Audio Engineering Society
Convention Paper 7516
Presented at the 125th Convention
2008 October 2–5 San Francisco, CA, USA

"Roughometer": Realtime Roughness Calculation and Profiling

Julián Villegas[1] and Michael Cohen[1]

[1] *University of Aizu, Aizu-Wakamatsu; Fukushima-ken 965-8580; Japan.*

Correspondence should be addressed to Julián Villegas (`julovi at yahoo · com`)

ABSTRACT

A software tool capable of determining auditory roughness in real-time is presented. This application, based on Pure-Data (Pd), calculates the roughness of audio streams using a spectral method originally proposed by Vassilakis. The processing speed is adequate for many realtime applications, and results indicate limited but significant agreement with an internet application of the chosen model. Finally, the usage of this tool is illustrated by the computation of a roughness profile of a musical composition that can be compared to its perceived patterns of 'tension' and 'relaxation.'

1. INTRODUCTION

Roughness is an acoustic phenomenon that produces continuous and quantitative changes in perception (prothetic sensation) associated with amplitude variations between 15 and 300 Hz approximately, as illustrated in Fig. 2 [1]. It has been linked to musical dissonance & consonance, 'tension' & 'relaxation,' and has been recognized as a technique for expression in composition and performance [2] [3] [4]. Fig. 1 shows the numerous factors contributing to musical consonance. In spite of this breadth, many researchers have focused on roughness to explain consonance and dissonance. This has prompted criticism from some others who maintain that the contribution of roughness to musical dissonance is overestimated, and that the abundance of research regarding the link between the two phenomena is due to the more accessible understandability and experimentability of roughness [6]. Nevertheless, there exists evidence indicating the de-

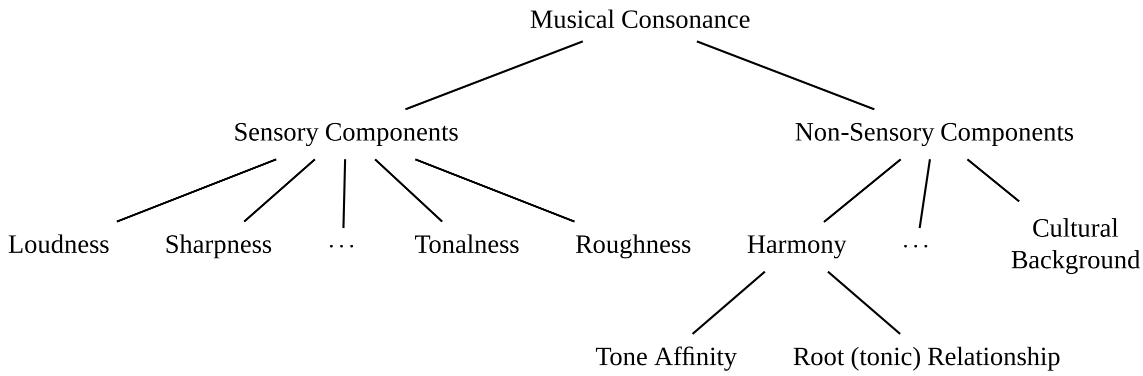


Fig. 1: Musical consonance taxonomy (incomplete). This modified version of Terhardt's hierarchy [5] includes harmony as a 'non-sensory' component. Tonalness (not to be confused with tonality) is the extent to which a sound is considered a tone or a noise.

pendence of musical dissonance on roughness. In various musical traditions (including non-western cultures), certain tuning systems yield musical renditions which are usually characterized as more consonant, and therefore preferred. These tuning systems are specified in a way that important notes of a given scale yield a minimal roughness [7] when rendered simultaneously (harmony). Also, experimenting with computer generated sounds, Pierce [8] shows how stretched scales played with stretched partial sounds are rated as consonant, whereas other combinations of stretched/unstretched scales and stretched/unstretched partial sounds were not (See Hotsuma et al. demonstrations 58, 59, 60, and 61). Finally, it has been shown that in *a cappella* SATB ensembles, singers change the intonation of tones in order to minimize roughness, producing in consequence tonal drift [9]. Regardless of the influence of roughness upon music, there are few ways to measure it in real-time. A tool that allows such quantification could be useful for analysis and synthesis of sounds. Sethares [7] describes other potential uses of a roughness-meter: dynamic (adaptive) tuning, construction of scales depending on timbre, construction of timbres based upon scales, and intonation monitoring. We present an alternative for such analysis consisting of a Pd object and patch.

