Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications

Pantelis N. Vassilakis*[a], Roger A. Kendall[b]
[a]Dept. of Audio Arts and Acoustics, School of Media Arts, Columbia College Chicago,
33 E. Congress Parkway, Room 601, Chicago, IL 60605-1996
[b]Dept. of Ethnomusicology, Program in Syst. Musicology, Music Perception and Cognition & Musical Acoustics Lab., UCLA, Schoenberg Hall, Box 951657, Los Angeles CA 90095-1657

ABSTRACT

The term "auditory roughness" was first introduced in the 19th century to describe the buzzing, rattling auditory sensation accompanying narrow harmonic intervals (i.e. two tones with frequency difference in the range of ~15-150Hz, presented simultaneously). A broader definition and an overview of the psychoacoustic correlates of the auditory roughness sensation, also referred to as sensory dissonance, is followed by an examination of efforts to quantify it over the past one hundred and fifty years and leads to the introduction of a new roughness calculation model and an application that automates spectral and roughness analysis of sound signals. Implementation of spectral and roughness analysis is briefly discussed in the context of two pilot perceptual experiments, designed to assess the relationship among cultural background, music performance practice, and aesthetic attitudes towards the auditory roughness sensation.

Keywords: Auditory roughness, roughness, sensory dissonance, dissonance, timbre, spectral analysis, computer modeling

1. INTRODUCTION

1.1 Auditory Roughness: Definitions

The term auditory roughness was introduced in the acoustics and psychoacoustics literature by Helmholtz[1] to describe the buzzing, harsh, raspy sound quality of narrow harmonic intervals. Within the Western musical tradition, auditory roughness constitutes one of the perceptual correlates of the multidimensional concept of dissonance, concept that has historical, cultural, and cognitive bases, along with physical and physiological ones[2-3]. The attribute of dissonance correlating best with auditory roughness has been termed sensory or tonal dissonance[4] or auditory dissonance[5], to mark its dependence more on physical and physiological, rather than cognitive, historical, or cultural considerations.

A familiar example of a signal corresponding to a rough sound would be the signal of a harmonic minor second performed, for instance, on two flutes. Although a harmonic minor second will sound rough regardless of the sound sources involved, steady state sources such as wind instruments, singing voices, or bowed strings (as opposed to impulse sources such as percussion, plucked strings, etc.) result in more salient roughness sensations[6-7]. At relatively low registers, wider intervals such as major seconds and minor thirds can also sound rough and, within the Western musical tradition, are usually avoided as dissonant. For example, the general practice in Western art music orchestration of spacing out harmonic intervals more at low registers than at high registers has its basis on roughness considerations.

More broadly, the term auditory roughness can be used to describe the buzzing sound quality of a variety of signals, beyond those of narrow harmonic intervals, such as the signals corresponding to fast trills, fast vibrato, percussive rolls, rattles, etc. Roughness is one of the perceptual manifestations of interference and, in the physical frame of reference it is usually described as a function of a signal's amplitude envelope (i.e. amplitude fluctuation rate and depth) and corresponding spectral distribution. As such, auditory roughness can also be considered an attribute of timbre.

The reason all complex signals, including the signals of chords, harmonic intervals, etc., exhibit amplitude fluctuations is physical and is related to the phenomenon of interference. The reason why some of these signals correspond to rough sounds is physiological and has to do mainly with the properties of the inner ear (full review in [2]).

*pvassilakis@colum.edu; phone 1 773 750 4874; www.acousticslab.org

Human Vision and Electronic Imaging XV, edited by Bernice E. Rogowitz, Thrasyvoulos N. Pappas,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7527, 75270O · © 2010 SPIE-IS&T · CCC code:
0277-786X/10/$18 · doi: 10.1117/12.845457

1.2 Signal Amplitude Fluctuation, Critical Band, and Auditory Roughness

Amplitude fluctuations describe variations in the maximum value (amplitude) of sound signals relative to a reference point and are the result of wave interference. The interference principle states that the combined amplitude of two or more vibrations (waves) at any given point in time and/or space may be larger (constructive interference) or smaller (destructive interference) than the amplitude of the individual vibrations (waves), depending on their phase relationship. In the case of two or more waves with different frequencies, their periodically changing phase relationship results in periodic alterations between constructive and destructive interference, giving rise to the phenomenon of periodic amplitude fluctuations.

