TONESCAPING: CORRELATING ACOUSTIC MEASUREMENTS WITH PERCEIVED TONAL DIMENSIONS – AN OVERVIEW

C. Bowkett University of Portsmouth
L. Ausiello University of Portsmouth

1 INTRODUCTION

The nature of timbre has been discussed for decades, with some contributions presented more than a hundred years ago. Although refinements in understanding how timbre could be defined have populated literature on a regular basis, manufacturers and retailers of musical instruments introduced a parallel vocabulary and jargon, which is not in line with the knowledge gathered by either the electroacoustic or the musical acoustic community [1, 2, 3]. By solely focusing on acoustic and classical guitar manufacturers, it is possible to find several pages of nouns and adjectives related to tone and timbre, in which none of the suggested perceptual attributes are backed by scientific and psychoacoustic research [4, 5]. This article aims at filling a methodological gap by drawing from two established fields of acoustics, namely the quantitative measurements and estimators which became standard according to ISO 3382 [6], and the Soundscaping approach, which is described in the ISO 12913 [7, 8], thus creating a new field of knowledge that we call Tonescaping

The process, which goes towards the creation of a standardized framework and language to describe the tone of a musical instrument, consists of several steps covering the whole spectrum of quantitative measurements and qualitative dimensions that can be correlated to the determination and perception of tone. In line with [7, 9], the terms *feature* and *indicator* are interchangeably used to refer to quantitative data (e.g. recorded samples, impulse responses), while the nouns *attribute*, *descriptor*, and *dimension* are used as synonyms to refer to subjective and perceptual data (e.g. Loudness, Brightness, Fullness, etc.). Figure 1 shows how Tonescaping is nested between acoustics (engineering), auditory perception and cognition (psychology), and music (art & humanities).

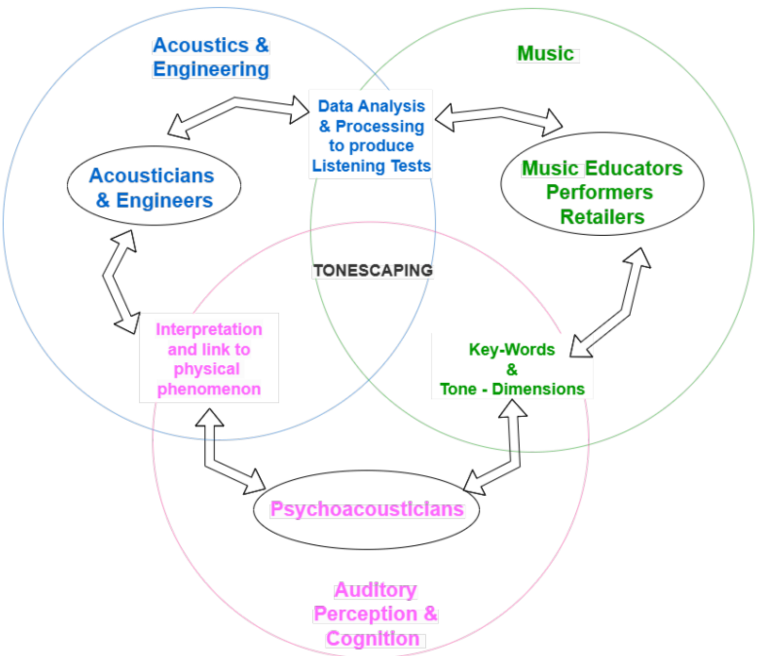

Figure 1: Three knowledge fields intersect to define Tonescaping as a new field of research.

This paper provides an overview of Tonescaping a as process. Section 2 recaps definitions of timbre, highlights one audio feature, which can be extracted from recordings of acoustic instruments (in specific guitars and pianos), and discusses how to prepare an objective estimator (cost function), in which several indicators are weighted. Section 3 shows how listening tests can be prepared to elicit an ordering function to be applied to recorded samples embedding different objective levels of a tonal dimension: *Brightness* will be used as an example. Section 4 describes future novel technologies which can be used to standardise the collection of samples of instruments. Section 5 summarizes the findings and suggests future research steps.

