A NONLINEAR METHOD FOR MANIPULATING WARMTH AND BRIGHTNESS

Sean Enderby
Digital Media Technology Lab
Birmingham City University
Birmingham, UK
sean.enderby@mail.bcu.ac.uk

Ryan Stables
Digital Media Technology Lab
Birmingham City University
Birmingham, UK
ryan.stables@bcu.ac.uk

ABSTRACT

In musical timbre, two of the most commonly used perceptual dimensions are *warmth* and *brightness*. In this study, we develop a model capable of accurately controlling the *warmth* and *brightness* of an audio source using a single parameter. To do this, we first identify the most salient audio features associated with the chosen descriptors by applying dimensionality reduction to a dataset of annotated timbral transformations. Here, strong positive correlations are found between the centroid of various spectral representations and the most salient principal components. From this, we build a system designed to manipulate the audio features directly using a combination of linear and nonlinear processing modules. To validate the model, we conduct a series of subjective listening tests, and show that up to 80% of participants are able to allocate the correct term, or synonyms thereof, to a set of processed audio samples. Objectively, we show low Mahalanobis distances between the processed samples and clusters of the same timbral adjective in the low-dimensional timbre space.

1. INTRODUCTION

The perception and manipulation of musical timbre is a widely studied aspect of sound production. This is because timbre, unlike pitch and loudness, is difficult to measure linearly along a single intuitive dimension. This means the focus of timbral research is often on the use of dimensionality reduction [1, 2] as a method of interpreting some complex representation in timbre space. In the study of musical timbre, natural language is often used to define perceptual dimensions [3]. This is a method of quantifying descriptions of sound, often through the use of correlation with audio features [4]. In music production, this descriptive vocabulary can also be used to define sound transformations [5]. This means we are able to control audio processing functions using parameters that are intuitive to the user as they represent high-level perceptual aspects of sound, as opposed to low-level algorithm parameters. For example, SocialEQ [6] allows participants to select a descriptive term, then to derive its spectral representation by rating a series of examples. Similarly, the SAFE Project [7] allows users to describe audio effects transformations directly within a Digital Audio workstation (DAW), which can then be processed and recalled to crowd-source processing parameters.

One of the most common perceptual dimensions in timbral research is the warmth / brightness dimension [8, 9]. This is because participants often tend to agree with confidence on the statistical representation of the two descriptive terms [4], and they are often considered to be orthogonal [10]. Because of this, a number of studies have focussed specifically on manipulating this dimension for musical purposes. For example, Stasis et al. [11, 10] provide a 2-dimensional interface, derived using machine learning, Zacharakis et al. [12] present a model using FM-Synthesis, and Williams and Brookes [13, 14] provide a timbre morphing system, capable of independently modifying warmth and brightness.

In this study, we identify the key audio features associated with warmth and brightness from a dataset of annotated musical transforms. We then propose an audio effect that combines linear and nonlinear modules in a way that allows us to manipulate the audio features directly associated with the perception of warmth and brightness in a sound source. To validate the performance of the effect, we then provide both objective and subjective validation.

2. PERCEPTUAL DIMENSIONS

To identify salient audio features associated with warmth and brightness, we compile a matrix of timbral data collected via the SAFE Project[1]. Here, annotated audio effect transformations are collected from within a DAW environment and uploaded anonymously. Each instance contains a set of plug-in processing parameters, an extensive feature set taken before and after processing has been applied, a string of descriptive terms, and a table of user metadata. As shown in [4], warmth and brightness tend to be related to changes in the spectral envelope associated with equalisation and distortion, so we therefore discard entries from compression and reverb effects. This leaves us with 1,781 semantically annotated transforms. As the tests are distributed over a wide network we do not have extensive data about the test participants, however it is assumed that each of the users of the system have a reasonable level of production experience. To capture the timbral modification applied by each of the transforms, we analyse the feature differences over a range of temporal, spectro-temporal and abstract statistical audio features.

