Considering Roughness to Describe and Generate Vertical Musical Structure in Content-Based Algorithmic-Assisted Audio Composition

Gilberto Bernardes
University of Porto,
Faculty of Engineering
g.bernardes@fe.up.pt

Matthew E. P. Davies
INESC Porto
mdavies@inescporto.pt

Carlos Guedes
NYU Abu Dhabi
carlos.guedes@nyu.edu

Bruce Pennycook
UT Austin
bpennycook@austin.utexas.edu

ABSTRACT

This paper examines the correlation between musical dissonance and auditory roughness—the most significant factor of psychoacoustic dissonance—and the contribution of the latter to algorithmic composition. We designed an empirical study to assess how auditory roughness correlates with human judgments of dissonance in natural musical stimuli on the sound object time scale. The results showed a statistically significant correlation between roughness and listeners' judgments of dissonance for quasi-harmonic sounds. This paper concludes by presenting two musical applications of auditory roughness in algorithmic composition, in particular to supervise the vertical recombination of sound objects in the software earGram.

1. INTRODUCTION

Composing can be seen as a decision-making process. Many choices have to be made during the creation of a musical piece from the macro down to the micro structural levels on both the horizontal (e.g., melodic) and vertical (e.g., harmony and melodic motive relationships) dimensions. Since the 1950s efforts have been made to understand and formalize organizing principles of both dimensions of musical structure in order to instruct computers to compose music. Today, computers constitute true assistants in several stages of the composer's tasks. Ultimately, composers can design algorithms that consequently "compose" musical pieces or provide large chunks of raw material, which can then be manipulated and assembled.

From the early days of computer music until recently, the most common, and almost exclusive, music representation used in computer-aided algorithmic composition systems was symbolic (e.g., MIDI). Despite the clean, robust, and discrete information provided by symbolic music representations, this type of data has limitations. For example, the MIDI specification does not include timbral information. Given the relevance of timbre formation and "harmonic" relationships between vertical musical structures in this study, our attention turned to an encoding format of the musical auditory experience, i.e. audio signals.

Audio signals are a precise, flexible, and rich representation of the auditory experience presenting new possibilities for music creation in comparison with those offered by symbolic music representations. Additionally, today, the most common music distribution format is digital audio rather than symbolic representations. Nonetheless, audio signals' low-level representation requires the use of algorithmic strategies—from the field of music information retrieval (MIR)—to attempt to obtain the same level of information provided by symbolic codes. Typical examples of such MIR strategies are (polyphonic) pitch detection, beat tracking, downbeat detection, and structural segmentation [1]. These algorithmic strategies not only offer an understanding of audio signals higher than its low-level (sample) representation, but also may effectively contribute to the process of music creation. A typical example of an MIR research topic that greatly combines most aforementioned tasks is automatic mashup creation, which attempts to identify, manipulate, and synthesize songs or musical excerpts that "fit" together. Goto [2] referred recently to automatic mashup creation as one of the grand challenges of MIR.

In this paper, we explore the reliability of a perceptually informed measure of (sensory) dissonance as a "general" measure of musical dissonance in the context of music mashup creation. Specifically, we aim to study the application of auditory roughness as an algorithmic composition strategy to control the "pleasantness" of sound objects vertical aggregates. This method was first proposed by Parncutt [3] and used as an algorithmic-assisted composition strategy by Strasburger [4] and Ferguson [5] in the symbolic music domain. The innovative aspects of this research is the use of audio signals as opposed to symbolic music representations and the possibility to process any type of sound independent of sources or causes—making the strategy suitable for analyzing and generating both soundscapes and polyphonic music.

It is important to note that the research presented here only deals with audio signals segmented at the sound object time scale and does not consider components of musical structure other than harmonic relationships between vertical musical elements. Many relevant elements that actively contribute to the quality of vertical musical structure such as rhythmic features will not be addressed.

