

Investigating Human Sensitivity to Modulated Wavetable Audio Samples

1st Sachin Subramanian
Center for Digital Music
Queen Mary University of London
London, United Kingdom
ec25057@qmul.ac.uk

2nd Johan Pauwels
Center for Digital Music
Queen Mary University of London
London, United Kingdom
j.pauwels@qmul.ac.uk

Abstract—Modulation is a fundamental component of sound synthesis, enabling synthesised sounds to evolve over time through controlled variation of synthesis parameters. Although modulation is widely used in music production and sound design, the perceptual consequences of changing different attributes of a modulation signal remain insufficiently characterised. In this work, amplitude and frequency refer to properties of the modulation signal itself, rather than to conventional amplitude-modulation (AM) or frequency-modulation (FM) synthesis.

This dissertation investigates human sensitivity to modulation in wavetable synthesis through a controlled listening experiment. Audio stimuli were generated using a Python-based wavetable-synthesis pipeline in which modulation type and modulation amount could be varied while the remaining synthesis parameters were held constant. Six representative wavetables were selected across three timbral characteristics — brightness, richness, and warmth — with one synthetic and one natural-source ("real") wavetable representing each characteristic. Three modulation-signal attributes — amplitude, frequency, and irregularity — were evaluated at multiple modulation amounts. Participants completed a WebMUSHRA listening experiment in which they rated the perceived difference between a minimally modulated reference and corresponding stimuli generated at greater modulation amounts, on a continuous scale from 0 (perceptually identical) to 100 (completely different).

Perceived difference increased strongly with modulation amount across the tested conditions. Changes to modulation-signal frequency generally produced the largest perceptual changes, followed by changes to modulation-signal amplitude, while changes to irregularity produced the smallest. Repeated-measures analysis indicated significant effects of modulation type, timbral characteristic, and modulation amount, while the magnitude of source and interaction effects was comparatively smaller. Linear regression further indicated that perceived difference increased approximately linearly with modulation amount over the range investigated.

These findings provide quantitative evidence that the perceptual effect of modulation depends on both the type and strength of the applied modulation. The work establishes a reproducible framework for studying time-varying synthesiser timbre and provides a basis for perceptually informed parameter scaling and sound-design interfaces.

to evolve through controlled variation of synthesis or effect parameters (Farnell, 2010; Zölzer, 2011). In its broadest form, modulation uses one signal or control trajectory to vary another parameter over time. Parameters such as amplitude, instantaneous frequency, filter cutoff, wavetable position, spatial position, and spectral content can therefore be manipulated to introduce movement, expression, and timbral complexity (Farnell, 2010; McAdams and Bigand, 1993).

Many familiar sound-design techniques rely on modulation. Tremolo varies amplitude, vibrato varies frequency, and other common processes include envelope control, dynamic filtering, wavetable scanning, chorus, flanging, phasing, panning, and spectral transformation (Farnell, 2010; Zölzer, 2011). Despite the prevalence of these techniques, comparatively less is known about how the perceptual magnitude of change differs across modulation types when the underlying sound source and synthesis conditions are controlled.

Previous perceptual work has shown that listeners can be sensitive to relatively subtle temporal variation in musical sounds (Henry, 2021; Hayes, Saitis and Fazekas, 2022). However, such studies often focus on a single modulation phenomenon or synthesis method, which makes it difficult to determine whether different modulation dimensions produce comparable perceptual changes, how perceived difference grows as modulation strength increases, and whether these relationships remain stable across contrasting source timbres.

This dissertation addresses that problem using controlled wavetable-synthesis stimuli and a reference-based listening experiment. The scope is deliberately centred on the perceptual consequences of varying three attributes of the modulation signal — amplitude, frequency, and irregularity — rather than on conventional AM or FM synthesis, or on the validation of computational audio-distance metrics. Objective perceptual metrics remain a relevant extension of the work and are considered in the Future Work section.

I. INTRODUCTION

Modulation is central to sound synthesis and audio production because it allows otherwise static sounds

A. Research Questions

The study addresses the following research questions:

- RQ1:** How does modulation type influence the perceived difference between more strongly modulated wavetable synthesiser sounds and a minimally modulated reference?
- RQ2:** How does increasing modulation amount affect perceived auditory difference?
- RQ3:** Does sensitivity to modulation vary across the timbral characteristics of brightness, richness, and warmth?
- RQ4:** Does modulation perception differ between natural-source (real) and synthetic wavetables?
- RQ5:** Are the effects of modulation type and amount consistent across timbral characteristic and wavetable source?

B. Contributions

The principal contributions of this dissertation are: (i) a controlled Python wavetable-synthesis framework for generating systematically modulated stimuli; (ii) a WebMUSHRA-based listening experiment for measuring perceived difference from a minimally modulated reference; (iii) a statistical comparison of modulation type, modulation amount, timbral characteristic, and wavetable source; and (iv) an empirical characterisation of the relationship between modulation strength and perceived auditory difference.

