Audio Engineering Society

Convention Paper 8420

Presented at the 130th Convention
2011 May 13–16 London, UK

An additive synthesis technique for independent modification of the auditory perceptions of brightness and warmth

Asteris Zacharakis[1] and Josh Reiss[1]

[1]*Centre for Digital Music, Queen Mary University of London, London, E1 4NS, U.K.*

Correspondence should be addressed to Asteris Zacharakis (asterios.zacharakis@eecs.qmul.ac.uk)

ABSTRACT

An algorithm that achieves independent modification of two low-level features that are correlated with the auditory perceptions of brightness and warmth was implemented. The perceptual validity of the algorithm was tested through a series of listening tests in order to examine whether the low-level modification was indeed perceived as independent and to investigate the influence of the fundamental frequency on the perceived modification. A Multidimensional Scaling analysis (MDS) on listener responses to pairwise dissimilarity comparisons accompanied by a verbal elicitation experiment, examined the perceptual significance and independence of the two low-level features chosen. This is a first step for the future development of a perceptually based control of an additive synthesizer.

1. INTRODUCTION

The task of associating adjectives and verbal terms that are used to describe musical timbre to low-level acoustic correlates has been investigated over the last thirty years. However, because of the complex and multidimensional nature of timbre this still remains an open field of research. This work constitutes only one part of a larger project that aims to achieve synthesis or modification of musical timbres in an intuitive way.

Brightness is regarded as the most significant semantic descriptor of musical sound and a great number of studies have shown its high possitive correlation with the spectral centroid [1], [2], [3], [4], [5]. The spectral centroid is the centre of gravity of the frequency spectrum and is calculated as shown in equation 1.

$$SC(n) = \frac{\sum_{k=1}^{K} f(k)A_n(k)}{\sum_{k=1}^{K} A_n(k)} \quad (1)$$

where $A_n(k)$ is the magnitude of the k^{th} coefficient of the DFT associated with the frame centered at time n, *f(k)* is the frequency associated with the k^{th} frequency component and K indicates the last frequency bin to be processed.

Warmth on the other hand, does not demonstrate such a commonly acceptable acoustic correlate and some studies have shown a lesser or greater amount of overlap between warmth and brightness. Recent work [6] has proposed a timbre morphing technique for achieving independent brightness-warmth modification using the SMS (spectral modelling synthesis) platform. In this work the acoustic correlate of warmth was defined as the relative percentage of energy in the first three harmonic partials (Eq. 2).

$$Warmth(n) = \frac{\sum_{k=1}^{3} A_n^2(k)}{\sum_{k=4}^{K} A_n^2(k)} \tag{2}$$

where $A_n(k)$ is the magnitude of the k^{th} harmonic partial associated with the frame centered at time n. Using the same definition for the acoustic correlate of warmth, our current work proposes a two-section additive synthesis algorithm for the independent modification of brightness and warmth.

A dissimilarity rating listening test featuring stimuli created with this algorithm was performed in order to examine the perceptual significance of the 'warmth' feature, the degree of independence between the perceptions of brightness and warmth and the influence of the fundamental frequency over the above. Multidimensional Scaling (MDS) analysis applied on the results constructed 2-D spaces that revealed the perceptual relationships among the stimuli. The test was completed with a verbal elicitation part which aimed at labelling the dimensions of these spaces.

2. ALGORITHM

The two-section algorithm that was utilized for the independent modification of the spectral centroid and the relative energy of the first three harmonics is described below. The following formula was used for calculation of the normalized version of the harmonic energy spectral centroid.

$$SC_{norm}(n) = \frac{\sum_{k=1}^{K} k \times A_n^2(k)}{\sum_{k=1}^{K} A_n^2(k)} \tag{3}$$

2.1. Brightness modification with constant warmth[1]

The modification of the spectral centroid position without affecting 'warmth' was achieved by altering the spectral distribution between the 4^{th} and the 30^{th} (last in our case) harmonics while preserving the overall energy in this region. For that purpose the above region is divided into two subgroups whose energy is altered according to the following procedure. The initial energies are given by Eq. 4, 5, 6.

$$E_{27} = E_1 + E_2 \tag{4}$$

$$E_1 = \sum_{k=4}^{r} A_k^2 \tag{5}$$

$$E_2 = \sum_{k=r+1}^{30} A_k^2 \tag{6}$$

where E_{27} is the overall energy of the last 27 partials and r is the rounded harmonic 50% roll-off point [2] for the spectral region of the last 27 partials. Thus, the initial energies are close to equal ($E_1 \simeq E_2$).