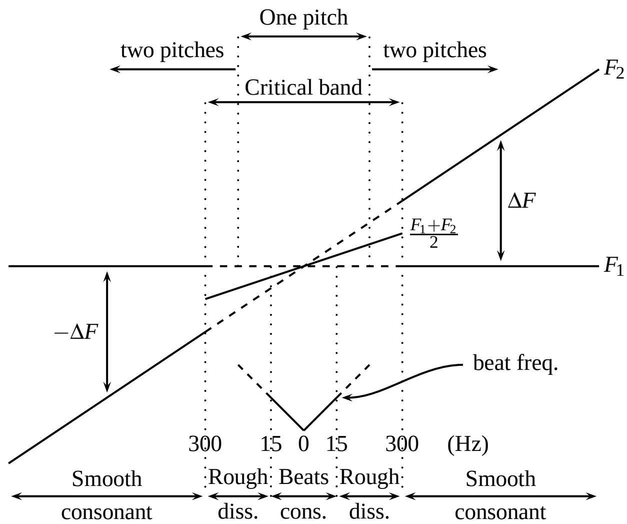


Fig. 2: The perception of two simultaneous pure tones with frequencies F_1 and F_2, juxtaposed against a critical band. The roughness region (between 15 and 300 Hz) grows with the center frequency of the critical band. Adapted from Roederer [10].

2. ROUGHNESS MODELS

Existing roughness models are derivations of pitch discrimination theories, and can be categorized into two groups: those asserting that the main cause of roughness is the presence of multiple frequency components within the same auditory filter, and those that claim that roughness is attributable to the variations in time of the amplitude envelope in each auditory filter. Table 1 summarizes the main differences between the two approaches.

To date, there is no single, unifying theory that explains all aspects of the pitch discrimination performed by the inner ear [11]. As a result, many roughness models have been proposed, some of which are incompatible. Several of the most-cited roughness models are: Helmholtz [12], Fastl & Zwicker [1], Kameoka & Kuriyagawa [13], Hutchinson & Knopoff [14], Aures [15, 16], and Daniel & Weber [17].

Recently, a model based upon the spectral approach was proposed by Vassilakis [3]. This model does not account for the contribution of the phases nor the temporal asymmetries of the sound waves, but it does include the effect of loudness, which had been missing in previous models (e.g., Plomp & Levelt [18] and Sethares [19]). More importantly, Vassilakis' model corrects the common misconception that the AM modulation index *m* is equivalent to the relative amplitude fluctuation *h*.

2.1. Pressnitzer's model

Pressnitzer found that in addition to the modulation frequency and index *m*, the phase and the temporal asymmetries of a sound wave also contribute to the perception of roughness [20]. However, he notes that in free field conditions, the effect of the phase differences is diminished by the multiple sound reflections and propagations as reported by Risset [21]. Pressnitzer opines that phase shouldn't be included in the list of acoustic parameters of a roughness computation model, but a model based on the simulation of the hearing system includes all roughness dependencies in a natural manner [2]. The spectral and temporal models account for the pitch discrimination performed in the basilar membrane but only temporal models include other characteristics of the hearing system.

Pressnitzer's model differs from other temporal models (particularly those of Aures and Daniel & Weber) in that the decomposition of the audio signal is performed by a bank of gammatone filters (lineal) [22] as opposed to the critical band filterbank (non-lineal) [23]. Also, Pressnitzer's model explicitly includes an RMS calculation stage instead of an inter-correlation stage, as illustrated by Fig. 3. According to Pressnitzer, the inter-correlation stage is unable to reproduce the effect of the envelope phase for wide frequency separations [2].

Pressnitzer acknowledges discrepancies between the results obtained with his and other temporal methods, but observes that they produce the same general predictions [20].