II. BACKGROUND

A. Modulation in Sound Synthesis

Technical work on modulation has established the signal-processing foundations of effects such as amplitude modulation, frequency modulation, vibrato, and other time-varying audio processes (Martínez Ramírez, Benetos and Reiss, 2019; Hyrkäs and Smyth, 2024; Chowning, 1973; Cook, 2002). These conventional AM and FM synthesis techniques are distinct from the manipulation used in the present study: here, amplitude and frequency describe attributes of a separate modulation signal that is subsequently used to modulate wavetable behaviour, rather than directly modulating output amplitude or carrier frequency. More broadly, the digital audio effects literature describes modulation as a central mechanism for creating temporal evolution in synthesised sound (Farnell, 2010; Zölzer, 2011).

B. Perception of Modulated Sound

Perceptual studies have investigated individual modulation phenomena and relationships between synthesis parameters and timbral attributes. Henry (2021), for example, examined the perception of amplitude-modulated vibrato, while Hayes, Saitis and Fazekas (2022) investigated semantic relationships in FM synthesis. Foundational psychoacoustic work has also demonstrated that sensitivity to temporal modulation depends strongly on modulation rate (Viemeister, 1979), while Søk and Moore (1995) showed that modulation detection and discrimination are influenced by factors including carrier frequency, modulation

rate, and modulation waveform. Earlier work by Lee and Bacon (1997) further established that listeners' ability to discriminate amplitude-modulation depth depends on stimulus duration, indicating that sensitivity to a modulation attribute cannot always be treated as constant across presentation conditions. Together, these studies demonstrate that modulation parameters can have systematic perceptual consequences, but none of them directly compares the perceptual effects of varying the amplitude, frequency, and irregularity of a modulation signal within the same controlled wavetable framework.

C. Timbre and Wavetable Synthesis

Wavetable synthesis provides a useful experimental framework because a source waveform can be controlled while other synthesis parameters are manipulated systematically. Recent work has also explored semantic control and navigation of wavetable spaces (Lee et al., 2024; Yutani et al., 2024). In the present study, brightness, richness, and warmth are used as timbral categories for selecting contrasting source wavetables rather than as interchangeable definitions of timbre itself. For each category, one synthetic and one natural-source wavetable were selected to provide variation in source character while maintaining a balanced design.

D. Perceptual Evaluation of Audio

Subjective listening experiments remain an important method for measuring perceptual audio differences. MUSHRA provides a reference-based framework in which listeners compare multiple stimuli within a common trial (International Telecommunication Union, 2015). Related work has demonstrated its use for controlled comparisons of audio systems and processing conditions (Berg et al., 2013). The present experiment adapts this reference-comparison principle to measure the perceived difference between a minimally modulated reference and more strongly modulated wavetable sounds.

E. Research Gap

Existing literature provides foundations in modulation synthesis, wavetable generation, and subjective audio evaluation, but there remains limited work directly comparing listeners' sensitivity to multiple modulation types while independently varying modulation amount and controlling the underlying source sound. This dissertation addresses that gap by examining changes to the amplitude, frequency, and irregularity of the modulation signal across multiple wavetable timbres and source types.

III. METHODOLOGY

A. Experimental Design

The experiment used a repeated-measures design in which the principal factors were modulation type,

timbral characteristic, wavetable source, and modulation amount. Three modulation-signal attributes were tested: amplitude, frequency, and irregularity. Three timbral categories were represented: brightness, richness, and warmth. Each timbral category contained one synthetic and one natural-source wavetable, producing six source sounds in total. Modulation amount was varied systematically within each modulation condition.

The experiment comprised 18 base conditions corresponding to every combination of modulation type (3), timbral characteristic (3), and wavetable source (2). Within each condition, listeners compared a minimally modulated reference with versions of the same source generated at progressively greater modulation amounts.

B. Wavetable Selection

Six representative wavetables were selected for the listening experiment. Synthetic wavetables were generated from analytically defined waveforms, whereas natural-source wavetables were derived from recorded acoustic material. The sources were chosen to represent brightness, richness, and warmth, with one synthetic and one natural-source example in each category. This balanced selection allowed modulation effects to be examined across contrasting timbral material without changing the overall factorial structure of the experiment (Yutani et al., 2024; McAdams and Bigand, 1993).

C. Stimulus Generation

Audio stimuli were generated using a wavetable-synthesis pipeline implemented in Python, following the approach of Python-based digital-audio-workstation bridging tools such as DAWDreamer (Braun, 2021). Wavetable synthesis was selected because it allowed the underlying source material and synthesis configuration to remain fixed while the behaviour of the modulation signal was manipulated independently (Farnell, 2010; Zölzer, 2011). This reduced variation from unrelated synthesis parameters and allowed changes in listener ratings to be interpreted primarily in relation to modulation.

1) *Modulation-Signal Amplitude*: In the amplitude condition, the amplitude of the modulation signal was varied while the remaining synthesis parameters were held fixed. Increasing the modulation amount therefore increased the magnitude of the modulation signal applied to the wavetable process. This should not be interpreted as conventional amplitude modulation (AM), in which the amplitude of an audio carrier is directly modulated.

2) *Modulation-Signal Frequency*: In the frequency condition, the frequency of the modulation signal was varied while the remaining synthesis parameters were held fixed. This parameter controlled how rapidly the modulation signal evolved before being applied to the wavetable process. It is therefore distinct from

conventional frequency-modulation (FM) synthesis, in which the instantaneous frequency of an audio carrier is directly modulated.

