Modifying Partials for Minimum-Roughness Sound Synthesis

Simon Schwär[1†], Meinard Müller[1], and Sebastian J. Schlecht[2]
[1] International Audio Laboratories, Erlangen, Germany
[2] Dept. of Information and Communications Engineering and Dept. of Art and Media, Aalto University, Espoo, Finland
[†] Corresponding author: simon.schwaer@audiolabs-erlangen.de

Discipline(s): Sound synthesis and design, Computational and AI insights and applications

Keywords: Sound synthesis, Adaptive timbre, Roughness minimization

Introduction

Auditory roughness is a psychoacoustic property that correlates with the perceived "pleasantness" of sounds for Western listeners (Terhardt, 1974; McDermott et al., 2016). It is an integral part of musical expression in terms of changing harmonies and consonances (Vassilakis, 2005; Berezovsky, 2019; Marijeh et al., 2022) and a multitude of models (e.g., Daniel and Weber, 1997; Vassilakis and Kendall, 2010) have been proposed to quantify the roughness sensation for tonal and noise-like sounds. The model by Plomp and Levelt (1965) is based on listening experiments with pairs of pure tones and allows to measure the perceptual dissonance between complex harmonic tones in a simple and differentiable way (Schwär et al., 2021). Sethares (1998) applied this model to adaptive tuning by finding an "optimal" fundamental frequency that minimizes roughness for a given musical scale degree in real time. Furthermore, he introduced a global optimization procedure for finding a fixed timbre (in terms of frequency and amplitudes of their partials) that minimizes the average roughness of a fixed set of intervals.

In this work, we introduce and compare two methods to *adaptively* modify the partials of simultaneously sounding synthesized tones to minimize roughness. By changing their amplitude and/or frequency over time (e.g., for each chord separately), it is possible to dynamically control the timbre of a polyphonic sound in real time. This introduces an additional parameter for sound synthesis that may allow for changing the roughness of a sound without modifying other perceptual attributes of the individual tones, like their fundamental frequency (F0) or loudness. We draw inspiration from choir singers, who may not only dynamically adapt their pitch, but also control their vocal formants (i.e., the prevalence of certain partials) as an additional means to facilitate intonation and voice blending between musicians.

Method

We consider a polyphonic sound that consists of T individual tones with N partials each. The vector $\mathcal{P} \in (\mathbb{R}_+ \times \mathbb{R}_+)^M$ with $M = TN$ contains a tuple $\mathcal{P}_\mathrm{m} = (f_m, a_m)$ with frequency f_m in Hz and amplitude a_m for all partials $m \in [1{:}M]$ of all simultaneously sounding tones. Using a parametrized model by Berezovsky (2019), the overall roughness $D(\mathcal{P})$ is then given by

$$D(\mathcal{P}) = \frac{1}{2}\sum_{i=1}^{M}\sum_{j=1}^{M} \min\left(a_i, a_j\right) \cdot exp\left(-\,ln^2\left(\frac{\left|log_2\left(f_i/f_j\right)\right|}{w_\mathrm{c}}\right)\right)$$

where w_c is a parameter that controls the frequency distance of maximal roughness (see Figure 1a). Roughness between a pair of partials is high when their frequency is similar but not equal, denoted as a "clashing pair" in the following.

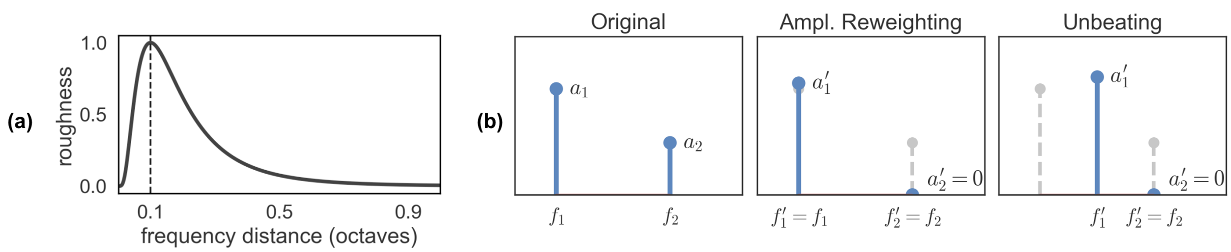

Figure 1: (a) Roughness curve for a single pair of partials w.r.t. the distance between their frequencies ($w_c = 0.1$, marked by the dotted line). (b) Visualization of the adaptation of a clashing pair of partials with amplitude reweighting and unbeating.

