VISUALIZATION AND CALCULATION OF THE ROUGHNESS OF ACOUSTICAL MUSICAL SIGNALS USING THE SYNCHRONIZATION INDEX MODEL (SIM)

Marc Leman

IPEM - Dept. of Musicology, Ghent University
Blandijnberg 2, B-9000 Ghent, Belgium
Tel:+32 9 2644125, Email:Marc.Leman@rug.ac.be

ABSTRACT

The synchronization index model of sensory dissonance and roughness accounts for the degree of phase-locking to a particular frequency that is present in the neural patterns. Sensory dissonance (roughness) is defined as the energy of the relevant beating frequencies in the auditory channels with respect to the total energy. The model takes rate-code patterns at the level of the auditory nerve as input and outputs a sensory dissonance (roughness) value. The synchronization index model entails a straightforward visualization of the principles underlying sensory dissonance and roughness, in particular in terms of (i) roughness contributions with respect to cochlear mechanical filtering (on a Critical Band scale), and (ii) roughness contributions with respect to phase-locking synchrony (=the synchronization index for the relevant beating frequencies on a frequency scale). This paper presents the concept, and implementation of the synchronization index model and its application to musical scales.

1. INTRODUCTION

Since its introduction by H. v. Helmholtz [1], the terms *sensory dissonance* and *roughness* have both been used to characterize the texture of a sound in terms of *impure* or *unpleasant* qualities. The sensation is associated with the physical presence of beating frequencies.

Following v. Helmholtz, many researchers have used the term *sensory dissonance* when speaking about tone relationships but it is nowadays more appropriate to use the term *roughness* [German: 'Rauigheit', French: 'Rugosité'] because this term is more general and can be applied to characterize impure or unpleasant qualities of all kinds of sounds, including noises. (In dealing with the texture of noisy sounds it is indeed rather awkward to use the term sensory dissonance.)

In the past decades, several models of sensory dissonance and roughness have been proposed. A distinction can be made between curve-mapping models and auditory models. The curve-mapping models (e.g. [2, 3, 4, 5]) perform a mapping of the frequency component pairs of a sound onto a psychoacoustical curve which expresses the dissonance value of the presented pair. The auditory models rely on auditory processing and provide an explanatory model for sensory dissonance. The approach is more general in that any kind of sound, including noise, can be dealt with. Daniel and Weber [6] have recently optimized the model of Aures [7] which calculates roughness as a sum of the 'energy' of the beating frequencies in auditory channels. Pressnitzer [8] uses a similar model but takes into account the effects of co-modulation in different auditory channels.

In what follows, we propose an auditory model of roughness which calculates the roughness using the energy of the beating frequencies. But rather than estimating the modulation depth or using a temporal filter for estimating the energy of the beating frequencies, our method uses a Fourier analysis of the neural discharges in the auditory channels. This approach relates the notion of roughness to the notion of *synchronization*, a important concept in neurophysiology which indicates the degree of a neuron's total firing rate that is phase-locked to the corresponding stimulus component [9, 10].

Hence the model is called the *synchronization index model* (SIM). It allows a straightforward visualization of roughness along two different scales: (i) roughness contributions with respect to cochlear mechanical filtering on a Critical Band scale, and (ii) roughness contributions with respect the synchronization indices for the relevant beating frequencies on a frequency scale.

2. FACTS AND MODELS OF SENSORY DISSONANCE AND ROUGHNESS

Studies of sensory dissonance and roughness have been conducted from both an experimental and computational point of view.

