VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

IDENTIFYING THE BASIS OF AUDITORY SIMILARITY IN CONCATENATIVE SOUND SYNTHESIS USERS: A STUDY BETWEEN MUSICIANS AND NON-MUSICIANS

Noris Mohd Norowi[1], Eduardo Reck Miranda[2] and Hizmawati Madzin[3]
[1]Human Computer Interaction Research Group, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia
[2]Interdisciplinary Centre for Computer Music Research, University of Plymouth, PL4 8AA, United Kingdom
[3]Computer Graphics, Vision and Visualization Research Group, Universiti Putra Malaysia, 43400, Serdang, Selangor, Malaysia
E-Mail: noris@upm.edu.my

ABSTRACT

This paper identifies the basis of auditory similarity in concatenative sound synthesis users. Concatenative sound synthesis (CSS) system is an existing approach to create new sounds based on a user supplied audio query. Typically, the audio is synthesised based on the least distance between the query sound unit and the available sound units in the database. However, sounds synthesised through this approach often times result in a mediocre level of satisfaction within the users as confusion between various audio perception attributes during the CSS system's matching process causes mismatches to occur. This study aims to determine the dominant perceptual attribute that humans base their judgment of sound similarity on. The study also looks at two categories of CSS system's users: musicians and non-musicians, and observes whether there is a significant difference in the subjective judgments between the two groups with regards to sound similarity. Thirty-eight participants were subjected to the listening test, where six pairwise comparisons from four different audio perceptual attributes (melody, timbral, tempo and loudness) were compared. In general, it was found that the majority of users in the Musicians group (73.3%) based their sound similarity on timbre attribute, whilst the majority of the users in the Non-musicians group (78.3%) based their sound similarity on the melody attribute. This information may be used to help CSS system cater to the expectations of its users and generate the sounds with the closest matching audio perceptual attribute accordingly.

Keywords: auditory similarity, concatenative sound synthesis, audio perception, pairwise comparison.

INTRODUCTION

In its simplest form, the physics of simple sound can be described as a function of frequency, amplitude and phase. Generally put, two sounds are similar if the values of these three criteria are the same. However, sounds very rarely exist in this simple form and oftentimes the Fourier analysis is used to break down complex sounds into a series of simple sound to achieve this. The psychology of sound, on the other hand, is based on the human perception of these criteria and also the time factor, giving rise to other sound elements such as pitch, intensity, timbre and rhythm, among others.

Usually, human listeners have a well-developed feeling whether two songs sound similar or whether they do not [1]. It is thus very important for any system that relies finding similar sounds such as the audio information retrieval system or sound similarity matching system to determine what these auditory characteristics are. Earlier works at the Muscle Fish research group in content based audio retrieval have described the ways in which humans may describe similar sounds – simile, acoustical or perceptual features, subjective features and onomatopoeia; all of which have been used individually or in combination, as a query mechanism for many sound similarity-based multimedia applications such as audio classification, audio retrieval and audio search engine [2].

Similarly, a Concatenative Sound Synthesis (CSS) system – an art of producing new sounds from a composite of many small snippets of audio by matching the target or query sound from the user to the available sounds in the database - would have to have the same capability of tackling of all the variability described above. Unfortunately, due to its extreme complexity, this level of perfection is yet to be accommodated, especially since audio perception is a vast subject.

As an example of how audio perception may affect the output of a CSS system, consider a query sound of an A4 note played on a piano, which of the two segments that are available in the database – an A4 note played on a string instrument or a C4 note played on a piano – will be considered as most similar to the target sound? Which attribute does human find to be more dominant than others (if any)?

There are different attributes that can be the basis of sound similarity, the basics being elements such as pitch, rhythm, tempo, timbre and loudness. Moreover, combinations of these elements then give rise to higher-order concepts such as meter, key, melody and harmony [3]. Identifying the perceptual audio attributes that influence sound similarity in humans may reveal the audio feature sets that are more likely to extract relevant information from sounds, which can possibly return perceptually closer matching segments from the database. Determining which the audio attributes are more dominant may be the key to improving similarity in sounds generated by CSS systems.