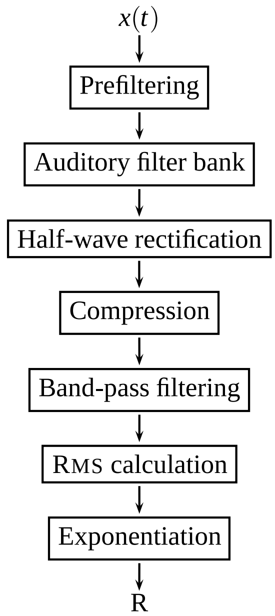


Fig. 3: Roughness calculation using Pressnitzer's model.

2.2. Vassilakis' model

Vassilakis observed that because of the confusion of modulation parameters *m* with *h*, Terhardt's experimental results on the perception of periodic sound fluctuations [25] must be reinterpreted. The modulation index is the coefficient that indicates the extent to which an AM signal is modulated. It was erroneously assumed that changes in *m* were equivalent to changes in *h*, e.g., fixing $m = 50\%$ was assumed to produce a crest-trough difference of 50% of the peak-value, when *h* is in fact about 67%. Fig. 4b illustrates this case.

Spectral Approach	*Temporal Approach*
Roughness can be calculated by analyzing sound spectra.	Roughness can be calculated by analyzing the temporal envelope of the signals in each aural filter across the hearing spectrum.
For a sinewave dyad, roughness disappears when the frequency difference is greater than the size of a critical band centered at the geometric mean.	For a 1 kHz AM tone with modulation index $m = 1$, roughness tends to disappear for modulation frequencies $f_m < 15$ Hz or $f_m > 300$ Hz.
Maximum roughness is produced when the frequency difference of a sinewave dyad is a constant fraction (about a quarter) of the centered critical band.	For a 1 kHz AM tone with $m = 1$, maximum roughness is achieved for a modulation frequency $f_m = 70$ Hz.
This approach contradicts the acoustic uncertainty principle ($\Delta f \Delta t =$ constant). There is no evidence that the frequency difference limen decreases inversely to the stimulus duration. Refer to demonstration 29 of Hotsuma et al. [24].	This approach doesn't explain the pitch-shift effect. As intensity increases, high and low frequency pitches are perceived higher and lower respectively. Refer to demonstrations 27 and 28 of Hotsuma et al. [24].

Table 1: Some differences between explanations of roughness and pitch discrimination [2] [11].

According to Terhardt, roughness r is related to modulation index m by the power law expression

$$r = c m^{2\cdot}, \tag{1}$$

where c is a constant. Vassilakis [3] shows that

$$m = \frac{h}{2-h} \tag{2}$$

and analyzes Terhardt's data to build his roughness model. For Vassilakis, the roughness r of a dyad $\langle f_1, a_1\rangle, \langle f_2, a_2\rangle$ is

$$r = \frac{(a_1 a_2)^{0.1}}{2}\left(\frac{2\min(a_1,a_2)}{a_1+a_2}\right)^{3.11} Z, \tag{3}$$

where

$$Z = (e^{-3.5F} - e^{-5.75F}), \tag{4}$$

$$F = S(\min(f_1, f_2))\ |f_1 - f_2|, \tag{5}$$

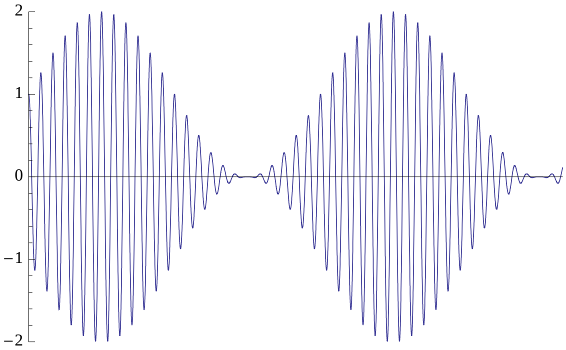
and

$$S(f) = \frac{0.24}{0.0207 f + 18.96}. \tag{6}$$

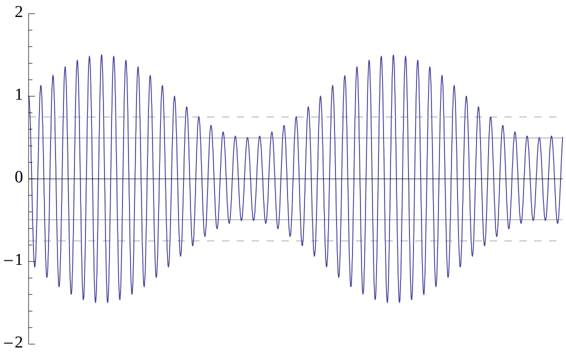
Eq. 3 accounts for the influence of the intensity (first term), degree of amplitude fluctuation (second term), and frequency separation vs. critical band (third term) on roughness [26].