Then the modification factors are calculated according to Eq. 7, 8 and 9.

$$E_{27} = E_1 + E_2 = a^2 E_1 + b^2 E_2 \tag{7}$$

where a and b are the factors that multiply every harmonic amplitude in each subgroup. Based on Eq. 7, b is expressed as a value of a as shown in 8.

$$b = \sqrt{\frac{E_1 + E_2 - E_1 \times a^2}{E_2}} \tag{8}$$

The square root of Eq. 8 introduces the following limitation for a.

$$a \leq \sqrt{\frac{E_1 + E_2}{E_1}} \tag{9}$$

[1] Whenever the terms brightness and warmth are used in the text instead of their acoustic correlates they will refer to the intended brightness and warmth.

[2] Mid-point of energy

It must also be stated that the above calculation does not require that E_1 be equal with E_2. However, the subgroups were divided based on the 50% roll-off point since in this way a more even modification of the spectral centroid around its initial value is achieved. Since both regions preserve their initial energies, this method does not alter the 'warmth' ratio while changing the position of the spectral centroid.

2.2. Warmth modification with constant brightness

The method used for warmth modification implemented a transformation of an existing signal using a modifying signal in the frequency domain. This transformation keeps the spectral centroid constant while altering the relative energy of the first three partials. The modifier signal has the same spectral centroid as the original as shown in Eq. 10.

$$SC_{org} = \frac{\sum_{n=1}^{K} n \times A_n^2}{\sum_{n=1}^{K} A_n^2} = SC_{mod} = \frac{\sum_{n=1}^{K} n \times X_n^2}{\sum_{n=1}^{K} X_n^2} \quad (10)$$

where A_n are the harmonic amplitudes of the original and X_n are the harmonic amplitudes of the modifier. Based on the fact that

$$\frac{a}{b} = \frac{c}{d} \Rightarrow \frac{a}{b} = \frac{a+c}{b+d} \quad (11)$$

we can construct a modified signal featuring the same spectral centroid as the original

$$\frac{\sum_{n=1}^{K} n \times B_n^2}{\sum_{n=1}^{K} B_n^2} = \frac{\sum_{n=1}^{K} n \times (A_n^2 \pm X_n^2)}{\sum_{n=1}^{K} (A_n^2 \pm X_n^2)} = SC_{org} = SC_{mod} \quad (12)$$

where $B_n = \sqrt{A_n^2 \pm X_n^2}$ are the harmonic amplitudes of the modified signal. Consequently, the above transformation provides a way of changing the spectral content of a signal without altering its spectral centroid. For the purpose of this work the modifier X_n (where n is the rounded normalized SC) consists of three harmonics that in essence create a formant centred around the normalized SC. X_{n-1} given X_n and X_{n+1} is calculated by Eq. 13.

$$X_{n-1} = \sqrt{\frac{SC \times (X_{n+1} + X_n^2) + X_{n+1}^2 \times (n+1) + X_n^2 \times (n)}{n - 1 - SC}} \quad (13)$$

The effect of the algorithm over warmth is greater for signals having a normalized SC between 1.5 and 2.5 as in such a case it alters the first three partials of the sound.

3. LISTENING TEST

Two identical pairwise dissimilarity rating listening tests were performed in order to investigate the perceptual significance of the modifications that were applied by the algorithm. In addition, the tests examined the influence of fundamental frequency on the perception of warmth and brightness.

3.1. Stimuli

The stimuli were generated by the application of the above algorithm to a parent timbre with an additive synthesis engine built in Max/MSP. Their spectrum was absolutely harmonic and consisted of 30 harmonics. The duration of all stimuli was 1.6 seconds and the temporal envelope was the same for all samples (100 msec attack, 50 msec decay, 0.8 sustain level and 100 msec release) so that listeners could concentrate absolutely on spectral changes. Both rise and release were linear. Two groups of 12 stimuli were produced, differing only in fundamental frequency (220 Hz for the first and 440 Hz for the second group). The normalized SC of the parent timbre was selected to be 2.2 and was created using a brightness creation function that is presented in Eq. 14 [8].