This study intends to identify the dominant acoustic information on which judgements are based by humans when performing a sound similarity task. Results from this study will ascertain the dominant attribute

involved when humans perceive sounds to be similar. By applying this attribute into the CSS system, it is envisioned that the sounds generated will be able meet more of the users' expectation and satisfaction.

This paper is arranged as follows: Section 2 presents the technical overview of a CSS system. Section 3 discusses several studies which are related to this paper. Section 4 describes the procedures of the auditory similarity experiment, while Section 5 analyses and discusses the results. The paper concludes its finding in Section 6.

CONCATENATIVE SOUND SYNTHESIS

The research involving CSS has been inspired by the art of mosaicing. Mosaic arts first appeared over five thousand years ago, in Abra, Mesopotamia where an assemblage of small pieces of coloured glass had been used to create larger, whole images that were typically seen in many decorative paraphernalia and were also applied to the design of many significant cultural and spiritual erections (Figure-1).

Figure-1. An example of a Roman mosaic.

Through the same concept, mosaicing was then applied to digital images and digital audio, and hence are referred to as 'Photomosaicing' and 'Musaicing' (musical mosaicing) respectively. In photomosaicing, small tiles of images are assembled together to compose a bigger overall picture [4], as illustrated in Figure-2. Likewise, musical mosaicing assembles a large number of unrelated sound segments together according to specifications given by an example sound, to form a coherent, larger, sound framework.

Like any other Information Systems, CSS is composed of humans (generally musicians, audio engineers, regular users) and computers, which then process or interpret information (e.g. musical information) to complete tasks such as organizing data, delivering knowledge and digital product, or in this instance, automatically composed sound file. CSS system uses sound as a target or sometimes referred to as the query, is decompose it into smaller sound segments, and having its spectral and other auditory content analysed. When the criteria of the target segments or the query has been completed, a unit selection matching process then takes place to search for a matching sound segments in the database. Once found, the segments which closely match those of the query segments are then concatenated together in sequence, and are then resynthesized to produce new sounds that are based on the original sound entered. Figure-3 illustrates the general mechanism of a CSS system.

Figure-2. An example of a photomosaic.

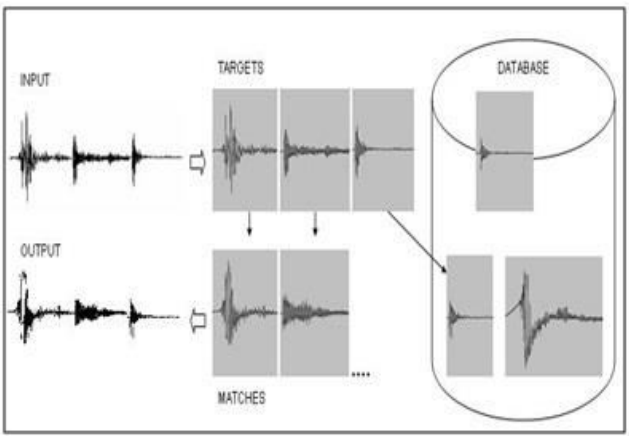

Figure-3. The general mechanism of a CSS system.

A typical CSS system has two major components; analysis and synthesis. During the analysis phase, both the original sounds (target) and the sounds in the database (source) are segmented into smaller sound snippets. Following segmentation, relevant information from these sound snippets is then extracted. In the synthesis phase, sound snippets in the database that match closely with the targets are selected and concatenated together forming a long string of sound, which are then synthesized (Figure-4).