In a polyphonic sound, the value of $D(\mathcal{P})$ is governed by the number of clashing pairs and it can be reduced by either decreasing the amplitude of those partials or by changing their frequency. While both adaptations naturally alter the timbre as they change the relation between partials, different adaptation strategies may have different implications on timbral attributes and perception. We introduce two possible methods for roughness minimization by adapting a given clashing pair of partials (f_1, a_1) and (f_2, a_2) to their new frequency and amplitude values (f'_1, a'_1) and (f'_2, a'_2), respectively. The effects of the methods on a clashing pair are illustrated in Figure 1b.

Method 1: Amplitude Reweighting – Given $a_1 > a_2$, set $a'_1 = (a_1 + a_2)/\sqrt{2}$ and $a'_2 = 0$. By setting the smaller of the two amplitudes to zero, it is effectively removed from the synthesis, while the loudness remains constant due to the reweighting of a'_1. The frequencies of the partials remain unchanged ($f'_1 = f_1$ and $f'_2 = f_2$).

Method 2: Unbeating – High roughness coincides with the occurrence of beating between two partials (Terhardt, 1974). This amplitude modulation of a "carrier" frequency can be shown by the trigonometric sum-to-product identity (Jeffress, 1968). For the two corresponding sinusoidals with time t in seconds, we can write

$$a_1 \sin(2\pi f_1 t) + a_2 \sin(2\pi f_2 t) = c_t \sin\left(2\pi \frac{f_1 + f_2}{2} t + \xi_t\right),$$

where

$$c_t = \left(a_1^2 + a_2^2 + 2a_1 a_2 \cos\left(4\pi \frac{f_1 - f_2}{2} t\right)\right)^{1/2}$$

is the time-varying amplitude modulation and

$$\xi_t = \arctan\left(\frac{a_1 - a_2}{a_1 + a_2} \tan\left(2\pi \frac{f_1 - f_2}{2} t\right)\right)$$

is a time-varying carrier phase modulation. To reduce beating, we can set

$$c_t = c_0 = (a_1^2 + a_2^2)^{1/2}$$

to be constant over time, removing the amplitude modulation (i.e., "unbeating" the sound). As a result, we set $a'_1 = c_0$, $a'_2 = 0$, $f'_1 = (f_1 + f_2)/2 + \xi_t/(2\pi t)$ and $f'_2 = f_2$. Optionally, we can additionally remove the phase modulation, setting $\xi_t = 0$.

Experiments

For an initial exploration of the introduced methods, we consider the use case of reducing roughness that is induced by inaccurate intonation. To simulate this situation, we use additive synthesis, where we have full control over all the partials of each voice. We synthesize example sounds that consist of three different chords with four voices: "D" (D5, A4, F#4, D4), "Am" (A4, E4, C4, A3), and "E7" (G#4, D4, B3, E3) and introduce slight random detuning (± 20 cents) of the F0 of each voice to introduce additional roughness in the original sounds. For each chord, we use two different timbres with $N = 12$ partials: A Sawtooth waveform and a "clarinet-like" timbre, where partial amplitudes and frequencies are estimated from a recorded clarinet tone. Finally, we compare six experimental conditions:

- *Original* – detuned without adaptation
- *Amplitude Reweighting*
- *Unbeating I* – without phase modulation ($\xi_t = 0$)
- *Unbeating II* – including phase modulation ξ_t
- *12-TET* – original partials but without detuning of the F0s, i.e., using 12-tone equal temperament
- *Optimized F0* – original partials, but the F0s are adapted to minimize roughness, as proposed in Schwär et al. (2021), yielding results comparable to a just intonation tuning