2.1. Features

To identify the most salient features associated with the *warmth / brightness* dimension, we apply dimensionality reduction to the dataset using Principal Component Analysis (PCA) and identify the most highly correlated features with the Principal Components (PCs) that explain the highest variance. This is demonstrated in Figure 1, in which the first two PCs of the timbre space describing the feature manipulations are shown. Here, the audio feature differences of each described transform is projected into two dimensional space, and the centroid for each term is found. The size of the term indicates its relative confidence (a measure which is inversely proportional to the variance within a cluster). These confidence scores for entries in the dataset are given in Table 1. Additionally, Figure 2 shows the isolated transforms described as either

[1] Data and plug-ins available at: www.semanticaudio.co.uk

warm or bright. This shows the distribution of points from each class are separable along PC 2. In the other PCs, these descriptors occupy very similar ranges, suggesting that the distinction between warm and bright is heavily loaded onto the second PC. To identify the salient features associated with each dimension, we correlate each feature vector with the first two PCs. The audio features with correlations which satisfy $|r| > .7$ and $p < .001$ are show below.

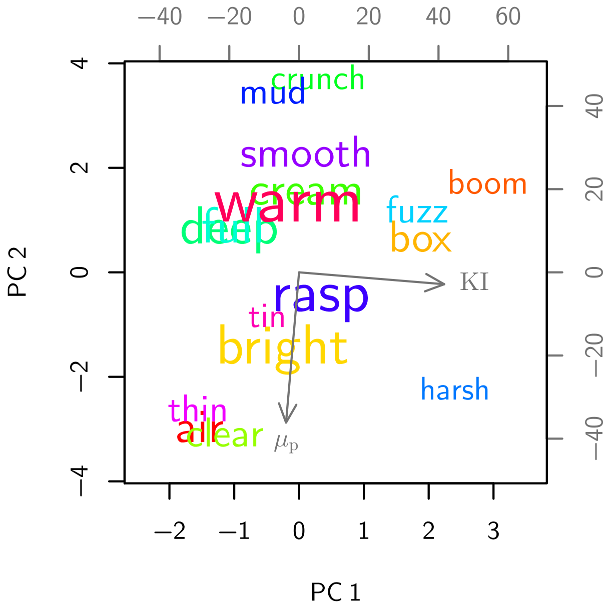

Figure 1: *A biplot of the feature difference timbre space, where μ_p and KI represent the peak spectral centroid and spectral irregularity projected into low-dimensional space.*

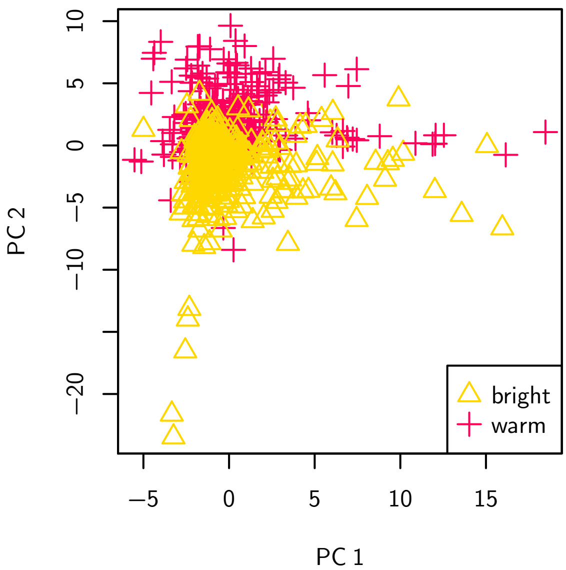

Figure 2: *Transforms labelled with warm and bright in the feature difference timbre space.*

Term	C	Term	C	Term	C
warm	1.5	box	0.7	boom	0.4
deep	1.4	clear	0.5	tin	0.4
bright	1.1	thin	0.5	crunch	0.3
full	0.9	mud	0.4	harsh	0.2
air	0.8	fuzz	0.4	smooth	0.1
cream	0.8	rasp	0.4		

Table 1: *The confidence scores (C) for terms in the feature difference timbre space.*

PC 1: Irregularity ($r = 0.985$), Irregularity$_p$ ($r = 0.964$), Kurtosis$_p$ ($r = 0.935$), Skew$_s$ ($r = 0.929$), Irregularity$_h$ ($r = 0.927$), Kurtosis$_h$ ($r = 0.919$), Skew$_p$ ($r = 0.890$), Std ($r = 0.873$), RMS ($r = 0.873$), Skew$_h$ ($r = 0.865$), Kurtosis$_s$ ($r = 0.835$), Var ($r = 0.812$).

PC 2: Centroid$_p$ ($r = -0.855$), Centroid$_h$ ($r = -0.853$), Rolloff$_s$ ($r = -0.852$), Std$_h$ ($r = -0.845$), Std$_p$ ($r = -0.834$), Centroid$_s$ ($r = -0.817$), Slope$_s$ ($r = -0.771$).

