A Psychophysical Investigation of Timbre

Mark R. Hurst ’24 and Marcus H. Pendergrass

Department of Mathematics and Computer Science, Hampden-Sydney College, Hampden-Sydney, VA 23943

Abstract

A study was conducted in an effort to relate descriptions of perceived timbre with characteristic patterns of harmonics found through the Fourier transform with a focus on the terms Bright, Warm, and Hollow. A listening survey was conducted with a response section of 28 stimuli in which participants ranked each of the three terms (Bright, Warm, and Hollow) by how well they describe the timbre of the stimulus. Participant responses were analyzed by comparing them to their distance from a predicted response for each stimulus which is based on its power spectrum. Bright timbres had the most agreement between participant responses and the predicted responses, while Hollow timbres had the least agreement. However, the results were inconclusive due to confounding variables in the survey which are the result of the limitations imposed by conducting the survey online.

Introduction

Timbre is defined as the quality of sound aside from pitch, duration, and volume. Among musicians or informed listeners, timbre is usually described metaphorically using terms like ‘bright’ and ‘warm.’ Such terms still allow for accurate descriptions and an understanding of timbre among informed listeners despite their subjective nature (Pendergrass 2014). In searching for an objective method of describing timbre, multiple studies have found that timbre is a multidimensional attribute of sound best described using three dimensions; however, the characteristics of those dimensions are still a point of contention. Zacharakis (2011) found three timbral dimensions associated with the semantic descriptors of “Volume/Wealth, Brightness-Density, and Texture/Temperature” through factor analysis of survey data. Grey (1977) used a multidimensional scaling algorithm for survey results based on the pairwise relationships of a set of stimuli, and the stimuli were found to be separated in dimensions related to the attack or the distribution of energy over time. Rather than modeling timbre as a space with multiple dimensions, Pendergrass (2014) modeled individual timbral qualities as partial orders that accounted for potential disagreement among listeners (Pendergrass 2014). This use of partial orders resulted in an ordering of stimuli relative to a how much of a specific timbral quality they possessed, and this model also allowed for the analysis of different qualities of timbre.

Concepts from these previous works and other literature on timbre motivated the approach of this study. Firstly, there are words to describe timbre that are distinct in meaning relative to others as seen by the three separate semantic dimensions described by Zacharakis (2011), and this idea influenced the choice of the three descriptors used in this study: Bright, Warm, and Hollow. Secondly, the qualities of the perceived timbre of a sound can be associated with its energy spectrum as seen from the common characteristics of energy distribution among dimensions from Grey (1977). Thirdly, Pendergrass (2014) showed mathematically that sounds could be quantified based on how well the timbre represents a specific timbral quality relative to other sounds, or, in other words, a sound could be described to represent different timbral qualities in varying magnitudes. For this study, spectral analysis of sound waves through the Fourier transform was used to classify sounds based on how well they represent each of the timbral qualities Bright, Warm, and Hollow. These classifications based on spectral analysis were then compared to the classifications from a set of listeners by survey.

Fourier analysis is based on the applied concepts of the Riesz-Fischer Theorem and Carleson’s Theorem. Together, these theorems imply that any physically realizable periodic waveform is completely described by its Fourier series. For a Fourier analysis, the sound being analyzed is broken down into its component frequencies, each associated with a specific amplitude and phase, in such a way that summing these components reconstructs the original sound. These component frequencies are known as “overtones,” and there are special overtones called “harmonics” that are integer multiples of the fundamental frequency. These harmonics are frequencies with higher amplitudes or more energy than the other overtones; thus, the classifications used in this study were derived from the presence of specific patterns in these harmonics seen through the power spectrum of the sound.

Bright timbres were associated with the second or third harmonic having a higher amplitude than the fundamental or an increased influence of higher harmonics. Warm timbres were associated with a decreased influence of higher harmonics. Hollow timbres were associated with a section of harmonics that starts and ends with harmonics that have a greater amplitude than the harmonics in between them.