For $A_k = B^{-k}$, where A_k is the amplitude of the k^{th} harmonic, the normalized energy SC is calculated as follows:

$$SC_{norm} = \frac{\sum_{k=1}^{K\to\infty} k(B^{-k})^2}{\sum_{k=1}^{K\to\infty} (B^{-k})^2} \simeq \frac{B^2}{B^2 - 1} \quad (14)$$

and for a known SC_{norm}, B is calculated from Eq. 15

$$B = \sqrt{\frac{SC_{norm}}{SC_{norm} - 1}} \quad (15)$$

The reason for selecting the SC position in 2.2 is because it was desired for the warmth modification algorithm to affect only the amplitude of the first three harmonics and at the same time to have a reasonably bright sound. The positions of the twelve stimuli in the warmth – brightness feature space is shown in figure 1.

The choice of the stimuli number was made in order to enable the MDS analysis to position them to up to a 3–D space according to the empirical rule of four stimuli per

dimension [7], while keeping the duration of the pairwise dissimilarity listening test relatively small. All stimuli were loudness equalized according to the experimenter's ear and only one out of 20 subjects reported difference in the perceived loudness level between them. The stimuli were stored in PCM Wave format, at 16 Bit, 44.1Khz, in Mono.

3.2. Listening Panel

Twenty subjects (aged 23–40, 5 female) participated in a pairwise dissimilarity listening test. None of them reported any hearing loss and all of them were critical listeners and had been practising music for 18 years on average (ranging from 8 to 30). Ten of them listened to the 220 Hz group of stimuli and the rest listened to the 440 Hz stimuli.

3.3. Pairwise Dissimilarity Listening Test

A pairwise dissimilarity experiment was performed in order to test the perceived dimensionality of the stimuli set. The experiment was conducted from a Macbook Pro laptop with the AKG K 217 MK II circumaural headphones, in a small acoustically isolated listening room. The interface of the experiment, part of which is presented in Figure 2, was built in Max/MSP.

Initially the listeners were presented with a familiarisation stage which consisted of the random presentation of the stimuli set in order for them to get a feel of the timbral range of the experiment. After that they performed a training stage that consisted of five dissimilarity ratings so as to get used to the procedure. Finally, they undertook the complete dissimilarity test where they were presented with all 78 combinations of pairs within the set in random order. Comparisons of pairs with themselves were included as a measure of the validity of each listener. Listeners were rating the differences of each pair using a hidden continuous scale with end–points marked as 'the same' and 'most dissimilar' as shown in Figure 2. They were also allowed to repeat each pair as many times as they wanted before making their judgement.

3.4. Verbal Elicitation Test

The experiment ended with a verbal elicitation stage where listeners were presented with four selected pairs of stimuli in random order. The pairs used for this reason were the two diagonals of the pseudo–rectangular feature space (1–12 and 4–9) as well as one pair on the 'warmth axis' (2–10) and one on the 'brightness axis' (5–8). The task of the listeners was to insert spontaneously

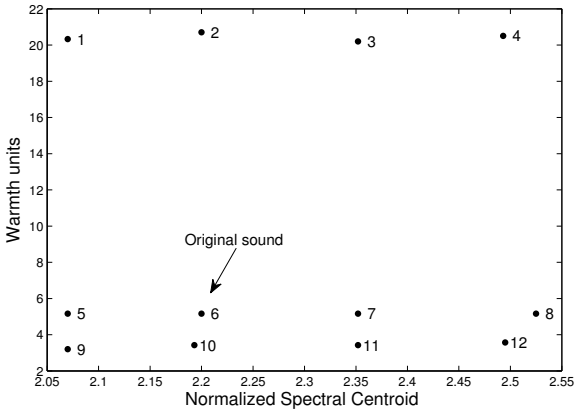

Fig. 1: Feature space of the twelve stimuli.

up to three verbal terms that could describe how the second sound in the pair was different from the first (Figure 2). Again each pair could be played back as many times as the listener required prior to submitting a description. The consistency of each listener's responses was tested by including two identical pairs of sounds in the test thus increasing the overall number to six.