Amplitude fluctuations can be placed in three overlapping perceptual categories related to the rate of fluctuation. Slow amplitude fluctuations (~ ≤15 per second) are perceived as loudness fluctuations referred to as beating. As the rate of fluctuation is increased, the loudness appears to gradually become constant and the fluctuations are perceived as "fluttering," "buzzing," or roughness. As the amplitude fluctuation rate is increased further, the roughness reaches a maximum strength and then gradually diminishes until it almost disappears (~ ≥75-150 fluctuations per second, depending on the frequency of the interfering waves)[2, 6, 8-9].

Assuming the ear performs a frequency analysis on incoming signals[1-6, 8], the perceptual manifestations of amplitude fluctuation can be related directly to the bandwidth of the hypothetical analysis-filters, depending upon and defining what Zwicker[9] termed critical bandwidth. For example, in the simplest case of amplitude fluctuations resulting from the addition of two sine signals with frequencies f_1 and f_2, the fluctuation rate is equal to the frequency difference between the two sines $|f_1-f_2|$, and the following statements represent the general consensus:

a) If the fluctuation rate is smaller than the critical bandwidth, then a single tone is perceived either with fluctuating loudness (beating) or with roughness.

b) If the fluctuation rate is larger than the critical bandwidth, then a complex tone is perceived, to which one or more pitches can be assigned but which, in general, exhibits little or no beating or roughness.

Psycho-physiologically, the roughness sensation can be linked to the inability of the auditory frequency-analysis mechanism to resolve inputs whose frequency difference is smaller than the critical bandwidth and to the resulting instability or periodic "tickling"[10] of the mechanical system (basilar membrane) that resonates in response to such inputs.

Along with amplitude fluctuation rate, the next most important signal parameter related to roughness is amplitude fluctuation degree[2, 7], that is, the level difference between peaks and valleys in signals with non-flat envelopes. The degree of amplitude fluctuation depends on the relative amplitudes of the components in the signal's spectrum, with interfering components of equal amplitudes resulting in the highest fluctuation degree and the highest roughness degree.

1.3 Auditory Roughness as Means of Musical Expression

The sensation of roughness has been creatively explored more than any other perceptual manifestation of amplitude fluctuation and by numerous musical traditions, a practice that has only recently been documented and researched[2-3]. Manipulating the degree and rate of amplitude fluctuation helps create the buzzing sound of the Indian tambura drone and the rattling effect of Bosnian ganga singing, resulting in a sonic canvas that becomes the backdrop for further musical elaboration. It permits the creation of timbral variations (e.g. Middle Eastern mijwiz playing) and rhythmic contrasts (e.g. ganga singing) through gradual or abrupt changes among roughness degrees. Whether such variations are explicitly sought after, as in ganga singing and mijwiz playing, or are introduced more subtly and gradually, as may be the case in the typical chord progressions/modulations of Western music, they form an important part of a musical tradition's expressive vocabulary. Other examples include the Quechua Haraui songs of Peru, with their frequent use of narrow harmonic intervals, and the performance of the taqara flutes of the Xingu river in Brazil, where sonic effects similar to those produced with the mizwij are produced by two or more simultaneous performers.

2. ROUGHNESS CALCULATION MODEL & SPECTRAL ANALYSIS METHOD

2.1 Background

Models that can systematically and reliably quantify the roughness degree of a given sound permit the empirical testing of hypotheses that link the roughness sensation to musical variables and concepts. For example, a reliable roughness calculation model may be used to experimentally examine claims that link auditory roughness to (a) dissonance within the Western musical tradition, (b) patterns of tension and release in Near Eastern or North Indian musical pieces (as intended by performers and/or perceived by listeners), or (c) rhythmic/timbral effects in Balkan folk songs.

Numerous roughness calculation models have been proposed over the last ~100-150 years [e.g. 1, 4-5, 11-13]. They have been employed in studies that attempt to link auditory roughness to auditory/sensory dissonance [e.g. 14-16], demonstrating a relatively low degree of agreement between calculated and experimental data. Surprisingly for post-1960 models, the two principal studies[6-7] that have systematically examined the relationship between a signal's amplitude fluctuation degree and roughness have been overlooked. All the above models (a) overestimate the contribution of sound pressure level (*i.e.* absolute amplitude values of the interfering signals) to roughness, (b) underestimate the contribution of the degree of amplitude fluctuation (*i.e.* relative amplitudes values of the interfering signals) to roughness, and (c) often misrepresent the relationship between roughness and register (review in [2]).