The remainder of this paper is structured as follows. In Section 2 we review algorithmic music strategies for the generation of vertical musical structure. In Section 3 we introduce two important concepts of this research—consonance and dissonance—constrain their application in the context of the current research, and present an algorithm for computing auditory roughness. In Section 4 we detail an experiment that aims at investigating how roughness correlates with human judgments of dissonance. In Section 5 we present and discuss the experiment results. In Section 6 we demonstrate how the current research has been applied in the software earGram [6] for recombining sound objects into soundscapes and polyphonic music. Finally, in Section 7 we state conclusions and future work.

2. VERTICAL DIMENSION OF MUSIC: AN ALGORITHMIC APPROACH

The vertical dimension of music is related to the relationship between simultaneous events (e.g., a piano chord or a contrapuntal texture), or the sonic matter (e.g., spectrum of a violin tone), which can occur at several layers of musical structure. For example, at the macro and meso levels of musical structure, a possible strategy to shape the vertical dimension of musical structure is to orchestrate its musical events. On the sound object temporal scale—the structural level of interest here—typical examples of vertical structures are chords and timbre formation/modulation.

The vertical dimension of music structure has been a research topic revisited since the early days of computer-assisted algorithmic composition in the late 1950s. The study of vertical musical structure generation is rather evident in algorithmic music strategies for style imitation, i.e. the branch of algorithmic composition that focuses on the formalization of principles extracted from music theory, particular works, or a body of works to generate music that resembles at some level the analyzed music. Some of the topics that have been continuously revisited within this line of research are: the generation of species counterpoint [7, 8]; functional harmony as used in Western music from the 17th to 19th centuries [9, 10]; the automatic generation of polyphonic rhythms, namely in the context of interactive music systems [11]; and the exploration of serial music operations [12, 13].

Despite the considerable body of knowledge on algorithmic strategies for generating vertical musical structures, very little research on this domain deals with musical events encoded as audio signals or even addresses musical representations other than symbolic music codes. Additionally, most algorithms presented in this domain cannot deal with the low-level representation of audio signals and only process clean and discrete data, in particular the pitch and duration of overlapping events. Despite the accuracy and robustness of pitch detection algorithms for monophonic audio signals, state-of-the-art algorithms for polyphonic pitch detection are not yet very reliable [14]. Therefore, the above-mentioned algorithms cannot consistently manipulate most music encoded as audio signals due to its predominantly polyphonic nature.

An exception to the prevailing use of symbolic representations in algorithmic composition is the recent work in MIR, which has been gradually expanding its area of action towards music creation [14]. One such emerging topic is mashup creation, which makes use of content-based analysis to retrieve "mashable" material from large databases according to particular audio features like harmonic compatibility [15, 16], or even automatically generate song remixes/mashups [17, 18]. Despite recent efforts, so far, results focus on simple harmonic models, whose matching criteria happens in chroma space (i.e., 12 dimensions) that does not address spectral/timbral properties. Our approach focuses on the study of a model for harmonic incompatibility between vertical sound events rather than the presence of high harmonic similarity, thus offering a broader range of musical possibilities.

3. CONSONANCE AND DISSONANCE: AUDITORY ROUGHNESS

In music, the terms "consonance" and "dissonance" are subject to various misconceptions, confusions, and disagreements as may be shown by their inconsistent definitions in dictionaries, harmony textbooks and books on musical acoustics [19]. Tenney [19] has also shown that both concepts refer to different phenomena depending on historical, cultural (tradition), and musical (composer's idiom or stylistic features) contexts. Additionally, while striving to clarify the semantics of what he calls the "consonance/dissonance-concept" (CDC), Tenney examined the roots and developments of the terms in western musical culture and presented the following five categories in which the terms are addressed distinctly: (1) melodic: distinguish degrees of "affinity, agreement, similarity, or relatedness" between melodic intervals; (2) diphonic: sonorous character of simultaneous dyads; (3) contrapunctual: consonance/dissonance defined by role in counterpoint (the important aspect is the context in which it occurs, not the physical properties of the sound); (4) chordal/functional: CDC applied to individual tones in a chord; and (5) timbral: equated with "roughness".

Due to computational limitations, namely the robustness of polyphonic pitch detection algorithms, Tenney's CDC 1-4 will not be considered in this study, because their organization relies on discrete characterization of notes. Our work will focus on timbral CDC because it can be readily measured and its computation measurement is well established.