2 TIMBRE DEFINITION, OBJECTIVE FEATURES, AND BRIGHTNESS COST FUNCTION

2.1 Short review of timbre definitions

Timbre is an important aspect of human perception of sounds, and is a debated topic in psychoacoustics due to its multidimensional nature [10]. One of its earliest definitions was presented in 1885 by Hermann von Helmholtz, who stressed on the importance of harmonic content [11]. Building on this, Carl Seashore in 1938 highlighted the significant roles of harmonic structure and phase relationships [11]. Licklider expanded upon these ideas and advocated for a broader consideration of overtone complexity and its relationships with pitch and loudness [12].

Schouten proposed that timbre can be expressed through five key parameters: tonal to noise character, spectral envelope, temporal envelope, spectral evolution, and the initial transient (prefix) of the sound [13], thus highlighting the dynamic temporal characteristics of timbre, going beyond purely spectral definitions. The Acoustical Society of America defines timbre as "that attribute of auditory sensation which enables a listener to judge that two nonidentical sounds, similarly presented and having the same loudness and pitch, are dissimilar" [14]. This definition has been criticized with authors suggesting that “This is, of course, no definition at all” [15]. Recently, Grey emphasized timbre’s temporal dynamics, and reinforced the complexity of timbre, highlighting spectral centroid, attack time, and spectral flux as perceptually salient dimensions [10].

By reflecting on all these perspectives, this investigation adopts the following: *Timbre is a time-frequency auditory quality which allows us to distinguish between sounds with identical pitch and loudness; it is influenced by both spectral features (e.g. harmonic structure, spectral distributions), and temporal features (e.g. attack, sustain and decay), and the dynamic evolution of the relationship of such features over time.*

2.2 Objective measurement: an initial setup

Tonescaping entails quantifying timbre first, which means extracting audio features from Impulse Responses (**IR**) and audio samples; those indicators must then be combined in a cost function weighting how important specific features are (thus becoming *instrument dependent* for example). The last step is the correlation between the cost function and the perceptual dimensions, which is validated by mean of listening tests. This must be done for each tonal descriptor, hence for example acoustic guitar *Brightness* having a specific cost function, different from acoustic guitar *Punch*, or piano *Brightness*.

We assume that robust methods to objectively measure an acoustic instrument are available. Literature is vast, but referring to acoustic guitars Ausiello et Alii presented several papers in which IRs of acoustic guitars can be obtained by using small exciters and sine sweep signals (Farina’s method) [16, 17, 18], while other methods are typically used for violins and pianos (see [16] and references therein). Other approaches involve the use of recordings performed within anechoic rooms during which professional musicians are asked to repeat chord sequences, or specific parts, trying their best to keep tempo and loudness constant [19].

For the purpose of this introduction the authors assume that both IRs and chords samples are available in the form of high-quality audio files (at least 24bit 48kHz). Chords are saved with normalised amplitude and duration, while IRs are recorded with a standardised power used to excite the instruments and a fixed gain in the recording chain to allow relative loudness comparisons.

From the corresponding audio files, standardized signal processing algorithms contained in the Timbre Toolbox [20], and MIRToolbox [21] become useful tools to extract comprehensive features categorized into temporal, spectral and harmonic domains. The Timbre toolbox notably differentiates between global features, which provide a single value for an entire sound event, or time-varying features, providing a value that changes over time. The latter ones help to enable dynamic representation of timbral qualities.

In general, in order to study the impact on a perceived tonal dimension of playing intensity, technique, and position, the selected instruments can be recorded under a plethora of conditions, as shown in Table 1 and 2 in the case of guitar and piano, for example.

Variable	Conditions
Playing Intensity	Soft, Medium, Hard
Playing Style	Fingerstyle, Pick
Pickup Position	Neck, Bridge
Chord Selection	A, A Minor, C, D, D Minor, E, E Minor, G

Table 1: Recording Conditions for a guitar Tonescaping study

Variable	Conditions
Playing Intensity	Soft, Medium, Hard
Playing Style	Muted, Normal
Chord Selection	A, A Minor, C, D, D Minor, E, E Minor, G

Table 2: Recording Conditions for Piano Tonescaping study

In post-production, the levels of all samples can be normalised to a pre-defined level (e.g. -14 LUFS), thus allowing fair comparison of timbral features without loudness bias. By controlling these variables and recording techniques, the experiment ensures that any differences in audio features are attributed to the intended changes in playing style and position rather than other factors.

