DIFFERENCE IN POLYPHONIC TIMBRE PERCEPTION BETWEEN MUSICIANS AND NON-MUSICIANS

Vivek Mathur [1] **Shri Vidhatri M M** [2] **Ayappa Siddik Naraharisetti** [3]
Vinoo Alluri
[1] 2020113002, CCNSB, IIIT-H
[2] 2019113006, CCNSB, IIIT-H
[3] 2020101089, MLL, IIIT-H
vivek.m@research.iiit.ac.in, shrividhatri.m@research.iiit.ac.in
siddik.naraharisetti@students.iiit.ac.in

1. INTRODUCTION

Polyphonic timbre, which refers to the overall timbral mixture in a music signal, is essential to understanding sound quality. Timbre is essential to music genre classification, generation and many music-related applications.

Unlike pitch, loudness or rhythm, timbre is more perceptual than quantitative. Due to this perceptual nature, timbre can be understood and analysed using different descriptors, constituting a multidimensional timbre space. Polyphonic timbre is the understanding of this multidimensional timbre space.

Learning music is believed to affect people in a certain way regarding aspects such as memory, reasoning, visualisation and many more. Thus, it is a reasonable hypothesis that the minds of musicians analyse music differently from non-musicians. It can also be hypothesised that musicians have a higher value of inter-category correlation than non-musicians since the learning of music(instrumental or vocal) trains people to listen to sounds universally.

2. LITERATURE REVIEW

(Alluri,2010) paper regarding perceptual and acoustic correlates of timbre is one of the primary papers on which this study is based. This paper details an experiment designed to make people rate Indian musical excerpts on perceptual descriptors. Later, the rating is correlated with acoustic features to identify the spectral features that best describe timbre.

The experiment described in this paper utilises 12 bipolar scales of perceptual descriptors, which were then narrowed down to 8 scales based on inter-scale correlation values. The final eight values selected in the experiment were those with the least squared multiple correlations, which are mathematically understood to be the most unique based on the data.

The participants were then made to listen to 100 excerpts of songs and rate them in these eight perceptual scales. The results of these ratings were analysed for their Cronbach alpha values for consistency and mean inter-subject correlation values for then independency.

The inter-subject correlations were also used to group the perceptual descriptors together to understand what real value in the quality of sound they refer to, such as activity or fullness.

The excerpts were then analysed computationally for features such as zero-crossing rate, spectral centroid, high energy-low energy ratio, entropy, spectral roll-off, MelFrequency Cepstral Coefficients, sub-band flux and roughness to be extracted.

The perceptual correlates were then compared with the acoustic features extracted. Descriptors such as brightness were correlated with zero-crossing rate & high energy-low energy ratio, fullness with high fluctuation in the lower end of the spectrum

3. AIM OF THE STUDY

This study aims to utilise perceptual descriptors and correlate them with their acoustic features to understand the difference between the timbral perception of musicians and non-musicians. By identifying the most important auditory qualities and exploring their inter-scale correlations, the study can aid in the understanding of how non-musicians and musicians perceive polyphonic timbre differently.

4. EXPERIMENT

The experiment consisted of 46 participants who were asked to listen to 20 Indian music excerpts of 1-3 seconds each. Of the 47 participants, 21 were musicians, and 25 were non-musicians. It was ensured that the audio quality was high enough for the participants to rate with the highest accuracy possible. In the non-musicians category, 60% of the participants were male and 40% were female. In the musicians' category, 66% were male, whereas 33% were female participants.

The participants were asked to complete a Google form with the basic gender and age questions. The form also contained questions regarding musical experience and the number of years of musical experience.
Eg: How warm do you think the song is
Very warm 10 9 8 7 6 5 4 3 2 1 Very cold

Similar questions were asked for all the five descriptors and the audience could choose values on a 10-point scale

Definitions of the characteristics were given in the form above each question for reference of the subject.

Any participant with more than one year of practicing musical experience was considered a musician. The participants were provided with the excerpts in the form of a YouTube video with a black background with the provision to rewind and listen to the excerpt multiple times.