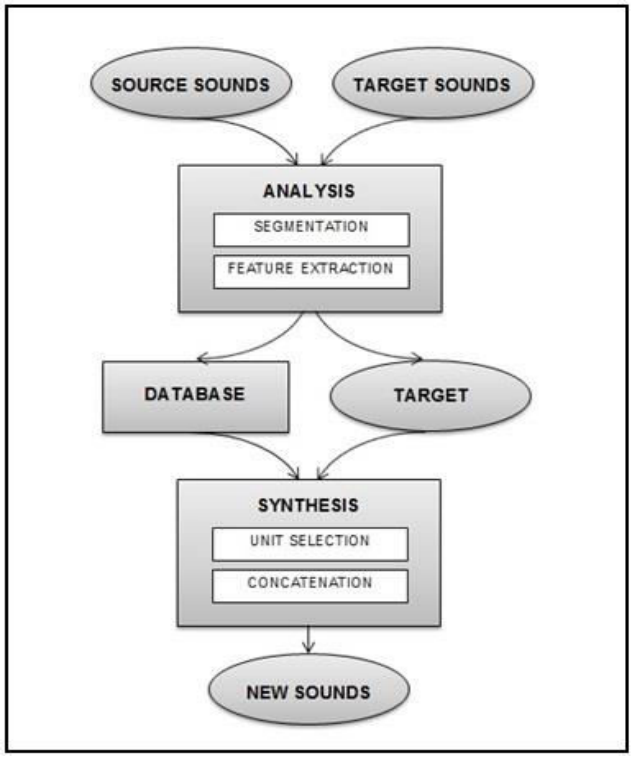

Figure-4. Components in a typical CSS system.

CSS has extensively used in speech synthesis, where speech are generated from actual recordings of human speaker. The speech synthesized through CSS typically produces a more natural sounding speech than those generated from rule-based synthesis [5]. Through the very same process, it was later found that CSS was also equally useful to be applied for synthesising music and other sounds [6]. However, several challenges were quickly discovered with the use of CSS in music which was not present when synthesising speech, such as the basis unit for segmentation and the time factor. In concatenative speech synthesis, phonemes are the basis unit for segmentation, whereas for music, the units are usually segmented according to musical notes or events. This requires a more complex analysis than using phonemes. Secondly, timing is crucial in music synthesis as it is needed to ensure that the rhythm and tempo is in place, but does not have the same effect in speech synthesis. This becomes the motivation for this study, as auditory similarity in this instance could mean that a sound is similar in terms of their spectral content (without regard for the time information), or similar in the melodic content (with respect to the time information).

This perceptual issue which needs to be brought into attention is a rather subjective subject, but is a crucially fundamental matter to the question of 'what makes humans perceive two sounds as similar?' The technical issues may have undergone many improvements, but unless the above question is answered, CSS systems may be generating sounds that are far from the expectation of its users.

RELATED WORKS

Over the past decade, various CSS systems have emerged and the research is continually progressing. Several systems such as the AudioQuilt [7], EarGram [8], CataRT [6], SoundSpotter [9], MATConcat [10] and Musical Mosaic [11] all have various strengths and weaknesses, in the form of their input format, segmentation modes, features selection, search methods, use of concatenation distance and transformation, visual display options and also real-time capability [12].

For instance, the most popular input formats are WAV, AIFF, or MP3, which are accepted almost all existing CSS systems. Some systems accept additional formats such as MIDI files, which gives users a wider option of source and target sounds to work with, as seen in Mosievius [13]. CataRT [6] advanced a step further by not only accepting both audio and MIDI files, but also pre-processed segmentation markers, i.e. SDIF and ASCII files that can be piped directly from other programs. Additionally, in some systems the input sounds are not restricted to the use of audio recordings only, but can also include live input, e.g. from a microphone, as seen in Soundspotter [9] and CataRT [6]. Similarly, during segmentation, several CSS systems offers users to decide between performing segmentation at note level or sub-note level, depending on whether they require longer or shorter sound segments respectively [13]. The most flexible so far are Caterpillar [14] and CataRT [6], where both system permit users to choose between four different segmentation modes - note or phone segmentation, symbolic data segmentation, steady pulse segmentation, and manual segmentation.