2.3 Audio features extractions and analysis

As mentioned before, a wide range of audio features can be derived to characterise the tonal qualities of an instrument. A summary of the capabilities of the Timbre Toolbox [20] is presented in Table 3.

	Audio descriptor	Units	Abbreviation	Input representation
Global descriptors	Attack	s	Att	Temporal Energy Envelope
	Decay	s	Dec	
	Release	s	Rel	
	Log-Attack Time	log(s)	LAT	
	Attack Slope	a/s	AttSlope	
	Decrease Slope	$\log(a)/s$	DecSlope	
	Temporal Centroid	s	TempCent	
	Effective Duration	s	EffDur	
	Frequency of Energy Modulation	Hz	FreqMod	
	Amplitude of Energy Modulation	a	AmpMod	
Time-varying descriptors	Autocorrelation (12 coefficients)	-	AutoCorr	Audio Signal
	Zero Crossing Rate	s^{-1}	ZcrRate	
	RMS-Energy Envelope	a	RMSEnv	Temporal Energy Envelope
	Spectral Centroid	F	SpecCent	STFTmagnitude (STFTmag) STFTpower (STFTpow) ERBfft (ERBfft) ERBgammatone (ERBgam) Harmonic
	Spectral Spread	F	SpecSpread	
	Spectral Skewness	-	SpecSkew	
	Spectral Kurtosis	-	SpecKurt	
	Spectral Slope	F^{-1}	SpecSlope	
	Spectral Decrease	-	SpecDecr	
	Spectral Rolloff	F	SpecRollOff	
	Spectro-temporal variation	-	SpecVar	
	Frame Energy	I	FrameErg	
	Spectral Flatness	-	SpecFlat	STFTmag, STFTpow, ERBfft, ERBgam
	Spectral Crest	-	SpecCrest	
	Harmonic Energy	a^2	HarmErg	Harmonic
	Noise Energy	a^2	NoiseErg	
	Noisiness	-	Noisiness	
	Fundamental Frequency	Hz	F0	
	Inharmonicity	-	InHarm	
	Tristimulus (3 coefficients)	-	TriStim	
	Harmonic Spectral Deviation	a	HarmDev	
	Odd to even harmonic ratio	-	OddEveRatio	

Table 3: Summary of Audio Features found in literature [20]

However, not all indicators are necessary to model each specific perceptual dimension. The challenge then lies in identifying which features are the most informative for any given tonal descriptor. Accordingly, based on existing literature, a selection of features can be extracted and combined to form a **cost function**. This can be used to empirically represent a specific tonal dimension of an instrument such as a guitar or a piano. Also, features can overlap or describe similar aspects of the sound, so to reduce complexity, an optimised cost function should use a minimum number of audio features.

As an example, Attack is used here to show how quantitative measurements capture different aspects of playing style, which end up having an impact on the perception of tone. Attack is defined as per Table 3, and corresponds to the duration it takes for a note's amplitude to rise from its onset to its peak value. It is normally measured in seconds, and is a useful indicator to quantify the initial properties of a sound in terms of tone [26, 27].

The attack time of a guitar chord sampled and recorded can be extracted using the Timbre Toolbox (using the temporal energy envelope representation, HARMrep). Table 4 and figure 2 summarise the measured attack times for an A Major Chord played in two different position (neck and bridge) with two different styles (fingerpicking and strumming).

Recording Condition	Attack Time (s)
A Bridge Pick	0.2093
A Bridge Fingers	0.3152
A Neck Pick	0.1974
A Neck Fingers	0.2682

Table 4: Attack Time Values of A Chord Recordings

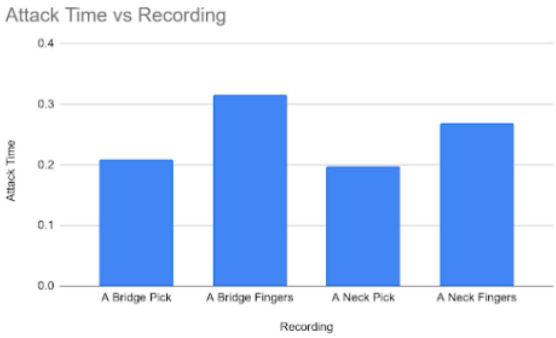

Figure 2: Bar chart showing attack times for A chord recordings.