Using the inter-scale correlation values and the SMC values given in the paper (Alluri,2010), out of the eight scales, five scales were selected for this experiment, which are:

$\rightarrow Warm - Cold$
$\rightarrow Acoustic - Synthetic$
$\rightarrow Colourful - Colourless$
$\rightarrow Empty - Full$
$\rightarrow HighEnergy - LowEnergy$

4.1 Mean Inter-subject corretation

The table below displays the percentage difference of mean-inter-subject correlations(r-value) for the five scales between musicians and non-musicians.

Descriptors	% difference of r value
Warm	20.27%
Acoustic	31.29%
Colourful	20.30%
Full	29.41%
Energetic	3.53%

Table 1. Percentage difference of mean inter-subject correlation values between musicians and non-musicians

The percentage difference between the mean inter-subject correlation of musicians and non-musicians reveal a higher correlation for musicians as compared to non-musicians

This indicates that there is a higher consistency/agreement in the perceptual description of timbre amongst musicians than non-musicians

This consistency varies across various descriptors, with energy being equally consistent and acoustic/synthetic nature and fullness being highly consistent in musicians

4.2 Cronbach alpha values

Cronbach Alpha values were calculated to understand the overall internal consistency of the data. The values for musicians and non-musicians are given below.

The Cronbach alpha values of all the perceptual descriptors(musicians and non-musicians) is higher than 0.75, indicating a high reliability and internal consistency in the data

The Cronbach alpha values are generally higher for musicians than non-musicians, leading to an understanding that the data from musicians have higher reliability and consistency.

Descriptors	Non-Musicians	Musicians
Warm	0.82	0.86
Acoustic	0.79	0.87
Colourful	0.83	0.86
Full	0.89	0.82
Energetic	0.81	0.89

Table 2. Percentage difference of mean inter-subject correlation values between musicians and non-musicians

4.3 Statistical analysis on the data

The tables below show the inter-scale correlations for the perceptual scales. Many pairs show a high correlation, implying that the descriptors might be related to the same perceptual dimension. Contrary to previous studies, fullness was highly correlated with three other scales: acoustic, warmth and colour. Energy resulted in being the least correlated perceptual scale. Warmth and Acoustic had high mean correlation values, implying that they were referring to the same perceptual dimension.

*	Warmth	Acoustic	Colour	Fullness
Acoustic	0.81	1	x	x
Colourful	0.78	0.71	1	x
Full	0.89	0.80	0.86	1
Energetic	0.02	0.18	0.51	0.35

*	Warm	Acoustic	Colourful	Full
Acoustic	0.85	1	x	x
Colourful	0.59	0.84	1	x
Full	0.62	0.75	0.86	1
Energetic	0.12	0.40	0.72	0.63

Table 3. Inter scale correlations for Musicians(top) and non-musicians(bottom)

Between musicians and non-musicians, there was a higher interscale correlation between energy and other scales for musicians than non-musicians. There is a higher correlation between colour and Acoustic descriptors in musicians than in non-musicians. The correlation between energy and other descriptors increased significantly in musicians compared to non-musicians.

4.4 Feature extraction

As mentioned in previous studies '(Alluri,2010), multiple features were extracted from the audio samples for the acoustic feature analysis. These features were MFCCs, brightness, roughness, spectral roll-off, Chromas, bpm and sub-band flux. The feature extraction was done in MATLAB using the MIR Toolbox.

The acoustic features with the highest correlation with the perceptual descriptors are given below for musicians and non-musicians.