Harmony resulting from roughness measures largely relates to orchestration, and to a lesser extent to harmonic tonal syntax. Additionally, it is also relevant in electroacoustic music and connected to contemporary approaches to pitch.[1] Barlow was probably the earliest composer to

[1] Roughness also proved to be helpful in the analysis of contemporary,

use roughness in composition in his piece *Çoğluotobüsişletmesi* (1978). The research behind the aforementioned piece would later be incorporated into the algorithmic composition program Autobusk. Spectral music composers, in their journey of discovery for new sound organizations based on sounds inner structure, also paid attention to the roughness phenomenon. A typical example of spectral music that explores the roughness phenomenon is the opening section of Grisey's *Jour, Contre-Jour* (1979). More recently, we can cite the works of Strasburger [4] and Fergusson [5]. In non-western musical traditions the effect of roughness has also been explored (e.g. Indian *tambura* drone, Bosnian *ganga* singing, and Middle Eastern *mijwiz* and *ganga* singing) [21].

Even if roughness has raised some interest within the music and scientific communities, few current musical applications take advantage of this measure, in particular to analyze large amounts of music as they unfold in time and generate vertical musical structures. A possible reason for this fact is the disconnect between the dissonance models from music theory and psychoacoustics. Nonetheless, despite this disconnect, roughness measures innate and intrinsic human perception phenomena, which contributes for concepts of musical consonance and dissonance [20]. Empirical research has also reinforced and confirmed this relationship. For example, Miskiewicz [22] has shown a strong correlation between how musical dyads are understood in sensory terms and in common tonal syntax. Nevertheless, "musical" dissonance embeds idiosyncrasies such as explicit and implicit rules or schemata that go beyond physics or physiology [20].

Our use of sensory dissonance departs from Terhardt's [23] psychoacoustic theory, which defines the phenomenon as a combination of the three following sound features: (1) sharpness (also addressed as brightness), (2) roughness, and (3) tonalness. Notwithstanding the phenomenon of sensory dissonance being regulated regulated by three factors, we will simply addressed it by its most prominent factor, which is the roughness of a sound, because there isn't a model that describes the interaction of the aforementioned psychoacoustic factors [24].

The roughness of a sound is the physical correlate of amplitude fluctuations [21] (also addressed as "beatings") produced when two frequencies are a critical bandwidth apart, which is approximately one third of an octave in the middle range of human hearing [23]. The sensation of "roughness" or "fast beats" occurs when the rate of two frequency amplitude fluctuations are over 20 Hz up to a critical bandwidth. Dissonant sounds within this approach produce "fast beats", and consonance is the absence of such beating sensation.

Timbre can also affect our subjective experience of musical dissonance and harmonic progression [24]. In particular, partials of complex tones can also produce a beating sensation when the same conditions are met, i.e., when they are a critical bandwidth apart. As a result, the timbre of complex tones can affect our experience of roughness. This evidence was concluded since the early experiments on this domain; however, only recently research on this domain started tackling this issue more systematically, i.e., investigating and developing algorithms to measure roughness between sonorities, taking into account the effects of timbre and microtonal inflection [24, 25]. Still, the latest most significant experiments on this domain rely on "artificially" created sounds (synthesized sounds with highly controlled parameters) or simplistic examples (e.g., the monophonic instruments sounds). To our knowledge, empirical studies on auditory roughness have not addressed natural and complex musical stimuli and do not represent the variability that can be present in natural music listening situations, which differ from "synthetic" ones in a number of ways, such as amplitude and phase of the partials, attack cues, etc. Consequently, despite the unpredictable factors associated with natural and complex musical stimuli, no clear knowledge exists about the correlation between natural musical stimuli and human judgments of dissonance as understood in tonal music syntax, as we study here. Before delving into the experiment, we should clarify the roughness measure used in the current study.