Notes being played with a pick have consistently shorter attack times than those played with fingers. A faster and steeper attack contribute to the perceptual difference in tone [27].

2.4 An example of Tonescaping Cost Function: Brightness

In this section the subjective descriptor *Brightness* is used as an example of cost function resulting from a combination of multiple audio features; this comes from the understanding that *Brightness* is multidimensional as discussed in sec. 2.1. Figure 3 shows how temporal, spectral and harmonic features are extracted, normalised and combined into a feature matrix, which can then be used to train a linear regression model.

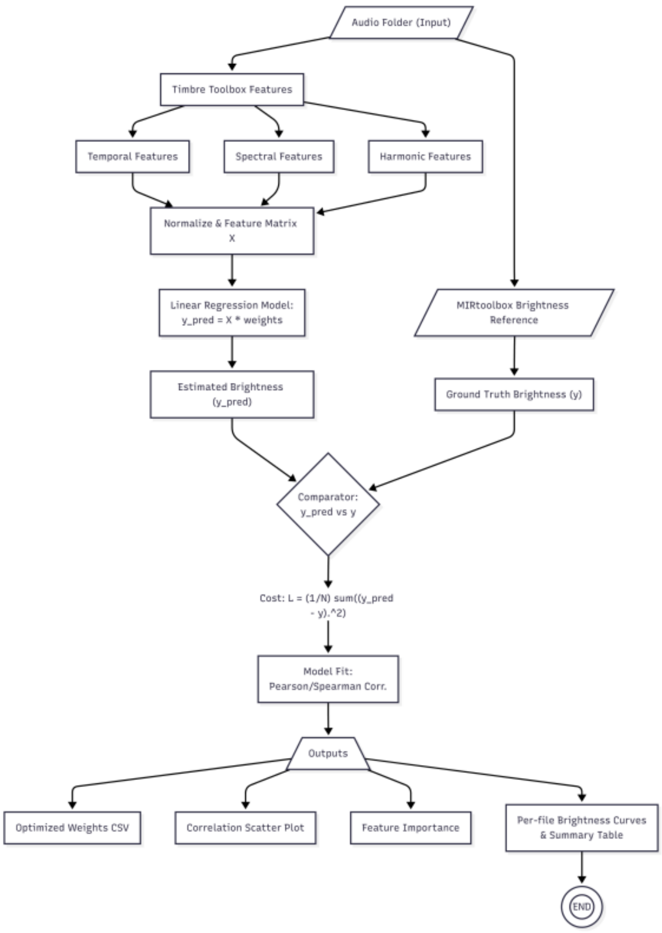

Figure 3: flow chart of the creation of a Tonescaping cost function (*Brightness*)

3 LISTENING TESTS AND ORDERING METHODOLOGY

The method suggested to validate the correlation between a cost function modelling a tonal descriptor on the "objective side", with the perceptual dimension of timbre on the "subjective side" is via listening tests. These tests aim at gathering human judgements that can be compared statistically with estimations obtained from a cost function as introduced in 2.4. Participants are needed, and statistically significant numbers are necessary. Amongst the participants there must be a mix of casual listeners, musicians, audio professionals, and all the potential stakeholder involved in perceiving or dealing with a musical instrument and its tone.

In the case of acoustic guitars for example, the tests must be circulated amongst designers and manufacturers, retailers, demonstrators, educators, professional and amateur musicians, and general public, who'd most certainly be part of an audience when such instrument is played. Keeping the focus on acoustic guitar tone, and referring to section 2.2 and 2.4, participants are to be presented with a page containing multiple samples, e.g. 4, under which a slider is located (see figure 4).

