Daniel Muzzulini: News from future timbre spaces (2006)

Daniel Muzzulini, Alemannengasse 35, CH-4058 Basel, ++41 61 691 59 16

Introduction

The present paper outlines possible areas of future timbre research, which are likely to improve our knowledge on an aspect of sound perception, which is sometimes named a ‘multidimensional dimension’. [Handschin 1948; Licklider 1951]

The notion of space usually used in psychometric investigations on sound quality should be rigorously scrutinized. The idea of arranging timbres in multidimensional contexts, in so-called timbre spaces, seems to reveal a close connection between psychology and mathematical concepts: perception psychologists use Euclidean vector spaces in order to bring perception objects into consistent spatial configurations. This works fine as long as there are any detectable polarity axes within such configurations or as long as relationships can be established, which connect physical and psychological polarities. However, some hearing phenomena are evidently linked to cyclic aspects. Examples of this kind are the octave sensitivity for pitches and for musical intervals, which have entered and deeply influenced the western theory of harmony, as well as phenomena connected to the ear’s phase sensitivity in complex sounds. In the latter case, the cyclic element is due to mathematical reasons. Given that genuinely cyclic structures should not be reduced to (or embedded in) Euclidean vector spaces, it is highly desirable to have tools which analyse given data not only in a linear way by finding their best Euclidean representation, but also would detect cyclic dimensions or combinations of linear and cyclic dimensions.

Furthermore, the stability of spatial timbre perception is to be questioned. In order to free the concept of timbre space from statically linking timbres with positions in fixed spaces, a spectral analysis algorithm for time-variable dimensionality is sketched out, which works in real time provided that real-time FFT-data is available. Some timbre space investigations emphasize sharpness as a consistent dimension of timbre, and as the single one that directly correlates with the spectral distribution of complex sound objects. There are, however, musical situations conceivable where the spectral analysis of our hearing system does not reduce to a single linear sharpness dimension. By dynamically extending sharpness to a multidimensional property, this possibility is taken into account by the proposed algorithm.

Fred Lehrdahl [Lehrdahl 1987] proposes a generative grammar system applied to timbres by means of a reduction of the universe of timbres to a few prototypes living in three- or four-dimensional timbre spaces. He claims that only on the basis of a few clearly distinct tokens a sophisticated musical language can be developed. And, he presumes sharpness to be a two-dimensional aspect of timbre. The use of cyclic or mixed linear/cyclic representations of prototypes is a possible answer to his remark concerning the missing octave periodicity in timbre spaces [Lehrdahl 1987, p. 147] and could serve as a test case for non-Euclidian scaling algorithms. Structures of this kind might provide for adequate transferences of approved composition paradigms, such as consonance/dissonance or functional harmonics, into timbres much more than ordinary vector spaces.

Distance estimations and psychometric investigations

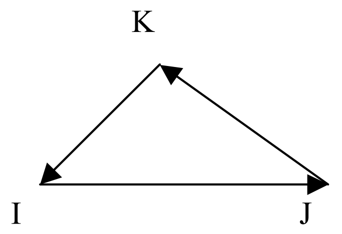

Figure 1: Distance estimations of three persons I, J, K.

Three friends Irena, Jan and Karin are standing motionless in a landscape (see Figure 1). Irena sees Jan and thinks that he is 7 meters away from her. Jan looks at Karin and assumes that she is 3 meters away from him. And Karin watches Irena and estimates the distance to her at 2 meters. Euclid is informed about their distance estimations and tries to figure out their relative positions in a plane. Taking Irena's estimation 7 as given, Karin would be within the intersection of the two circles with radiuses 3 and 2 respectively and centers Irena and Jan with distance 7.

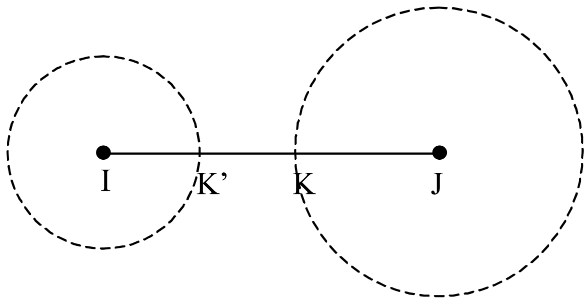

Figure 2: Violation of the triangular condition.

