MODULATION TYPE AND AMOUNT SHAPE HUMAN PERCEPTION OF TIMBRAL CHANGE IN WAVETABLE SYNTHESIS

Sachin Subramanian [♭] **Iran R. Roman** [♭] **Johan Pauwels** [♭]
Joshua D. Reiss [♭] **Christopher Mitcheltree** [♭]
[♭] Centre for Digital Music, Queen Mary University of London, UK
s.a.subramanian@se25.qmul.ac.uk

ABSTRACT

Modulation is central to sound synthesis, yet its perceptual salience remains poorly characterized. We investigate how amplitude, frequency, and irregularity of a modulation signal driving wavetable-position scanning affect perceived timbral change. In a MUSHRA-style listening test, participants rated perceived difference between a minimally modulated reference and four increasingly modulated versions, across a wide array of timbres. Perceived difference increases linearly with modulation amount for all three modulation types. A repeated-measures ANOVA confirms significant main effects of modulation type and amount, and their interaction, with frequency modulation producing the largest perceptual changes, followed by amplitude, and irregularity the smallest; an ordering that holds consistently at both low and high modulation amounts. Our results provide quantitative evidence for perceptually informed modulation design, and a reproducible framework for studying time-varying synthesizer timbre.

1. INTRODUCTION

Modulation allows synthesized sounds to evolve over time through controlled variation of synthesis parameters [1–3]. Amplitude modulation, frequency modulation, filter modulation, and wavetable scanning are widely used to shape timbre and introduce movement [4,5] but their relative perceptual salience has received limited attention [6, 7]. This question also matters for audio processing and machine-learning systems, where time-varying effects and synthesis remain active research areas [8–11]. Recent work uses embedding-based distances to evaluate perceptual and semantic similarity [12–14]. However, such objective measures may not reflect human perceptual judgements [15].

We therefore investigate how perceived difference scales with the amplitude, frequency, and irregularity of the modulation signal, hypothesizing that it increases with modulation amount, regardless of timbre (Fig. 1).

© S. Subramanian, I. R. Roman, J. Pauwels, J. D. Reiss, and C. Mitcheltree. Licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). **Attribution:** S. Subramanian, I. R. Roman, J. Pauwels, J. D. Reiss, and C. Mitcheltree, " Modulation Type and Amount Shape Human Perception of Timbral Change in Wavetable Synthesis ", in *Extended Abstracts for the Late-Breaking Demo Session of the 27th Int. Society for Music Information Retrieval Conf.*, Abu Dhabi, UAE, 2026.

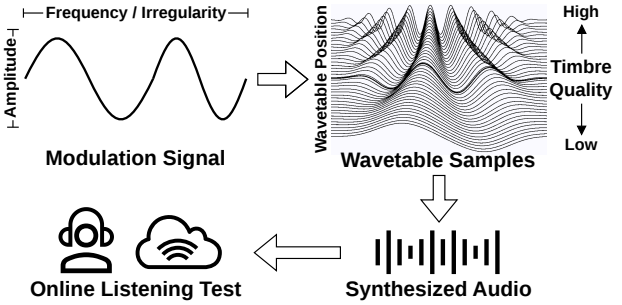


Figure 1. A modulation signal can vary in amplitude, frequency, and irregularity to scan through a wavetable that spans a range of timbre qualities. The resulting audio was rated by listeners for perceived difference from a minimally modulated reference. We hypothesized that perceived difference correlates with modulation amount across all three properties, independent of timbre.

2. APPROACH

2.1 Stimuli

Stimuli were generated with a wavetable synthesis pipeline (4 sec. long) [1] . Six wavetables were used, two per timbre quality (warmth, brightness, richness), each sweeping linearly along its quality axis (see Fig. 1). For each wavetable,

[1] Complete stimuli generation details, timbre quality definitions, code, audio samples, ethics approval, and listening test data provided at: https://christhetree.github.io/mod_perception

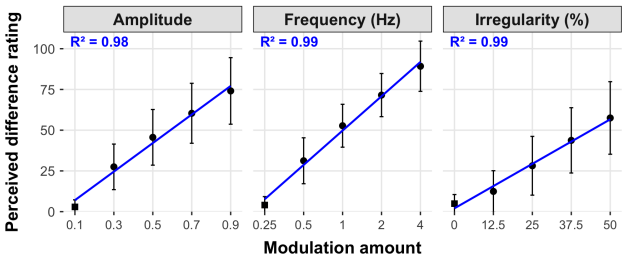


Figure 2. Perceived-difference ratings between the minimally modulated reference (■) and versions with increasing modulation (●) (N = 52). Ratings range from 0 ("identical") to 100 ("completely different"); error bars show SD. Linear fits (blue) are strongly positive ($R^2 \geq 0.98$), showing perceived difference increased with modulation across all three properties. However, perceptual salience differed: at the highest level ratings reached ∼75 for amplitude and ∼90 for frequency, but only ∼57 for irregularity.