The overall listening test lasted approximately 35 minutes and participants were advised to take breaks if they felt signs of fatigue.

4. RESULTS

4.1. MDS Analysis

The average dissimilarity matrices that were produced by the listener responses for both two different fundamental frequencies were analysed by the Multidimensional Scaling (MDS) ALSCAL algorithm in SPSS [1]. This analysis attempted to shed some light on the perceptual effect of the brightness–warmth algorithm by investigating the dimensionality of the perceptual space and by positioning the stimuli in it. The 'measures-of-fit' calculated by the MDS analysis were examined in order to determine the optimal number of dimensions for this set of data. Tables 1 and 2 show the squared correlation factors (RSQ) and the s-stress tests for up to a 3-D solution.

As the number of dimensions increases, RSQ will naturally be increasing while s-stress will be decreasing. It is up to the researcher to decide when the further increase and decrease of these measures should be attributed to

[1] All twenty subjects rated identical pairs as being identical (0 value) so none was excluded from the analysis.

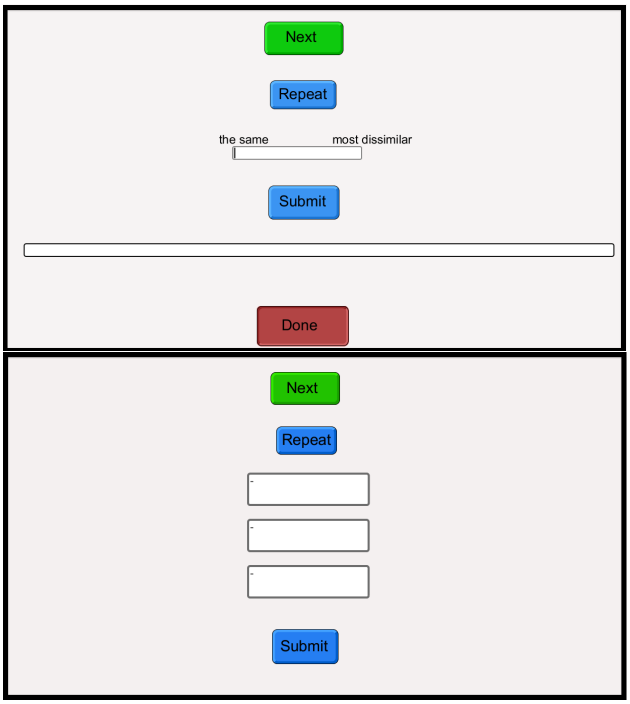

Fig. 2: Sections of the listener interface. Pairwise dissimilarity test where listeners were asked to rate the similarity between a pair of stimuli using the horizontal slider (Top). Verbal elicitation test where listeners were asked to input up to three verbal descriptors that were appropriate for characterizing the difference between selected pairs of stimuli (Bottom).

modelling noise in the data rather than to modelling the actual structure of the data. For the F0 = 220 Hz case the movement from 1-D to 2-D solution results to an increase on the order of 0.1654 for the RSQ as well as to a significant decrease of the s-stress (Table 1). Adding a 3rd dimension brings a negligible improvement to the measures (improvement < 0.05 for the RSQ) that could be as well attributed to noise. Thus, the optimal fit to the data appears to be the 2-D solution. The same can be supported also for the 440 Hz case, however with slightly worst results for the 'measures-of-fit' values (Table 2).

The 2-D MDS spaces that are produced are shown in Figure 3. S_1-S_{12} represent the twelve stimuli (S_6 is the original stimuli) and the arrows suggest an interpretation of the perceptual space. Indeed, S_1-S_4 change only in terms of the spectral centroid and S_1-S_5-S_9 change only in terms of the relative energy of the first three partials (see Figure 1). In the 220 Hz case the position of

Table 1: Measures-of-fit for the MDS solution of the 220 Hz (Top) and the 440 Hz (Bottom) pairwise dissimilarity tests. The scree plots would have a 'knee' on the 2-D solution both for the RSQ and the s-stress values which is a good indication that a 2-D space will offer the optimal fit for this set of data. Maximum possible RSQ improvement at 3-D is given by 1-(3-D RSQ).