Although these technical specifications plays a major role in determining the output of the concatenated sound synthesis [15], little work has been conducted on determining the effect which audio perception has on matching the CSS system's output with the users requests. Researchers in the field of image similarity have generally agreed that there are four major low-level attributes that influence this, which are colour, texture, shape and spatial constraint [16-18]. Unfortunately, such clear cut low-level attributes cannot be applied to audio similarity, mainly due to the nature of music itself. For instance, music similarity deals with issues which do not concern that of image similarity, such as the issue of monophonic or polyphonic, with or without singing, tonal or atonal, lyrics and meaning, etc.

Moreover, human beings hear music in a "non-linear" way. Studies in music perception and cognition have found that many of these perceptual attributes such as pitch, loudness, timbre and duration are as cleanly separable as they first appear. For example, very short notes can be heard as being a little less loud than notes of the very same tone and loudness, but presented in longer durations [19-21]. Also, when melody is involved, the changes in timbre can be a little less obvious.

Although many studies in musical perception have been carried out, there is still no single consensus over which perceptual attribute is the most utilized by

humans. Some argues that the durational values may outweigh pitch values in facilitating music recognition [22], some suggests that the pitch contour or melody, has the most effect [23] whilst others believe that it is a weighted distribution between pitch, rhythm, timbre and dynamics [20]. It is therefore the aim of this study to determine which perceptual attributes of the sound works best in finding the closest sound match in a CSS system.

EXPERIMENTAL SETUP

The objectives of this study are two-fold: (1) to identify the dominant perceptual attribute that humans base their judgment of sound similarity on; and (2) to observe whether there is a significant difference in the subjective judgments between musicians and non-musicians with regards to sound similarity.

The sound attributes that are included in this test are Melody (the linear succession of musical notes that gives the tune of a musical piece), Timbre (the quality and texture of sound that distinguishes an instrument from another, including information such as the relative brightness or brashness of a sound), Loudness (the way in which humans perceive the amplitude of sound) and Tempo (the speed or pace of music, indicating how slow or fast a sound, usually music, is played). These four attributes, when placed in a pairwise comparison against one another, resulted in a total of six comparison pairs (Table-1). The aim of this experiment was to observe which attribute from each pair is favoured most often.

Table-1. Six comparison pairs from the four perceptual attributes of melody, Timbre, Tempo and Loudness.

Pairs	Melody	Timbre	Tempo	Loudness
Melody		--	--	--
Timbre	Timbre vs. Melody		--	--
Tempo	Tempo vs. Melody	Tempo vs. Timbre		--
Loudness	Loudness vs. Melody	Loudness vs. Timbre	Loudness vs. Tempo	

Participants

Thirty eight healthy participants with self-declared normal hearing, aged between 21–60 years old were asked to participate in this study on a voluntary basis. The subjects were comprised of 21 females and 17 males. Participants were divided into two groups – musicians and non-musicians. In this test, the term 'musicians' were defined as those who have received formal musical training for four years and above, or have been or are currently employed in the music industry, e.g. performer, music researcher, music lecturer, tuner, etc. All participants were asked to detail any formal musical training they had had and the number of years that they had been trained for before the start of the test. The intended ratio between the two groups was at 1:1, so as not to create any bias in the results. However, the number of non-musician participants was slightly larger (23 non-musicians to 15 musicians.) A Chi-squared test was done to determine if the dataset was biased in terms of sex and musical training. At $\chi2$ (1) = 1.421, $p<0.5764$, it was found that there was no gender bias within these participants. Similarly, was found that there was no musical background bias within these participants ($\chi2$ (1) = 1.684, $p<0.1944$). No other demographics effect was studied.