Where, the subscript s denotes features extracted from a magnitude spectrum, p denotes features taken from a peak spectrum, and h denotes features taken from a harmonic peak spectrum. Features with no subscript are either temporal or abstract spectral features. Spectral irregularity in this study is calculated using the method described by Krimpoff et al. [15].

These results indicate that the dimension along which warmth and brightness can be considered separable (PC 2), is highly correlated with spectral centroid, spectral standard deviation and spectral roll off. Negative values of PC 2 correspond to an increase in these features and positive values to a decrease. Signals can therefore be made warmer by reducing the spectral centroid and 'brighter' by increasing it. This is most often done by introducing more energy at low or higher frequencies respectively. These findings are aligned with similar studies of musical timbre such as [8, 2], in which the spectral centroid of a sound source is demonstrated to be a salient low-level descriptor in the perception of both warmth and brightness.

2.2. Synonyms

Given that the dataset has a large number of descriptive terms, we apply agglomerative clustering to the feature space in order to identify potentially synonymous terms in the dataset. This allows us to judge the relative distances between data points in this space, thus providing a method of evaluating dissimilarities during subjective evaluation. Terms with less than 4 entries are omitted for readability and the distances between data points are calculated using Ward criterion [16]. The clusters are illustrated in Figure 3, where the prefix *E:* represents transforms taken from the equaliser and *D:* represents transforms taken from the distortion. The position, $\mu_{d,k}$ of a term, d, in the k^{th} dimension of the audio feature space is calculated as the mean value of feature k across all N_d transforms labelled with that descriptor, given in Eq. 1.

$$\mu_{d,k} = \frac{1}{N_d} \sum_{n=1}^{N_d} \bar{x}_{d,n,k} \qquad (1)$$

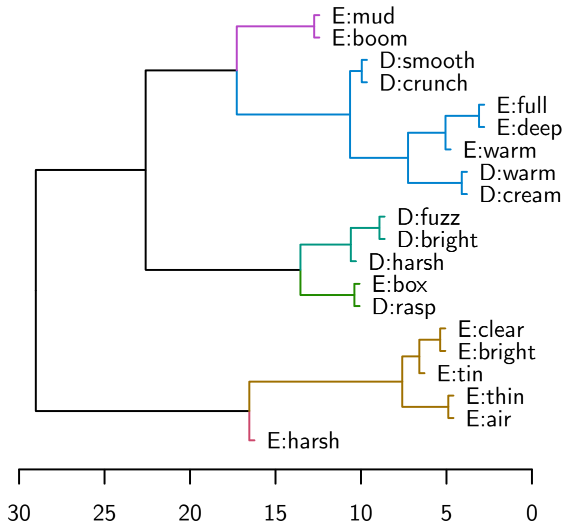

Figure 3: *Clustering of descriptors from the both the distortion and equaliser.*

The agglomerative clustering process demonstrates that the data points tend to cluster based on relative spectral energy. For example, terms that typically describe signals with increased high frequency energy such as *bright*, *thin*, *tin* and *air* all have low cophenetic distances to each other. Similarly, terms typically associated with increased low frequency energy such as *boom*, *warm*, *deep* and *full* fall within the same cluster. In this case, the cluster containing warm and the cluster containing bright are clearly separated, with a minimum cophenetic distance of 22.6.

3. PARAMETERISATION OF SPECTRAL CENTROID

Given the correlation between the *warmth* / *brightness* dimension (PC 2) and spectral centroid, we investigate methods for manipulating the feature directly, thus being able to increase *brightness* by increasing the centroid, or increase *warmth* by lowering the spectral centroid. A primitive method for moving the centroid towards a given frequency (μ) is to increase the spectral energy at that frequency, either by introducing a sinusoidal component to the signal or to use a resonant filter, centered around μ. Whilst this works conceptually, it is destructive to the original structure of the spectrum. As the centroid is moved towards the desired frequency the spectrum is dominated by a sinusoid or resonance at μ.

Less destructive methods include that used by Zacharakis et al [12], where the spectrum is split into two bands, one above and one below the spectral centroid. The relative amplitudes of these bands can then be altered to manipulate the frequency of the spectral centroid. This more accurately preserves the original signal's structure, as no additional spectral components are introduced and the relative levels of partials within each band remain the same. Using this method the new spectral centroid will lie somewhere between the respective centroids of the two bands. The relative gains of the two bands required to reproduce a given spectral centroid μ, can be calculated using Equation 2. To facilitate precise control of the spectral centroid these bands should not share any frequency components.

$$\frac{\sum_{n=c+1}^{N} a_n}{\sum_{n=1}^{c} a_n} = \frac{\mu_l - \mu}{\mu - \mu_u}, \quad \mu_l \leq \mu < \mu_u \quad \text{or} \quad \mu_u < \mu \leq \mu_l \tag{2}$$

Where μ_l and μ_u are the spectral centroid of the lower and upper bands, and c is the highest frequency spectral component in the lower band.