Below, we outline a new roughness calculation model that has been implemented in SRA[17], a Web-based tool dedicated to the spectral and roughness analysis of sound signals. Perceptual experiments testing the model indicate that it reliably and validly represents the perception of roughness, performing better than previous roughness calculation models[2-3].

2.2 Outline of the Proposed Roughness Calculation Model

The roughness **R** of a signal whose spectrum has two sinusoidal components with frequencies f_1, f_2 and amplitudes A_1, A_2, where $f_{\min} = \min(f_1, f_2)$, $f_{\max} = \max(f_1, f_2)$, $A_{\min} = \min(A_1, A_2)$, and $A_{\max} = \max(A_1, A_2)$, is[2]:

$$\mathbf{R} = \mathbf{X}^{0.1} * 0.5(\mathbf{Y}^{3.11}) * \mathbf{Z} \quad (1)$$

where:

$$\mathbf{X} = A_{\min} * A_{\max} \quad (1a)$$

$$\mathbf{Y} = 2A_{\min} / (A_{\min} + A_{\max}) \quad (1b)$$

$$\mathbf{Z} = e^{-b1s(f\max - f\min)} - e^{-b2s(f\max - f\min)} \quad (1c)$$

[where $b1 = 3.5$; $b2 = 5.75$; $s = 0.24/(s_1 f_{\min} + s_2)$; $s_1 = 0.0207$; $s_2 = 18.96$]

The term $\mathbf{X}^{0.1}$ in (1) represents the dependence of roughness on intensity (related to the amplitude of the added sines). It is based on previous experiments[7], adjusted[2, 18] to account for the quantitative difference between modulation depth, used in the previous experiments, and amplitude fluctuation degree, the signal parameter influencing roughness.
The term $\mathbf{Y}^{3.11}$ in (1) represents the dependence of roughness on amplitude fluctuation degree (related to the amplitude difference of the added sines). It, too, is based on previous experiments[7], adjusted to account for the quantitative difference between modulation depth and amplitude fluctuation degree[2, 18].
The term **Z** in (1) represents the dependence of roughness on amplitude fluctuation rate (frequency difference of the added sines) and register (frequency of the lower sine). It is based on Sethares's[13] modeling of the roughness curves in Figure 1, which have been derived from multiple perceptual experiments examining the roughness of sine-pairs[4-5, 11-13].

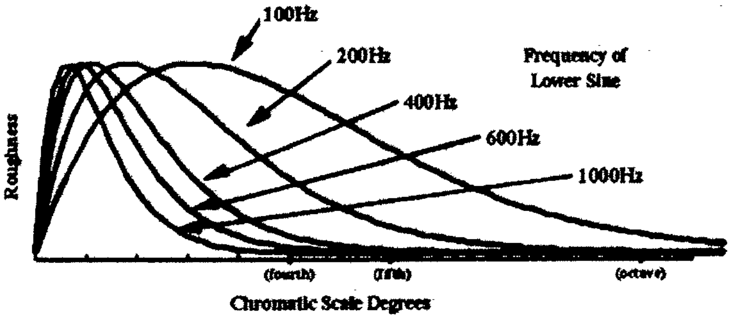

Figure 1. Roughness curves plotting observed roughness (arbitrary measure, *y* axis) of a pair of equal-amplitude sines, as a function of frequency separation (*x* axis) and frequency of the lower sine[2, 13].

The roughness of signals corresponding to spectra with more than two sine components is calculated by summing the roughness of all sine-pairs in the spectrum. Although it has been argued that, depending on the relative phase of the respective amplitude fluctuations, the total roughness can be less than the sum of the roughness values for individual sine-pairs[6], several studies[7, 19] and pilot experiments[2] indicate otherwise. More specifically, past studies[19] have concluded that the total roughness is summed over all auditory filters. In addition, since roughness modeling is meaningful to roughness comparisons among multiple signals, rather than to roughness calculations of isolated signals, any potential signal-envelope phase effects are more likely to be diffused within the signals of interest, the more complex the signals.