The roughness computation used in our experiments and in the musical applications detailed and discussed in the reminder sections of this paper uses Porres's implementation [24] of Parncutt's roughness (ρ) measure [26]:

$$\rho = \sum_{j=0}^{n} \sum_{k=1}^{n-1} \frac{a_j \cdot a_k \cdot g(f_{cb})}{a_j^2} \tag{1}$$

where a_j and a_k are the amplitudes of two frequencies being compared; f_{cb} is the distance between the frequencies in critical bandwidths (Bark); and $g(f_{cb})$ is a "standard curve" developed by Parncutt (equation 4) that models experimental data of Plomp and Levelt [27]. To convert a frequency f from Hz to Bark, we use the equation proposed by Barlow [28], which merges Terdardt and Traunmüller:

$$z = \begin{cases} 13.3 \cdot \operatorname{atan}\left(3 \cdot \frac{f}{4000}\right), & f < 219.501 \\ \left(\frac{26.81 \cdot f}{1960+f}\right) - 0.53, & f > 219.501 \end{cases} \tag{2}$$

Traunmüller's equation (equation 2, lower row) has an added correction factor for values of $z > 20.1$:

$$z' = z + 0.22 \cdot (z - 20.1) \tag{3}$$

$$g(f_{cb}) = \begin{cases} \left(e\left(\frac{f_{cb}}{0.25}\right) \cdot e^{\left(-\frac{f_{cb}}{0.25}\right)}\right)^2, & f_{cb} < 1.2 \\ 0, & f_{cb} > 1.2 \end{cases} \tag{4}$$

We used Pure Data's external *sigmund~* developed by Puckette to extract pairs of frequency and amplitude of the 50 most prominent peaks of the spectra.

non-tonal and non-western music and performance where traditional analytical systems fail, and for the exploration of arbitrary musical scales or tunings other than the 12 temperate scale [20].

4. EXPERIMENT

An experiment was carried on to assess how well auditory roughness can be applied as a "general" measure of musical dissonance. The experiment consisted of a listening test, which aimed to evaluate the relationship between human judgment of dissonance and roughness, with the hypothesis that human judgment would be correlated with roughness. Specifically, we expect a negative correlation between both variables, because, to simplify the experiment, the scale of the human ratings was inverted in relation to the measure of auditory roughness. Additionally, we conjecture that the presence of non-pitched sonorities may bias the established hypothesis.

We created 3 datasets for the listening experiment, each with 150 musical stimuli with duration between 1-2s, resulting from the overlap of different sound events. The 3 datasets encompass the following sound types: (1) quasi-harmonic sounds (clarinet notes and piano chords); (2) quasi-harmonic and non-pitched percussion sounds (clarinet notes, guitar motives, and drumbeats); and (3) environmental sounds (field recordings of a park and a forest). Then, for each stimulus, we calculated its roughness using the algorithm described at the end of Section 3. We then sorted the values of each dataset in an ascending order, divided the entire range of values in five equal parts, and randomly selected three stimuli from each part in order to guarantee that the musical stimuli used in the experiment covered the entire range of auditory roughness per dataset. In total, each participant was asked to rank 45 musical stimuli—15 musical stimuli per dataset.

The experiment was run as follows: for each new excerpt the participants were asked to rate the degree of dissonance of each stimuli on a 1-5 scale, with 1 being very dissonant and 5 very consonant. The three datasets were evaluated separately, and the order of the stimuli was randomly selected. To allow the participants to get familiar with the experiment there was a short training phase prior to starting the main experiment.

In total, 41 participants were recruited to take the experiment (22 males and 19 female, with ages ranging from 18 to 27 years old). Since musical training could affect the type of judgments, we restricted the participants to classically trained music students undergoing a bachelor's or master's degree. The participants were not paid for taking part in the experiment.

5. RESULTS

To examine the results of the listening test we first computed the mean values of all participants' dissonance ratings for each stimulus and then, for each corpus, we computed the Pearson correlation coefficient between the mean values of the human dissonance ratings and roughness.

The results indicate a statistically significant negative relationship between roughness and user judgments for quasi-harmonic sounds (dataset 1), and no significant relationship for the two remaining sets (Table 1 presents the Pearson correlation coefficient results for the three datasets and their statistical significance and Figure 1 depicts in a scatter plot the relationship between the experiment variables for the 3 datasets). The negative correlation observed in datasets 1 and 2 results from the fact that the human ratings and roughness scales are inverted, i.e. the most consonant sounds are values close to zero according to the roughness measure used and correspond to the maximum value (5) in the human ratings scale (1-5).