2.1. Experimental Evidence

In a contribution to sensory dissonance, Plomp [11] concludes that his experiments largely confirm v. Helmholtz's findings [1]. In particular, when two tones with small frequency differences are simultaneously presented, we hear one frequency with a beating of the amplitude. But at larger frequency differences it becomes impossible to follow rapid successions of beats and the sound gets a rough and unpleasant character, which is called (sensory) *dissonance*. When the frequencies widen even more, we hear two (or perhaps even more) pitches. The maximum for dissonance was postulated by v. Helmholtz to be in the order of 30-40 Hz but that statement could not be confirmed by the experiments of Plomp. Based on experimental findings, he extracted a rule of thumb suggesting that the maximum dissonance corresponds to an interval of about 25% of the critical bandwidth (Fig. 1), i.e. about 25 Hz in the range below 500 Hz (taking roughly 100 Hz as the bandwidth), and about 4-5 % of the frequency in the range above 500 Hz. Plomp's findings, in other words, imply that the frequency of the beats generating maximum dissonance increase with increasing carrier frequency, rather than being constant over the whole frequency range as v. Helmholtz suggested.

Later studies, based on amplitude modulated (AM) tones and noises, confirm Plomp's results, but provide additional details. In

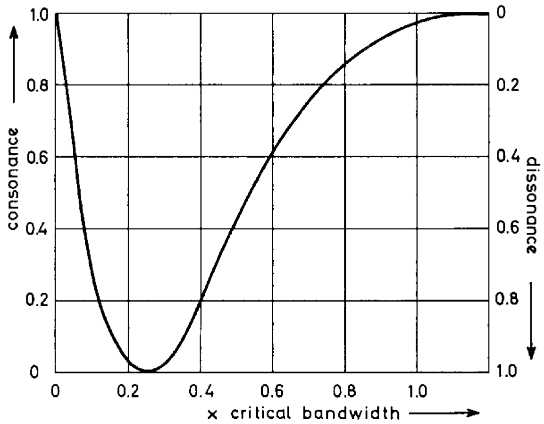

Figure 1: *Psychoacoustical curve representing the degree of sensory dissonance (roughness) in function of a critical band (CB). The maximum roughness is at .25 CB.*

working with AM tones, the term *roughness*, rather than sensory dissonance, is then often used. Zwicker and Fastl [12, p.234] relate roughness to three attributes: (i) the degree of amplitude modulation (or the modulation index) m, (ii) the frequency of the modulation f_m, and (iii) the center frequency of the sound f_c.[1]

- The dependency of the roughness R on the degree of modulation m has been expressed as a power law $R \sim m^n$. Depending on the experimental setup, n can slightly vary from 1.2 up to about 2 [6, 13].
- For a 1000 Hz tone with $m = 1$ the maximum roughness is found at f_m 70 Hz. This corresponds to 0.4375 of the critical band (of about 160 Hz), which appears to be higher than Plomp's suggestion to take 0.25 of the critical band. However, a close inspection of the data[2] reveals that the discrepancy between Plomp's data and Zwicker and Fastl's is perhaps less than would appear when applying the rule of thumb ("0.25 of the critical band").
- With respect to f_c, the roughness curves seem to follow an attenuated pattern in that the curve for f_c 1000 Hz is high, while the roughness curves for higher and lower f_c are typically attenuated. A further characteristic is that for $f_c \geq 1000$ Hz , maximum roughness is at 70 Hz (or even slightly higher), while it is lower for $f_c < 1000$ Hz.
- Recent results by Pressnitzer [8] show a dependency of roughness on phase. In particular, the form of the amplitude envelope may play a role in the perception of roughness and there are clear effects of co-modulation of phase. The latter means that roughness is normally higher when the synchronization in auditory channels is in phase. It is lower, when the synchronization is not in phase. Hence, the total roughness is not a simple linear addition of the roughness in the auditory channels. In fine-tuning roughness models, phase effects should be taken into account.

[1] The parameters correspond to the parameters found in the classical expression of an AM signal: $s(t) = (1 + m.sin(2\pi f_m t)).sin(2\pi f_c t)$

[2] Thanks to D. Greenwood (personal communication, and Auditory List)

2.2. Computational Models

A first category of computational models of sensory dissonance is based on a *curve-mapping* principle, i.e.: the mapping of all frequency intervals or frequency component pairs present in the spectrum of the sound to a psychoacoustical curve of sensory dissonance. The dissonance of a complex tone is then defined to be equal to the sum of the dissonances generated by each pair of adjacent frequency components. A family of computational models (e.g. [2, 3, 4, 5]) calculate sensory dissonance according to the following steps:

1. determination of the frequencies of the partials
2. matching of each possible interval of partials onto the psychoacoustical curve of Fig. 1
3. summation of all dissonance values.