Warm	Acoustic	Colourful
Skewness -0.20	MFCC8 0.21	Brightness 0.41
Flatness -0.21	SBF5 -0.22	Flux 0.47
Roughness -0.30	Skewness -0.23	Entropy 0.49
Centroid -0.37	Roll off 0.32	Roll off 0.57

Full	Energy
Flux 0.50	Brightness 0.59
Entropy 0.52	Roll off 0.69
ZCR 0.56	Flux 0.70
Roll off 0.72	Entropy 0.76

Table 4. Acoustic Features Exhibiting the Highest Correlation with the Perceptual Descriptors for Non-musicians

Warm	Acoustic	Colourful
Roughness -0.26	SBF9 -0.39	ZCR 0.39
Centroid -0.32	Roll off 0.41	Brightness 0.40
MFCC8 0.36	Skewness -0.49	Entropy 0.45
Skewness -0.39	RMS 0.50	Roll-off 0.54

Full	Energy
MFCC 5 -0.30	Centroid 0.57
ZCR 0.32	Roll off 0.62
Entropy 0.35	Flux 0.67
Roll-off 0.49	Entropy 0.71

Table 5. Acoustic Features Exhibiting the Highest Correlation with the Perceptual Descriptors for Musicians

5. DISCUSSION

5.1 Inferences from Acoustic correlates

Non-musicians correlate warmth with spectral roughness and centroid higher than musicians. Since these represent energy in a higher frequency of the spectrum, the centroid being more correlated in non-musicians implies the association of warmth with frequency more than musicians.

The acoustic nature of audio is more associated with skewness in musicians than in non-musicians. Skewness represents the shape of the spectrum using the direction of outliers. A higher correlation of skewness among musicians indicate a higher capability to identify the shape/quality of sound

Colourfulness is almost similarly related to acoustic features between musicians and non-musicians

When it comes to fullness, non-musicians associated it with those with a higher spectral roll-off, ZCR and entropy than non-musicians. This implies that non-musicians have associated fullness with features that relate to a higher brightness and warmth than musicians.

This can also be seen in the inter-scale correlation of non-musicians, where fullness was highly correlated with warmth and colourfulness.

The energy was higher with brightness in non-musicians, whereas it was higher with spectral entropy and flux in musicians.

6. CONCLUSION

This study was conducted to understand the differences in the perception of polyphonic timbre between musicians and non-musicians. Musicians, in this study, were defined as people practising music with at least a year of professional music experience.

The mean inter-subject correlation values indicate a higher consistency among musicians than non-musicians, and this correlation is higher for dimensions like acoustics and fullness. The Cronbach alpha values are higher than 0.75 for all dimensions, with the values being generally higher for musicians than non-musicians, indicating higher reliability and consistency.

The inter-scale correlation values show that energy is the most independent scale. They also show warmth & Acoustic had a high correlation, implying that they were referring to the same perceptual dimension. Correlation values between energy and other scales increased significantly in musicians than in non-musicians.

The acoustic correlates with perceptual descriptors implied the association of warmth with frequency more in non-musicians than musicians. Musicians were also more capable of identifying the shape/quality of sound as understood from skewness. Non-musicians have associated fullness with features that relate to a higher brightness and warmth than musicians.

7. REFERENCES

[1] Alluri, V., & Toiviainen, P. (2010). Exploring Perceptual and Acoustical Correlates of Polyphonic Timbre. Music Perception, 27(3), 223–242. doi:10.1525/mp.2010.27.3.223

[2] Alluri, V., & Toiviainen, P. (2009). In search of perceptual and acoustical correlates of polyphonic timbre. In ESCOM 2009: 7th Triennial Conference of European Society for the Cognitive Sciences of Music.

[3] Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. Journal of experimental psychology: human perception and Performance, 20(5), 976.

[4] Gaser C, Schlaug G. Brain structures differ between musicians and non-musicians. J Neurosci. 2003 Oct 8;23(27):9240-5. doi: 10.1523/JNEUROSCI.23-27-09240.2003. Erratum in: J Neurosci. 2013 Sep 4;33(36):14629. PMID: 14534258; PMCID: PMC6740845.

[5] Kaplan, E. C., Wagner, A. E., Toffanin, P., & Başkent, D. (2021). Do musicians and non-musicians differ in speech-on-speech processing? Frontiers in Psychology, 281.

[6] F. A. de Leon and K. Martinez, "Music genre classification using polyphonic timbre models," 2014 19th International Conference on Digital Signal Processing, Hong Kong, China, 2014, pp. 415-420, doi: 10.1109/ICDSP.2014.6900697.