Vassilakis agrees with Pressnitzer in that phases and temporal asymmetries of audio signals are in many cases random and hence their influence is not accounted in Eq. 3, however, Vassilakis indicates the needed steps for including phases in his model if needed [27].

The exponent 3.11 in the second term is the result of reinterpreting Terhardt's data. It has been shown that the exponent in Eq. 1 varies from 1.4 to 2.0, depending on the experimental method used [28]. Terhardt used paired comparisons, whereas Guirao & Garavilla [29] used direct magnitude estimation and found an exponent of 1.4. Vassilakis' model is based on Terhardt's findings, and adjustments of Eq. 3 to reflect other exponents are not contemplated.



(a) For $m = 100\%$, the relative peak-valley difference in amplitude (h) is 100%.



(b) For $m = 50\%$, $h = 2/3$ of the maximum amplitude. The solid and dashed lines indicate 33% and 50% of the maximum amplitude respectively.

Fig. 4: Differences between m and h in an AM tone $x(t) = 1 + m\cos(\omega_m t)\sin(\omega_c t)$.

Following the approaches taken by Plomp & Levelt [18] and Sethares [19], Vassilakis assumes that the roughness of a complex tone is the accumulation of the partial contribution of each dyad. Therefore, for a complex tone (or a set of complex tones), the total roughness R is given by the expression

$$R = \sum_{h=0}^{n}\sum_{i=0}^{n}\sum_{j=0}^{p}\sum_{k=0}^{p} r(a_{hj}, a_{ik}, f_{hj}, f_{ik}), \quad (7)$$

where n is the number of audio streams, p is the number of partials considered in the calculation, x_{uv} means the amplitude (a) or frequency (f) of the v-th frequency component of the u-th audio stream, and r is as defined earlier in Eq. 3.

In general terms what Vassilakis is proposing is that given the frequency components of a sound, the total roughness can be calculated by adding the roughness of all constituent dyads. The roughness of the dyads depends on their frequencies, amplitudes, and spectral separation. Therefore, loud complex tones with many frequency components and a low fundamental frequency are expected to yield a great roughness value.

Vassilakis' model can be considered an improvement of previous spectral methods, but it is too focused on the pitch discrimination performed by the basilar membrane, disregarding other possible causes. There is still a lot to investigate about roughness but we have incorporated Vassilakis' model into the present work in an attempt to use it to better understand other musical phenomena related to roughness.

3. PREVIOUS WORK

Of the surveyed implementations of roughness models, none were suitable for realtime computation. Three of the most relevant applications are 'Psysound3' [30], Spectral and Roughness Analysis (SRA) [27] [26], and 'Audition' [31]. Psysound3 implements the models of Hutchinson & Knopoff and Daniel & Weber. SRA implements Vassilakis' model. Both programs only process sound files, and therefore perform a postmortem (offline) analysis of roughness. Additionally, it is unclear how to extend these applications, or how to use them with multi-track audio. Audition (a Pd library), developed by Gnansia & Pressnitzer, realizes a variation of Pressnitzer's model in real-time (replacing the RMS stage by a low-pass filtering) [32]. Currently, it is not possible to duplicate their results concerning roughness because at least one object present in their publication is not present in their library. The library can be downloaded from Gnansia's website [31], and our Mac OS implementation is available at Villegas' website [33].

4. IMPLEMENTATION

Our application to measure roughness in real-time was created in Pure-data (Pd), a graphical programming environment developed by Miller Puckette. Details on the Pd paradigm, programming model, and extension possibilities are documented by its active user community [34].

