Influence of phase effects on roughness modeling.

Daniel Pressnitzer (1)
Stephen McAdams (1,2)
(1) Institut de Recherche et de Coordination Acoustique Musique (IRCAM),
pressnitzer@ircam.fr
(2) Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes,
smc@ircam.fr

Abstract

Experiments in tonal as well as in non-tonal contexts have shown the potential relevance of a proper model of roughness estimation in the framework of computer-aided composition environments. Two types of models are currently available that operate in either the spectral or the temporal domain. An experiment has been designed in which the respective influences of different cues have been pitted against one another using phase differences. Significant variation in roughness appeared between sounds with different temporal envelopes and roughness was generally linked to the rms value of the envelope. However, significant differences also appeared between sounds with the same spectrum and the same envelope but different temporal fine structure. This calls for a new modeling approach based on peripheral auditory processing.

1 Introduction

Roughness is a sound attribute related to the perception of amplitude fluctuation in the range of 20-200 Hz. Two simultaneous sounds that produce partials in the same frequency region, like a minor second played in the medium register of the piano, can produce rough beats.

Roughness was originally defined by Helmholtz [6] to provide a sensory basis for musical consonance. A basic feature of Western tonal harmony is to distinguish within the ensemble of possible intervals obtainable with the chromatic scale those that are considered consonant from those that are considered dissonant. This distinction is of course by no mean rigid, as shown by the progressive shift from dissonance to consonance of some intervals throughout the history of music. However, as has been noted since Pythagoras, intervals made of simple integer ratios are situated at one extreme of the scale (consonant) whereas the more complex ratios are considered dissonant. Numerous theories have been proposed to account for this enigma, and Helmholtz provides one of them: as the ratio between two sounds involves more complex integers, more partials will be “mistuned” and create roughness. Since the dissonance of Western tonal intervals corresponds to their roughness, this attribute may be proposed as a sensory basis upon which the complex rules of harmony were built. Cross-cultural studies involving American and Japanese listeners [3] did show some agreement on judgments of simple isolated intervals according to their consonance.

Along with other factors such as melodic motion or implicit knowledge of tonal pitch space hierarchy, dissonance is considered to play a part in the expression of an important feature of tonal music: tension and release movements. A dissonant chord is considered unstable, calling for a resolution on a more consonant chord. As a potential source of sensory dissonance, roughness was shown [2] to play a part in the expression of these schemas in tonal contexts.

A correspondence between a sensory feature and a musical notion takes on a new interest when listeners are confronted to music not leaning on internalized syntactic rules. In such a case, a previous study [10] showed that tension movements could nevertheless clearly be perceived by musician or non-musician listeners. Roughness appeared as strongly correlated to the tension judgments. As the material tested had no harmonic tonal function, its roughness value could provide a relevant indication for the composer. Naturally, music is (generally) more than a juxtaposition of sound events and a relative roughness value of an isolated chord will not determine a harmonic function in the time course of a piece. Nevertheless, roughness estimation of musical material can be, in
our view, a relevant parameter in the framework of computer-aided composition environments.

2 Roughness models

Like loudness and pitch, roughness has no simple physical correlate. Therefore, computational models derived from experimental data have been proposed. Experimental studies that have sought to quantify roughness first considered the effects of the frequency composition of stimuli [9]. The presence of components within the limits of a critical band was considered to be the source of beats that produced the percept. Consequently, a first class of models estimates roughness on the basis of a sound’s spectral composition ([7] for instance).

A different approach consisted in studying the influence of temporal parameters, by means of amplitude-modulated stimuli. Terhardt [11] demonstrated a dependence of roughness on the frequency and depth of the modulation. He also found that differences in roughness can be measured for sounds having the same amplitude spectrum but different waveform envelopes. The hypothesis he advanced is that roughness is determined by the envelope fluctuations of the signal within an auditory filter. Models have therefore been developed in which roughness estimates are based on the magnitude of temporal fluctuations of the signal envelope after auditory filtering ([4] for instance).

In the general case, the spectral and temporal cues are strongly covariant. An experiment has therefore been designed to oppose them in order to provide quantitative experimental data relevant to a discussion of the concurrent approaches. In addition, some authors claimed that roughness differences could be heard for sounds having the same envelope and frequency composition. This point, still lacking experimental evidence, will be addressed as well.

3 Experiment

3.1 Stimuli and method

Let us consider three spectral components with frequencies \(\{f_c - f_m, f_c, f_c + f_m\}\), with relative amplitudes \(1/2, 1, 1/2\) and with initial phases of \(\{0, \phi, 0\}\). For \(\phi = 0\), a 100% sinusoidally amplitude modulated sound with center frequency \(f_c\) and modulation frequency \(f_m\) is obtained. By varying \(\phi\), a family of sounds that have the same amplitude spectrum but different temporal structures is obtained. We will refer to these as “pseudo-AM” (pAM) stimuli. If \(\phi\) changes, the shape of the temporal envelope usually varies. It is also possible to obtain the same amplitude spectrum and the same envelope, but a different temporal fine structure by creating pAM signals corresponding to \(\phi = +\Phi\) and \(\phi = -\Phi\) with \(\Phi \in [0, \pi/2]\).

Seven series of pAM stimuli were created in which a given center frequency and “modulation” frequency were associated. In each case, \(f_m\) was chosen to produce maximum roughness for a pure tone of frequency \(f_c\) modulated sinusoidally at \(f_m\). The phase \(\phi\) of the central component was varied from \(-\pi/2\) to \(+\pi/2\) in steps of \(\pi/6\). For each series, seven stimuli were thus produced with four different envelopes corresponding to the four absolute phase differences (characterized hereafter by their rms value).