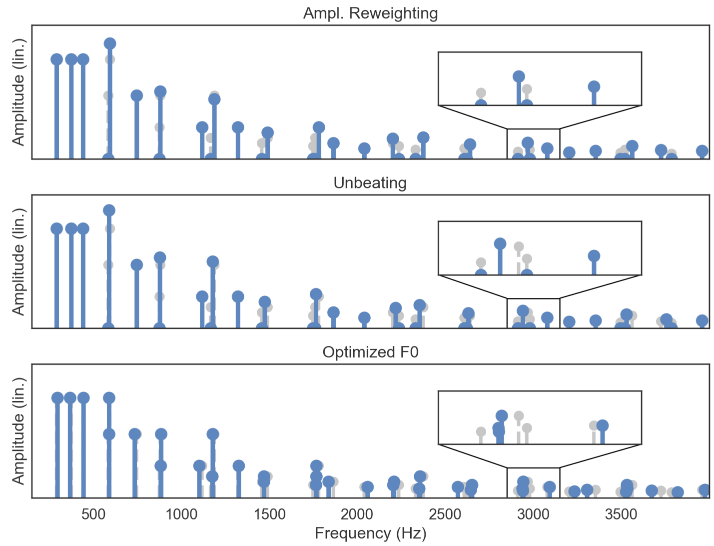

Figure 2: Comparison of partials for the "D" chord in different experimental conditions (grey: original partials, blue: adapted partials).

In total, this results in 36 unique sound items (2 timbres, 3 chords, 6 conditions), which we can compare in terms of their objective and subjective properties. The discussed sound examples are available on a supplemental website: https://audiolabs-erlangen.de/resources/2023-TIMBRE-AdaptiveTimbre

Throughout the experiments, we treat a pair of partials as clashing when the distance between frequencies is larger than one cent and smaller than 50 cents. While different definitions of clashing partials are possible, for example based on critical bands or a threshold of the pure tone roughness, this simple strategy can easily be used to adjust the effect strength and it is musically interpretable. To remove clashing pairs in $\mathcal{P}$, we iterate over all partials in the set and apply the respective adaptation method when another partial is closer than the threshold. Figure 2 shows how this algorithm and chosen distance threshold affects the partials of the "D" chord compared to the *Optimized F0* condition.

For each sound item, we calculate $D(\mathcal{P})$ as a measure of roughness, and the frame-wise spectral centroid of the synthesized audio as a simple measure of how the methods affect the spectral envelope of the sound. In addition, we conduct a listening test, where participants were asked to judge the perceived "pleasantness" of the sound items using a best-worst scaling method (Louviere et al., 2015). The participants were presented with four sound items in each trial and asked to select the most and least pleasant item. The comparisons within a single trial were limited to one type of timbre and chord, so that no rankings of the pleasantness of chords or timbres relative to each other can be obtained. Each item was presented twice, resulting in three trials per timbre/chord combination. The listening test was completed by eight participants with varying degrees of musical listening experience. From the participants' answers, we calculate a relative ranking between the experimental conditions using the *bwsample* Python package (Hamster, 2021).

Results

The mean roughness $D(\mathcal{P})$ over all chords and timbres is reduced by the proposed methods, and it is also lower compared to the *12-TET* and *Optimized F0* conditions (*Original:* 3.38, *12-TET:* 3.23, *Optimized F0:* 3.16, *Amplitude Reweighting:* 2.28, *Unbeating I & II: 2.30*). A reduction similar in relative magnitude can be observed for the roughness model by Vassilakis and Kendall (2010). While the absolute reduction in $D(\mathcal{P})$ naturally depends on chord and timbre, all three proposed methods yield similar results for the considered combinations. Regarding spectral envelope, *Amplitude Reweighting* and *Unbeating I* result, on average, in a higher deviation from the spectral centroid compared to the *Original* condition (*12-TET:* 26.4 Hz, *Optimized F0:* 27.0 Hz, *Amplitude Reweighting:* 91.5 Hz, *Unbeating I:* 85.8 Hz, *Unbeating II:* 13.9 Hz). Figure 3 shows how *Amplitude Reweighting* and *Unbeating I* yield the same spectral centroid that is constant over the entire sound, but different compared to the other conditions.

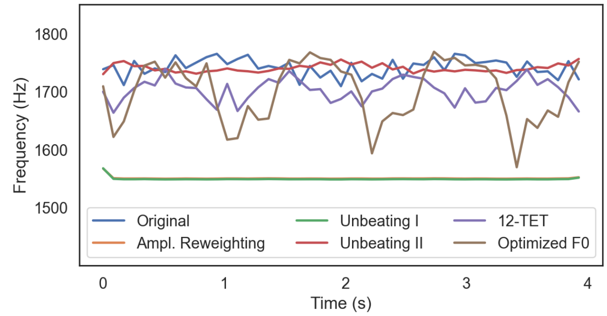

Figure 3: Spectral centroid fluctuations over time for the "D" chord with Sawtooth timbre.