Unfortunately, the intersection of the two circles is empty and Euclid realizes that at least one of the three estimations must be wrong (see Figure 2). Therefore, he decides to adapt all of the three values by the same amount. The least correction, which provides a possible geometric configuration in the plane, is $\frac{2}{3}$. With the corrected distances $6\frac{1}{3}$, $2\frac{2}{3}$ and $3\frac{2}{3}$ the three persons would be positioned on a straight line (see Figure 3), which can be excluded, since, in this case, Jan would be hidden from Irena's sight by Karin making a distance estimation impossible.

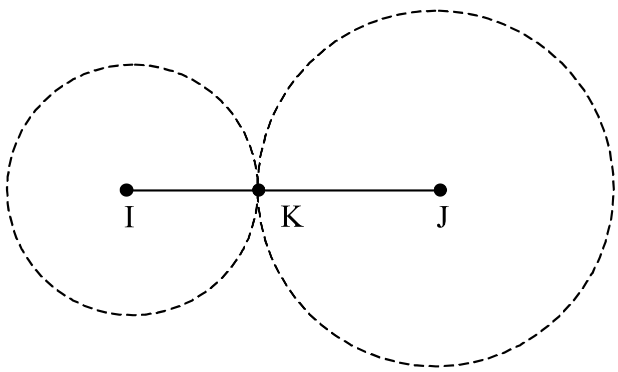

Figure 3: Minimally adapted distance estimations.

An over-all-correction by 1 yields the distances 6, 3 and 4 and to two symmetric configurations on triangles *IJK* and *IJ'K*, as shown in Figure 4.

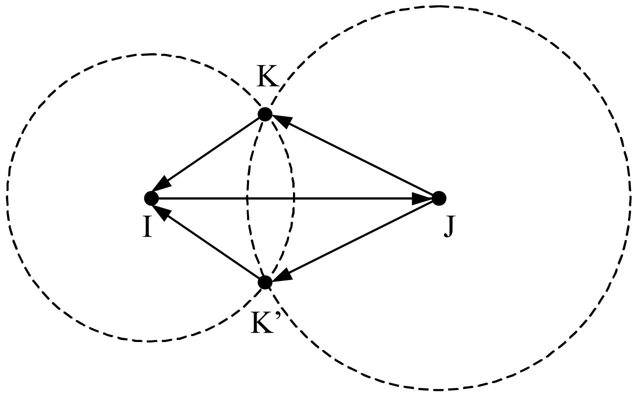

Figure 4: Two symmetric configurations with non-minimal corrections.

The example plays with the triangle condition in Euclidean geometry, the fact that for any two sides of a triangle the sum of their lengths is longer than the third. Points fulfilling this condition are said to be in *general position*, which means that they build up a triangle and thereby determine a plane within a three (or higher) dimensional space.
Figure 3 shows the case where the triangle degenerates in a straight line. Depending on the amount of correction, there are zero, one or two possible configurations.

Psychometric investigations try to determine a geometric point configuration, which best fits with a given set of distance estimations, i.e. data stemming from dissimilarity judgments. Even small sets of estimation data can have no corresponding exactly fitting geometric configurations, as is proved by the above example.
Algorithms used to calculate such configurations are named MDS-algorithms (multidimensional scaling). An MDS-algorithm calculates a vector space containing a point configuration, which correlates best with the estimation data. Within a configuration calculated in this way the global error is minimized. The notion *vector space* generalizes our three-dimensional world to an arbitrary number of dimensions.
Some MDS-programs allow the user to choose between different *metrices*, i.e. the underlying notions of distance. Usually, the Euclidean metrics, which calculates distances by making use of the theorem of Pythagoras, is applied.
The algorithm itself determines the dimension number n of the optimal solution, which depends on the data set. For m perception objects to be arranged, the maximal dimension n needed in order to represent them as points is $n = m–1$, which can only occur when the objects are in general position.
In most of the psychometric investigations on the timbre of sounds of equal pitch, duration and loudness, MDS-algorithms establish so called *timbre spaces* of dimension 3 or 2, with individual differences. Grey [1975] for instance arranges sixteen examined sound objects in a three-dimensional space.