Procedure

Participants were told to rank the three terms Bright, Warm, and Hollow by how well they described the timbre of each sound sample; thus, each participant's answers were collected as a permutation of Bright, Warm, and Hollow. These permutations from the participant answers were compared with a predicted permutation based on the same classifications using the power spectrum of each stimulus. The data from the participant responses was then categorized into distances from the predicted permutation based on similarity. With six different possible permutations of Bright, Warm, and Hollow, the distances went from 0 to 5 with 0 being the same permutation as the predicted response and 5 being the permutation that is the reverse or opposite of the predicted response.

Materials and Methods

Stimuli

The set of 34 stimuli used in this study consisted of acoustic samples and synthesized sounds, and each of the stimuli were at A4 (440 Hz) with a normalized amplitude.

9 acoustic samples were used in the survey: Trumpet, Bass Clarinet, Oboe, French Horn, Piano, Tenor Trombone, Eb Clarinet, Flute, and Soprano Saxophone.

Of the synthesized sounds, 12 were synthesized in MATLAB with the intent of recreating the power spectra of acoustic samples. These synthesized sounds are designated by the prefix "Im_" followed by the acoustic instrument it aims to recreate.

3 of the synthesized sounds were an ideal Sawtooth wave, Square Wave, and Triangular Wave.

8 stimuli (Bright_1, Bright_2, Hollow_1, Hollow_2, rearrangeHarmonic_1, rearrangeHarmonic_2, Warm_1, and Warm_2) were synthesized in pairs using 4 MATLAB scripts with adjustable parameters for different patterns of manipulating Fourier Coefficients. The patterns associated with three of the four scripts were related to a different classification of timbre, and the stimulus labeled "1" in each of these pairs was determined to better represent its corresponding timbre classification. For the last pair (rearrangeHarmonic_1 and rearrangeHarmonic_2), the pattern used did not appeal directly to the pattern of harmonics associated with either Bright, Warm, or Hollow.

The last 2 stimuli (exampleHollow_5 and exampleWarm_2) were synthesized using patterns of Fourier coefficients that were not derived from the manipulation of parameters.

Participants

26 subjects participated in the listening test. The selection of participants was limited to those with at least two years of some manner of formal musical experience be it performing, teaching, or learning. This limitation was put in place to ensure that the listeners were familiar with listening critically to musical tones. The participants had musical experience that ranged from 5 to 55 years with an average of 23 years. The survey was conducted online, but every participant was requested to use headphones or earbuds before participating.

Listening Survey

The listening test consisted of an introductory section and a response section. The introductory section used 9 acoustic samples with the purpose of familiarizing the participants with the three classifications of timbre used in the survey: Bright, Warm, and Hollow. The samples in this introductory section were classified as follows: Trumpet, Bass Clarinet, and Oboe as Bright; French Horn, Piano, and Tenor Trombone as Warm; and Eb Clarinet, Flute, and Soprano Saxophone as Hollow. The response section consisted of a total of 28 sounds presented in a randomized order for each participant. All 25 of the synthesized stimuli described earlier were used, and 3 of the acoustic samples (Bass Clarinet, French Horn, and Eb Clarinet) were included as well to act as a control for each of the three timbre classifications used in the study.