Dimensionality	RSQ	RSQ improvement to next dimension up	s-stress
1-D	0.78239	0.16545	0.2842
2-D	0.94784	0.01147	0.1274
3-D	0.95931	0.04069 (max)	0.09722

Dimensionality	RSQ	RSQ improvement to next dimension up	s-stress
1-D	0.81298	0.07454	0.299
2-D	0.88752	0.02622	0.1875
3-D	0.91374	0.08626 (max)	0.1348

these two groups of stimuli resembles the feature space quite closely as they appear orthogonal in the perceptual space. Orthogonality is becoming weaker for stimuli with higher spectral centroids which are also perceived as being lower in the warmth dimension (S_8 and S_{12}). Additionally, for sounds with higher SC a decrease in warmth is also perceived as an increase in brightness (for example S_2-S_6 and S_3-S_7). This is an indication that for sounds with higher spectral content the modification of the SC and the warmth feature does not have a totally independent perceptual effect. Furthermore, the warmth feature relationship with perception seems to be a logarithmic one as the perceptual distances among S_1-S_5-S_9 are almost equal while in the feature space the S_1-S_5 distance is roughly ten times larger than S_5-S_9. Finally, a widening of the perceptual space for sounds with higher spectral centroid is obvious as S_1-S_9 appear closer than S_4-S_{12} even though they are equidistant in the feature space.

For the 440 Hz case the matching of the feature space to the perceptual space is not that close. The S_1-S_5-S_9 group is again positioned somewhat independently from the the S_1-S_2-S_3-S_4 group but the angle between them is certainly less than 90°. Sounds with higher spectral centroid such as S_7-S_8-S_{10}-S_{11} are clustered together in the high brightness, medium warmth region. However, the space is still expanded for higher SCs (S_4-$S_{12} > S_1$-S_9).

The Pearson's correlation coefficients between the per-

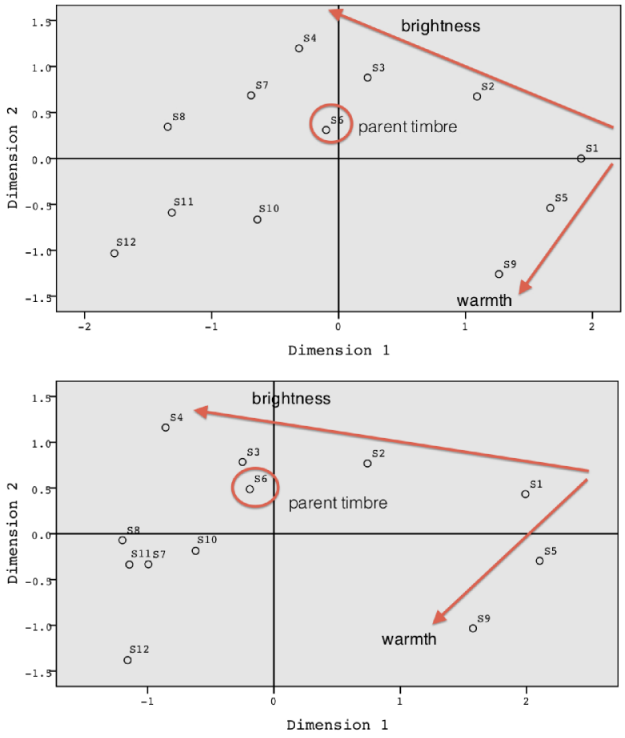

Fig. 3: The two perceptual spaces created by the MDS analysis. The 220 Hz stimuli have a closer matching to the feature space (Top) than the 440 Hz (Bottom). The brightness arrow shows the direction of SC increase and the warmth arrow shows the direction of warmth decrease.

ceptual space dimensions and some spectral features extracted from the sounds are shown in Tables 3 and 4. T_1, T_2 and T_3 stand for Tristimulus 1, 2 and 3 which are shown at Eq. 16, 17 and 18 [9]:

$$T_1 = \frac{A(1)}{\sum_{k=1}^{K} A(k)} \tag{16}$$

$$T_2 = \frac{\sum_{k=2}^{4} A(k)}{\sum_{k=1}^{K} A(k)} \tag{17}$$

$$T_3 = \frac{\sum_{k=5}^{K} A(k)}{\sum_{k=1}^{K} A(k)} \tag{18}$$

where $A(k)$ are the amplitudes of the k harmonics of a harmonic sound.