Audio Dataset

The audio dataset for this test were comprised recordings from natural sounds (animals and environmental) and also music. The lengths of audio tracks varied from 1 to 10 seconds, as in some cases, longer audio tracks were necessary in order to allow information to be amply presented and identified by subjects, i.e. melody or tempo. Sound similarity between the target and the source tracks were decided through the use of several sound analysis programs such as MARSYAS and Praat for information on the timbre and loudness respectively. The tempo information was obtained at different websites over the internet that provided ground truth on the beat per minute (BMP) of a particular track. Information on the melodic similarity was also obtained over several websites that compared or surveyed melodic similarity manually . Since this information was submitted by humans and are open to preconception, the tracks' melodic contours were then compared visually in Praat to confirm similarities.

Procedure

Tracks were delivered to the participants via headphones at a comfortable loudness level. Three sound tracks were presented; one of which was a target track, and two of source tracks. Participants were required to first listen to the target track, followed by the source tracks. They were then asked, in a forced choice manner, to make a selection between the two tracks, based on which tracks they felt were more similar to the target, e.g. 'Which of these two sounds do you feel match more closely to the target sound?'. The test was designed so that each source tracks in the pair would correspond to a different attribute that was being compared. For example, in a melody versus timbre pair, one source track would be melodically similar to the target, whilst the other would be closer in terms of timbral similarities, whilst other perceptual attributes that were not being compared were kept constant. This this information was not revealed to the participants so as to allow selection to be made at will, since no basis of similarity or perceptual attribute was specified. Each participant was presented with twelve of these sets, and re-playing of the tracks was allowed. The average time taken to complete this test was roughly ten to fifteen minutes.

RESULTS

Figure-5 shows the result of all six pairwise comparisons, for the combined average between all participants (musicians and non-musicians). The average

between both groups is given in percentage values on top of each bar in bold.

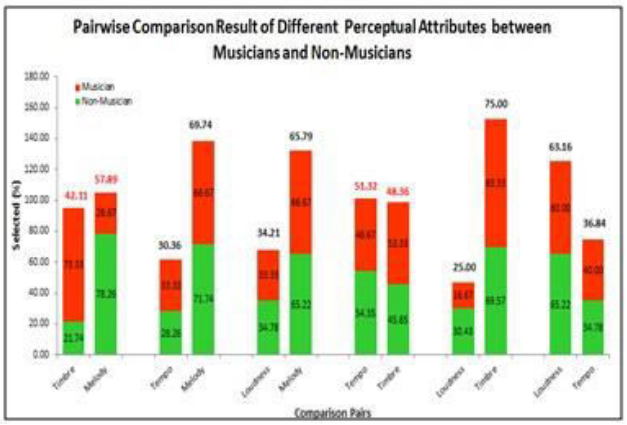

Figure-5. Result of Pairwise Comparisons.

From the test, it was found out that Melody showed a striking pattern of domination, where out of the six comparison pairs, three which had involved Melody went unchallenged by other attributes, i.e. in pairs Timbre-Melody, Tempo-Melody and Loudness-Melody. It was also found that in general, Timbre appeared to be more dominant than Loudness, and Loudness more dominant than Tempo. However, in the Tempo-Timbre pair, no dominant attribute can be conclusively derived.

To ensure that the results of these pairwise comparisons were not biased, the significance of each result from the pairs was determined through the use of Chi-squared test. Results from four pairs (Tempo-Melody, Loudness-Melody, Loudness-Timbre and Loudness-Tempo) were all found to be statistically significant; indicating the slight difference in the number of participants from the two groups (Musician and Non-Musician) had not introduced bias into the result. Thus, the results are considered valid and it can be accepted that Melody were more dominant than Tempo and Loudness, whilst Timbre was more dominant than Loudness and Loudness dominated over Tempo in such relationship as Melody = Timbre > Loudness > Tempo.