Any additional point on a triangle is either contained in the two-dimensional space spanned up by the triangle, or three dimensions are needed to describe their mutual position. Similarly, in three dimensions at most for points can be generally positioned, so that they form the corners of a tetrahedron. Any additional fifth point is either located in the three-dimensional space spanned up by the tetrahedron, or four dimensions are needed to describe the mutual position of the five points consistently.

Compared with our initial example, real psychometric investigations have to deal with many more distance estimations than the expected dimension of their spatial configuration, since no valid statements going beyond the current random-sample can be deduced from too sparse data. The possibility, however, that a given set of distance estimations will become incompatible with an exact spatial representation obviously grows with the size of the sample.

After having algorithmically determined a spatial arrangement of the sound objects under consideration in a timbre space, the next step is to define therein a system of coordinates and to interpret its axes in psycho-acoustical terms. This means that the axes of the timbre spaces are mapped on both, the corresponding psychological polarities and the physical properties of the sound objects, in order to establish a one-to-one relationship between the physical and the psychological representation of sounds.
Within a given spatial representation, the selection of the origin and the definition of an orthogonal system of coordinates are arbitrary, since configurations calculated by the algorithm are uniquely determined only up to translations and orthogonal transformations. It makes sense to determine the axes by a descending diameter strategy. Maximal diameters are likely to represent stable inter-subjective psychological polarities.

Cyclic aspects of music perception

Suppose now the Euclidean axioms are no longer valid in the world of Irena, Jan and Karin, so that space and sight are strongly curved, even a one-dimensional geometric interpretation can be given which fits exactly with the above estimation data, as given in Figure 5 [For the notion of the curved space see Govers 2002, pp. 104–108].

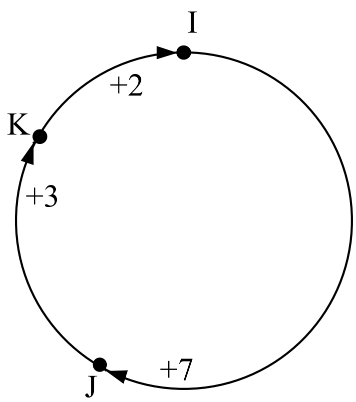

Figure 5: Distance estimations in the curved space.

This kind of spatial interpretation could not be detected by a classical MDS-algorithm, which calculates one linear dimension as in Figure 2, since all two-dimensional configurations correlate less with the estimation data.
Cyclic structures are meaningful in music perception. A prominent example is the octave identity, which seems to be an almost universal cultural phenomenon.
Transferring the three distance estimations into pitch classes of the 12-tempered tuning gives: *g* is one fifth above *c*, *b-flat* is one minor third above *g*, and *c* a major second above *b-flat*. The three sounds build up a seventh chord with a missing third.
From the perspective of piling up thirds to define chord structures, as in western music theory, there is nothing unusual about saying that *g* and *c* are separated by seven semitone-steps, although the shortest distance in the group of pitch classes is 5.

There are always two connections between two given points on a circle line. Only when they form the opposite end of a diameter are the two connections of equal length. Piling up thirds to define meaningful chord structures is a paradigm in western music theory. This construction principle abstracts from the octave position in order to obtain chords with similar musical meaning. Both pitches and intervals are hereby reduced modulo octave, and they are

interpreted as cyclic aspects of the musical system. The harmonic system is empirically and psychologically justified by the octave identity as a universal.
The ambiguity in measuring the distance of two points on a circle line can be resolved, when it is done in a predefined direction on ordered pairs of pitches. If the pitches *c*, *c-sharp*, *d*, … , *b* are ordered in clockwise direction, the ordered pair (*c*; *g*) has distance 7, and the ordered pair (*g*; *c)* has distance 5 in semitone-metrics.