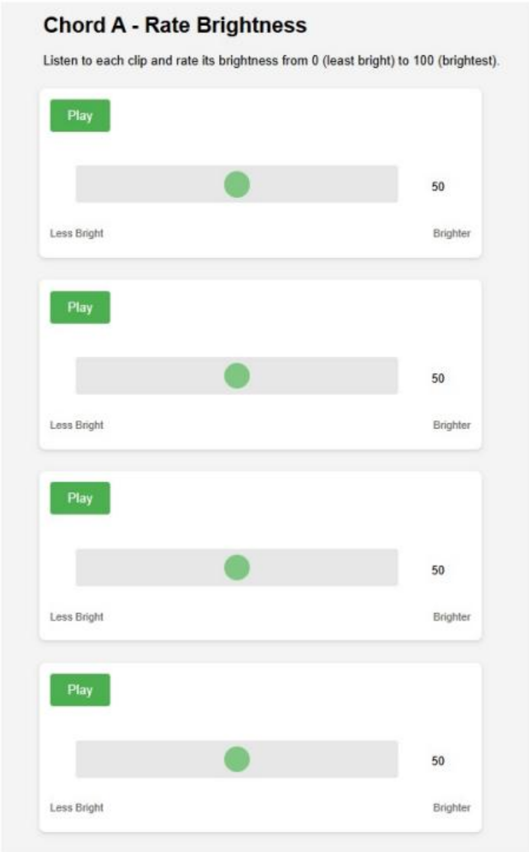

Figure 4: Layout for Tonescaping listening test page. All samples are the same chord, presented at the same loudness (normalised LUFS), while containing differences with respect to other possible variable, such as playing style or position of playing, in the case of an acoustic guitar experiment.

Each sample contains the same chord; participants can play one sample at the time, as many times they want. The samples in a page differ along one or more of the remaining variables, i.e. playing style (fingers or pick), position (bridge or neck), and original playing intensity (soft, medium, hard) despite the fact that when played back all samples are normalised in Loudness to a specific LUFS.

Participants are asked to use the slider located below each sample in such way to order them according to the perceived magnitude of a tonal dimension, which in this example corresponds to *Brightness*. In order not to confuse participants on the semantic, the slider shows a scale ranging from [0 - 100], and the two extremes are labelled "less bright" to "brighter".

The responses of each participant are stored and aggregated to derive a perceptual ordering function, representing the average human assessment of the tonal attribute under investigation. This helps providing a perceptual reference that can be compared to the cost function. Using statistical analysis techniques such as ANOVA, the relationships between objective features and perceptual dimension can be identified. Figure 5 summarises all phases of Tonescaping.as a process.