Music theory, when dealing with sensory dissonance, traditionally draws on this model because it allows a straightforward symbolic and hence highly abstract calculation of sensory dissonance that often suits the representation of music as a score. The curve-mapping models moreover give a good fit with the tone perception experiments because they are directly based on it. Sethares [5] shows that curve-mapping models can go beyond the symbolic input. In particular, the Fast Fourier Transform (FFT) can be applied to sounds, providing lists of amplitude-frequency pairs which can then be processed with the curve-mapping model.

Although it was realized that sensory dissonance had to be explained on the basis of temporal properties of sounds [3, p.1457], the early models relied on calculations using the mapping method. As a result, the curve-mapping models are unable to handle different important details such as amplitude (e.g. partials with different amplitude), and phase effects, which are not dealt with at all. Also noises cannot be handled, although many sounds in popular as well as modern music include noises. Apart from their useful applications in music theory, curve-mapping models thus have a limited scope both in view of theoretical grounding and practical application.

A second class of models is based on auditory modeling. Aures [7] has formulated an *auditory* model of sensory dissonance and roughness that was intended to deal with the above mentioned shortcomings. The model has recently been optimized by Daniel and Weber [6]. In what follows, we propose an alternative view based on the idea that roughness may be accounted for in terms of the energy provided by the neural synchronization to beating frequencies rather than a direct estimation of the modulation depth or 'energy' of fluctuations. Our method uses Fourier analysis to estimate the modulation depth. This allows a visualization of the beating components that contribute to sensory dissonance and roughness.

3. NEURAL SYNCHRONIZATION AND SENSORY DISSONANCE

3.1. The Synchronization Index Model

The synchronization index model consists of two parts: (a) a module of the auditory periphery, called APM, and (b) a synchronization index module, called SIM. In the present model, APM is the acoustical front end taken from [14], which is similar to the model of Daniel and Weber [6], and Pressnitzer [8]. It simulates the cochlear mechanical filtering using an array of overlapping band-pass filters. The module provides rate-code patterns of neural discharge at the level of the auditory nerve, representing the amount of neural excitation during short time intervals.

A most important feature of the APM – on which our modeling approach is built – is that temporal fluctuations, or beats, are introduced as effective beating frequencies into the spectrum of the neural rate-code patterns. Recall that an amplitude modulated sound with carrier frequency f_c and modulation frequency f_m shows only three frequencies in the spectrum, in particular: f_c, $f_c - f_m$, and $f_c + f_m$. Due to wave rectification, a property of the frequency encoding in the cochlea which is due to a polarization of the stereocillia [15, 16], the lower part of the modulated signal is cut off, and as a result new frequencies are introduced of which the most important ones correspond with the beating frequency f_m and its multiples. As a matter of fact, neurons may synchronize with these modulation frequencies provided that they fall in the frequency range where synchronization is physiologically possible [9] thus forming a basis for the detection of beats and hence, the sensation of roughness.

According to Javel et al. [9, p. 520-521], the synchronization index represents the degree of phase-locking to a particular frequency that is present in the neural pattern. The index can be mathematically expressed as the (normalized) Fourier Series coefficient at the frequency of interest. Normalization implies the division of the Fourier amplitude by the DC or overall discharge rate, which gives a rate-independent measure of neural synchronization. In a similar way, Sach et al. [10] define the synchronization index as the amplitude of the Fourier transform normalized by the average rate. It shows the fraction of a unit's total rate that is phase-locked to the corresponding stimulus component.