Results

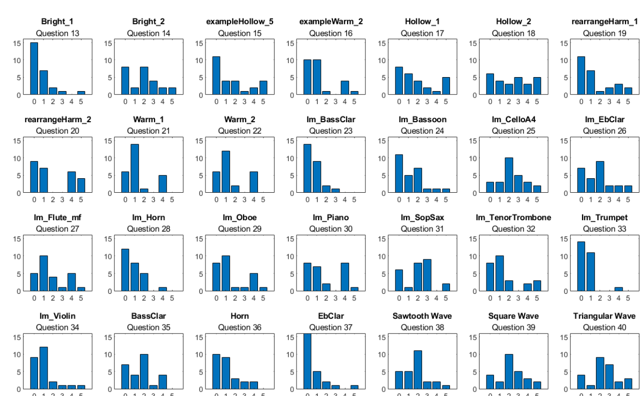

Figure 1: Number of Responses per Response Distance

Question #	SoundSample	First	Second	Third
13	Bright_1	Bright	Warm	Hollow
14	Bright_2	Bright	Warm	Hollow
15	exampleHollow_5	Hollow	Warm	Bright
16	exampleWarm_2	Warm	Bright	Hollow
17	Hollow_1	Hollow	Warm	Bright
18	Hollow_2	Hollow	Bright	Warm
19	rearrangeHarm_1	Bright	Warm	Hollow
20	rearrangeHarm_2	Bright	Warm	Hollow
21	Warm_1	Warm	Bright	Hollow
22	Warm_2	Warm	Bright	Hollow
23	Im_BassClar	Bright	Hollow	Warm
24	Im_Bassoon	Warm	Hollow	Bright
25	Im_CelloA4	Warm	Bright	Hollow
26	Im_EbClar	Hollow	Warm	Bright
27	Im_Flute_mf	Warm	Hollow	Bright
28	Im_Horn	Warm	Hollow	Bright
29	Im_Oboe	Bright	Warm	Hollow
30	Im_Piano	Warm	Bright	Hollow
31	Im_SopSax	Hollow	Bright	Warm
32	Im_TenorTrombone	Warm	Bright	Hollow
33	Im_Trumpet	Bright	Warm	Hollow
34	Im_Violin	Bright	Hollow	Warm
35	BassClar	Bright	Hollow	Warm
36	Horn	Warm	Hollow	Bright
37	EbClar	Hollow	Warm	Bright
38	Sawtooth Wave	Bright	Warm	Hollow
39	Square Wave	Hollow	Bright	Warm
40	Triangular Wave	Hollow	Warm	Bright

Figure 2: Table of Predicted Responses. Timbral qualities (Bright, Warm, and Hollow) are listed for each stimulus as either First, Second, or Third with First being the most prevalent timbral descriptor for the stimulus and Third being the least.

Question #	SoundSamples	0 to 1	2 to 3	4 to 5
13	Bright_1	84.6%	11.5%	3.8%
14	Bright_2	38.5%	46.2%	15.4%
15	exampleHollow_5	57.7%	19.2%	23.1%
16	exampleWarm_2	76.9%	3.8%	19.2%
17	Hollow_1	53.8%	23.1%	23.1%
18	Hollow_2	38.5%	30.8%	30.8%
19	rearrangeHarm_1	69.2%	11.5%	19.2%
20	rearrangeHarm_2	61.5%	0.0%	38.5%
21	Warm_1	76.9%	3.8%	19.2%
22	Warm_2	69.2%	7.7%	23.1%
23	Im_BassClar	88.5%	11.5%	0.0%
24	Im_Bassoon	61.5%	30.8%	7.7%
25	Im_CelloA4	23.1%	57.7%	19.2%
26	Im_EbClar	42.3%	42.3%	15.4%
27	Im_Flute_mf	57.7%	19.2%	23.1%
28	Im_Horn	76.9%	19.2%	3.8%
29	Im_Oboe	69.2%	7.7%	23.1%
30	Im_Piano	57.7%	7.7%	34.6%
31	Im_SopSax	26.9%	65.4%	7.7%
32	Im_TenorTrombone	69.2%	11.5%	19.2%
33	Im_Trumpet	96.2%	0.0%	3.8%
34	Im_Violin	80.8%	11.5%	7.7%
35	BassClar	42.3%	42.3%	15.4%
36	Horn	73.1%	19.2%	7.7%
37	EbClar	84.6%	11.5%	3.8%
38	Sawtooth Wave	38.5%	50.0%	11.5%
39	Square Wave	23.1%	57.7%	19.2%
40	Triangular Wave	19.2%	61.5%	19.2%

Figure 3: Percentage of Participant Responses per Distance Range.

For participant answers between distances 0 and 1, the participant agreed with the term ranked first in the predicted ranking. For distances 2 to 3, the participant thought that the second-ranked term in the predicted response was the most important. For distances 4 to 5, the participant thought that the third-ranked term in the predicted response was the most important. For 19 of the 28 stimuli, a majority of participants answered within distances 0 and 1. For 5 of the 28 stimuli, a majority of participants answered within distances 2 and 3.