3) *Modulation-Signal Irregularity*: In the irregularity condition, the temporal regularity of the modulation signal was manipulated rather than its nominal rate. Temporal irregularity provides an important auditory cue, and introducing temporal jitter into otherwise periodic signals can disrupt their periodic structure and associated perceptual representation (Bendor, Wang and Bartlett, 2010). In the implementation used here, the modulation signal’s period was perturbed on a cycle-by-cycle basis by an amount drawn proportionally to the irregularity setting, so that increasing irregularity introduced progressively greater departure from a perfectly regular modulation pattern while preserving the broader temporal envelope of the stimulus. This condition was included to test listener sensitivity to temporal predictability and regularity alongside changes to modulation-signal amplitude and frequency.

4) *Modulation Amount*: For each modulation type, stimuli were generated at multiple increasing modulation amounts while the remaining synthesis parameters were held fixed. The reference was itself modulated at the minimum modulation setting rather than being strictly unmodulated, and therefore served as a minimally modulated baseline against which progressively greater modulation amounts were judged. For inferential comparisons of modulation strength, non-reference amounts were additionally grouped into low (amounts 1–2) and high (amounts 3–4) categories. This grouping enabled modulation amount to be included as a within-subject factor while retaining sufficient observations within each level; it is used alongside, and not as a replacement for, the continuous regression analysis reported in Section IV.C.

D. Timbral Feature Analysis

Five audio descriptors were computed during stimulus development to characterise the generated sounds and support wavetable selection: spectral centroid, spectral flatness, loudness, warmth, and richness.

Spectral centroid measures the centre of mass of the magnitude spectrum and was used as an acoustic indicator of brightness. Higher centroid values generally correspond to a greater concentration of energy at higher frequencies and therefore to brighter timbres (Yutani et al., 2024).

Spectral flatness describes the extent to which a spectrum is tonal or noise-like by comparing the geometric and arithmetic means of the power spectrum. Tonal sounds typically exhibit lower spectral flatness, whereas broadband or noise-like spectra produce higher values (Peeters, 2004).

Loudness was estimated using integrated Loudness Units relative to Full Scale (LUFS), following the ITU-R BS.1770 recommendation (International Telecommunication Union, 2023). Loudness normalisation was

subsequently applied to reduce playback-level differences between stimuli, allowing perceptual comparisons to focus primarily on changes introduced by modulation.

Warmth was characterised using a harmonic power-based descriptor informed by previous work on semantic timbre control (Yutani et al., 2024). The implementation measured the relative contribution of harmonic energy, with the DC component excluded from the overall energy calculation.

Richness represented the complexity and distribution of spectral content. Following previous work on semantic wavetable synthesis (Yutani et al., 2024), it was estimated from the variation of spectral energy across the spectrum, providing an indication of spectral density.

These descriptors were calculated throughout each wavetable sweep, allowing changes in spectral and timbral characteristics introduced by modulation to be inspected and visualised over time. They were used as analysis and stimulus-development tools rather than as substitutes for the perceptual ratings collected in the listening experiment.

E. Stimulus Visualisation and Loudness Normalisation

Animated visualisations were produced during stimulus development to inspect the waveform and the evolution of extracted audio descriptors over time. Several loudness-normalised versions of the stimuli were also produced using LUFS-based normalisation following ITU-R BS.1770 and the pyloudnorm implementation (Steinmetz and Reiss, 2021; International Telecommunication Union, 2023). Pilot testing was used to select a playback-normalisation target that maintained comfortable listening levels while preserving perceptual differences between modulation conditions. A target of -18 LUFS was selected, providing sufficient headroom to reduce the risk of peak clipping while maintaining an appropriate listening level. This target was applied consistently across all experimental stimuli to minimise loudness as a confounding factor.

F. Listening Experiment

Subjective evaluation was conducted using WebMUSHRA, based on the reference-comparison principles of the MUSHRA methodology (International Telecommunication Union, 2015). A total of 53 participants took part in the listening experiment, all of whom were over 18 years of age. Participants represented a range of musical backgrounds, including both listeners with prior musical experience and those with no formal musical experience. Musical experience was recorded as participant metadata but was not used as an inclusion criterion.

Each trial presented a minimally modulated reference together with versions generated from the same wavetable at progressively greater modulation amounts. Participants rated the perceived difference

between each stimulus and the reference using a continuous scale from 0 to 100, where 0 represented "perceptually identical" and 100 represented "completely different."

The listening experiment contained 18 trials, covering every combination of modulation type, timbral category, and wavetable source. This allowed responses to be collected across listeners with a broad range of musical backgrounds. A five-participant pilot was conducted before the main study to evaluate the interface, stimulus presentation, and experimental procedure.

G. Data Screening

Responses were exported from WebMUSHRA and processed in R. Quality-control criteria were applied before statistical analysis to improve the reliability of the dataset. Responses were excluded if any of the following conditions were satisfied:

- The participant reported using an unsuitable listening device, such as built-in laptop, MacBook, computer, or phone speakers
- The hidden reference received a rating greater than 25
- All stimuli within a trial received identical ratings
- The total completion time for a trial was less than 24 seconds

The final criterion ensured that participants had sufficient time to listen to every stimulus before submitting a response, reducing the likelihood of inattentive or incomplete evaluations. These screening criteria were intended to reduce the influence of technically unsuitable or inattentive responses before subsequent statistical analysis.