In what follows, we give an analysis of this concept in view of auditory modeling and aspects related to roughness. As a first approximation, we could say that the degree of roughness can be defined as the sum of the normalized magnitudes or 'energies' of the relevant beating frequencies in the Fourier spectrum. Obviously, in view of the above facts, a more sophisticated account will be needed.

3.2. Formal Description, and Definitions

In this section, a formal description is given of possible ways of how to deal with the neural discharge patterns from the point of view of signal analysis. In this section, however, we describe some general definitions and observations of synchronization in terms of energy contributions. In a next section these concepts are then adapted to roughness modeling.

3.2.1. Signal Decomposition and wave-rectification

The decomposition into frequency sub-bands is formally expressed as:

$$APM: \; s(t) \rightarrow \tilde{d}(t) \; = \; < d_c(t) >_{c=1 \ldots C} \qquad (1)$$

where the acronym APM stands for *auditory peripheral module*, and $d_c(t)$ specifies the probability of neural discharge (i.e. the so-called *rate*-code) at intervals of 0.4 ms of the auditory channel or nerve fiber c. The auditory channel is considered a filter with a bandwidth equal to the critical bandwidth. C such filters are considered. The APM is based on [14] but any other functional equivalence model of the auditory periphery could be used. What is needed is the output at the level of the auditory nerve fibers in terms of rate-code, that is, the probability of neural firing for a given time instance.

3.2.2. Analysis of Synchronization

A frequency analysis of the neural discharge patterns in the auditory channels gives information about the synchronization of the neurons to particular frequencies. In this model we take for granted that the synchronization properties have been defined in the APM. The short-term spectrum of the neural synchronization in channel c is:

$$D(t,f,c) \; = \; \int_{t'=-\infty}^{+\infty} d_c(t) w(t'-t) e^{-j2\pi f t'} \, dt' \qquad (2)$$

where $d_c(t)$ is the neural pattern in channel c, and $w(t'-t)$ is a (hamming) window. This formula calculates the amount of synchronization for each frequency. The *magnitude* spectrum is then defined as $\mid D(t,f,c) \mid$, and the *phase* spectrum as $< D(t,f,c)$.

The synchronization index, according to the specifications given above [9, 10], is defined as:

$$\mid I(t,f,c) \mid \; = \; \left| \frac{D(t,f,c)}{D(t,0,c)} \right| \qquad (3)$$

where $D(t,0,c)$ is the DC component or the average range of the signal.

Different analysis approaches are now possible, given the fact that the synchronization information is represented along three dimensions: time, frequency, and auditory channel. In adopting the general modeling strategy that the concept of a model should be as simple as possible we focus on two straightforward approaches which both assume that roughness is somehow related to the energy of the neural synchronization. The first model (Model I) is based on the idea that the energy of neural synchronization is first computed in each individual auditory channel, and that the total energy is then the sum of these channel energies. The second approach (Model II) assumes that the neural synchronization over all channels is first combined and that the total energy is based on the neural synchronization accumulated over all channels. In practice it turns out that both approaches work reasonably well.

3.2.3. Model I

First consider the idea that the total energy of neural synchronization is the sum of channel energies. The following definitions give an account of different concepts used:

- The short-term *energy spectrum* of neural synchronization in channel c is:

$$\mathbf{D}(t,f,c) \; = \; \mid I(t,f,c) \mid \,^2 \qquad (4)$$

  Each frequency component f gives the amount of synchronization in terms of energy to this particular frequency. Each component is called a *synchronization index*.

- The short-term energy spectrum of neural synchronization *over all channels* c is:

$$\mathbf{D}(t,f) \; = \; \sum_{c=1}^{C} \mathbf{D}(t,f,c) \qquad (5)$$

  where C is the total number of auditory channels.