	Pearson correlation coefficient (r)	Statistical significance (p)
Dataset 1	-0,7754	< 0.001
Dataset 2	-0,4571	0.09
Dataset 3	0,0863	0.76

Table 1. Correlation between human judgments of dissonance and roughness for three different datasets of sound stimuli (see section 4).

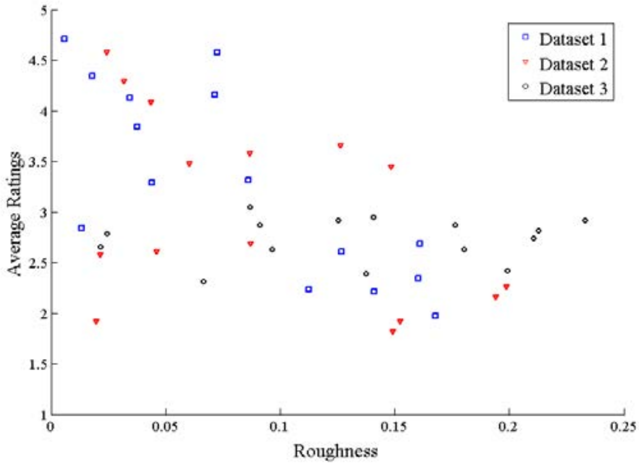

Figure 1. Scatter plots exposing the correlation between roughness and human judgments of dissonance by trained musicians for three datasets of sound stimuli.

Despite being non-significant, dataset 2 still presents some degree of correlation between roughness and human judgments of dissonance, which is unverified in dataset 3. This phenomenon may result from the increasing level of inharmonicity of dataset 3 in comparison with dataset 2. However, additional research is necessary to verify this hypothesis. Additionally, the participant's ratings of dataset 3 may suffer from some inconsistency due to a lack of understanding of the concept of dissonance/consonance in environmental sounds. We believe that this fact is due to the lack of exposure of the participants to these types of sounds in an analytical manner given their musical background (which typically does not consider environmental sounds as "musical sounds").

Summarizing, the experiment results show a high degree of correlation between human judgments of dissonance/consonance and auditory roughness for quasi-harmonic sounds and no significant relationship for, or in the presence of, non-pitched sounds.

6. MUSICAL APPLICATIONS

The use of auditory roughness in computer music has a large range of applications in musical analysis and composition, in particular to describe and/or generate vertical musical structures. In terms of analysis, the use of roughness may provide some insights about the organization of the vertical dimension of music at specific times, or provide a curve that exposes the temporal evolution of the roughness of a particular composition. Auditory roughness is not a guaranteed measure of musical dissonance—which is a subjective and context-dependent concept—nevertheless, there’s a strong correlation between the two concepts, which makes roughness a good measure to analyze music where no score is available, for music outside of the Western music tonal vocabulary for which strict rules are known in advance, or to automatically analyze large amounts of music. In terms of composition, the most strikingly aspect of roughness is the possibility to systematically organize non-harmonic sonorities according to a “timbral grammar.” This includes two important areas that lack a systematic approach to the task: (1) all possible tunings related to timbres, (2) extending sound-objects *solfège* with a sort of “tonal” vocabulary, taking over the role of pitch and harmonic syntax in Western music.

In the context of our work auditory roughness was used to regulate the quality of vertical musical layers of sound objects in earGram [6], a concatenative sound synthesis (CSS) software for content-based algorithmic-assisted audio composition. Even if CSS deals primarily with the horizontal dimension of music, i.e., the generation of musical sequences, current practice expands the technique to the synthesis of overlapping units [29, 30]. Despite the popularity of this new approach, the resulting sound quality of the vertical superposition of audio units has been overlooked. Specifically, roughness was used in earGram to regulate the dissonance of overlapping audio units in two “playing modes” of the software: shuffMeter and soundscapeMap. ShuffMeter was designed to recombine sound objects into phrases characteristic of a user-assigned meter and soundscapeMap the manipulation and synthesis of soundscapes. Both methods allow the generation of several concurrent vertical layers by superimposing sound objects. A detailed description of both algorithmic strategies and particularly how they apply roughness to guide vertical musical structure follows. Both algorithmic strategies rely on a corpus of structurally segmented-analyzed/described sound objects to generate musical sequences. For a comprehensive explanation of the foundations and implementation of the software and in particular to the analytical modules of the system please refer to [6].