Table 2: Fundamental of 220 Hz. Pearson's correlation coefficients between SC, warmth feature and Tristimulus 1, 2, 3 and the dimensions of the rotated MDS space. D_1 is parallel to the direction $S_1 \rightarrow S_5 \rightarrow S_9$ and D_2 parallel to the direction $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$.

Dimensions	SC	Warmth	T1	T2	T3
D1	-0.28	**0.80**	-0.59	**0.935**	**-0.81**
D2	**0.91**	-0.176	-0.44	-0.58	**0.83**

Table 3: Fundamental of 440 Hz. Pearson's correlation coefficients between SC, warmth feature and Tristimulus 1, 2, 3 and the dimensions of the rotated MDS space. D_1 is parallel to the direction $S_1 \rightarrow S_5 \rightarrow S_9$ and D_2 parallel to the direction $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_4$.

Dimensions	SC	Warmth	T1	T2	T3
D1	-0.30	**0.82**	-0.57	**0.90**	**-0.77**
D2	**0.87**	-0.49	-0.53	-0.38	**0.79**

The two MDS spaces were rotated clockwise by 60° and 72° in order to achieve an alignment between the warmth and brightness axes with dimensions 1 and 2 correspondingly. It is clear that D_2 is highly correlated to the spectral centroid. D_1 on the other hand seems to have a significant correlation with the warmth feature but is more correlated to T_2. T_3 has both a positive correlation with D_2 and a negative correlation with D_1. Similar results hold for both fundamentals, except for the correlation of warmth with D_2 which is -0.49 for F0 = 440 Hz when only -0.176 for F0 = 220 Hz. This shows that there is some dependency between D_2 and warmth for F0 = 440 Hz that was not present for F0 = 220 Hz.

4.2. Verbal Elicitation

The Tables 5 to 12 show the results of the verbal elicitation part of the test for the two different fundamental frequencies.

All twenty subjects were consistent with their verbal judgements between identical pairs. They usually did not use the exact same verbal descriptors for both cases but the context was always the same. The groupings were made based on semantic relevance and according to the groupings in [6], [10], [11]. Words in bold indicate the word with higher frequency of appearance within the group.

The pair S_1-S_{12} represents a difference from the maxi-

Table 4: Verbal elicitation results for the S_1-S_{12} pair (F0 = 220 Hz).

Clusters of adjectives used to describe how S_1 differs from S_{12}	**Number of occurrences**	**Percentage of the total number of answers.**
bright, clear, trebly	8	40%
fuzzy, crackly, buzzy, harsh, robotic, less round	6	30%
small, thin, tight	3	15%
various	3	15%

Table 5: Verbal elicitation results for the S_4-S_9 pair (F0 = 220 Hz).

Clusters of adjectives used to describe how S_4 differs from S_9	**Number of occurrences**	**Percentage of the total number of answers.**
dull, gloomy, damp, muted, closed	7	28%
soft, smooth	6	24%
warm, round	5	20%
full, dense	3	12%
various	4	16%

Table 6: Verbal elicitation results for the S_2-S_{10} pair (F0 = 220 Hz).

Clusters of adjectives used to describe how S_2 differs from S_{10}	**Number of occurrences**	**Percentage of the total number of answers.**
bright, treble	9	56%
big, full, open	3	19%
harsh, buzzy	2	12.5%
warm, less pleasant	2	12.5%

Table 7: Verbal elicitation results for the S_5-S_8 pair (F0 = 220 Hz).

Clusters of adjectives used to describe how S_5 differs from S_8	**Number of occurrences**	**Percentage of the total number of answers.**
bright, nasal, clear, treble	11	55%
thin	3	15%
various	6	30%

Table 8: Verbal elicitation results for the S_1-S_{12} pair (F0 = 440 Hz).

Clusters of adjectives used to describe how S_1 differs from S_{12}	**Number of occurrences**	**Percentage of the total number of answers.**
bright, **sharp**, less muffled, nasal, edgy	13	65%
ring, harsh, metallic	3	15%
thin, reedy, less brass	3	15%
less warm	1	5%
full	1	5%

Table 9: Verbal elicitation results for the S_4-S_9 pair (F0 = 440 Hz).