- The short-term *energy* of neural synchronization in channel c is defined as the sum of the energy of the neural synchronization indices:

$$E_{\mathbf{D}}(t,c) \;=\; \int \mathbf{D}(t,f,c)\,df \tag{6}$$

From the previous definitions it follows that the total amount of synchronization energy over all channels can be expressed in different ways:

$$\begin{aligned} E_{\mathbf{D}}(t) &= \int \mathbf{D}(t,f)\,df &&= \int \textstyle\sum_{c=1}^{C} \mathbf{D}(t,f,c)\,df \\ &= \textstyle\sum_{c=1}^{C} E_{\mathbf{D}}(t,c) &&= \textstyle\sum_{c=1}^{C} \int \mathbf{D}(t,f,c)\,df \end{aligned} \tag{7}$$

Equation 7 implies that is possible to visualize two aspects of the energy as represented in the spatio-temporal representation at the level of the auditory nerve. Visualization of the energy spectrum $\mathbf{D}(t,f)$ gives the energy of the neural synchronization along a frequency axis. Visualization of the energy along the auditory channels $E_{\mathbf{D}}(t,c)$ gives the energy of neural synchronization along the critical band scale. This double viewpoint will be used for the representation of roughness.

3.2.4. Model II

The approach of Model I does not take into account possible effects of phase over all channels because the synchronization is first calculated per channel, and afterwards summed. Model II first sums the spectra:

- The short-term spectrum of neural synchronization over all channels is:

$$D(t,f) \;=\; \sum_{c=1}^{C} I(t,f,c) \tag{8}$$

The short-term *energy spectrum* of neural sychronization is:

$$\mathbf{D}(t,f) \;=\; \left|\, \sum_{c=1}^{C} D(t,f) \,\right|^{2} \tag{9}$$

and the (short-term) *energy* of neural synchronization is:

$$E_{\mathbf{D}}(t) \;=\; \int \mathbf{D}(t,f)\,df \tag{10}$$

The quantities obtained in Equation 10 are different from those of Equation 7. Furthermore, in Model II it is not possible to visualize *two* aspects of the energy of neural synchronization because the signal is summed over all channels.

4. ROUGHNESS MODELING

The SIM approach calculates roughness in terms of the 'energy' of neural synchronization to the beating frequencies. The 'energy' refers to a quantity which we derive from the magnitude spectrum. Since the beating frequencies are contained in the lower spectral area of the neuronal pattern d_c, the spectral part we are interested in is defined as:

$$B(t,f,c) \;=\; F(f,c)D(t,f,c) \tag{11}$$

where $F(f,c)$ is a filter whose magnitude spectrum is depending on the channel c. In order to be able to reproduce the psychoacoustical data, the filters should become more narrow at auditory channels whose center frequency is below 800 Hz. And the filters should be attenuated for high center frequencies as well (see Sect. 5). In general, however, we can say that $B(t,f,c)$ represents the *spectrum* of the neural synchronization to the beating frequencies in channel c. The synchronization index of the beating frequencies is given by:

$$|\,I(t,f,c)\,| \;=\; \left|\, \frac{B(t,f,c)}{D(t,0,c)} \,\right| \tag{12}$$

where $D(t,0,c)$ is the DC-component of the whole signal. We now assume that roughness is related to the 'energy' of this normalized Fourier transform.

In analogy with the previous section we consider two models:

4.0.5. Model I

Roughness is calculated in the individual channels and the total roughness is the sum of the channel roughnesses. Given the above discussion, it will be possible to visualize the contribution of the synchronization energy along the axis of the auditory channels, as well as along the axis of the beating frequencies. The short-term *'energy' spectrum* of the neural synchronization to beating frequencies in a particular auditory channel c is defined as:

$$\mathbf{B}(t,f,c) \;=\; |\,I(t,f,c)\,|^{\;\alpha} \tag{13}$$

where α is a parameter which can be related to the power law mentioned in Sect. 2.1 ($1 < \alpha < 2$). In analogy with Equation 7 we then obtain the following relationships for the calculation of roughness:

$$\begin{aligned} R_{\mathbf{B}}(t) &= \int \mathbf{B}(t,f)\,df &&= \int \textstyle\sum_{c=1}^{C} \mathbf{B}(t,f,c)\,df \\ &= \textstyle\sum_{c=1}^{C} E_{\mathbf{B}}(t,c) &&= \textstyle\sum_{c=1}^{C} \int \mathbf{B}(t,f,c)\,df \end{aligned} \tag{14}$$

This expression entails a proper visualization along the axis of auditory channels and along the axis of the (beating) frequencies.