Matthias Hauer [1918] proposes a cyclic interpretation of the nature of timbre as follows (see Figure 6): He takes the octave-reduced interval between the most prominent overtone and the root of a complex sound as the determining property of its timbre. By means of this strong reduction he establishes a one-to-one correspondence between the colours of the light spectrum and the musical timbres, both of them arranged on circle lines. Since Newton, similar correspondences between the colours of the spectrum on the one hand and the pitch classes or the interval classes respectively on the other have been postulated in a rather speculative way [for an overview see Helmholtz 1867, pp. 269–270; Jewanski 1999].

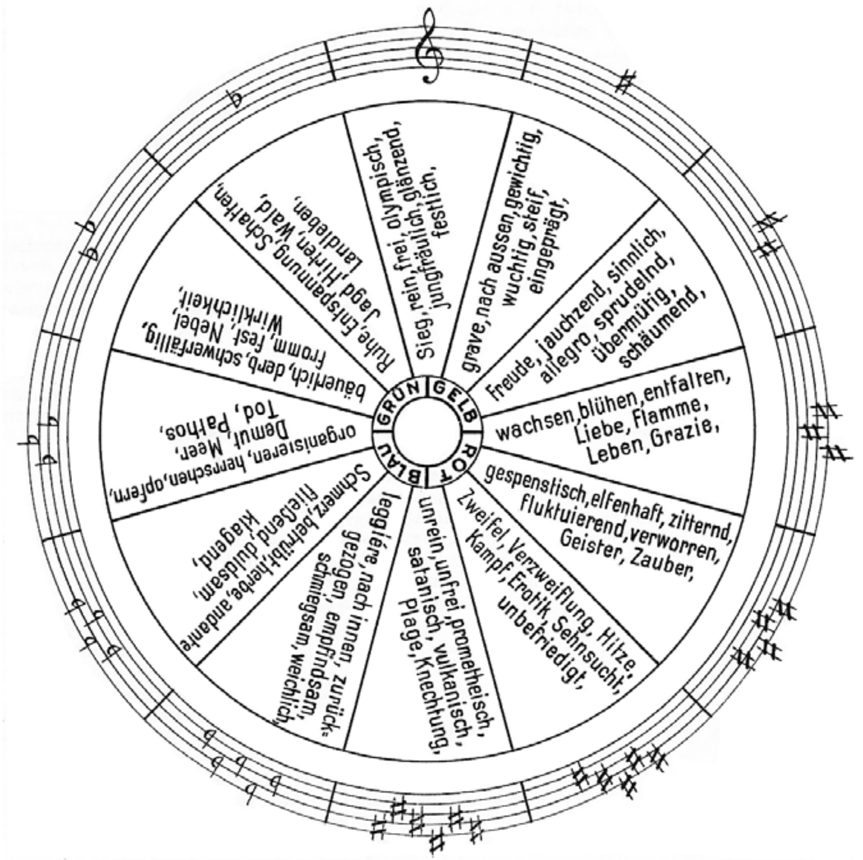

Figure 6: Matthias Hauer’s correspondence between diatonic scales, emotions and colours [source: Hauer 1918, p. 10]. The 12-tempered cycle of fifths (indicated by the accidents on the outermost circle) is mapped onto the cycle of spectral colours. The four colors on the innermost circle are arranged in the same way in the CIELAB-color-model, wherein the pairs yellow-blue and red-green of complementary colors form the main axes. The ‘Hauer colour’ of a given timbre is calculated by means of its characteristic interval, which is the interval between tonics of diatonic scales with the reference C = 0, e.g. a timbre with a characteristic major third (due to a strong fifth harmonic) corresponds with E major (4 #), *gespenstisch* (eerie) and red, whereas a characteristic fifth (caused by a dominating third harmonic) is associated with G major (1 #), *grave* and yellow.

Hauer’s onset, which treats timbres as if they were intervals, turns out to be a timbre theory that takes into account – independently of the frequency of the root – only the shape of the amplitude spectrum. The timbre theory of Helmholtz has often been discussed in this respect. According to Hauer, a linear stretching of a sound spectrum in the frequency domain does not alter the associated emotions or 'Hauer colour'. Sounds with the same formants, however, such as the same vowel sung at different pitches, would result in different colours: by augmenting the pitch of a vowel the interval between the root and its most prominent overtone becomes smaller and, therefore, the Hauer colour of the vowel changes.