The developed prototype relies mainly on the output of '`sigmund~`,' a built-in Pd object that performs spectral analysis and pitch tracking of either an incoming audio stream or audio samples stored in arrays. `Sigmund~` reports, among other things, a specified number of frequency components ordered by amplitude ('peaks') or arranged in continuous voices ('tracks'). A complete explanation of the features of this object is beyond the scope of this paper, but is provided in its Pd help patch.

Best results are obtained using `sigmund~` 'peaks' output. The frequency components reported by this object are routed to '`roughometer`,' a roughness calculation object created for the purposes of this research. `Roughometer` estimates the roughness of concurrent sounds by applying Vassilakis' model. The number of audio streams is specified as a parameter of the object. `Roughometer` automatically creates the necessary number of inlets according to this specification.

Limitation of the number of streams that can be analyzed depends on the underlying hardware and a trade-off between the number of frequency components considered and the specified delay time in the audio interface. The number of frequency components is automatically inferred from the list received at each inlet. This list observes the convention used by `sigmund~` output (component number, frequency, peak amplitude, cosine and sine parts).

The first inlet of `roughometer` receives '`bang`' messages that cause the calculation to be performed and sent to the output. The same inlet also receives an amplitude threshold that excludes small amplitude frequency components from the roughness calculation. The number sent to this inlet ins divided internally by 1,000. This 'gate' feature lessens the influence of background noise in live situations, poor quality recordings, false components reported by `sigmund~`, etc.

In Pd, audio signals are represented by floating point numbers normalized between ± 1, and `sigmund~` peak amplitudes are reported as positive floating point numbers, so we set the default amplitude threshold to 0.001 in order to preserve the relevant information but exclude possible noise.

The values reported by `roughometer` are scaled by 100. Fig. 4 shows a screenshot from a patch created using `roughometer`. `Roughometer` source code, libraries for Mac OS and MS Windows, documentation, and example patches, are available at the first author's website [33].

5. RESULTS

The application was tested on a MacBook with 2 GB of RAM running Mac OS X v.10.5.3. Version 0.40.3-extended-20080531 of Pd was connected to Jack OS X audio server v.0.8.6. The sampling rate of the patch was set to 48 kHz, and the buffer in the audio device to 3 ms. This latency was determined empirically so that no audio drops occur.

In order to test the application, the roughness profile of Bach chorale BWV 264 was calculated twice and compared to the profile obtained from SRA. The first time, each of the four chorale voices was synthesized independently in GarageBand v.4.1.2 using the default piano sound. In the second, a mixdown of the four channels was analyzed. By means of these two renditions, it was possible to register the behavior of the application for separate sound sources and ensembles.

In both exercises, roughness was computed every 250 ms, and `sigmund~` was parametrized to report 'peaks,' to use a window of 4096 samples with 256 samples between analyses, and to exclude frequencies greater than 100 kHz. For the mixdown version, `sigmund~` was configured to use the first 40 frequency components, and `roughometer` threshold was set to 0.0025. In the case of the four independent files, only the first 16 frequency components were considered in the analysis and the `roughometer` amplitude threshold was set to 0.001. Table 2 summarizes the obtained results.

In parallel to the roughness profile, a delay profile was computed. The delay was measured between