4.0.6. Model II

Model II is similar but based on the idea that the channels are first combined so that the phase can be taken into account in a direct way. The spectrum of the neural synchronization to beating frequencies over all channels is:

$$B(t,f) \;=\; \sum_{c=1}^{C} B(t,f,c) \tag{15}$$

The synchronization index of the beating frequencies is then defined as:

$$|\,I(t,f)\,| \;=\; \left|\, \frac{B(t,f)}{D(t,0)} \,\right| \tag{16}$$

where $D(t,0)$ is the DC-component summed over all channels, i.e. $D(t,0) \;=\; \sum_{c=1}^{C} D(t,0,c)$.

The short-term *'energy' spectrum* of the overall neuronal synchronization is:

$$\mathbf{B}(t,f) \;=\; |\,I(t,f)\,|^{\;\alpha} \tag{17}$$

and *roughness* is defined as the sum of these energies:

$$R_{\mathbf{B}}(t) \; = \; \int \mathbf{B}(t,f)\, df \qquad (18)$$

The addition of the neural synchronization spectra automatically takes into account certain effects related to the phase. For example, due to the phase delay in the lower auditory channels, it is predictable that tones with a low carrier frequency will get less roughness. The spread of beating energy over auditory channels is somehow countered by phase delays in these channels, so that the beating becomes less prominent. Expression 18 allows for a visualization in terms of the beating frequencies. Additional constrains, such as those mentioned by Pressnitzer [8, p.165] could be taken into account for dealing with the effect of phase but more research is needed to fine-tune these aspects.

5. IMPLEMENTATION

The SIM for roughness has been implemented in Matlab on top of the APM of Van Immerseel and Martens [14]. This APM has been adapted for musical purposes. In particular, synchronization in all channels is allowed up to 1250 Hz. Both APM and SIM form part of the IPEM Toolbox for auditory-based musical analysis [17].

Given the general concept as described above, the implementation requires a specification of the filters $F(f,c)$. In the present modeling, we choose elegance as an important criterium to describe the filters mathematically. The form of the filter $F'(f,c)$ has first been defined according to the following formula's:

$$F'(f,c) \; = \; e^{-8\frac{f}{f_B}} \left[1 - cos(2\pi \frac{f}{10 f_B}) \right] w(c) \qquad (19)$$

where f_B is the frequency range of the filter and f is running from 1 to f_B, $w(c)$ is a weight depending on the channel c. The weight takes into account a linear decrease of the impact of the filter depending on the auditory channel number:

$$w(c) \; = \; 1 - \frac{0.55c}{C}) \qquad (20)$$

with c running from 1 to C (=40). The frequency range f_B is defined such that it is narrow for the lower auditory channels below 800 Hz, broad at about 1000 Hz and again slightly more narrow at frequencies higher than 1500 Hz.

A sigmoid function has been defined according to

$$S \; = \; \frac{(\frac{c}{C})^{.2}}{0.04 + (\frac{c}{C})^{2.45} - 0.007c} \qquad (21)$$

The sigmoid function is used to define the frequency range as:

$$f_B(c) = 10 + (300 * \frac{S}{S_{max}}); \qquad (22)$$

It means that the frequency range of the filter in the lowest auditory channel is 10 Hz, and that this range is at most 310 Hz (around 1000 Hz). The values of S are normalized so that the maximum value of $\frac{S}{S_{max}}$ is equal to one.

Given the set of filters $F(f,c)$, the sigmoid function (21) has also been used to define where the maximum of the filter should be located. Data indicate that the maximum around 1000 Hz is at about 70 Hz. It is lower for low auditory channels.

$$f_M(c) = 20 + (52 * \frac{S}{S_{max}}); \qquad (23)$$

This function starts at 20 Hz and the maximum is obtained at 72 Hz.