6.1 ShuffMeter

ShuffMeter relies on music theory knowledge to guide the generation of musical sequences that reflect a user-assigned meter. The generation of patterns characteristic of a given meter result from the stochastic recombination of units with different stresses given by a metrical template generated by Barlow’s metrical indispensability algorithm [31]. We ascribed the template representation to two audio descriptors: loudness and spectral variability, because spectral and loudness changes are most likely to occur on stronger metrical accents [32]. The template may be altered during performance to regulate the smoothness and loudness of the generated phrases by regulating the clusters’ color position on interface (see Figure 2). ShuffMeter also allows the creation of up to 8 synchronized vertical layers, each assigned to a sub-space of the corpus. The corpus is automatically divided into groups that expose common characteristics by clustering algorithms. Although the algorithm may adopt any “type” of temporal unit, it conveys better results when using units segmented on a beat basis.

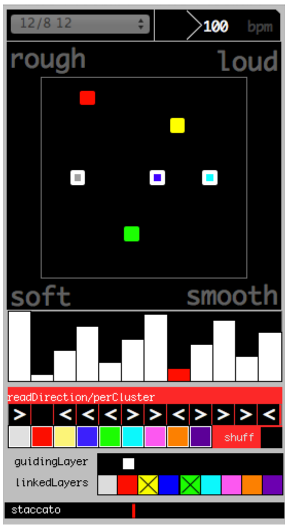

Figure 2. ShuffMeter’s interface.

Roughness was used in shuffMeter to supervise the quality of overlapping of units. Prior to generation the user must define a guiding layer and all remaining layers that must conform to it. This user input is mostly necessary because, to achieve better results, one must exclude from the roughness quality assessment non-pitched sounds (as shown by the experiment results). At each iteration, from the set of units that have a spectral variability and loudness corresponding to a particular metrical accent, signed layers will weight the decision of the best matching unit according to the minimum roughness values between the candidate units that will be overlapped with the guiding layer.

6.2 SoundscapeMap

SoundscapeMap defines target phrases to be synthesized by navigating in a two-dimensional plane, whose axes are assigned to musical features that control the density and the "sharpness" of the sound events (see Figure 2). Smoothness (x-axis) controls the stability (amplitude, pitch, and timbre changes) of the synthesis and is assessed by the (non-normalized) spectral flux of the audio units. Density (y-axis) regulates the number of units played simultaneously and ranges from one to five.

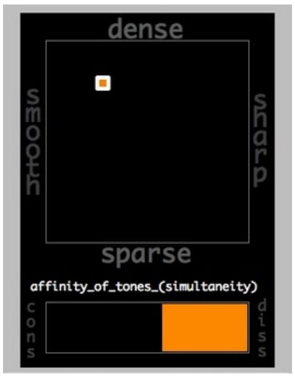

Figure 3. SoundscapeMap's interface.

Despite the disconnect between "musical dissonance" and roughness for non-pitched sounds, roughness still measures important auditory phenomena. Therefore, in soundscapeMap the user can control the roughness' degree of overlapping units. In order to do so, the user must define a region of roughness in which the units should preferably fall on a slider. Consequently, at each iteration, the algorithm restricts the corpus to units that have roughness values that fall within the selected range in relation to the last played unit. If the algorithm does not retrieve any units, it searches for the closest unit to the specified range of sensory dissonance.

7. CONCLUSION & DISCUSSION

In this paper we detailed an experiment that aimed at evaluating the correlation between auditory roughness and the Western concept of musical dissonance assessed by the empirical judgment of trained musicians. The experiment results showed a statistically significant correlation between the two variables for quasi-harmonic sounds. In addition, the relationship between the variables appears to show a decrease in correlation when sound inhamonicity increases.