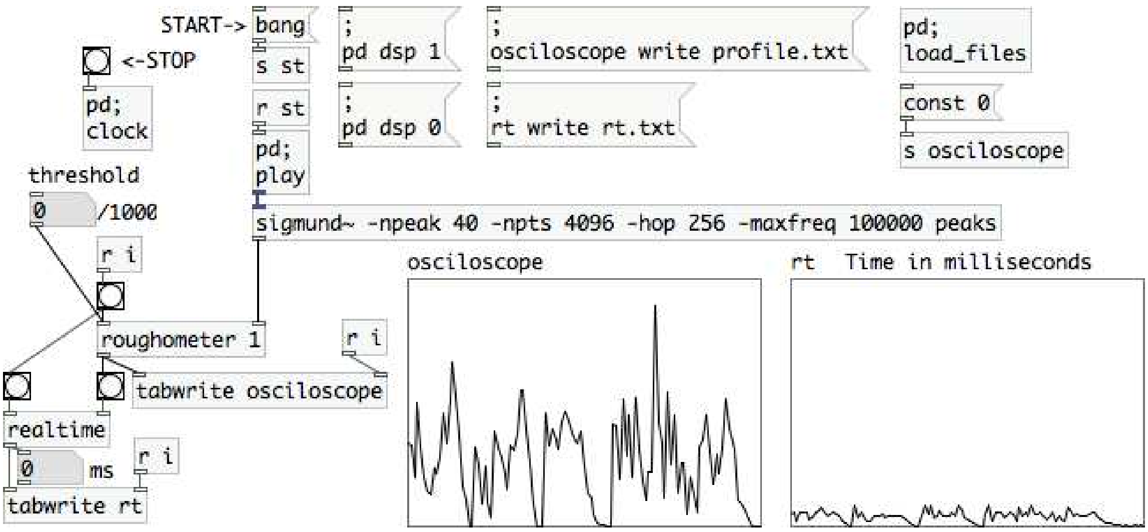


Fig. 5: A screenshot of a patch including `sigmund~` and `roughometer` objets. In this example, only one audio source is being analyzed. It could easily be modified to allow analysis of concurrent audio streams by connecting `adc~` objects to an equal number of `sigmund~` objects, and connecting the latter to a single `roughometer`.

	roughometer		
	four files	*mix-down*	SRA
avg. delay	0.8 ms	0.15 ms	N/A
correl. with SRA	60%	61%	(100%)

Table 2: A comparison between the results obtained using `roughometer` and SRA.

the arrival of each ‘`bang`’ message, originated in a `clock` subpatch, and its corresponding output. The delay profile for the mix-down version is shown in Fig. 5. The average computation time for the four independent voices (the most demanding condition) was 0.8 ms with a standard deviation of 0.55 ms and a maximum value of 2.19 ms. Although these time values may change for different inputs, it is expected that they would remain in the same range for the same platform. In the same vein, the delay was measured with a `realtime` object. It is known that the accuracy of this object differs across different CPUs, yet its approximate estimate is sufficient for illustrating the time demands of `roughometer`.

The roughness profile obtained with `roughometer` for the mixed-down version was compared against that obtained with SRA. The two profiles are shown in Fig. 7. Their correlation was 0.61. This rather low value is explained by the differences between the Pd and SRA implementations. One of the most important differences is that `sigmund~` uses an FFT with a Hann window whereas SRA uses short-term Fourier transformation (STFT) and automatic spectral peak-picking [27]. Also, SRA uses information from a time window centered at the analysis point in time; this window includes part of the spectra beyond the time being analyzed. For causality reasons, it is impossible to perform such analysis in a realtime implementation, and therefore the values reported by `roughometer` can be considered a compromise between accuracy and speed. Finally, notice that neither of the applications consider the absolute roughness value (in *Aspers* units)

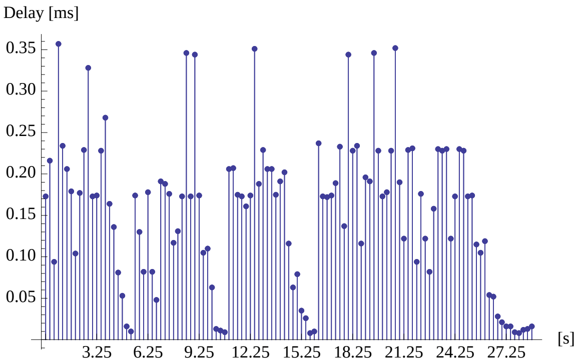
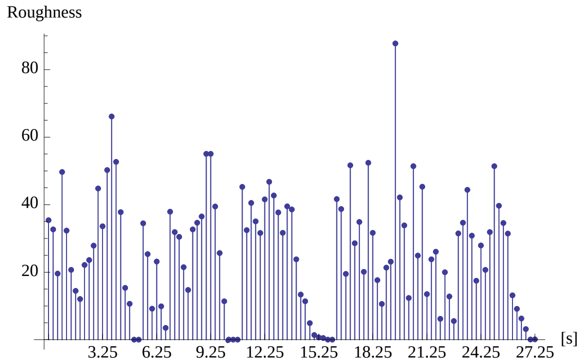
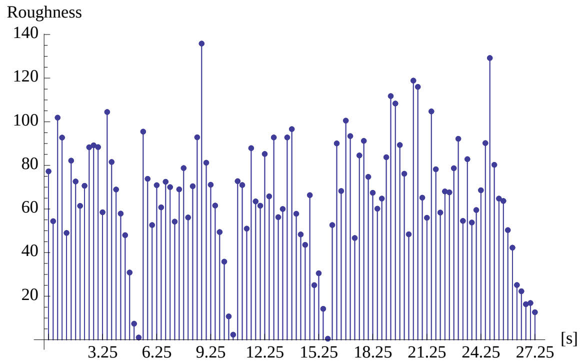