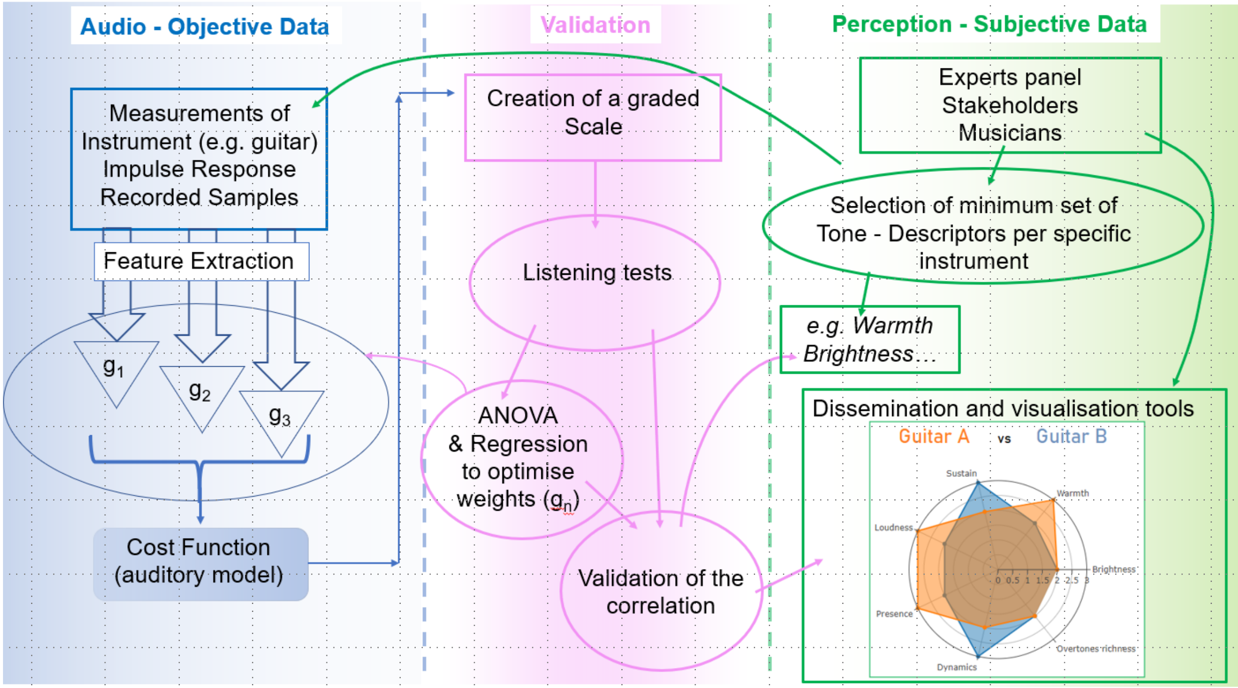

Figure 5: Tonescaping as a process. Multiple interactions and relationships are visible amongst different stakeholders and research fields.

4 NOVEL TECHNOLOGY FOR REPEATABLE INSTRUMENT SAMPLING

Let's reflect on Tonescaping as a process; section 2.3 suggests that despite the nature of the instrument, audio features can be extracted from audio samples and recordings. Extreme care and detail must be spent to gather the initial audio data. Accordingly, some limitation of the initial setup discussed in section 2.2 can be discussed: a musician is required to sample an instrument, although this seems to involve some unavoidable level of uncertainty and lack of repeatability in the process. To address the latter in particular, the use of robots and robotic arms appears to be a viable option. For example, the use of silent robotic arms attached to a head-and-torso simulator for example, would provide researchers with two benefits.

Firstly, repeatable movements can be programmed and executed as many times as needed without imposing stress and discomfort on players participating a study, secondly the opportunity to record the instrument's output also from the musician perspective via the binaural recording (and not only from the audience point of view) opens new possibilities in the investigation of tone and its perception. The problem though seems merely shifted to "*how to program a robotic musician*"? Despite the progress of contemporary robotics, the authors could not find an example of silent and proficient robotic arm or hand capable of replicating the nuances and skills of humans when playing an instrument. Nonetheless, considering that Tonescaping is aimed at providing a standardised timbre vocabulary rooted in quantitative measurements, we can accept that some aspects of playing can be approximated and standardised. In the case of a guitar, for example, it seems reasonable to start collecting samples of strummed chords, thus nor requiring fingers articulation, or movements. For pianos, one or two arms can be given the task of pushing keys on the keyboard by mean of pre-shaped planks which will hit only certain notes, hence playing standardised voicing of a certain number of chords (the reader can get a glimpse of this idea by watching this video at about 1:05 - 1:35 - *https://www.youtube.com/watch?v=ifKKlhYF53w*).

Even accepting such demanding restrictions, how to program the moving arms and hands remains a complex task. Commercial technology used to capture motion for creating animation and video-games is available, which includes special gloves to record hands' movements. By adding accelerometers, gyroscopes, velocity and position sensors, a new type of "musical glove" could be designed in such way to record all the physical quantities needed to program a robotic substitute.

This would also open the chance to record musicians whose talent or "tone" is coveted and described as inspiring, which happens to be the case in the guitar world for example. The authors tentatively experimented with a simplified robot holding a pick and strumming an acoustic guitar on a test bench. No substantial differences in the audio files produced by a real professional musician could be detected in the initial listening test, but for the presence of higher background noise caused by the cooling fan of the robotic arm. This is encouraging, and suggests the use of motion sensors applied to human hands and arms as part of the Tonescaping research.