The phase of a signal's spectral components is not included as a parameter in the roughness calculation model. According to a previous study[20], the relative phase of the components of a three-component spectrum influences the complex signal's overall envelope shape and/or amplitude fluctuation degree, consequently influencing the signal's roughness, especially when three or more sine components fall within the same critical band. In spite of this observation, the absence of the phase parameter from the model does not significantly distort the model's calculations. For the types of signals submitted to the calculator (synthetic signals, where the phase relationship of the components can be controlled and remain the same for all, or natural signals from polyphonic passages, where the phase relationships are more likely to be random than systematic), differences in the roughness phase effects among the signals to be examined are either controllable or defused. This supports valid comparisons of the resulting relative roughness values.

As is the case with all roughness calculation models, the absolute roughness values calculated by the model are arbitrary and are only useful for roughness comparisons among signals that have been analyzed using consistent analysis parameters. The roughness calculations of the above model correlate very well (r = 0.98) with roughness ratings obtained in a set of perceptual experiments[2-3], better than predictions by previous models (r = 0.73 [1] and r = 0.87 [5]).

2.3 Spectral Analysis Method

The roughness model calculates the roughness of sound signals based on spectral information (frequency and amplitude values of a signal's spectral components). In its SRA[17] implementation, the model uses spectral information obtained from an improved Short-Time Fourier Transform (STFT) algorithm, which is based on reassigned bandwidth-enhanced modeling[21-25], and incorporates an automatic spectral peak-picking process to determine which frequency analysis bands correspond to spectral components of the analyzed signal.

Frequency reassignment[26] works differently from traditional Fast Fourier Transform (FFT) and has more in common with phase vocoder methods. For example, as in traditional FFT, frequency resolution of 10Hz will not be able to resolve frequency components laying less than 10Hz apart. But, unlike traditional FFT, the precision of the frequency values returned will not be limited by this 10Hz "bandwidth," since the frequency band boundaries are floating rather than fixed. This (a) fine-tunes the frequencies reported and (b) practically eliminates spectral smearing, since the method ensures that the standard assumption of all energy being located at the high-frequency end of an analysis band can, for the most part, be fulfilled.

Similarly, as in traditional FFT, a given analysis window length determines the length of the shortest signals that can be reliably analyzed. But, unlike traditional FFT, the temporal resolution of a signal's spectral (and therefore roughness) time-profiles is not limited by this "window length," since the frequency and amplitude estimates are not time-window averages but instantaneous at the time-window's center. This (a) pin-points time with much higher precision than implied by the window length and (b) practically eliminates temporal smearing, since the spectra estimated through time-window overlaps do not involve averaging over the entire analysis windows[21-25].

In practical terms, spectral analysis results are fine-tuned through the incorporation of a dual STFT process. Frequency values reported correspond to the time derivative of the argument (phase) of the complex analytic signal representing a given frequency bin. Similarly, time values reported correspond to the frequency derivative of the STFT phase, defining the local group delay and applying a time correction that pinpoints the precise excitation time.

3. POTENTIAL APPLICATIONS OF ROUGHNESS CALCULATION

Spectral analysis of sound signals and calculation of auditory roughness date back to the 19th century. SRA[17] performs spectral analysis on submitted sound signals and calculates their auditory roughness, either as a single roughness value or in the form of a roughness time-profile.

The spectral analysis portion of the application uses an improved STFT algorithm, based on reassigned bandwidth-enhanced modeling. It can pinpoint the instantaneous frequency and amplitude of a signal's spectral components with minimization of spectral and temporal smearing.

The roughness calculation portion incorporates a new roughness calculation model, which represents perceptual roughness more reliably and validly than previous auditory roughness models.

The implementation of spectral and roughness analysis in SRA has already found application in studies addressing a wide variety of research topics, ranging from timbre analysis[29-30] and sound synthesis[31] to dissonance[3] and musical tension[27-28] (see, for example, Figure 2). The reliability and validity of the spectral and roughness time-profiles returned by SRA can support work in acoustics, music and speech perception, voice pathology, or cross-cultural music research, and address questions that could not previously be tackled in a systematic manner.