Fig. 6: The delay profile for the Bach chorale. This profile resembles the roughness profile. For long notes, the natural decay of the sound produces lower amplitudes and therefore lower roughness values. Some of these amplitudes are smaller than the computation threshold, easing in consequence the overall calculation.



(a) Roughmeter



(b) SRA

Fig. 7: The roughness of Bach chorale BWV 264 ‘Als der gütige Gott vollenden wollt sein Wort’ calculated by the two tools.

as they are concerned only with relative differences of roughness.

6. CONCLUSIONS AND FUTURE WORK

We have introduced `roughometer`, a Pd roughness calculator based on Vassilakis’ model. The time requirements are suitable for many realtime applications, and its limited accuracy can be used as a coarse indicator of roughness when time constraints are severe.

The accuracy of our implementation is attributable to the frequency components extraction. We have used `sigmund~`, which was available in Pd, for illustrating the use of `roughometer`, but it could be replaced by a better algorithm in the future.

We are encouraged by the results obtained with `roughometer` to study roughness-related problems. We are currently working on an application that minimizes the roughness of an ensemble. This system could be used as an psychoacoustic alternative for auto-tuning systems.

Finally, Hartmann [11] has suggested that pitch determination might depend on spectral cues for high frequencies and time cues for low frequencies. Such dependency could be reflected in the perception of roughness. We are interested in integrating `roughometer` with Pressnitzer’s model, so users can analyze audio with either of those methods in a single application.

7. REFERENCES

[1] H. Fastl and E. Zwicker, *Psychoacoustics: facts and models*, vol. 22 of *Springer series in information sciences*. Berlin: Springer, 3rd ed., 2007.

[2] M. D. Pressnitzer, *Perception de Rugosite Psychoacoustique: D’un Attribut Elementaire de*

L'Audition a L'Ecoute Musicale. PhD thesis, Universite Paris VI, 1998.

[3] P. N. Vassilakis, *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance*. PhD thesis, University of California — Los Angeles, 2001.

[4] B. Muszkalska, "Vos desafinada — the 'out-of-tune voice' in Portuguese polyvocal songs," *Revista Transcultural de Música*, vol. 5, 2000. ISSN 1697-0101.

[5] E. Terhardt, "Ein psychoakustisch begründetes konzept der musikalischen konsonanz," *Acustica*, vol. 36, pp. 121–137, 1976.

[6] R. Parncutt, "Commentary on Keith Mashinter's 'Calculating sensory dissonance: Some discrepancies arising from the models of Kameoka & Kuriyagawa, and Hutchinson & Knopoff'," *Empirical Musicology Review*, vol. 1, no. 4, pp. 201–204, 2006.

[7] W. A. Sethares, *Tuning, timbre, spectrum, scale*. London: Springer, 2nd ed., 2005.

[8] J. Pierce, *Music, Cognition, and Computerized Sound*, ch. Consonance and Scales. Cambridge, Masachusetts: The MIT Press, 2001. ISBN 0-262-53190-9.

[9] D. Howard, "A cappella SATB quartet in-tune singing: Evidence of intonation shift," in *Proc. Stockholm Music Acoustics Conf.*, vol. 2, pp. 462–466, SMAC-03, August 2003.

[10] J. G. Roederer, *Introduction to the physics and psychophysics of music*, vol. v. 16. London: English Universities Press, 1973.