Also in the four pairs that were found to be significant, both showed that the two groups tend to agree on the same dominant attributes, e.g. when the majority of musicians thought the dominant attribute were Loudness in the Loudness-Tempo pair, non-musicians thought the same. However, there were two cases in which this agreement was not found to be true – the Timbre-Melody and the Tempo-Timbre pairs. The average selection percentages of these two cases are highlighted in red ink in previous chart (Figure-3).

Interestingly, Chi-squared test found that the result of these two pairs to be statistically insignificant too. At $\chi2(1) = 1.895$, $p<0.1687$ for the former pair and $\chi2(1) = 0.053$, $p<0.8185$ for the latter, the null hypothesis must be rejected, suggesting any pattern that might be present occurred only by chance. Hence, it cannot be accepted that Melody is more dominant than Timbre, nor can it be said that Tempo is more dominant than Timbre, as the values obtained from this test were not significant enough to deduce this.

Perhaps it was difficult to conclusively agree on the dominant perceptual attributes as the percentage of selection between the two attributes compared are split in the middle between the Musician and Non-Musician group. Looking closely at isolated charts of these two pairs in Figure-6 and Figure-7, this was indeed the case. A 2x2 Contingency Table of Chi-squared Test for Independent was done on both pairs to verify whether there was the case.

In the Timbral-Melody pair, the test of independence had found an extremely significant association between preferred perceptual attribute and participants' musical background ($\chi2(1)=19.829$, $p<0.0001$). Referring again to the graphs in Figure-4, it can be clearly seen that in the Timbral-Melody pair, Melody was only found to be dominant amongst the vast majority of non-musicians, whereas more than 70% of the musicians selected Timbre.

However, when a similar test of independence was performed on the Tempo-Timbre pair, it was not found to be statistically significant ($\chi2(1) = 0.429$, $p<0.2563$). This means that unlike the previous pair, the different musical background of participants did not play a part in their decision between the Tempo-Timbre pair.

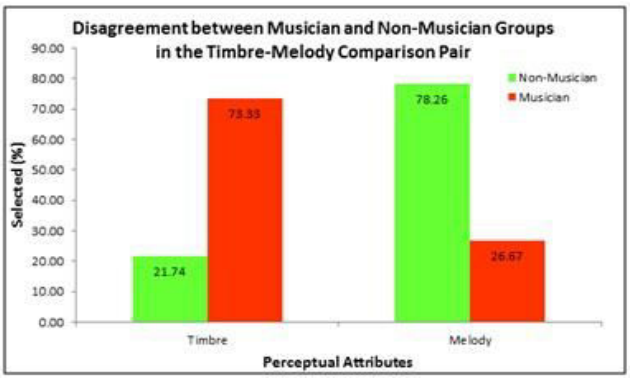

Figure-6. Timbre-Melody Result Disagreement Between Musician and Non-Musician Groups

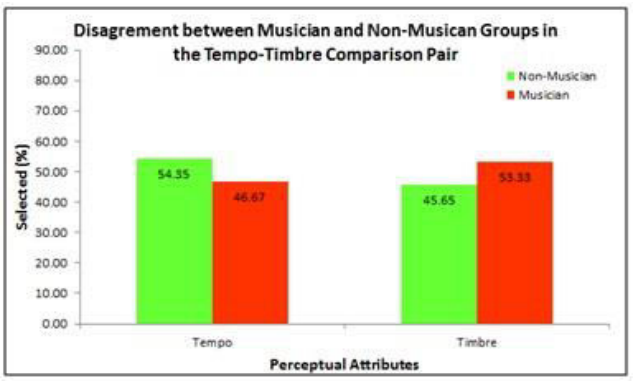

Figure-7. Tempo-Timbre Result Disagreement Between Musician and Non-Musician Groups.

Perhaps this was due to the flaw in the sound selections in the test design for this pair, or that the number of sound stimuli and size of participants were too small to effectively solve this. Unfortunately, for such test, it must be remembered that it is difficult to obtain a large number of volunteers, especially which required participation of those with a specific expertise on the subject (Musician group). Moreover, in a listening test like this, there can only be a limited number of stimuli presented to the participants before it becomes too long for them to manage.