The precise location of the filter shape $F'(f,c)$ on the frequency axis is then determined by placing its maximum at $f_M(c)$. The obtained set of filters $F(f,c)$ are shown in Fig. 2. Small changes in parameters do not have a dramatic effect on the performance of the model. In what follows, we give some examples of

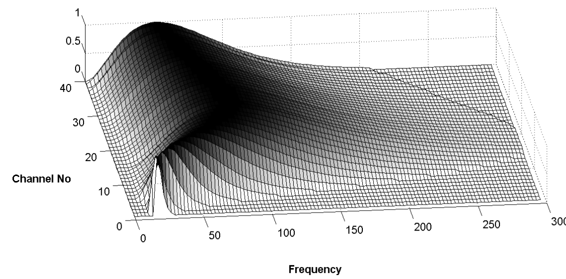

Figure 2: *Filters become more narrow in the lower auditory channels.*

the output of the model, using the calculations according to Model I. Model II is only slightly different.

6. APPLICATION TO PSYCHOACOUSTIC DATA

Figure 3 shows the curves for amplitude modulated sounds having a different f_c. The modulating frequency f_m changes from 0 to 250 Hz, while the modulation index m is 1. The figures shows a

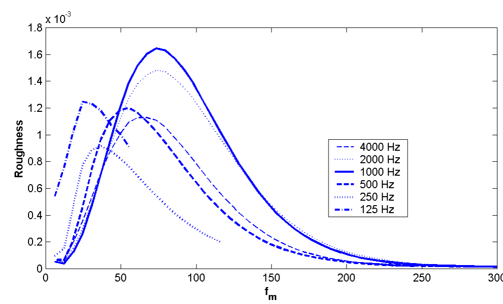

Figure 3: *Roughness in function of modulation frequency on different carrier frequencies.*

good similarity with the figure in [12, p.232] but some more fine-tuning is needed to bring down the roughness of the 125 Hz carrier.

7. APPLICATION TO MUSIC

Below we apply SIM to a tone complex that defines different musical intervals of a timbre over one octave. The harmonic tone complex consists of a fundamental (f_0) at 500 Hz and 5 harmonics with equal amplitude. This tone is played together with a pitch shifted copy. The shift over 5 seconds is linear in frequency up to the upper octave (f_0 1000 Hz). The roughness as calculated with the synchronization index model can be compared the model of [5], which uses the curve-mapping method of Plomp. Sethares' model takes the frequency-amplitude values as input (no sound!)

and calculates the curve using the psychoacoustical curve shown in Fig. 1.

SIM, apart from its good agreement with this theoretical model provides an additional cue in showing the 'spectral' (=excitation in the auditory channels) as well as 'temporal' (=synchronization index) factors that contribute to roughness, as shown in Fig. 4. The upper panel shows how the energy is distributed over the auditory channels, and the middle panel shows how the energy of the beating frequencies contributes to the roughness curve shown in the lower panel. According to the model, both the upper and middle panel lead to the same curve. The points of minimal roughness or sensory dissonance indicate a hierarchical order of intervals in terms of roughness. This hierarchy can be musically exploited as the points of minimum roughness may indicate candidates for a musical scale.