[11] W. M. Hartmann, "Pitch, periodicity, and auditory organization," *J. Acoust. Soc. America*, vol. 100, no. 6, pp. 3491–3502, 1996.

[12] H. v. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, ch. VIII, On the beats of simple tones. Dover Publications, II english ed., 1954.

[13] A. Kameoka and M. Kuriyagawa, "Consonance theory, part II: Consonance of complex tones and its calculation method," *J. Acoust. Soc. America*, vol. 45, no. 6, pp. 1460–1469, 1969.

[14] W. Hutchinson and L. Knopoff, "The acoustic component of western consonance," *Interface*, vol. 7, no. 1, pp. 1–29, 1978.

[15] W. Aures, "Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale," *Acustica*, vol. 59, pp. 130–141, 1985.

[16] W. Aures, "Der sensorische wohlklang als funktion psychoakustischer empfindungsgrössen," *Acustica*, vol. 58, pp. 282–290, 1985.

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

[18] R. Plomp and W. Levelt, "Tonal consonance and critical bandwidth," *J. Acoust. Soc. America*, vol. 38, no. 4, pp. 548–560, 1965.

[19] W. Sethares, "Adaptive tuning for musical scales," *J. Acoust. Soc. America*, vol. 96, no. 1, pp. 10–18, 1994.

[20] D. Pressnitzer and S. McAdams, "Two phase effects in roughness perception," *J. Acoust. Soc. America*, vol. 105, no. 5, pp. 2773–2782, 1999.

[21] J.-C. Risset, "Perception, environnement, musiques," *InHarmoniques*, no. 3, Musique et perception, 1988.

[22] R. D. Patterson, J. Nimmo-Smith, J. Holdsworth, and P. Rice, "An efficient auditory filterbank based on the gammatone function," in *Proc. Meeting of the IOC Speech Group on Auditory Modelling at RSRE*, 1987.

[23] E. Zwicker, G. Flottorp, and S. Stevens, "Critical bandwidth in loudness summation," *J. Acoust. Soc. America*, vol. 29, no. 5, pp. 548–557, 1957.

[24] A. J. M. Houtsma, T. D. Rossing, and W. M. Wagenaars, "Auditory demonstrations," 1987. Philips compact disc No. 1126-061.

[25] E. Terhardt, "On the perception of periodic sound fluctuations (roughness)," *Acustica*, vol. 30, pp. 201–213, 1974.

[26] P. N. Vassilakis and K. Fitz, "SRA: A web-based research tool for spectral and roughness analysis of sound signals," 2008. [Online; accessed 9-July-2008]: `www.acousticslab.org/roughness`.

[27] P. N. Vassilakis, "A web-based research tool for spectral and roughness analysis of sound signals," in *Proceedings SMC'07, 4th Sound and Music Computing Conference*, pp. 319–325, July, 2007. Lefkada, Greece.

[28] W. D. Baene, A. Vandierendonck, and M. Leman, "Roughness perception of amplitude modulated tones is context dependent," 2004. [Online; accessed 9-July-2008]: `http://ecow.engr.wisc.edu/cgi-bin/get/ece/401/2sethares/homework/psychophysical.pdf`.

[29] M. Guirao and J. M. Garavilla, "Perceived roughness of amplitude-modulated tones and noise," *J. Acoust. Soc. America*, vol. 60, no. 6, pp. 1335–1338, 1976.

[30] D. Cabrera, S. Ferguson, F. Rizwi, and E. Schubert, "Psysound3: Sound analysis software," 2008. [Online; accessed 9-July-2008]: `http://psysound.wikidot.com`.

[31] D. Pressnitzer and D. Gnansia, "The audition library for pure data: a platform for real-time auditory modelling," 2008. [Online; accessed 9-July-2008]: `http://lumiere.ens.fr/Audition/tools/realtime`.

[32] D. Gnansia, "Modèle auditif en temps réel," Master's thesis, Pierre & Marie Curie University, 2005.

[33] J. Villegas, "Julián Villegas website," 2008. [Online; accessed 9-July-2008]: `www.julovi.net`.

[34] PD-community, "The Pure Data Portal," 2008. [Online; accessed 9-July-2008]: `http://puredata.info`.