DISCUSSIONS

From this study, it can be agreed that based on the average selection percentage, Melody seems to be the most dominant perceptual attribute for audio. This could be because Melody is perceptually grouped as part of the same event unfolding over time, based on the Gestalt's principles of perceptual organisation such as similarity, proximity and good continuation. As humans conform to these principles, Melody tends to be preferred over attributes such as Tempo or Loudness [24].

Nevertheless, it is evident that the human's musical background also affects the judgment in finding the dominant attribute as musical training alters the way music is perceived by humans. The human brain is divided into two hemispheres, the left lies the more logical and calculative thinking and the right handles the more intuitive feelings. Musicians tend to use the left hemisphere of the brain to a larger extent when listening to music because they possess an analytical knowledge of it and thus approach music more intellectually. In comparison, those with no musical background mostly perceive music in the right hemisphere because they are not analysing, but are simply experiencing the music [25].

This test shows that musicians generally are more tuned to selecting sounds that are similar timbrally than they are melodically, whereas the reverse is true for non-musicians. Again, this is possibly owing to their analytical behaviour in listening to music, where experienced musicians can be very sensitive in assessing similarities based on the quality of musical expressions rather than the actual melody.

In the context of a CSS system, different attributes can be applied during the unit selection stage, depending on the intended user. For instance, based on the findings from this experiment, the low-level audio features which correspond to the timbral attribute e.g. spectral centroid, spectral rolloff, spectral flux, MFCCs, etc., can be applied for users with musical training. On the other hand, for non-musically sound users, audio features such as pitch which represents the melodic information can be used instead. Such implementation can increase the chances of synthesising sounds which are in line with the expectation of its users.

As small-scaled as this test was, it demonstrated that the basis of sound similarity is a very wide and complex area. It would be very difficult to develop a working CSS system that can cater all these perceptual attributes that affect the way humans listen and judge sound similarity on. Nevertheless, since the system is primarily targeted for musicians, and since it was found that the most dominant perceptual attribute that musicians are more prone to base their sound similarity on is timbre, this study will incorporate the audio features that correspond to the timbral attributes in the framework of the new CSS system.

CONCLUSIONS

A listening test involving human participants were conducted to identify the dominant perceptual attribute which humans most often use to pass their sound similarity judgment on. Whilst the area of sound similarity was indeed vast and complex, the tests revealed that sound similarity in humans was affected by their musical background. Non-musicians generally regarded sound similarity in terms of melody, whilst musicians tended to base their similarity judgment on the timbral quality. It was deduced from the results of these tests that by customising the basis of similarity according to the respective target user, the human-computer sound similarity misinterpretation could be minimised. This promotes higher satisfaction amongst the user of the CSS system, as the system tries to synthesise sounds which correctly matches the perceptual expectation of its users.

ACKNOWLEDGEMENTS

We wish to thank Professor Judy Edworthy, from the School of Psychology, Faculty of Science and Technology, University of Plymouth, United Kingdom, for the fruitful discussions over the topic of audio similarity, which tremendously helped with this study. Our thanks also go to all the volunteered participants who made it possible for this study to be tested and validated.

REFERENCES

[1] Allamanche, E., Herre, J., Hellmuth, O., Kastner, T., & Ertel, C. (2003, October). A multiple feature model for musical similarity retrieval. In Proc. ISMIR.

[2] Wold, E., Blum, T., Keislar, D., & Wheaten, J. (1996). Content-based classification, search, and retrieval of audio. MultiMedia, IEEE, 3(3), 27-36.

[3] Mitrović, D., Zeppelzauer, M., & Breiteneder, C. (2010). Features for content-based audio retrieval. Advances in computers, 78, 71-150.

[4] Tran, N. (1999). Generating photomosaics: an empirical study. In Proceedings of the 1999 ACM symposium on Applied computing (pp. 105-109). ACM.