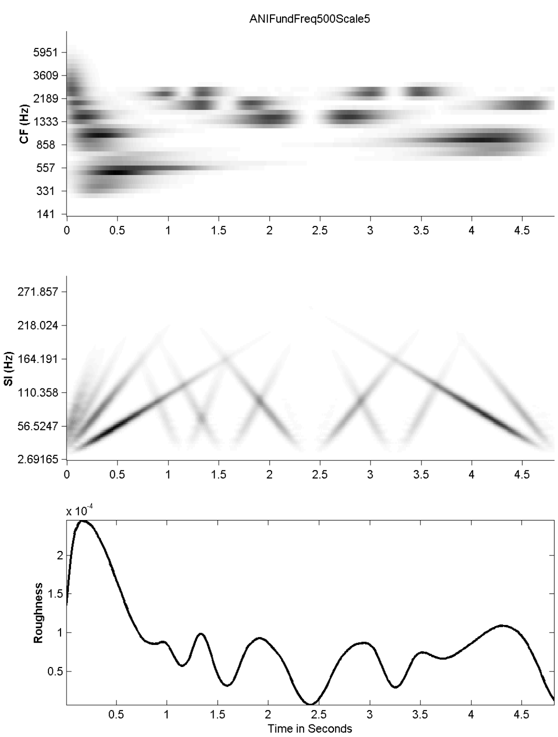

Figure 4: *Roughness of a harmonic tone complex.*

8. CONCLUSION

In this paper, we introduced a new model of roughness based on the concept of *synchronization index*, that is, the amount of neural activation that is synchronized to the timing of the amplitudes of the beating frequencies in the stimulus. The visualization provides interesting cues for analyzing the factors that contribute to sensory dissonance and roughness.

9. REFERENCES

[1] H. von Helmholtz, *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, Georg Olms Verlagsbuchhandlung, Hildesheim, 1863/1968.

[2] R. Plomp and W. Levelt, "Perception of tonal consonance," in *Studies in Perception*, M.A. Bouman, Ed. Institute for Perception, Soesterberg, 1966.

[3] A. Kameoka and M. Kuriyagawa, "Consonance theory Part I: Consonance of dyads," *The Journal of the Acoustical Society of America*, vol. 45, pp. 1451–1459, 1969.

[4] W. Hutchinson and L. Knopoff, "The acoustic component of western consonance," *Interface - Journal of New Music Research*, vol. 7, pp. 1–29, 1978.

[5] W. Sethares, *Tuning, Timbre, Spectrum, Scale*, Springer-Verlag, Berlin Heidelberg, New York, 1998.

[6] P. Daniel and R. Weber, "Psychoacoustical roughness: Implementation of an optimized model," *Acustica*, vol. 83, pp. 113–123, 1997.

[7] W. Aures, "Ein Berechnungsverfahren der Rauhigkeit," *Acustica*, vol. 58, pp. 268–281, 1985.

[8] D. Pressnitzer, *Perception de rugosité psychoacoustique: d'un attribut élémentaire de l'audition a l'écoute musicale*, Ph.D. thesis, Université Paris 6, 1999.

[9] E. Javel, J. McGee, J.W. Horst, and G.R. Farley, "Temporal mechanisms in auditory stimulus coding," in *Auditory function: neurobiological bases of hearing*, G.M. Edelman, W. Gall, and W. Cowan, Eds. John Wiley and Sons, New York, 1988.

[10] M. Sachs, R. Winslow, and C. Blackburn, "Representation of speech in the auditory periphery," in *Auditory function: neurobiological bases of hearing*, G.M. Edelman, W. Gall, and W. Cowan, Eds. John Wiley and Sons, New York, 1988.

[11] R. Plomp, *Experiments on tone perception*, Van Gorcum, Assen, 1966.

[12] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, Springer-Verlag, Berlin, Heidelberg, 1990.

[13] E. Terhardt, *Akustische Kommunikation*, Springer, Berlin, Heidelberg, 1997.

[14] L. Van Immerseel and J. Martens, "Pitch and voiced/unvoiced determination with an auditory model," *The Journal of the Acoustical Society of America*, vol. 91, pp. 3511–3526, 1992.

[15] J. Pickles, *An Introduction to the Physiology of Hearing*, Academic Press, London, 1982.

[16] H.-P. Zenner, *Hören – Physiologie, Biochemie, Zell- und Neurobiologie*, Georg Thieme Verlag, Stuttgart, 1994.

[17] M. Leman, M. Lesaffre, and K. Tanghe, *A Toolbox for Perception-Based Music Analysis (manuscript)*, IPEM - Dept. of Musicology, Ghent University, Ghent, 2000.