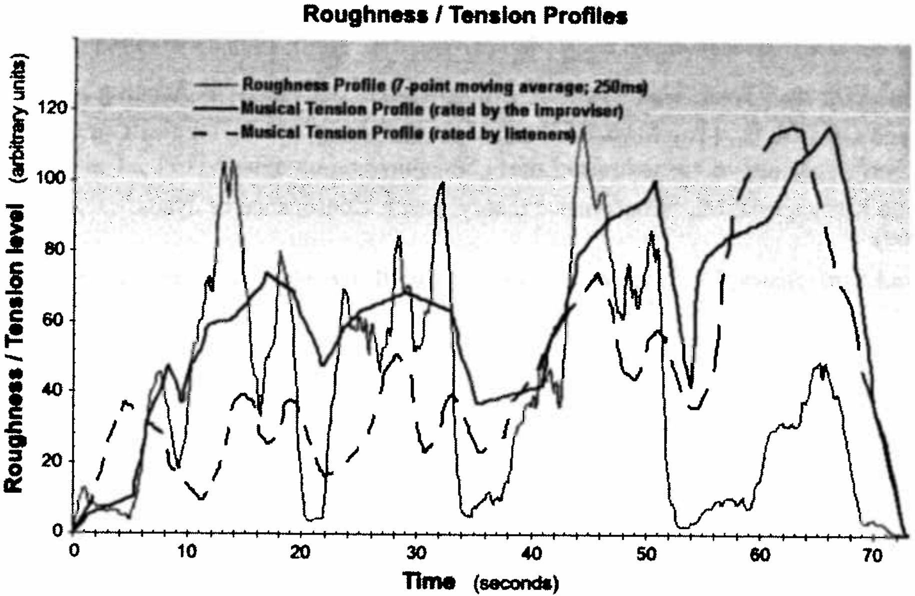

Figure 2. Roughness profiles obtained by SRA in numeric form can be easily represented graphically. The above results are adapted from a study examining the relationship among the roughness profile (calculated by SRA) and musical tension profiles (indicated by participants in a perceptual experiment) of an improvisation on the Middle Eastern mijwiz[27].

We are currently in the process of examining if/how cultural learning influences use of and emotional responses to auditory roughness. Based on SRA, we calculated the roughness time-profiles of a stylized improvisation on the Middle-Eastern mijwiz and of a traditional Bosnian ganga song and compared them to musical tension/release patterns indicated by (a) a Lebanese mijwiz player and scholar, (b) a Bosnian ganga singer and scholar, and (c) Western-trained musicians. Based on our initial data, the tension/release patterns indicated by the Lebanese and Bosnian musicians correlate well the respective calculated auditory roughness patterns, suggesting that roughness is closely related to the non-Western musicians' sense of musical tension. The patterns indicated by the Western-trained musicians indicate that roughness is just one of the cues guiding musical tension judgments, often overridden by tonal and temporal cues, and/or by expectations of tension/resolution raised by such cues. The observed differences between the non-Western performers' expressive intent and the Western-trained listeners' interpretation support understanding musical tension/release as culture-specific concepts, guided by the equally culture-specific musical cues employed in their organization and experience.

REFERENCES

[1] Helmholtz, H. L. F., [On the Sensations of Tone as a Physiological Basis for the Theory of Music], 2nd ed., A.J. Ellis, Trans., Dover Publications, Inc., New York (1885/1954).

[2] Vassilakis, P. N., [Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance], Doctoral Dissertation, University of California, Los Angeles, Los Angeles (2001).

[3] Vassilakis, P. N., "Auditory roughness as a means of musical expression," *Selected Reports in Ethnomusicology*, 12 (Perspectives in Systematic Musicology), 119-144 (2005).

[4] Plomp, R. and Levelt, W. J., "Tonal consonance and critical bandwidth," *J. Acoust. Soc. Am.*, 38(4), 548-560 (1965).

[5] Hutchinson, W. and Knopoff, L., "The acoustic component of Western consonance," *Interface*, 7, 1-29 (1978).

[6] von Békésy, G., [Experiments in Hearing], Acoustical Society of America Press, New York (1960/1989).

[7] Terhardt, E., "On the perception of periodic sound fluctuations (roughness)," *Acustica*, 30(4), 201-213 (1974).

[8] Plomp, R., "The ear as a frequency analyzer," *J. Acoust. Soc. Am.*, 36(9), 1628-1636 (1964).

[9] Zwicker, E., "Subdivision of the audible frequency into critical bands," *J. Acoust. Soc. Am.*, 33(2), 248-249 (1961).

[10] Campbell, M. and Greated, C., [The Musician's Guide to Acoustics], 2nd ed. Oxford University Press, New York (1994).