H. Statistical Analysis

Descriptive statistics were first calculated for each experimental condition. Distributional assumptions were inspected using Shapiro–Wilk tests. Because the low/high amount distributions contained several departures from normality, normality results are reported transparently in Appendix A rather than used as a central substantive result.

To quantify the relationship between modulation amount and perceived difference, linear and exponential regression models were fitted to each modulation–feature–source condition. Model fit was compared using the coefficient of determination (R^2). The linear model provided the stronger fit across the tested conditions and is therefore used to summarise the modulation–amount relationship.

Inferential analysis used repeated-measures ANOVA to examine within-participant effects of modulation type, timbral characteristic, wavetable source, and low/high modulation amount. Significant main effects were followed by Bonferroni-adjusted paired comparisons. Generalised eta-squared (η^2_G) was reported as a measure of effect size for the ANOVA. Statistical significance was assessed at $\alpha = 0.05$.

IV. RESULTS

A. Participant Screening

After applying the pre-defined quality-control criteria, 53 participants contributed usable responses across the 18 modulation–feature–source conditions. Because the four-way repeated-measures ANOVA (Section IV.D) requires a complete set of observations across every factorial condition for each participant, listwise deletion was applied for that analysis; 21 participants had fully complete data across all 18 conditions and were retained for the ANOVA reported in Table II. The descriptive statistics, regression analysis (Section IV.C), and figures reported below draw on the full set of 53 screened participants at the condition level, since these analyses do not require complete within-participant coverage of every condition. This distinction is noted here to avoid ambiguity between the sample sizes reported across analyses.

B. Effect of Modulation Amount

Perceived difference increased consistently as the modulation amount increased. When amounts were grouped into low (levels 1–2) and high (levels 3–4) conditions, high-modulation stimuli received larger difference ratings across modulation types. This indicates that modulation strength was a major determinant of perceived departure from the minimally modulated reference.

C. Relationship Between Modulation Amount and Perceived Difference

Across all 18 modulation–feature–source conditions, the linear regression model provided a better fit than the exponential alternative. Figure 2 shows the positive relationship between modulation amount and perceived difference. The modulation-signal frequency condition generally exhibited the strongest linear relationships, the amplitude condition showed intermediate relationships, and the irregularity condition showed weaker relationships.

The condition-level coefficients of determination are shown in Table I. The modulation-signal frequency condition ranged from $R^2 = 0.732$ to 0.897 , the amplitude condition from $R^2 = 0.582$ to 0.716 , and the irregularity condition from $R^2 = 0.381$ to 0.631 .

D. Effects of Modulation Type, Timbre, Source, and Amount

A four-way repeated-measures ANOVA examined modulation type, timbral characteristic, wavetable source, and modulation amount group as within-subject factors, using the 21 participants with complete observations described in Section IV.A. The results are reported in Table II. They indicate significant main effects of modulation type, timbral characteristic, source, and modulation amount. Significant interactions were also found for feature by source, modulation by amount, and the four-way interaction, while the remaining interactions were not significant at $\alpha = 0.05$.

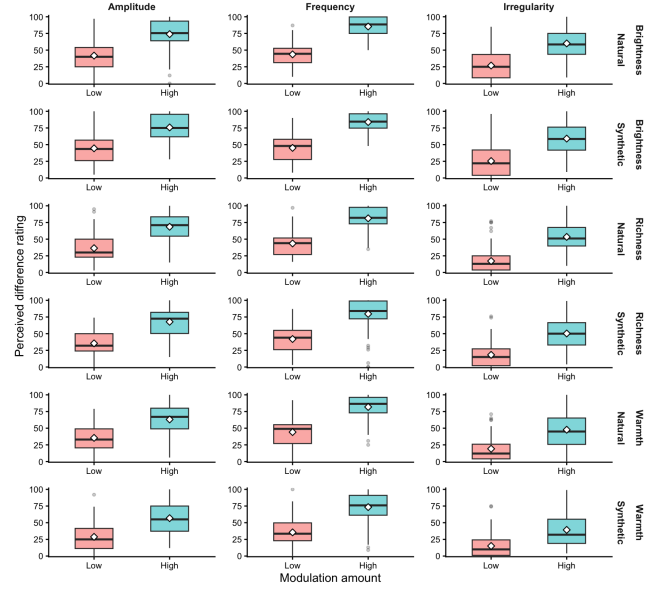


Fig. 1. Distribution of perceived-difference ratings for low (amounts 1–2) and high (amounts 3–4) modulation groups across the experimental conditions.

E. Post-hoc Comparisons: Modulation Type

Bonferroni-adjusted paired comparisons indicated significant differences between each pair of modulation types (Table III). The modulation-signal frequency condition produced higher perceived-difference ratings than the amplitude condition, and both produced higher ratings than the irregularity condition.

F. Post-hoc Comparisons: Timbral Characteristic

Post-hoc comparisons between timbral characteristics also indicated significant differences between brightness, richness, and warmth (Table IV). Brightness produced higher perceived-difference ratings than richness and warmth, and Table IV additionally reports a significant richness- warmth difference.