Clusters of adjectives used to describe how S_4 differs from S_9	**Number of occurrences**	**Percentage of the total number of answers.**
warm, round	7	30%
dark, **dull**, less nasal	7	30%
muffled, smooth, less harsh	5	22%
thick, more body	2	9%
various	2	9%

Table 10: Verbal elicitation results for the S_2-S_{10} pair (F0 = 440 Hz).

Clusters of adjectives used to describe how S_2 differs from S_{10}	**Number of occurrences**	**Percentage of the total number of answers.**
bright, less dull, nasal	10	53%
rich, full, more harmonics	3	17%
thin	2	10%
harsh, punchy	2	10%
various	2	10%

Table 11: Verbal elicitation results for the S_5-S_8 pair (F0 = 440 Hz).

Clusters of adjectives used to describe how S_5 differs from S_8	**Number of occurrences**	**Percentage of the total number of answers.**
bright, nasal, penetrating, sharp, edgy	11	55%
harsh	2	10%
less round, less warm	2	10%
various	5	25%

mum warmth and minimum brightness to the maximum brightness and minimum warmth. The most prominent group of answers is the one that includes the descriptor 'bright'. This is more clear for F0 = 440 Hz and is an indication that the perceptual effect for the simultaneous increase of SC and decrease of the warmth feature is an increase in brightness. Only one out of forty one answers was 'less warm' even though the warmth feature had roughly decreased by 80% of its initial value and SC had only increased by 25%.

The S_4-S_9 pair provides even more revealing results. The movement for this pair is from maximum brightness and warmth to minimum brightness and warmth. The results for both F0 do not suggest a unique prominent group but rather three groups of descriptors that have the highest frequency of appearance. Sound S_9 is generally rated as being warmer, duller or darker and softer or smoother. This fact implies that the perception of brightness overshadows the perception of warmth, and that warmth might be the perceptual antonym of brightness. Indeed, no-one rated S_9 as being less warm. On the contrary many participants actually described it as being warmer. This shows a discrepancy between the suggested warmth feature and the actual perception of warmth and a high level of overlapping between brightness and warmth.

The S_2-S_{10} pair represents a movement from maximum to minimum warmth having a constant SC position. Tables 7 and 11 show that the brightness group predominates with very similar results for both F0s. This result reveals that participants rated S_{10} as brighter, even though the position of the SC was exactly the same with S_2. This agrees with the MDS spaces that position S_{10} away from S_2 both in warmth and brightness direction. Despite the fact that distance in warmth is greater than distance in brightness, it is the latter that is spontaneously verbalized. The fact that no-one responded 'less warm' or 'colder' needs to be highlighted and contributes to the hypothesis of warmth being a perceptual antonym for brightness.

Finally, the S_5-S_8 pair represents an increase of the SC while keeping the warmth feature constant. The results are again quite similar for both F0s and indicate that the brightness group is the more important one but at the same time there is a significant number of non grouped responses. This is a quite expected result that confirms previous research regarding the perceptual effect of the spectral centroid's position [1], [2], [3], [4], [5].

5. CONCLUSIONS

An algorithm for the independent modification of the spectral centroid and the relative energy of the first three partials of a harmonic sound was designed and implemented on a Max/MSP additive synthesis engine. The perceptual validity of these two features together with the possible influence of the fundamental frequency were tested through two pairwise dissimilarity listening tests.

The 2-D spaces that were produced by the Multidimensional Scaling Analysis show a relatively good matching between the feature space and the perceptual space for F0 = 220 Hz. It is also clear that for low spectral centroids the modification of these two features is perceived independently and that for higher spectral centroids there seems to be a degree of overlap between them. For F0 = 440 Hz the matching between the two spaces worsens significantly but there still is evidence of perceptual independence for lower SCs. The correlation between the rotated axis of the space (so that they coincide with what seems to be the basic directions of movement on the MDS space) and some spectral features are calculated. A strong correlation between D_2 and the SC and a correlation between the warmth feature and D_1 are revealed for both fundamentals. D_2 demonstrates some correlation with the warmth feature for the 440 Hz case which, together with the difference between the MDS spaces, suggests that fundamental frequency might have some influence on the perception of this particular modification. However, the level of differences noticed, as well as the range of frequencies used, are not sufficient to support this finding. Further research needs to be done on this direction. Here it must be stated that Tristimulus 2 features the strongest correlation with D_1 and also that Tristimulus 3 features a quite strong negative correlation with D_1. That is a sign that T_2 or/and T_3 might be more influential on the listeners' judgements than the warmth feature.