Alternatively, Brookes et al. [13] employ a spectral tilt to modify spectral centroid, applying gain to the partials of a signal as a linear function of their frequency. This allows the spectral centroid to be altered in frequency, whilst still retaining the frequency content of the signal. A disadvantage of this method is that the change in centroid cannot be easily parameterised as it depends on the content of the signal being processed.

3.1. Proposed Model

We propose a more flexible method for directly manipulating the spectral centroid of the input signal using a nonlinear device (NLD). The effects of an NLD are more easily predicted for sinusoidal inputs, generating a series of harmonics of the input frequency. To ensure a sinusoidal input to the NLD, the system is restricted to processing only tonal signals with a single pitch, which can be represented by their fundamental frequency (f_0). A low-pass filter is applied to isolate the f_0 which is then processed with an NLD generating a series of harmonics relevant to the signal. This is then high-pass filtered, leaving a band that consists solely of generated harmonics. A second band is generated by low-pass filtering the signal at the spectral centroid and the relative levels are then adjusted in order to manipulate the frequency of the centroid μ_s. Separating the bands at the spectral centroid in this way ensures that their respective spectral centroids sit either side of the input's centroid after processing has been applied. In this instance, we generate harmonics in the high frequency band by applying Full Wave Rectification (FWR) to the isolated fundamental, this ensures the system is positive homogeneous. A schematic overview of the system is presented in Figure 4.

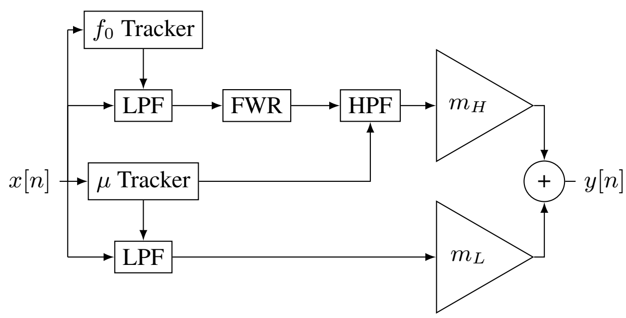

Figure 4: *The system employed in the warmth / harshness effect.*

This is conceptually similar to the two band method proposed by [12], but allows the second band to contain frequency content which was not in the original signal. This method has advantages over manipulating the amplitudes of two bands using only linear filters. For example, the nonlinear device can still be used to reconstruct the high frequency portion of the signal and the relative gains adjusted similarly to if two filters were used. Alternatively, the properties of the NLD can be altered to change the upper band's centroid. This provides more flexibility allowing the centroid to be changed independently of some other features. For example, changing the gains of two bands will change the spectral slope of the signal. If instead additional partials are introduced to the upper band, with amplitudes which are determined by the signal's current slope, the centroid can be changed, whilst the slope is unaltered.

The effect is controlled using a single parameter p, which ranges from 0 to 1 and is used to calculate the respective gains m_L and m_H, applied to the low and high frequency bands using Equation 3.

$$\begin{aligned} m_H &= p^3 \\ m_L &= 1 - m_H \end{aligned} \tag{3}$$

When $p = 0$ the output is a low pass filtered version of the input signal resulting in a lower spectral centroid than the input. This corresponds to transforms described as *warm* in the SAFE dataset. When $p = 0.5$ additional harmonic energy is introduced into the signal, meaning the transform should be perceived as *bright*. To achieve this, the exponent in p^n was set experimentally, so that the Mahalanobis distance between the *bright* cluster and the transform's feature differences is minimised. When $p = 1$ the output signal consists primarily of high order harmonics resulting in an extreme increase in spectral centroid. This is perceived as Harshness, which in the SAFE dataset is defined as an increase in spectral energy at high frequencies, as shown in Figure 5.