5 CONCLUSIONS

This article presented a summary of a new field of research aimed at correlating quantitative measurements of musical instruments, in the form or impulse responses and recorded samples, with the perceptual space of their timbre, in the form of graded scales describing tone dimensions. This new approach is called Tonescaping. The authors hope to see it becoming a new research topic leading to a standardized vocabulary of musical instruments' timbre similar to the acoustical parameters found in the ISO 3382 norm used to characterise acoustic spaces. The paper introduces ideas to prepare a psychoacoustical model containing several weighted audio features extracted from audio recordings, which is called a cost function. A method to prepare listening tests and use them to correlate a Tonescaping cost function with a perceptual graded scale is also presented.

Acknowledgments
The authors want to thank Davide Tomassone of Tomassone S.r.l for his indispensable contribution and ideas about Tonescaping as a concept. Also, special thanks to Charalampos Saitis, Hyunkook Lee, Chris Barlow, and Sebastian Duran for the ongoing discussions about the definition of tone, its correlation with psychoacoustical models, and Tonescaping terminology.

REFERENCES

[1] Premier Guitar. Tone terminology, 2007.
[2] Breedlove Guitars. Tone glossary, 2019.
[3] Yamaha Corporation. Acoustic guitar tonality, 2022.

[4] Martin & Co. Acoustic guitar wood types: A guide, 2025.
[5] Taylor Guitars. Taylor tonewoods, 2021.
[6] ISO. 3382-2:2008, 06 2008.
[7] ISO. 12913-1:2014 acoustics — soundscape, 09 2014.
[8] R Murray Schafer. The Soundscape : the Tuning of The World. Destiny Books, USA, 1994.
[9] Francesco Aletta, J. Kang, and Ö. Axelsson. Soundscape descriptors and a conceptual framework for developing predictive soundscape models. Landscape and Urban Planning, 149:65–74, 2016.
[10] John M. Grey. Multidimensional perceptual scaling of musical timbres. The Journal of the Acoustical Society of America, 61:1270–1277, 05 1977.
[11] Carl E Seashore. Psychology of Music. Dover Publications, 1967.
[12] J. C. R. Licklider. A duplex theory of pitch perception. Experientia, 7:128–134, 04 1951.
[13] J.F Schouten. The Perception of Timbre. The 6th International Congress on acoustics, Tokyo, Japan, 1968.
[14] Ansi/asa s1.1-2013 (r2020) - acoustical terminology, 2020.
[15] Albert S Bregman. Auditory Scene Analysis : the Perceptual Organization of Sound. Cambridge, Mass. Mit Press, 1994.
[16] Ludovico Ausiello, Michele Ducceschi, Sebastian Duran, and Benjamin Morrison. Affordable wide-band measurement ecosystem for musical acoustics based on electro-dynamic transducers. Acta Acustica, 8:53– 53, 01 2024.
[17] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In Proceedings of the AES 108 Convention, pages 1–24, Paris, France, 19-22 February 2000.
[18] A. Farina. Advancements in impulse response measurements by sine sweeps. In Proceedings of the AES 122 Convention, pages 1–21, Vienna, Austria, 5-8 May 2007.
[19] Sebastian Merchel, M. Ercan Altınsoy, and David Olson. Perceptual evaluation of bracewood and soundboard wood variations on the preference of a steel-string acoustic guitar. The Journal of the Acoustical Society of America, 146:2608–2618, 10 2019.
[20] VincentPerreault0. Github - vincentperreault0/timbretoolbox: A toolbox for extracting audio descriptors in matlab., 2018.
[21] Mirtoolbox - file exchange - matlab centralfile exchange - matlab central, 12 2011.
[22] Emery Schubert and Joew Olfe. Does timbral brightness scale with frequency and spectral centroid?, 2006.
[23] Savvas Kazazis, Philippe Depalle, and Stephen McAdams. The timbre toolbox user's manual, 2022.
[24] Aidan. Pick attack: what is it & why does it matter? - happy bluesman, 10 2018.
[25] Sandra Carral. Plucking the string: The excitation mechanism of the guitar.
[26] Paweł Bielski and Marcin Kujawa. Nonlinear modelling in time domain numerical analysis of stringed instrument dynamics. AIP Conference Proceedings, 1863:020003–020003, 01 2017.
[27] John W. Gordon. The perceptual attack time of musical tones. The Journal of the Acoustical Society of America, 82(1):88–105, 07 1987.