G. Results Summary

Perceived difference was strongly affected by modulation amount and modulation type. Ratings increased with modulation strength, and this increase was approximately linear over the tested range. The modulation-signal frequency condition generally produced the largest perceived changes, the amplitude condition produced intermediate changes, and the irregularity condition produced the smallest. Timbral characteristic also influenced ratings, while source and interaction effects were smaller in magnitude than the dominant modulation and amount effects.

V. DISCUSSION

A. Perceptual Salience of Modulation Type

The clear differences between modulation conditions suggest that equal numerical changes in different attributes of the modulation signal should not

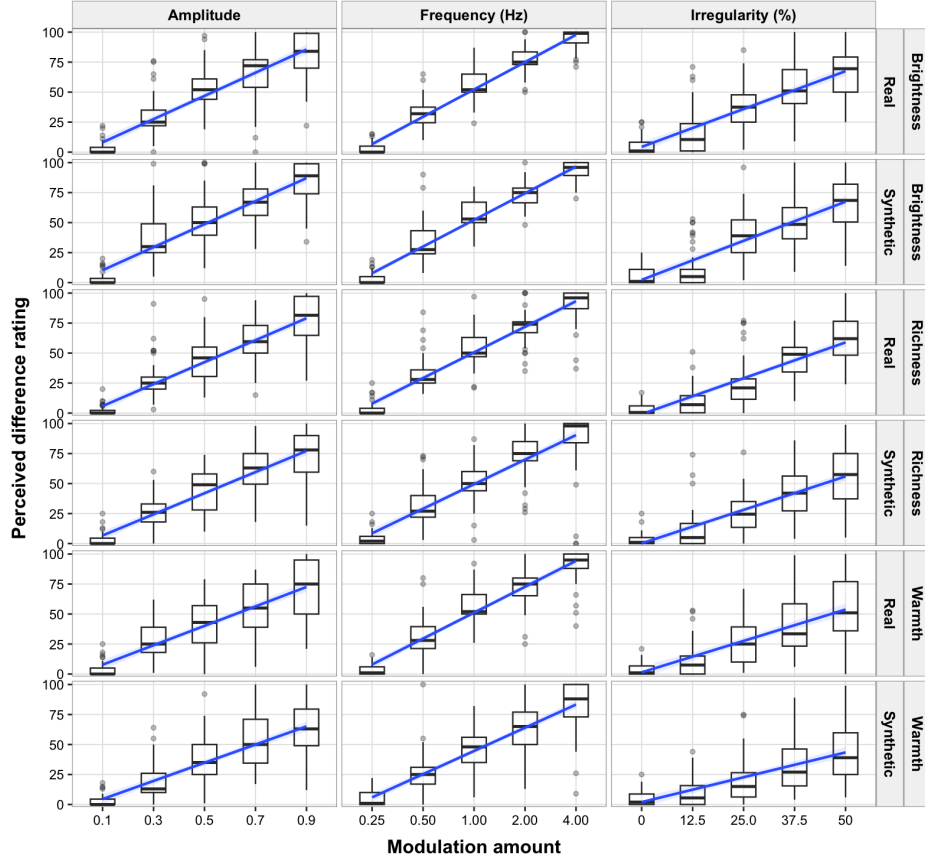


Fig. 2. Relationship between modulation amount and perceived-difference ratings for each modulation type, timbral characteristic, and wavetable source. Solid lines show least-squares linear fits with 95% confidence intervals.

TABLE I
COEFFICIENT OF DETERMINATION (R^2) FOR THE LINEAR
REGRESSION MODEL IN EACH EXPERIMENTAL CONDITION.

Modulation	Feature	Source	R^2
Frequency	Brightness	Real	0.897
Frequency	Brightness	Synthetic	0.885
Frequency	Richness	Real	0.841
Frequency	Richness	Synthetic	0.732
Frequency	Warmth	Real	0.818
Frequency	Warmth	Synthetic	0.748
Amplitude	Brightness	Real	0.695
Amplitude	Brightness	Synthetic	0.689
Amplitude	Richness	Real	0.716
Amplitude	Richness	Synthetic	0.690
Amplitude	Warmth	Real	0.602
Amplitude	Warmth	Synthetic	0.582
Irregularity	Brightness	Real	0.602
Irregularity	Brightness	Synthetic	0.601
Irregularity	Richness	Real	0.631
Irregularity	Richness	Synthetic	0.574
Irregularity	Warmth	Real	0.471
Irregularity	Warmth	Synthetic	0.381

be assumed to produce equal perceptual changes. Changes to modulation-signal frequency were especially salient in the present experiment, while changes to modulation-signal amplitude produced intermediate ratings. These results describe the perceptual consequences of the tested modulation-signal parameters

and should not be interpreted as a comparison between conventional FM and AM synthesis techniques.

Changes to modulation-signal irregularity generally produced the smallest perceived differences. One interpretation is that listeners were less sensitive to the tested changes in temporal regularity than to changes in the frequency or amplitude attributes of the modulation signal. This conclusion is specific to the modulation ranges and implementation tested here; the experiment does not establish a universal ordering for all possible irregularity, amplitude, and frequency modulation settings.

Importantly, the differences between modulation types were also present at both low and high modulation amounts. Pairwise comparisons showed significant differences between the amplitude, frequency, and irregularity conditions within both amount groups. The relative differences between modulation types therefore emerged even at comparatively low modulation amounts rather than being driven only by the most extreme parameter settings, providing further evidence that modulation type itself influenced perceptual salience within the ranges tested.