The stimuli with an average distance less than 1 include Bright_1 (Avg = 0.7307), Im_Bass Clarinet (Avg = 0.6153), Im_Horn (Avg = 0.8461), Im_Trumpet (Avg = 0.5769), and Eb Clarinet (Avg = 0.6538). The stimuli with an average distance greater than 2 include Hollow_2 (Avg = 2.3846), Im_CelloA4 (Avg = 2.3076), Im_Soprano Saxophone (Avg = 2.0769), Square Wave (Avg = 2.2692), and Triangular Wave (Avg = 2.4230). The predicted response for the Im_Trumpet had the most accuracy with 96% of participants responding within distances 0 and 1, and the Im_Trumpet had the lowest average response distance among all of the stimuli (Avg = 0.5769). The predicted response for the Triangular Wave had the least accuracy with the highest average response distance among all of the stimuli (Avg = 2.4230).

Discussion

For the stimuli with the least predicted response accuracy with an average response distance greater than 2, all except Im_CelloA4 were predicted to have “Hollow” ranked in the first position. The average response distance greater than 2 shows that participants did not agree with the classification of “Hollow” used in the introductory section.

Each of the timbre classifications are represented in the stimuli with the most predicted response accuracy or those with an average response distance less than 1. Four of the five stimuli with an average distance less than 1 were synthesized except for the acoustic Eb Clarinet sample. Despite the predicted response accuracy seen with the Eb Clarinet, there is insufficient evidence to validate the classification of “Hollow” used in the study.

For the three acoustic samples with their corresponding synthesized imitation (Bass Clarinet, French Horn, and Eb Clarinet), the synthesized imitations of the Bass Clarinet (Im_BassClar) and the French Horn (Im_Horn) have a lower average response distance than their corresponding acoustic samples. The opposite is true for the Eb Clarinet pair in that the acoustic sample has a lower average response distance than its corresponding synthesized imitation (Im_EbClar). The stimulus with the lower average response distance in each pair has an average distance less than one, which supports the accuracy of the predicted results. The parity in the predicted response accuracy between the imitation and acoustic stimuli, however, questions the validity of the classifications used in the introductory section of the survey.

In future experiments, an algorithm or script should be used to classify and predict responses for the stimuli. Despite identifying a pattern of harmonics to classify Bright, Warm, and Hollow timbres, the method used to classify the stimuli was influenced by subjectivity, especially for ranking the three timbre classifications. The audio quality and fidelity of the participants' devices also presented confounding variables due to conducting the survey online, and audio quality and fidelity are especially important when differences in timbre could be subtle. Another limitation of this study is in the validity of the classifications in the introductory section. If all of the acoustic samples used in the introductory section were to be included in the response section, then it would allow for more conclusive analysis between the imitation and acoustic stimuli.

Overall, the classifications of Bright, Warm, and Hollow timbres used were fairly accurate in predicting participant answers with 19 of the 28 stimuli having a majority of participant responses between distances 0 and 1. Out of the three classifications (Bright, Warm, and Hollow), Hollow had the least agreement among participants with all of the stimuli classified as mostly Hollow except for the acoustic Eb Clarinet stimulus having an average distance greater than 1.5. On the other hand, the predicted responses for Bright and Warm were more accurate. Due to limitations in conducting the survey and confounding variables, there is insufficient evidence to draw conclusions about the validity of the classifications used in the study. However, the results of this study do show that patterns of harmonics seen through the Fourier transforms of sounds are a promising method of objectively defining dimensions of timbre.

REFERENCES

[1] Zacharakis, K. Pastiadis, G. Papadelis, & J.D. Reiss (2011). An investigation of musical timbre: Uncovering salient semantic descriptors and perceptual dimensions. In *12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, pages 807-812, Miami, USA, 24-28 October 2011.

[2] Grey, J.M (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, 61: 1270-1277, 1977

[3] Pendergrass, M (2014). Two musical orderings. *Journal of Mathematics and Music: Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, 8:1, 59-72, DOI: 10.1080/17459737.2013.871357