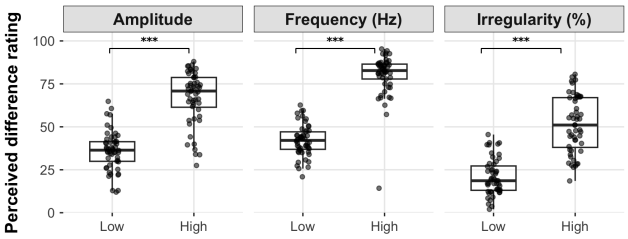


Figure 3. Perceived-difference ratings for low and high modulation amounts, by modulation type. Each point shows one participant's mean rating within the *Low* or *High* grouping; boxplots show the median and interquartile range (IQR), with whiskers extending to 1.5× IQR.

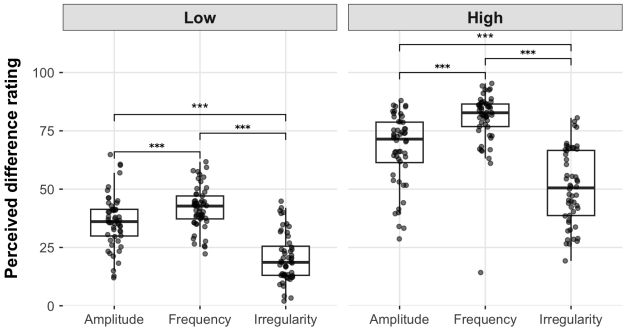


Figure 4. Perceived-difference ratings by modulation type, shown separately for low and high modulation amounts (same data as Fig. 3, regrouped by amount level). Each point shows one participant's mean rating for a given modulation type and amount level; boxplots show the median and IQR, with whiskers extending to 1.5× IQR. Paired t-tests indicated significant differences between all three modulation types within both amount levels (all $p < .001$), with frequency eliciting the highest ratings and irregularity the lowest at both amounts.

three modulation types drove the wavetable-position scan: amplitude (0.1–0.9), frequency (0.25–4.0 Hz), and irregularity (0–50%, i.e. aperiodicity), each tested at five levels with the other two dimensions held fixed. The lowest level served as reference: for amplitude and frequency this retained slight modulation (0.1 depth, 0.25 Hz); for irregularity it was fully periodic (0%).

2.2 Procedure and Participants

The online test followed the MUSHRA standard [16]. Each of 18 trials (3 modulation types × 3 timbre qualities × 2 wavetables) presented five stimuli from one source wavetable and modulation type, rated for perceived difference from the reference on a 0 ("identical") to 100 ("completely different") scale. 53 participants completed the test. Trials were flagged if the reference itself was rated above 25, all five stimuli received identical ratings, or completion took under 24 sec. One participant was therefore excluded, yielding $N = 52$. The listening test complies with and was approved by the Queen Mary Research Ethics Committee.

2.3 Analysis

Ratings were averaged within participant, modulation type, and amount level (Fig. 2). The four non-reference levels were grouped into *Low* (1–2) and *High* (3–4), and a two-way repeated-measures ANOVA tested modulation type × amount group on these averages, with significant effects followed by Bonferroni-corrected paired t-tests.

3. RESULTS, DISCUSSION & CONCLUSION

Figure 2 shows perceived-difference ratings as a function of modulation amount for all three modulation types, with linear fits capturing a positive trend ($R^2 \geq 0.98$).

To formally test the effects observed in Figure 2, we ran a two-way repeated-measures ANOVA with modulation type (amplitude, frequency, irregularity) and amount group as within-subjects factors. For each modulation type, the two lowest tested levels were grouped as *Low* and the two highest as *High* (the minimally modulated reference was excluded). Both main effects were significant: modulation type, $F(2, 102) = 115.76$, $p < .001$, $\eta_g^2 = .430$, and amount group, $F(1, 51) = 1044.89$, $p < .001$, $\eta_g^2 = .659$, confirming that perceived difference increased substantially from low to high modulation amounts across all three parameters, consistent with the linear trends in Figure 2. A modulation type × amount group interaction was also significant, $F(2, 102) = 21.04$, $p < .001$, but accounted for a comparatively small share of variance ($\eta_g^2 = .027$), indicating that the low-to-high increase differed somewhat in magnitude across modulation types without displacing the dominant main effects. [2]

Figures 3 and 4 help interpret this interaction. The low-to-high increase was significant for every modulation type individually (Fig. 3), confirming that the amount effect was not driven by any single type. The three modulation types also differed significantly from one another at both amount levels, in the same order (frequency > amplitude > irregularity) at low amounts (amplitude vs. frequency, $p = 1.07 \times 10^{-4}$; amplitude vs. irregularity, $p = 5.51 \times 10^{-12}$; frequency vs. irregularity, $p = 2.87 \times 10^{-15}$) and at high amounts (amplitude vs. frequency, $p = 1.67 \times 10^{-10}$; amplitude vs. irregularity, $p = 1.79 \times 10^{-10}$; frequency vs. irregularity, $p = 6.50 \times 10^{-19}$; Fig. 4). Because this ordering is preserved across amount levels, the interaction reflects differences in the *size* of the low-to-high increase across modulation types rather than a reordering of their relative perceptual salience.