B. Sensitivity to Modulation Amount

The strong effect of modulation amount demonstrates that listeners' ratings tracked the magnitude

TABLE II
FOUR-WAY REPEATED-MEASURES ANOVA. 53 PARTICIPANTS CONTRIBUTED USABLE DATA OVERALL; THE ANOVA USED THE 21 PARTICIPANTS WITH COMPLETE OBSERVATIONS ACROSS EVERY FACTORIAL CONDITION.

Effect	DFn	DFd	F	p	ges
Modulation	2	40	60.661	< 0.001	0.320
Feature	2	40	17.840	< 0.001	0.066
Source	1	20	13.662	0.001	0.007
Amount	1	20	599.469	< 0.001	0.597
Mod. $\times$ Feat.	4	80	1.670	0.165	0.011
Mod. $\times$ Src.	2	40	0.104	0.901	< 0.001
Feat. $\times$ Src.	2	40	3.842	0.030	0.006
Mod. $\times$ Amt.	2	40	7.931	0.001	0.013
Feat. $\times$ Amt.	2	40	0.157	0.855	< 0.001
Src. $\times$ Amt.	1	20	0.981	0.334	< 0.001
M $\times$ F $\times$ S	4	80	0.141	0.966	< 0.001
M $\times$ F $\times$ A	4	80	1.955	0.109	0.004
M $\times$ S $\times$ A	2	40	0.036	0.965	< 0.001
F $\times$ S $\times$ A	2	40	0.383	0.684	< 0.001
M $\times$ F $\times$ S $\times$ A	4	80	2.828	0.030	0.005

TABLE III
BONFERRONI-ADJUSTED POST-HOC COMPARISONS BETWEEN MODULATION TYPES.

Comparison	t(51)	Adjusted p
Amplitude vs Frequency	-6.95	1.99×10^{-8}
Amplitude vs Irregularity	8.64	4.50×10^{-11}
Frequency vs Irregularity	13.50	5.58×10^{-18}

TABLE IV
BONFERRONI-ADJUSTED POST-HOC COMPARISONS BETWEEN TIMBRAL CHARACTERISTICS.

Comparison	t(50)	Adjusted p
Brightness vs Richness	5.83	1.21×10^{-6}
Brightness vs Warmth	6.95	2.13×10^{-8}
Richness vs Warmth	3.41	0.004

of the applied manipulation. This pattern was consistent across modulation types, with perceived difference significantly increasing from the low- to high-amount condition for each modulation-signal attribute: amplitude, frequency, and irregularity. The effect of modulation amount was therefore not restricted to a particular modulation mechanism.

The approximately linear regression relationships are useful from a sound-design perspective, as they suggest that, within the tested ranges, increasing a modulation parameter produced a relatively predictable increase in perceived difference. The strength of this relationship nevertheless varied by modulation type, with the modulation-signal frequency condition showing the largest R^2 values and the irregularity condition the smallest. This distinction concerns not only the magnitude of perceived difference but also its predictability: higher R^2 values indicate that modulation amount accounted for a greater proportion of the variation in listeners' ratings. The modulation-signal frequency condition therefore exhibited a more consistent relationship between parameter amount and perceived difference, whereas the perceptual conse-

quences of increasing irregularity were comparatively less predictable.

C. Influence of Timbre and Wavetable Source

Differences across brightness, richness, and warmth indicate that the starting timbre can influence the perceptual consequence of modulation. A modulation process is therefore not perceived independently of the spectrum to which it is applied. At the same time, the comparatively small generalised effect sizes for several source and interaction terms suggest that the dominant ordering of modulation types was not restricted to a single wavetable category.

The lack of significant interaction effects is also informative. Although timbral characteristic influenced ratings overall, there was limited evidence that the effects of modulation type or modulation amount changed substantially depending on timbre or source. Within the selected stimulus set, the principal effects of modulation type and amount therefore appeared comparatively consistent across the different starting sounds. This should not be interpreted as evidence that modulation and timbre are universally independent, but rather that no strong dependency was detected within the present experimental design.

The source factor should be interpreted cautiously. "Natural-source" and "synthetic" describe the origin of the selected wavetables, but each category contains a small number of representative sounds. The experiment therefore supports conclusions about the selected source set rather than all natural and synthetic timbres in general.

D. Implications for Synthesiser Design

These results suggest that synthesiser control ranges could benefit from perceptually informed scaling. If changes to modulation-signal frequency produce a larger perceptual change across their parameter range than changes to another modulation dimension, identical linear interface mappings may not provide perceptually comparable control. A perceptually motivated

interface could instead map parameter ranges so that equal control movements produce more comparable changes in perceived sound.

The regression and inferential results together point to an important distinction between parameter space and perceptual space. Within each modulation type, increasing modulation amount produced an approximately systematic increase in perceived difference. Across modulation types, however, movement through their respective parameter ranges did not produce equivalent perceptual changes. Numerical parameter distance may therefore provide a useful indication of perceptual change within a particular modulation dimension without necessarily providing a comparable measure across different dimensions.