Stimuli were presented over headphones to listeners. Center frequencies were arranged in randomized separate blocks, and within a block all 42 pairs of non-identical stimuli in both orders were tested. For each pair, listeners were asked to decide which sound was rougher (2AFC). The notion of roughness was introduced with the help of amplitude-modulated sounds the modulation depth of which the listener could vary in a continuous fashion, thus changing concomitantly the amplitude spectrum and envelope. Two groups of 15 listeners were presented two different sets of \(f_c\)s.

3.2 Results

The Bradley-Terry-Luce (BTL) method was used to construct a roughness scale derived from the binary paired-comparison judgments. In order to test whether differences between scale values are significant, they must be compared to the standard deviation of the results. The standard deviations were estimated nonparametrically by the bootstrap technique. The results for the seven series of stimuli are presented in Figure 1.

The relatively small size of the standard deviations indicates a strong agreement across subjects. The most obvious factor contributing to roughness is the envelope rms, greater envelope rms corresponding to higher roughness values. This effect is similar in range across most \(f_c\)s tested although there is a decrease in range for low \(f_c\)s. However, for \(f_c\)s at or below 4 kHz there is also a strong effect of the sign of \(\phi\), indicating a significant contribution of phase for stimuli with the same overall envelope and the same amplitude spectrum. This effect disappears for all \(f_c\)s at \(\phi = \pm \pi/2\) and for all phases for \(f_c = 8\) kHz.
4 Discussion

4.1 Interpretation of the results

The global increase of roughness with envelope rms for a given \( f_m \) and \( f_c \) is in agreement with the temporal models of roughness. We interpret the overall range reduction of variation at lower frequencies in terms of a critical bandwidth effect, as the frequency span of the pAM is much greater compared to the critical band at these \( f_c \)s. This suggests that the effect of global envelope, when manipulated by phase relationships, is all the more important when the three components can interact.

The observed significant influence of phase for a given envelope and amplitude spectrum is not expected with the envelope fluctuation hypothesis. A spectral interpretation may be proposed to account for this result: combination tones (CTs) generated by the three components of the pAM stimuli may have played a role in the roughness judgments. Among these tones, the first order cubic difference situated at \( 2(f_c) - (f_c + f_m) = f_c - f_m \) interacts with the acoustic component at the same frequency. Its phase and amplitude change as a function of \( \phi \), which may explain the differences observed in perceived roughness [1]. This interpretation nevertheless raises two questions. First, the phase effect disappears at \( f_c = 8 \) kHz, whereas combination tones are likely to be still present. To accommodate this finding, one would have to accept that the phase of the CTs does not change with the phase of the primaries in this high frequency region. Second, a large intersubject variability is generally observed in the phase dependence of the combination tone on the phases of the primaries. Our results display on the contrary a remarkable agreement between listeners.

A second framework consists in trying to link the roughness judgments to the temporal fine structure differences between stimuli. To assess these differences, a computational model that mimics the auditory system was used. A large variety of these models exist that present very similar characteristics. We chose to generate “neural activity patterns” by means of compression, rectification and 2-dimensional adaptive thresholding of the output of a gammatone filterbank [8]. These patterns are intended to simulate the neural responses (along the cochlea) to the stimuli. Examination of the simulations indeed showed that the patterns produced by the opposite phases had different shapes. In the positive phase condition, the modulation shows a well defined maximum, that resembles the zero-lag condition, whereas the negative condition displays a “smoother” shape. This difference disappears for the \( \pi/2 \) condition, as well as for the highest carrier frequency where temporal fine structure following by the auditory system is impossible. A periodicity coding mechanism
based on detection of maxima would be affected by such a change. Recent results obtained in the field of comparative psychology [5] indicated that members of other species also seem to be sensitive to auditory roughness. Therefore, it may be plausible that roughness derives from ubiquitous physiological mechanisms such as periodicity coding. An alternative interpretation of the results could therefore be that the temporal fine structure of the modulation present at the level of primary auditory nerve fibers has an influence on roughness perception.

4.2 Possible application to complex sounds

In the light of these experimental results, the roughness models presented in the introduction can be reviewed. The spectral models are generally well adapted to musical notation, where chords can be represented by a list of component notes (frequencies) and are therefore helpful when working on a score [10]. However, they would predict exactly the same roughness for all the stimuli tested (within a $f_0$ condition). If one is interested in more subtle differences of recorded or synthesized sounds, temporal models will therefore be better suited. These models would reproduce the overall roughness variation with envelope $rms$ but would fail to reproduce the difference observed between positive and negative phase condition. It may be argued that in free field listening conditions the effect of phase differences will be smeared out by sound propagation and reflections. However, the study of the influence of phase effects in a laboratory-like situation was a means to try to gain some insights in the mechanisms involved in roughness perception, that may turn out as determinant when carrying on to understand roughness of complex musical sounds. Our claim is that phase per se should be considered as an extra parameter, but rather that only a model based on auditory peripheral mechanisms will be able to include in a natural way all the dependencies of roughness on various acoustical features.

5 Conclusion

Phase changes that affect a signal’s envelope were shown to have a clear effect on roughness perception, all the more as the bandwidth of the signal is small compared to the auditory critical band. Differences in roughness could also be exhibited for signals having the same envelope, and were linked to characteristics of the peripheral auditory system. The results obtained with laboratory-like stimuli open up paths of reflection for roughness modeling in the framework of computer-aided composition environments.

References