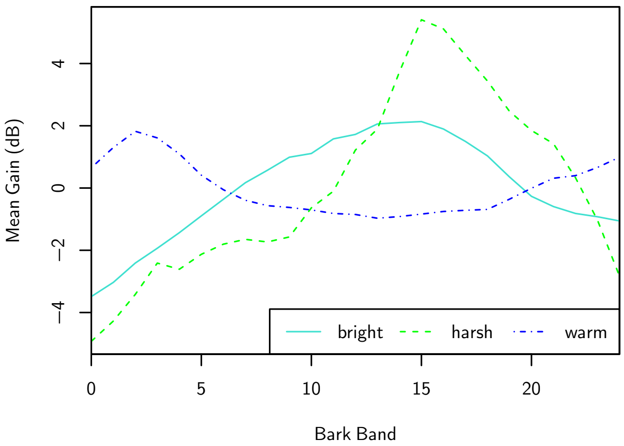

Figure 5: *Mean Bark spectra of transforms labelled warm, bright and harsh in the SAFE dataset. Here, Bark spectra are used to represent the spectral envelopes of the transforms as these are the only spectral representations collected in-full by the SAFE plug-ins due to data protection.*

When transforms are applied using a distortion, *bright* and *harsh* are considered to be very similar, with a low cophenetic distance of 10.6 (see Figure 3), however they exhibit some subtle differences in transforms taken from the equaliser, with a cophenetic distance of 16.5. This is demonstrated in Figure 1, where *harsh* sits below *bright* in PC 2.

4. MODEL VALIDATION

The performance of the effect is evaluated using a set of ten test signals, comprising two electric bass guitars (B1 and B2), a flute (F), two electric guitars (G1 and G2), a marimba (M), an oboe (O), a saxophone (S), a trumpet (T) and a violin (V). The signals were adjusted to have equal loudness prior to experimentation. Firstly the effects are evaluated objectively by comparing them to the analysis performed in section 2. Secondly the effects are evaluated subjectively through a series of perceptual listening tests.

4.1. Objective Evaluation

The performance of the effect is evaluated objectively by examining how it manipulates the features of the test signals. The effect is used to process each of the signals with its parameter p set to 0 (*warm*), 0.5 (*bright*) and 1 (*harsh*). The audio features of the unprocessed and processed signals in each of these applications are calculated in the same manner as in the SAFE plug-ins. These audio features are then compared to those taken from the SAFE dataset.

Each combination of descriptor and test signal is measured to find the distance between changes in the feature space caused by the effect, and points labelled with the descriptor in the feature difference timbre space. The performance of the effect, with a particular parameter setting on a particular test signal is measured by projecting the extracted audio features to a point in the timbre space. The Mahalanobis distance, $M(x, d)$, between this point x, and the distribution of transforms labelled with the relevant term d, is taken using Equation 4

$$M(x, d) = \sqrt{(x - \mu_d)^T \Sigma_d^{-1} (x - \mu_d)} \tag{4}$$

Where x is a column vector containing the coordinates of the point in the timbre space, μ_d is a column vector containing the mean coordinates of all transforms in the timbre space labelled with descriptor d and Σ_d is the covariance matrix of those transforms' coordinates in the timbre space. Where there are more than five transforms in the distribution, the coordinates in the first five PCs of the timbre space are used. Where the number of points in the distribution, N_d, is lower, only the first $N_d - 1$ coordinates can be used in order to avoid Σ_d being singular. Where the descriptor, d, is represented by two distributions of transforms, one from the distortion and one from the equaliser, the Mahalanobis distance from both distributions is taken and the minimum distance is considered the measure of performance.

4.2. Subjective Evaluation

To assess the performance of the effect subjectively, a series of perceptual listening tests were undertaken. For the purposes of testing, the *warmth / brightness* effect was implemented a DAW plug-in. Participants were presented with a DAW session containing a track for each of the test signals. The plug-in was used on each track with a simple interface labelled "Plug-In 1", shown in Figure 6. To mitigate influence of the plug-in's interface layout on the result of the experiment, the direction of the parameter slider was randomised for each participant. The order of the tracks in the DAW session was also randomised to mitigate any effect the order of tracks may have on results.

For each signal, participants were asked to first audition the effect to become accustomed to how changing the parameter value affects that particular input signal. Once they had investigated the operation of the effect they were asked to label the parameter slider at three positions (p is equal to 0, 0.5 and 1) with a term they

Figure 6: *The interface used for assessing the performance of the warmth / brightness effect.*

felt best described the timbre of the effect at that parameter setting. A list of available terms was provided in a drop down list at each of the 3 parameter values to be labelled, pictured in Figure 6. The available terms were *airy, boomy, boxy, bright, clear, creamy, crunchy, deep, full, fuzzy, harsh, muddy, raspy, smooth, thin, tinny and warm*. These were chosen for their confidence scores and number of entries in the SAFE dataset. For each combination of signal and parameter positions, there is an intended descriptor (those the effects were designed to elicit) and a descriptor provided by the participant.