The findings are also relevant to sound-design search and interpolation. Parameter-space distance is not necessarily equivalent to perceptual distance: two settings that are numerically close may differ substantially in perception when a highly salient modulation parameter is changed. The present study does not directly evaluate computational audio-distance metrics, but it provides perceptual data and an experimental framework that could support such comparisons in future work.

E. Limitations

Several limitations constrain the interpretation of the findings. First, only six wavetables were used, so the results should not be generalised to the full space of synthesiser timbres. Second, the experiment tested specific ranges and implementations of modulation-signal amplitude, frequency, and irregularity; different modulation-signal ranges, wavetable mappings, or modulation-waveform structures may produce different perceptual relationships. The modulation dimensions also represent different physical quantities and parameter ranges, meaning that their numerical values are not directly comparable; comparisons between modulation types should therefore be interpreted in terms of their behaviour across the tested ranges rather than as equivalent changes in physical parameter magnitude.

Third, the experiment was conducted online using participants' own listening equipment, and although unsuitable playback conditions were screened, the acoustic environment could not be standardised fully. Finally, the low/high grouping used for the four-way ANOVA simplifies the underlying continuous progression of modulation amount and should be interpreted alongside the regression analysis in Section IV.C.

VI. CONCLUSION

This dissertation investigated how listeners perceive changes to the amplitude, frequency, and irregularity of a modulation signal used in controlled wavetable synthesiser sounds. A Python-based synthesis pipeline was used to generate stimuli across six wavetables

representing brightness, richness, and warmth, and perceptual difference was measured using a Web-MUSHRA listening experiment.

The results show that both modulation type and modulation amount substantially influenced perceived difference. Perceived difference increased approximately linearly as modulation amount increased. Changes to modulation-signal frequency generally produced the largest perceptual changes, changes to modulation-signal amplitude produced intermediate changes, and changes to irregularity produced the smallest. Timbral characteristic also influenced ratings, indicating that modulation perception depends partly on the sound to which the modulation is applied.

The study therefore demonstrates that modulation parameters are not perceptually interchangeable: comparable numerical parameter changes can produce different magnitudes of auditory change. These findings contribute to the understanding of time-varying synthesiser timbre and provide a foundation for perceptually informed parameter mappings, sound-design interfaces, and future models of perceptual similarity for modulated audio.

VII. FUTURE WORK

Future work should extend the experiment to a larger and more diverse set of wavetables, modulation rates, depths, and synthesis methods. Detection-threshold experiments could determine the point at which each modulation type first becomes perceptible, while perceptual-matching experiments could test whether different modulation types can be adjusted to produce equivalent perceived differences.

Participant metadata such as musical training and sound-design experience could also be analysed to determine whether expertise affects sensitivity to modulation. Laboratory replication would provide greater control over playback level and acoustic conditions.

A further extension is to compare the collected perceptual ratings with objective audio representations and distance measures, including spectral, scattering-based, and learned perceptual metrics (Bruna and Mallat, 2013; Yamamoto, Song and Kim, 2020; Manocha et al., 2020). Such analysis would directly address whether computational similarity measures capture the modulation characteristics that listeners perceive as salient, without treating that comparison as a completed contribution of the present dissertation.

APPENDIX

Shapiro–Wilk tests were used as a diagnostic check on the low- and high-modulation distributions. The results show multiple departures from normality, particularly within several irregularity conditions. The complete table is retained here for transparency rather than presented as a central perceptual finding.

TABLE V
SHAPIRO–WILK NORMALITY TEST RESULTS FOR LOW AND HIGH MODULATION AMOUNTS.

Amount	Modulation	Feature	Source	W	p
Low	Amp	Brightness	Real	0.970	0.038
Low	Amp	Brightness	Synthetic	0.955	0.003
Low	Amp	Richness	Real	0.930	< 0.001
Low	Amp	Richness	Synthetic	0.962	0.007
Low	Amp	Warmth	Real	0.965	0.017
Low	Amp	Warmth	Synthetic	0.945	< 0.001
Low	Freq	Brightness	Real	0.978	0.110
Low	Freq	Brightness	Synthetic	0.965	0.010
Low	Freq	Richness	Real	0.943	< 0.001
Low	Freq	Richness	Synthetic	0.983	0.238
Low	Freq	Warmth	Real	0.978	0.116
Low	Freq	Warmth	Synthetic	0.975	0.055
Low	Reg	Brightness	Real	0.934	< 0.001
Low	Reg	Brightness	Synthetic	0.901	< 0.001
Low	Reg	Richness	Real	0.829	< 0.001
Low	Reg	Richness	Synthetic	0.889	< 0.001
Low	Reg	Warmth	Real	0.872	< 0.001
Low	Reg	Warmth	Synthetic	0.836	< 0.001
High	Amp	Brightness	Real	0.912	< 0.001
High	Amp	Brightness	Synthetic	0.932	< 0.001
High	Amp	Richness	Real	0.967	0.016
High	Amp	Richness	Synthetic	0.961	0.007
High	Amp	Warmth	Real	0.961	0.009
High	Amp	Warmth	Synthetic	0.968	0.022
High	Freq	Brightness	Real	0.903	< 0.001
High	Freq	Brightness	Synthetic	0.924	< 0.001
High	Freq	Richness	Real	0.907	< 0.001
High	Freq	Richness	Synthetic	0.819	< 0.001
High	Freq	Warmth	Real	0.897	< 0.001
High	Freq	Warmth	Synthetic	0.931	< 0.001
High	Reg	Brightness	Real	0.979	0.163
High	Reg	Brightness	Synthetic	0.973	0.049
High	Reg	Richness	Real	0.983	0.289
High	Reg	Richness	Synthetic	0.974	0.067
High	Reg	Warmth	Real	0.963	0.025
High	Reg	Warmth	Synthetic	0.923	< 0.001