In conclusion, our findings provide a starting point for understanding the main effects of human perception of modulated audio outside of frequency and amplitude modulation, and support the design of perceptually informed synthesizer interfaces [17–19]. We believe that comparing the perceived-difference ratings from this study with perceptual audio distance functions [20] and neural audio embeddings [21] are compelling directions for future work.

[2] Groups reflect each parameter's own tested range, not equivalent-sized manipulations across modulation types. The modulation main effect remains informative, but the interaction may partly reflect differences in tested range rather than intrinsic sensitivity to modulation amount.

4. ACKNOWLEDGMENTS

Christopher Mitcheltree is supported jointly by UK Research and Innovation (grant number EP/S022694/1) and Queen Mary University of London.

5. AI USAGE STATEMENT

Generative AI tools were used to assist with language editing, structural revision, and condensation of material from a longer initial version of this manuscript. The authors reviewed and revised all generated text and remain responsible for the study design, analysis, interpretation, and final manuscript.

6. REFERENCES

[1] D. Smalley, "Spectromorphology: explaining sound-shapes," *Organised Sound*, vol. 2, no. 2, pp. 107–126, 1997.

[2] A. Farnell, *Designing Sound*. Cambridge, MA, USA: MIT Press, 2010.

[3] U. Zölzer, Ed., *DAFX: Digital Audio Effects*, 2nd ed. Chichester, UK: Wiley, 2011.

[4] M. Henry, "A perceptual study of amplitude modulation vibrato," Master's thesis, McGill Univ., Montréal, Canada, 2021.

[5] J. Hyrkas and T. Smyth, "On vibrato and frequency (de)modulation in musical sounds," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Sep. 2024.

[6] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and Models*, 3rd ed., ser. Springer Series in Information Sciences. Berlin, Germany: Springer, 2007, vol. 22, ch. 10, pp. 247–256.

[7] S. M. Town and J. K. Bizley, "Neural and behavioral investigations into timbre perception," *Frontiers in Systems Neuroscience*, vol. 7, Nov. 2013.

[8] C. Mitcheltree, C. J. Steinmetz, M. Comunità, and J. D. Reiss, "Modulation extraction for LFO-driven audio effects," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Sep. 2023.

[9] C. Yu, C. Mitcheltree, A. Carson, S. Bilbao, J. D. Reiss, and G. Fazekas, "Differentiable all-pole filters for time-varying audio systems," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Sep. 2024.

[10] C. Mitcheltree, H. H. Tan, and J. D. Reiss, "Modulation discovery with differentiable digital signal processing," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2025.

[11] A. Carson, A. Wright, and S. Bilbao, "Gradient-Based Optimization of Modulation Effects," *Journal of the Audio Engineering Society*, vol. 74, no. 7/8, pp. 533–543, Aug. 2026.

[12] P. Manocha, Z. Jin, R. Zhang, and A. Finkelstein, "CDPAM: Contrastive learning for perceptual audio similarity," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Jun. 2021, pp. 196–200.

[13] B. Hayes, C. Saitis, and G. Fazekas, "Disembodied Timbres: A Study on Semantically Prompted FM Synthesis," *Journal of the Audio Engineering Society*, vol. 70, pp. Paper 373–391, May 2022.

[14] M. Tailleur, J. Lee, M. Lagrange, K. Choi, L. M. Heller, K. Imoto, and Y. Okamoto, "Correlation of Fréchet audio distance with human perception of environmental audio is embedding dependant," Aug. 2024.

[15] P. Manocha, A. Finkelstein, R. Zhang, N. J. Bryan, G. J. Mysore, and Z. Jin, "A differentiable perceptual audio metric learned from just noticeable differences," in *Proceedings of the International Speech Communication Association Conference (Interspeech)*, Oct. 2020.

[16] International Telecommunication Union, "Method for the subjective assessment of intermediate quality level of audio systems," International Telecommunication Union, Geneva, Switzerland, Recommendation ITU-R BS.1534-3, 2015.

[17] H. Lee, K. Kim, S. Lee, and K. Lee, "Wavespace: A highly explorable wavetable generator," 2024, presented at the IRCAM Forum Workshop. [Online]. Available: http://arxiv.org/abs/2407.19862

[18] J. Shier, C. Saitis, A. Robertson, and A. McPherson, "Real-time timbre remapping with differentiable DSP," Sep. 2024.

[19] T. Yutani, Y. Yamamoto, S. Nakatani, and H. Terasawa, "Wavetable synthesis using CVAE for timbre control based on semantic label," in *Proceedings of the APSIPA Annual Summit and Conference (APSIPA ASC)*, Dec. 2024.

[20] C. Vahidi, H. Han, C. Wang, M. Lagrange, G. Fazekas, and V. Lostanlen, "Mesostructures: Beyond spectrogram loss in differentiable time–frequency analysis," *Journal of the Audio Engineering Society*, vol. 71, no. 9, pp. 577–585, Sep. 2023.

[21] H. Tian, S. Lattner, and C. Saitis, "Assessing the alignment of audio representations with timbre similarity ratings," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Sep. 2025.