We compare the participant responses against the hierarchical clustering performed in Section 2.2. The dendrogram shown in Figure 3 provides information about how similar the transforms described by certain terms are. This information can be used as a metric describing the proximity of the users' responses to the intended response. The proximity of two descriptors is measured as the cophenetic distance between the clusters in which the two descriptors lie. Where a descriptor appears twice in the dendrogram (from both the distortion and equaliser) the combination of points which yield the lowest cophenetic distance is used. All listening tests were undertaken using circumaural headphones in a quiet listening environment. In total 22 participants took part in the listening tests, all of whom reported no known hearing problems. On average participants took 25 minutes to complete the test.

5. RESULTS / DISCUSSION

5.1. Objective Evaluation

The Mahalanobis distances between the test signals after being processed by the effect and the distributions of corresponding transforms in timbre space are shown in Figure 7. Here, each of the instrument samples is processed using $p = 0$ (*warm*), $p = 0.5$ (*bright*) and $p = 1$ (*harsh*).

The results show that overall, the *warmth* setting is timbrally very similar to the corresponding entries into the SAFE dataset, with a mean Mahalanobis distance of 1.03 ($\sigma = 0.53$). *Bright* samples are also very similar, with a mean distance of 1.36 ($\sigma = 0.36$). *Harshness* however is less similar to the distribution of terms in the dataset $\mu = 2.41$, $\sigma = 1.71$. This is potentially due to the term's ambiguity, and relatively low-confidence. *Harshness* for distortion and *harshness* for equalisation, for example, fall into different groups when hierarchical clustering is applied (see Figure 3). Also, due to the relative number of dataset entries for each term (*warm* = 464 entries, *bright* = 427 entries, *harsh* = 8 entries), smaller distances from the *harsh* distribution are deemed

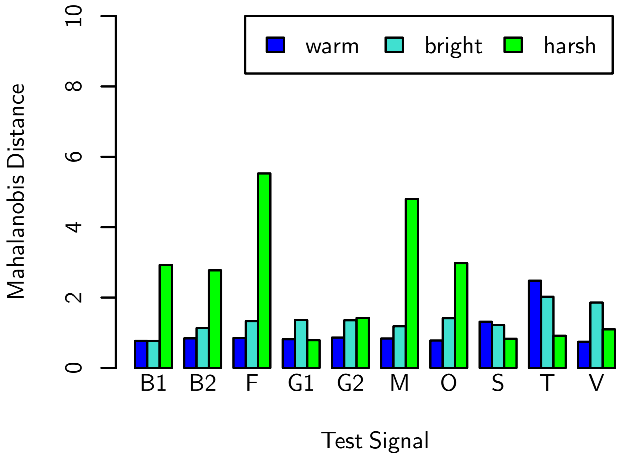

Figure 7: *Mahalanobis distances for the warmth / brightness effect.*

to be more statistically significant.

5.2. Listening test Results

The mean cophenetic distances between the participants' annotations of the effect's parameters and the descriptors *warm*, *bright* and *harsh* taken from the SAFE dataset are shown in Figure 8. Here the error bars represent the 95% confidence intervals for each mean. To show the performance of each instrument sample, markers on the y-axis indicate the cophenetic distances that correspond to the cluster heights for the groups containing *bright* from the distortion, *bright* from the EQ, and warm from both plug-ins, as per Figure 3.

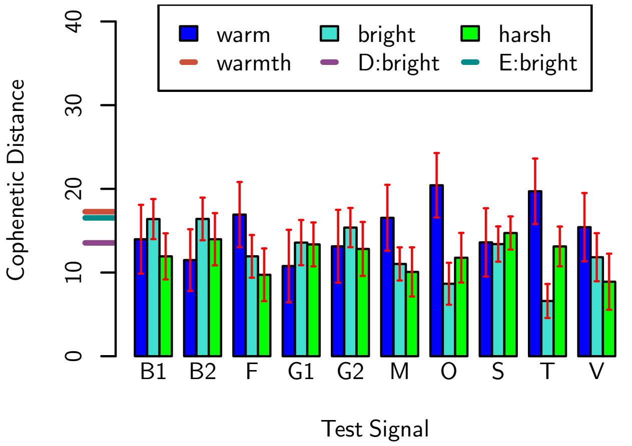

Figure 8: *Cophenetic distances for the warmth / brightness effect.*

The results show that almost all of the instrument samples have mean cophenetic distances that fall within the same cluster when the effect applies a *bright* or *harsh* transform. This means participants label parameter states with terms synonymous to the intended term. Samples processed to be *warm* also have similar sub-