REFERENCES

- Bendor, D., Wang, X. and Bartlett, P.J. (2010) 'Neural coding of irregularity in marmoset auditory cortex', *Journal of Neurophysiology*, 103(4), pp. 1809–1822.
- Berg, J., Bustad, C., Jonsson, L., Mossberg, L. and Nyberg, D. (2013) 'Perceived Audio Quality of Realistic FM and DAB+ Radio Broadcasting Systems', *Journal of the Audio Engineering Society*, 61(10), pp. 755–777.
- Braun, D. (2021) 'DAWDreamer: Bridging the Gap Between Digital Audio Workstations and Python Interfaces', *Late-Breaking Demo, 22nd International Society for Music Information Retrieval Conference (ISMIR)*.
- Bruna, J. and Mallat, S. (2013) 'Invariant Scattering Convolution Networks', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), pp. 1872–1886.
- Chowning, J.M. (1973) 'The Synthesis of Complex Audio Spectra by Means of Frequency Modulation', *Journal of the Audio Engineering Society*, 21(7), pp. 526–534.
- Cook, P.R. (2002) *Real Sound Synthesis for Interactive Applications*. Natick, MA: A K Peters.
- Farnell, A. (2010) *Designing Sound*. Cambridge, MA: MIT Press.
- Hayes, B., Saitis, C. and Fazekas, G. (2022) 'Disembodied Timbres: A Study on Semantically Prompted FM Synthesis', *Proceedings of the Conference on AI Music Creativity (AIMC 2022)*.
- Henry, M. (2021) *A Perceptual Study of Amplitude Modulation Vibrato*. Master of Arts thesis, McGill University, Montréal, Canada.
- Hyrkäs, J. and Smyth, T. (2024) 'On Vibrato and Frequency (De)Modulation in Musical Sounds', *Proceedings of the 27th International Conference on Digital Audio Effects (DAFx-24)*, Guildford, UK.
- International Telecommunication Union (2015) *Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems (MUSHRA)*. Recommendation ITU-R BS.1534-3. Geneva: ITU.
- International Telecommunication Union (2023) *Recommendation ITU-R BS.1770-5: Algorithms to Measure Audio Programme Loudness and True-Peak Audio Level*. Geneva: ITU.
- Lee, H., Kim, K., Lee, S. and Lee, K. (2024) 'Wavespace: A Highly Explorable Wavetable Generator', *arXiv*, arXiv:2407.19862.
- Lee, J. and Bacon, S.P. (1997) 'Amplitude modulation depth discrimination of a sinusoidal carrier: Effect of stimulus duration', *Journal of the Acoustical Society of America*, 101(6), pp. 3688–3693.

- Manocha, P., Jin, Z., Hwang, Z., Finkelstein, A., Zhang, R., Bryan, N.J. and Kumar, G. (2020) 'A Differentiable Perceptual Audio Metric Learned from Just Noticeable Differences', *Proceedings of Interspeech 2020*, pp. 529–533.
- Martínez Ramírez, M.A., Benetos, E. and Reiss, J.D. (2019) 'A general-purpose deep learning approach to model time-varying audio effects', *Proceedings of the 22nd International Conference on Digital Audio Effects (DAFx-19)*, Birmingham, UK.
- McAdams, S. and Bigand, E. (eds.) (1993) *Thinking in Sound: The Cognitive Psychology of Human Audition*. Oxford: Oxford University Press.
- Peeters, G. (2004) *A large set of audio features for sound description (similarity and classification) in the CUIDADO project*. Technical Report. Paris: IRCAM.
- Sek, A. and Moore, B.C.J. (1995) 'Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type', *Journal of the Acoustical Society of America*, 97(4), pp. 2468–2478.
- Steinmetz, C.J. and Reiss, J.D. (2021) 'pyloudnorm: A Simple Yet Flexible Loudness Meter in Python', *Proceedings of the 150th Audio Engineering Society Convention*.
- Viemeister, N.F. (1979) 'Temporal modulation transfer functions based upon modulation thresholds', *Journal of the Acoustical Society of America*, 66(5), pp. 1364–1380.
- Yamamoto, R., Song, E. and Kim, J.-M. (2020) 'Parallel WaveGAN: A Fast Waveform Generation Model Based on Generative Adversarial Networks with Multi-Resolution Spectrogram', *Proceedings of ICASSP 2020*, pp. 6199–6203.
- Yutani, T., Yamamoto, Y., Nakatani, S. and Terasawa, H. (2024) 'Wavetable Synthesis Using CVAE for Timbre Control Based on Semantic Label', *Proceedings of APSIPA ASC 2024*.
- Zölzer, U. (ed.) (2011) *DAFX: Digital Audio Effects*. 2nd edn. Chichester: John Wiley & Sons.