[5] Hunt, A. J., & Black, A. W. (1996, May). Unit selection in a concatenative speech synthesis system using a large speech database. In Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference

Proceedings., 1996 IEEE International Conference on (Vol. 1, pp. 373-376). IEEE.

[6] Schwarz, D. (2006). Concatenative sound synthesis: The early years. Journal of New Music Research, 35(1), 3-22.

[7] Fried, O., Jin, Z., Oda, R., & Finkelstein, A. AudioQuilt: 2D Arrangements of Audio Samples using Metric Learning and Kernelized Sorting.

[8] Bernardes, G., Guedes, C., & Pennycook, B. (2012). Eargram: an application for interactive exploration of large databases of audio snippets for creative purposes. In Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (pp. 265-277).

[9] Casey, M. A. (2005). Acoustic lexemes for organizing internet audio. Contemporary Music Review, 24(6), 489-508.

[10] Sturm, B. L. (2004). MATConcat: an application for exploring concatenative sound synthesis using MATLAB. Proceedings of Digital Audio Effects (DAFx), Naples, Italy.

[11] Zils, A., & Pachet, F. (2001, December). Musical mosaicing. In Digital Audio Effects (DAFx).

[12] Norowi, N. M., & Miranda, E. R. (2011, April). Order dependent feature selection in Concatenative Sound Synthesis using Analytical Hierarchy Process. In EUROCON-International Conference on Computer as a Tool (EUROCON), 2011 IEEE (pp. 1-4). IEEE.

[13] Lazier, A., & Cook, P. (2003, September). MOSIEVIUS: Feature driven interactive audio mosaicing. In Digital Audio Effects (DAFx).

[14] Schwarz, D. (2003, September). The caterpillar system for data-driven concatenative sound synthesis. In Proceedings of the COST-G6 Conference on Digital Audio Effects (DAFx) (pp. 135-140).

[15] Norowi, N.M., (2013). An Artificial Intelligence Approach to Concatenative Sound Synthesis.

[16] Gudivada, V. N., & Raghavan, V. V. (1995). Design and evaluation of algorithms for image retrieval by spatial similarity. ACM Transactions on Information Systems (TOIS), 13(2), 115-144.

[17] Chen, C., Gagaudakis, G., & Rosin, P. (2000, August). Similarity-based image browsing. In Proceedings of the 16th IFIP World Computer Congress. International Conference on Intelligent Information Processing.

[18] Laaksonen, J., Oja, E., Koskela, M., & Brandt, S. (2000, November). Analyzing low-level visual features using content-based image retrieval. In Proceedings of the 7th International Conference on Neural Information Processing (ICONIP'00), Taejon, Korea (pp. 1333-1338).

[19] Kirsteen M. Aldrich, Elizabeth J. Hellier & Judy Edworthy (2009): What determines auditory similarity? The effect of stimulus group and methodology, The Quarterly Journal of Experimental Psychology, 62:1, 63-83.

[20] Byrd, D., & Crawford, T. (2002). Problems of music information retrieval in the real world. Information processing & management, 38(2), 249-272.

[21] Miranda, E. R., Correa, J., & Wrights, J. (2000). Categorising complex dynamic sounds. Organised Sound, 5, 95–102.

[22] Hewlett, W. B., & Selfridge-Field, E. (Eds.).(1998). Melodic similarity: Concepts, procedures, and applications (Vol. 11). Mit Press.

[23] Downie, J. S. (1999). Evaluating a simple approach to music information retrieval: Conceiving melodic n-grams as text (Doctoral dissertation, The University of Western Ontario London).

[24] Gates, A., & Bradshaw, J. L. (1977). The role of the cerebral hemispheres in music. Brain and Language, 4(3), 403-431.

[25] Segalowitz, S. J. (1983). Two sides of the brain. Englewoocl Cliffs: Prentice Hall.