguère, C., and Woodland, P. C. (**1994**). ''A computational model of the auditory periphery for speech and hearing research. I. Ascending path,'' J. Acoust. Soc. Am. **95**, 331–342.

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hearing Res. **47**, 103–138.

Goldstein, J. L. (**1967a**). ''Auditory nonlinearity,'' J. Acoust. Soc. Am. **41**, 676–689.

Goldstein, J. L. (**1967b**). ''Auditory spectral filtering and monaural phase perception,'' J. Acoust. Soc. Am. **41**, 458–479.

Greenwood, D. D. (**1991**). ''Critical bandwidth and consonance: Their operational definitions in relation to cochlear nonlinearity and combination tones,'' Hearing Res. **54**, 209–246.

Hartmann, W. M. **(1997)**. *Signals, Sound, and Sensation* (AIP Press, Woodbury, New York).

von Helmholtz, H. L. F. **(1877)**. *On the Sensations of Tone as the Physiological Basis for the Theory of Music*, 2nd ed., translated by A. J. Ellis (1885), 4th ed. (Dover, New York, 1954; from German).

Hulse, S. H., Bernard, D. J., and Braaten, R. F. (**1995**). ''Auditory discrimination of chord-based spectral structures by European starlings (Sturnus Vulgaris),'' J. Exp. Psychol. **124**, 409–423.

Hutchinson, W., and Knopoff, L. (**1978**). ''The acoustic component of Western consonance,'' Interface **7**, 1–29.

Kameoka, A., and Kuriyagawa, M. (**1969**). ''Consonance theory Part I,'' J. Acoust. Soc. Am. **45**, 1451–1458.

Kohlrausch, A., and Sander, A. (**1995**). ''Phase effects in masking related to dispersion in the inner ear. II. Masking period pattern of short sound targets,'' J. Acoust. Soc. Am. **97**, 1817–1829.

Mathes, R. C., and Miller, R. L. (**1947**). ''Phase effects in monaural phase perception,'' J. Acoust. Soc. Am. **19**, 780–797.

Patterson, R. D. (**1994a**). ''The sound of a sinusoid: Spectral models,'' J. Acoust. Soc. Am. **96**, 1409–1418.

Patterson, R. D. (**1994b**). ''The sound of a sinusoid: Time-interval models,'' J. Acoust. Soc. Am. **96**, 1419–1428.

Patterson, R. D., Allerhand, M. H., and Giguère, C. (**1995**). ''Time-domain modelling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.

Plomp, R. **(1976)**. *Aspects of Tone Sensation* (Academic, London).

Plomp, R., and Levelt, W. (**1965**). ''Tonal consonance and critical bandwidth,'' J. Acoust. Soc. Am. **38**, 548–560.

Plomp, R., and Steeneken, H. J. M. (**1968**). ''Interference between two simple tones,'' J. Acoust. Soc. Am. **43**, 883–884.

Pressnitzer, D., and McAdams, S. (**1997**). ''Influence de la phase sur la perception de rugosité de sons complexes,'' Actes du 4ème Congrès Français d'Acoustique **1**, 535–538.

Pressnitzer, D., and McAdams, S. **(1998)**. ''Phase effects in roughness perception,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 286–292.

Schellenberg, E. G., and Trainor, L. J. (**1996**). ''Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: Test of adult and infant listeners,'' J. Acoust. Soc. Am. **100**, 3321–3328.

Smith, B. K., Sieben, U. K., Kohlrausch, A., and Schroeder, M. R. (**1986**). ''Phase effects in masking related to dispersion in the inner ear,'' J. Acoust. Soc. Am. **80**, 1631–1637.

Stilson, T., and Smith, J. **(1996)**. ''Alias-free digital synthesis of classic analog waveform,'' in *Proceedings of 1996 International Computer Music Conference*, Hong Kong, edited by L. Ayers and A. Norner.

Terhardt, E. **(1970)**. ''Frequency analysis and periodicity detection in the sensation of roughness and periodicity pitch,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. Smoorenburg (A. W. Sijthoff, Leiden), pp. 278–287.

Terhardt, E. (**1974**). ''On the perception of periodic sound fluctuation (roughness),'' Acustica **30**, 201–212.

Winter, I. M., and Patterson, R. D. **(1998)**. ''Comparisons of the responses from neurones in the cochlear nucleus and inferior colliculus encoding to ramped and damped sinusoids,'' in *Assoc. Res. Oroloryngol. Abs.*, edited by G. R. Popelka, St. Petersburg Beach, FL., p. 99.

Zurek, P. M., and Leishowitz, B. H. (**1976**). ''Measurement of the combination tones $f2$-$f1$ and $2f2$-$f1$,'' J. Acoust. Soc. Am. **60**, 155–168.

Zwicker, E., and Fastl, H. **(1990)**. *Psychoacoustics, Facts and Models* (Springer-Verlag, Berlin).