The results of the experiment helped refining the design of two algorithmic composition algorithms (shuffMeter and soundscapeMap) embedded in the software earGram that concatenate and layer short snippets of audio into musical phrases characteristic of a given meter and soundscapes. Specifically, we used auditory roughness to control the degree of dissonance of vertical musical structures resulting from the overlap of two or more audio units. Despite the poor results concerning the relationship between "musical consonance" and roughness for non-pitched sounds, roughness still measures important perceptual phenomena of environmental sounds (as used in soundscapeMap), which makes it suitable to regulate the generation of any audio signal independently of their cause and musical context. Nonetheless, users must be aware that the relationship with tonal musical syntax appears to decreases with increased inharmonicity. Both the software and several sound examples are available at: https://sites.google.com/site/eargram/.

Even if the results of the experiment detailed here enlighten the relationship between roughness and musical dissonance and although roughness shows great value for music analysis and composition by providing a quantified measure of (sensory) dissonance, its application in algorithmic composition needs ultimately to rely on human judgments to verify or adapt the harmonic syntax to the application context of the creative task at issue. Roughness alone does not guarantee good artistic results, just as consonant sounds are not necessarily preferred to dissonant. In fact, listeners tend to prefer a certain optimal amount of dissonance, complexity, or information flow [33]. Thus, more research is necessary to understand and formalize effective strategies for regulating the dissonance levels of the musical surface. This contrast forms one of the key ingredients of music composition, in which dissonant chords are used to create feelings of tension that are later released by consonant chords. In future work we intend to further study the application of roughness as an algorithmic composition strategy, mainly by understanding its relation with musical tension.

Acknowledgments

This research has been made possible by the funds of the Fundação para a Ciência e a Tecnologia (FCT) post-doctoral grant SFRH/BPD/88722/2012, and the Media Arts and Technologies project (MAT), NORTE-07-0124-FEDER-000061, financed by the North Portugal Regional Operational Programme (ON.2–O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF), and by national funds, through the Portuguese funding agency FCT. The presentation of this research at ICMC-SMC was partially supported by the Music Program at the New York University in Abu Dhabi.

8. REFERENCES

[1] A. Klapuri and M. Davy (Eds.), *Signal Processing Methods for Music Transcription*. Springer-Verlag New York, 2006.

[2] M. Goto, "Grand challenges in music information research," in M. Muller, M. Goto, and M. Schedl (Eds.), *Multimodal Music Processing*, pp. 217–225. Dagstuhl Publishing, 2012.

[3] R. Parncutt, "A psychoacoustical model of tonal composition," *Proc. of the Int. Computer Music Conference,* Montreal, 1991, pp. 210-213.

[4] R. Parncutt and H. Strasburger, "Applying psychoacoustics in composition: 'Harmonic' progressions of 'non-harmonic' sonorities," *Perspectives of New Music*, vol. 32, no. 2, pp. 1-42, 1994.

[5] S. Ferguson and R. Parncutt, “Composing 'In the Flesh': perceptually-informed harmonic syntax,” *Proc. of Int. Conf. Sound and Music Computing*, Paris, 2004, pp. 20-22.

[6] G. Bernardes, C. Guedes, and B. Pennycook, "EarGram: an Application for Interactive Exploration of Concatenative Sound Synthesis in Pure Data," in M. Aramaki, M. Barthet, R. Kronland-Martinet, and S. Ystad (Eds.), *From Sounds to Music and Emotions*, pp. 110-129. Springer-Verlag, 2013.

[7] R. Ovans and R. Davison, “An interactive constraint-based expert assistant for music composition,” in *Proc. of the 9th Canadian Conference on Artificial Intelligence,* pp. 76-81, 1992.

[8] M. Farbood and B. Schoner, “Analysis and synthesis of Palestrina-style counterpoint using Markov chains,” in *Proc. of the Int. Computer Music Conf.*, Havana, 2001, pp. 471-474.

[9] F. Pachet and P. Roy, “Musical harmonization with constraints: A survey,” *Constraints*, vol. 6, no. 1, pp. 7-19, 2001.