In the listening test, participants ranked the *12-TET* condition as most pleasant overall, followed by *Unbeating I, Optimized F0, Unbeating II, Amplitude Reweighting,* and *Original* as the least preferred condition. We did not observe large differences in the rankings between different chords/timbres, except that *Optimized F0* was ranked worse for the “Am” chord. Between listeners, rankings of *12-TET* as the most and *Original* as the least pleasant condition was mostly consistent, while the order of the proposed methods and *Optimized F0* varied. After the test, some participants reported that the *12-TET* condition (particularly with the Sawtooth timbre) sounded “harsher” to them than the proposed methods, but this did not influence their pleasantness rating.

Discussion

In the simple experimental setting of additive synthesis, the proposed methods consistently reduce measured roughness. Moreover, the reduction is stronger than that achieved by the *Optimized F0* approach due to the additional degrees of freedom afforded by changing individual partials. The spectral centroid is only a crude measure of overall spectral similarity, but it makes it possible to visualize some effects of the partial modifications. While all three methods remove time-variance in the spectral centroid by reducing the amplitude modulation induced by beating, retaining the phase modulation in *Unbeating II* leads to a spectral centroid much closer to the *Original* condition.

In the listening test, the proposed methods were consistently ranked more pleasant than the *Original* condition. *Amplitude Reweighting* was overall ranked lower than the *Unbeating* methods, even though the latter introduce some inharmonicity to the partials. The phase modulation in *Unbeating II* seems to reduce subjective pleasantness compared to *Unbeating I.* However, the varied order in rankings of individual participants hints at relatively small differences between the proposed methods and the *Optimized F0* condition. Somewhat surprisingly, most participants rated the *12-TET* condition as most pleasant, which does not correlate with the measured roughness. We suspect an influence of the familiarity of Western listeners with the 12-TET tuning that influences pleasantness ratings in this context. Other listening test methods, as for example proposed by Marijeh et al. (2022), may be more suitable to isolate such effects.

Overall, the experiments are encouraging to further investigate adaptive partials as an expressive musical tool through timbre modification. In future work, we intend to extend the proposed methods for real-time implementation, which would make it possible to dynamically change roughness as a synthesis parameter in response to other simultaneously sounding instruments. Furthermore, we would like to explore how methods from signal processing can be used to adapt the roughness of recorded signals. This would also enable further investigation of how the roughness reduction impacts perceptual aspects such as instrument recognition and intonation.

Acknowledgements

We thank the anonymous reviewers for their valuable feedback. The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. This project is supported by the German Research Foundation (DFG MU 2686/13-2).

References

Berezovsky, J. (2019). The structure of musical harmony as an ordered phase of sound: A statistical mechanics approach to music theory. *Science Advances*, 5(5):eaav8490.

Daniel, P. and Weber, R. (1997). Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica*, 83(1):113–123.

Hamster, U. A., (2021). bwsample: Processing Best-Worst Scaling data. *Journal of Open Source Software, 6(64), 3324.*

Jeffress, L. A. (1968). Beating Sinusoids and Pitch Changes. *The Journal of the Acoustical Society of America*, 43(6):1464–1464.

Louviere, J., Flynn, T., and Marley, A. (2015). *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press, Cambridge.

Marijeh, R., Harrison, P. M. C., Lee, H., Deligiannaki, F., and Jacoby, N. (2022). Reshaping musical consonance with timbral manipulations and massive online experiments. *bioRxiv 2022.06.14.496070 (preprint)*.

McDermott, J. H., Schultz, A. F., Undurraga, E. A., and Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, 535(7613):547–550.

Plomp, R. and Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America*, 38(4):548–560.

Schwär S., Rosenzweig S., and Müller, M. (2021). A Differentiable Cost Measure for Intonation Processing in Polyphonic Music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR): 626–633*.

Sethares, W. A. (1998). *Tuning, Timbre, Spectrum, Scale.* Springer, London.

Terhardt, E. (1974). On the perception of periodic sound fluctuations (roughness). *Acta Acustica united with Acustica*, 30(4):201–213.

Vassilakis, P. N. (2005). Auditory roughness as means of musical expression. *Selected Reports in Ethnomusicology, 12:119-144.*

Vassilakis, P. N. and Kendall, R. A. (2010). Psychoacoustic and cognitive aspects of auditory roughness: definitions, models, and applications. In *Proceedings of Human Vision and Electronic Imaging XV, 75270O, SPIE 7527.*