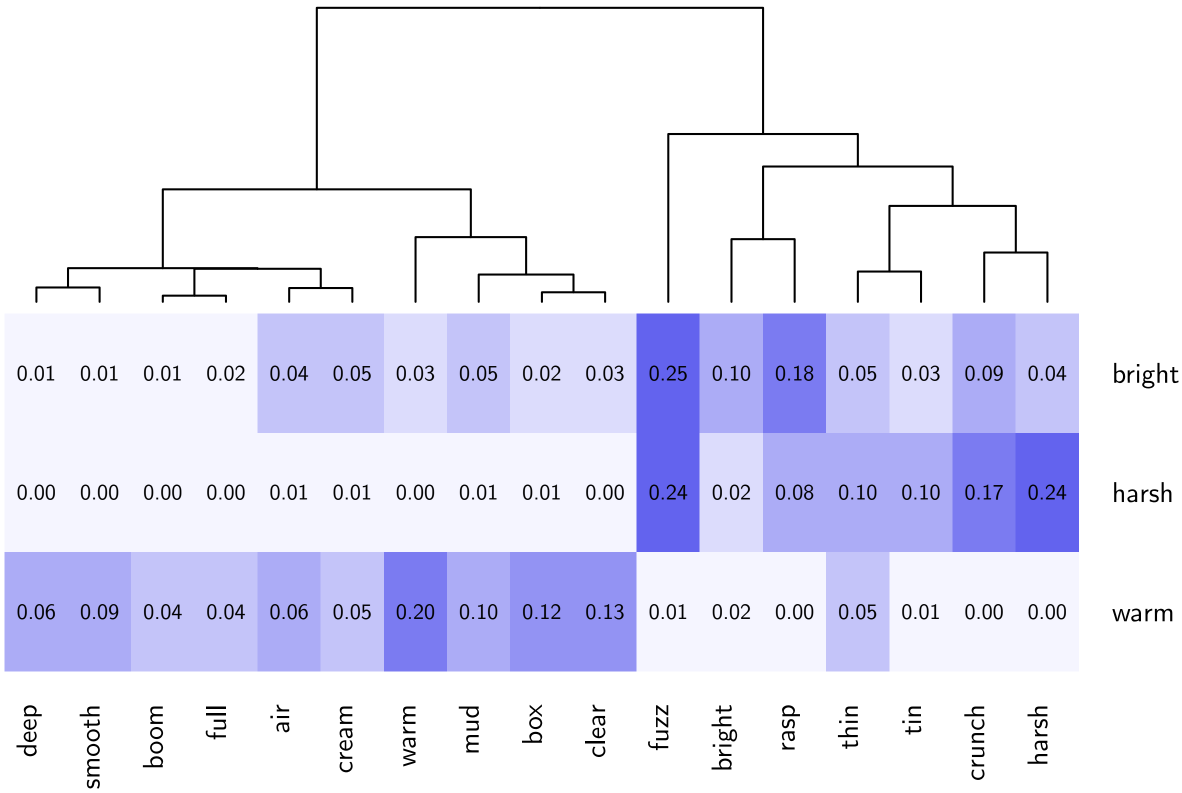

Figure 9: *A matrix showing subjective mis-classifications from the warmth / brightness effect, organised by the frequency of the mis-classifications across the effect's intended terms.*

jective accuracies, however two of the instrument samples (oboe and trumpet) have mean distances which are larger than the cluster heights, suggesting there is more ambiguity in their definition.

Figure 9, shows a mis-classification matrix comparing the usage of terms by test participants and the terms that the effects were designed to produce. Each cell in the matrix represents the frequency of which each of the available descriptors (bottom) was used to describe the corresponding timbral effect at a given parameter position. Above the figure is a dendrogram representing the clustering of terms based on their frequency of usage.