Although the MDS analysis shows that a degree of perceptual independence among sounds with different warmth and SC does exist the verbal elicitation experiment does not support this finding. 'Bright' was the most prominent semantic descriptor that was elicited through the free response test for describing an increase in the SC, decrease in warmth and a combination of the two. For the decrease in warmth and SC the terms varied and the three most prominent terms were dull, warm and soft.

The above results put under question the claim that the relative energy of the first three partials is an adequate

acoustic correlate for the auditory perception of warmth. Furthermore, they seem to agree with previous works that there exists a degree of overlap between the auditory perception of brightness and warmth [11], [12], [13]. Another interesting finding is the fact that sounds with the same spectral centroid are rated as differing in brightness implying that it is not merely dependent on the spectral centroid position.

Future work will attempt to examine the exact relationship between the auditory perceptions of brightness and warmth and to seek for a more reliable acoustic correlate for warmth. Possible candidates will be Tristimulus 1, inharmonicity, MFCC coefficients and even temporal characteristics like attack time and spectrotemporal ones like 'incoherence' and vibrato. The validity of the findings from listening tests will be tested through an additive synthesis engine that will enable manipulation of the target low-level features. Participants will be asked to modify a given sound based on a semantic description through a graphical user interface. The results should shed some light on the perceptual significance of certain low-level audio features.

6. ACKNOWLEDGEMENTS

The authors would like to thank the subjects who took part in the listening tests as well as Dr. Konstantinos Pastiadis and Dr. Georgios Papadelis from the Music Department of the Aristotle University of Thessaloniki for their contribution to the analysis of the results.

7. REFERENCES

[1] Lichte W. H., "Attributes of complex tones", Journal of Experimental Psychology 28, 455-480, 1941

[2] von Bismarck G., "Timbre of steady tones: A factorial investigation of its verbal attributes", Acustica 30, 146-159, 1974.

[3] McAdams S., Winsberg S., Donnadieu S., De Soete S ., Krimphoff J., "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes" Psychol Res: 58 pp 177-192, 1995.

[4] Williams D. and Brookes T., "Perceptually-motivated audio morphing: brightness", Presented at the 122th Convention, Vienna, Austria, May 5–8, 2007.

[5] Schubert E., Wolfe J. and Tarnopolsky A., "Spectral centroid and timbre in complex, multiple instrumental textures", Proceedings of the 8th International Conference of Music Perception and Cognition, Evaston, IL, 2004.

[6] Williams D. and Brookes T., "Perceptually-motivated audio morphing: warmth", Presented at the 128th Convention, London, U.K., May 22-25, 2010.

[7] Neher T., Brookes T. and Rumsey F., "A hybrid technique for validating unidimensionality of perceived variation in a spatial auditory Stimulus Set", Journal of Audio Engineering Society, Vol. 54, No. 4, April 2006.

[8] Jensen K., "Timbre Models of Musical Sounds", PhD thesis, University of Copenhagen, Denmark, 1999.

[9] Pollard H. F. and Jansson E. V., "A tristimulus method for the specification of musical timbre.", Acustica 51, 162-71, 1982.

[10] Moravec O. and Štepánek J., "Verbal description of musical sound timbre in Czech language", Proceedings of the Stockholm Music Acoustics Conference, SMAC03, Stockholm, Sweden, 2, 643–645, 2003.

[11] Howard David M., Disley Alastair C. and Hunt Andrew D., "Timbral adjectives for the control of a music synthesizer", 19th International Congress on Acoustics, Madrid, 2-7 September, 2007.

[12] Ethington R. and Punch B., "SeaWave: A system for musical timbre description", Computer Music Journal, Vol. 18, 1994.

[13] Pratt RL and Doak PE, "A subjective rating scale for timbre", Journal of Sound and Vibration, Vol. 45, 1976.