[10] T. Anders, E. Miranda, “A computational model that generalises Schoenberg’s guidelines for favourable chord progressions,” *Proc. of Int. Conf. Sound and Music Computing*, Porto, 2009, pp. 48-52.

[11] A. Eigenfeldt, “The evolution of evolutionary software intelligent rhythm generation in kinetic engine,” *EvoWorkshops, Lecture Notes in Computer Science*, volume 5484, 2009, pp. 498-507.

[12] K. Essl, “Lexikon-Sonate: An interactive realtime composition for computercontrolled piano.” *Proc. of the II Brazilian Symposium on Computer Music*, 1995.

[13] C. Ariza, “Navigating the landscape of computer-aided algorithmic composition systems: A definition, seven descriptors, and a lexicon of systems and research,” in *Proc. of the Int. Computer Music Conference*, Barcelona, pp. 765-772, 2005.

[14] X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E. Gómez, F. Gouyon, P. Herrera, S. Jordà, O. Paytuvi, G. Peeters, J. Schlüter, H. Vinet, and G. Widmer, *Roadmap for Music Information ReSearch*. Creative Commons BY-NC-ND 3.0 license, 2013.

[15] Mixed in Key. Mashup. [Online]. Available: http://mashup.mixedinkey.com.

[16] Fraunhofer, IDMT. Harmonic Mixing Tool. [Online]. Available: http://www.idmt.fraunhofer.de/en/Service_Offerings/products_and_technologies/e_h/harmonic_mixing_tool.html.

[17] M. Davies, P. Hamel, K. Yoshii, and M. Goto, M. (2013). “AutoMashUpper: An automatic multi-song mashup system,” in *Proc. ISMIR*, Curitiba, 2013, pp. 575-580.

[18] Echonest. The Infinite Jukebox. [Online]. Available: http://labs.echonest.com/Uploader/index.html.

[19] J. Tenney, *A History of “Consonance and Dissonance.”* Excelsior Music Publishing, 1988.

[20] W. Sethares, *Tuning, Timbre, Spectrum, Scale*. Springer-Verlag, 1999.

[21] P. Vassilakis, *Perceptual and Physical Properties of Amplitude Fluctuation and their Musical Significance.* PhD thesis, University of California, 2001.

[22] A. Miskiewicz, “Roughness and dissonance of musical dyads,” in *Proc. of the 5th Triennial ESCOM Conference*, Hanover, 2003, pp. 420-424.

[23] E. Terhardt, “On the Perception of Periodic Sound Fluctuations (Roughness),” *Acustica*, vol. 30, pp. 201-213, 1974.

[24] A. T. Porres, “Dissonance Model Toolbox in Pure Data,” in *Proc. of the 4th Pure Data Convention,* Weimar, 2011.

[25] J. MacCallum and A. Einbound, “Real-time analysis of sensory dissonance,” in *Computer Music Modeling and Retrieval. Sense of Sounds*. Springer-Verlag, 2008. p. 203-211.

[26] R. Parncutt, *Harmony: A Psychoacoustical Approach*. Springer-Verlag, 1989.

[27] R. Plomp and W. Levelt, “Tonal Consonance and the Critical Bandwidth,” *Journal of the Acoustical Society of America*, vol. 38, no. 4, pp. 548-560, 1965.

[28] C. Barlow, “Von der Musiquantenlehre,” *Feedback Paper*, vol. 34, 2008. Feedback Studio Verlag.

[29] D. Schwarz, “The sound space as musical instrument: Playing corpus-based concatenative synthesis,” in *Proc. of the New Interfaces for Musical Expression Conf.*, Ann Arbor, 2012.

[30] D. Schwarz and B. Hackbarth, “Navigating variation: Composing for audio mosaicing,” in *Proc. of the Int. Computer Music Conf.*, Ljubljana, 2012.

[31] C. Barlow, “Two essays on theory,” *Computer Music Journal*, vol. 11*, no.* 1, pp. 44-60, 1987.

[32] J. London, *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press, 2012.

[33] K. Smith, L. Cuddy, “The pleasingness of melodic sequences: Contrasting effects of repetition and rule-familiarity,” *Psychology of Music*, vol. 14, no. 1, pp. 17-32, 1986.