[11] Kameoka, A. and Kuriyagawa, M., "Consonance theory, part I: Consonance of dyads." *J. Acoust. Soc. Am.*, 45(6), 1451-1459 (1969).

[12] Kameoka, A. and Kuriyagawa, M., "Consonance theory, part II: Consonance of complex tones and its calculation method," *J. Acoust. Soc. Am.*, 45(6), 1460-1469 (1969).

[13] Sethares, W. A., [Tuning, Timbre, Spectrum, Scale], Springer-Verlag, London (1998).

[14] Bigand, E., Parncutt, R. and Lerdahl, F., "Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training," *Perception and Psychophysics*, 58, 125-141 (1996).

[15] Vos, J., "Purity ratings of tempered fifths and major thirds," *Music Perception*, 3(3), 221-258 (1986).

[16] Dibben, N., "The perception of structural stability in atonal music: The influence of salience, stability, horizontal motion, pitch commonality, and dissonance," *Music Perception*, 16(3), 265-294 (1999).

[17] Vassilakis, P. N. and Fitz, K., "SRA2.0: A Web-based Research Tool for Spectral and Roughness Analysis of Sound Signals," supported by a Northwest Academic Computing Consortium grant to J. Middleton, Eastern Washington University, 2009.
[http://musicalgorithms.ewu.edu/algorithms/roughness.html] or [http://www.acousticslab.org/roughness]

[18] Vassilakis, P. N., "Amplitude modulation depth versus degree of amplitude fluctuation: implementation error, adjustment and implications," *J. Acoust. Soc. Am.*, 108(5/2), 2597 (2000).

[19] Lin, J. Y. and Hartmann, W. M., "Roughness and the critical bandwidth at low frequency," *J. Acoust. Soc. Am.*, 97(5/2), 3274 (1995).

[20] Pressnitzer, D. and McAdams, S., "Two phase effects in roughness perception," *J. Acoust. Soc. Am.*, 105(5), 2773-2782 (1999).

[21] Fulop, S. A. and Fitz, K., "A spectrogram for the twenty-first century," *Acoustics Today*, 2(3), 26-33 (2006).

[22] Fulop, S. A. and Fitz, K., "Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications," *J. Acoust. Soc. Am.*, 119(1), 360-371 (2006).

[23] Fulop, S. A. and Fitz, K., "Separation of components from impulses in reassigned spectrograms," *J. Acoust. Soc. Am.*, 121(3), 1510-1518 (2007).

[24] Fitz, K. and Haken, L., "On the use of time-frequency reassignment in additive sound modeling," *J. Aud. Engin. Soc.*, 50(11), 879-893 (2002).

[25] Fitz, K., Haken, L., Lefvert, S., Champion, C. and O'Donnell, M., "Cell-utes and flutter-tongued cats: Sound morphing using Loris and the Reassigned Bandwidth-Enhanced Model," *Computer Music Journal*, 27(4), 44-65 (2003).

[26] Auger, F. and Flandrin, P., "Improving the readability of time frequency and time scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, 43, 1068-1089 (1995).

[27] Vassilakis, P. N., "An improvisation on the Middle-Eastern mijwiz; auditory roughness profiles and tension/release patterns," *J. Acoust. Soc. Am.*, 117(4/2), 2476 (2005).
Available online through the *Acoustical Institute of Physics Press Room* [http://www.aip.org/149th/Vassilakis.html]

[28] Vassilakis, P. N. and Kendall, R. A., "Auditory roughness profiles and musical tension/release patterns in a Bosnian ganga song," *J. Acoust. Soc. Am.*, 124(4/2), 2448 (2008).

[29] Bolger, D. and Griffith, N., "Multidimensional timbre analysis of shakuhachi *hohkyoku*," *Proc. Conference on International Musicology*, (2005).
[http://www.oicm.umontreal.ca/cim05/cim05_articles/BOLGER_D_CIM05.pdf]

[30] Hsu, W., "Managing gesture and timbre for analysis and instrument control in an interactive environment," *Proc. 2006 International Conference on New Interfaces for Musical Expression*, 376-379 (2006).

[31] D'Inca, G. and Mion, L., "Expressive audio synthesis: From performances to sounds," *Proc. 12th International Conference on Auditory Display*, 182-186 (2006).