From the figure, it is clear that warm is often correctly assigned to the intended transform, but summing the cells of the row shows that participants only used descriptors in the same cluster as *warm* 54% of the time. This suggests that the addition of low frequency energy to the signal does not necessarily invoke the use of synonyms of *warm*. When describing the effect, participants used a term related to the intended descriptor 74% of the time for *bright* and 80% of the time for *harsh*, suggesting that these transforms were perceptually more similar to the transforms in the dataset.

6. CONCLUSION

We first present empirical findings from our analysis of a dataset of semantically annotated audio effect transforms, and show that the *warmth / brightness* dimension can be loaded heavily onto a single PC. The variance of this PC is explained predominantly by the centroid of the peak, harmonic peak, and magnitude spectra. We then present a model for the manipulation of timbral features using a combination of filters and nonlinear excitation, and show that the model is able to manipulate the respective *warmth* and *brightness* of an audio signal. We verify this by performing objective and subjective evaluation on a set of test signals, and show that subjects describe the transforms with synonymous descriptors 54%, 74% and 80% of the time for *warmth*, *brightness* and *harshness* respectively.

By using a NLD component in the feature manipulation process, we are able to increase the flexibility of the timbral modifier. This is because other audio features can be preserved, whilst the spectral centroid is modified independently. Conversely, the algorithm is currently limited to pitched monophonic sound sources due to its reliance on tracking the f_0 of the input signal. This issue will be addressed in future work.

7. REFERENCES

[1] J. M. Grey, "Multidimensional perceptual scaling of musical timbres," *the Journal of the Acoustical Society of America*, vol. 61, no. 5, pp. 1270–1277, 1977.

[2] A. Zacharakis, K. Pastiadis, J. D. Reiss, and G. Papadelis, "Analysis of musical timbre semantics through metric and non-metric data reduction techniques," in *Proceedings of the 12th International Conference on Music Perception and Cognition (ICMPC12)*, 2012, pp. 1177–1182.

[3] M. Sarkar, B. Vercoe, and Y. Yang, "Words that describe

timbre: A study of auditory perception through language," in *Proceedings of the 2007 Language and Music as Cognitive Systems Conference*, 2007, pp. 11–13.

[4] R. Stables, B. De Man, S. Enderby, J. D. Reiss, G. Fazekas, and T. Wilmering, "Semantic description of timbral transformations in music production," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 337–341.

[5] T. Zheng, P. Seetharaman, and B. Pardo, "Socialfx: Studying a crowdsourced folksonomy of audio effects terms," in *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 2016, pp. 182–186.

[6] M. Cartwright and B. Pardo, "Social-eq: Crowdsourcing an equalization descriptor map.," in *ISMIR*, 2013, pp. 395–400.

[7] R. Stables, S. Enderby, B. De Man, G. Fazekas, and J. D. Reiss, "Safe: A system for the extraction and retrieval of semantic audio descriptors," 2014.

[8] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?," *Acta Acustica united with Acustica*, vol. 92, no. 5, pp. 820–825, 2006.

[9] J. Marozeau and A. de Cheveigné, "The effect of fundamental frequency on the brightness dimension of timbre," *Journal of the Acoustical Society of America*, vol. 121, no. 1, pp. 383–387, 2007.

[10] S. Stasis, R. Stables, and J. Hockman, "A model for adaptive reduced-dimensionality equalisation," in *Proceedings of the 18th International Conference on Digital Audio Effects, Trondheim, Norway*, 2015, vol. 30.

[11] S. Stasis, R. Stables, and J. Hockman, "Semantically controlled adaptive equalisation in reduced dimensionality parameter space," *Applied Sciences*, vol. 6, no. 4, pp. 116, 2016.

[12] A. Zacharakis and J. Reiss, "An additive synthesis technique for independent modification of the auditory perceptions of brightness and warmth," in *Audio Engineering Society Convention 130*, May 2011.

[13] T. Brookes and D. Williams, "Perceptually-motivated audio morphing: Brightness," in *122nd Audio Engineering Society Convention*. Audio Engineering Society, 2007.

[14] D. Williams and T. Brookes, "Perceptually motivated audio morphing: warmth," in *128th Audio Engineering Society Convention*, 2010.

[15] J. Krimphoff, S. McAdams, and S. Winsberg, "Caractérisation du timbre des sons complexes. ii. analyses acoustiques et quantification psychophysique," *Le Journal de Physique IV*, vol. 4, no. C5, pp. C5–625, 1994.

[16] J. H. Ward Jr, "Hierarchical grouping to optimize an objective function," *Journal of the American statistical association*, vol. 58, no. 301, pp. 236–244, 1963.