The order of presentation and asymmetric judgments

In comparisons of timbres by means of psychometric investigations, the sound objects are usually presented one after another, since a simultaneous presentation would affect the listener's ability to hear them as independent entities due to interacting partials.
It can happen that the distance estimation e_{ij}, where the sound i is followed by j, and the estimation e_{ji}, where j is followed by i results in different values, and it is not impossible that an object compared with itself results in an estimation e_{ii} different from 0.
The distance estimations e_{ij} of a set of comparisons with m objects can be arranged in a quadratic matrix E with row index i and column index j.

$$E = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} = \begin{bmatrix} * & 7 & ? \\ ? & * & 3 \\ 2 & ? & * \end{bmatrix}$$

Figure 7: Estimation matrix for Irena = 1, Jan = 2 and Karin = 3

The estimation matrix in Figure 7 is incomplete, for each person made only one of three possible estimations. The main diagonal marked by '*' contains the self-comparisons, and the symmetrically positioned estimations are marked with '?'. So e_{21}, the first entry in the second row, would hold the distance estimation of Jan about Irena.

The simplest application of a classical MDS-algorithm bases on symmetric judgments $e_{ij} = e_{ji}$ and on $e_{ii} = 0$. Deviations from symmetry of E indicate unsteadiness in the estimations and can serve as a measure for the expected uncertainty in the configuration calculated by an algorithm, which, for simplification, uses arithmetic means $\overline{e_{ij}} = \frac{1}{2} \cdot (e_{ij} + e_{ji}) = \overline{e_{ji}}$ instead of the original data.

As soon as cyclic aspects of sound perception are the focus of interest, a principal deviation from symmetry of E is to be expected. It is a hint to an asymmetry of this kind, when the estimations for $\overline{e_{ij}}$ are distributed around two accumulation values. A comparable example of asymmetric judgments in interval perception would be the confusion of the minor sixth (8 semitones) with a major third (4 semitones) caused by a prominent second partial.

The influence of phase relations upon timbre – a cyclic element

In contradiction with a common opinion about the phase rule of Ohm/Helmholtz, there are perceptible phase effects within periodic sounds consisting of two sinusoidal components with constant amplitudes. According to some authors the components involved do not even need to be transmitted within the same critical band [Craig et al. 1962].
If the lower component of a dyad is kept constant and the phase angle of the upper component is augmented by equal steps of $\pi/6 = 30°$ (related to its shortest period $2\pi = 360°$), an identical time signal will occur after twelve steps. This mathematical fact favors a cyclic interpretation

of this elementary case. (Depending on the frequency relationship, isomorphic signals are obtained already after six phase steps.) It can be asked if such a phase graduation could form the base of a consistent geometric interpretation as interval perception on pitch classes. By comparing phase effects with amplitude effects, Fleischer [1976] concludes that within a full phase modulation cycle up to 10 or 12 perceptible phase shifts can occur.

The above example of dyads suggests a three-dimensional modeling ($\mathbb{R}\times\mathbb{C}$), if not only the phases but also the amplitudes of the sound complex are being varied, whereby the frequency variable component is represented by polar coordinates within $\mathbb{C}$ and the constant component in orthogonal direction by $\mathbb{R}$. In this three-dimensional space, arbitrary time shapes of a given pair of constant frequencies can be represented as points. However, this simple geometrical model provides no reliable correspondence between sound objects and related distance estimations by theoretical reasons [Muzzulini 2004, pp. 336–337].
Furthermore, generalization of this additive modeling to sounds which consist of more than two sinusoidals would result in spaces with higher dimension than those of ordinary timbre spaces: with three components the dimension will be five, with four components it will be seven, and with n components $2n$–1. No correspondence between this mathematical description and spatial sound perception can be expected.
Already without taking phase relationships into account, neither the mathematical dimension given by the number of components nor the number of critical bands equals the low dimension of the timbre spaces currently discussed. Sounds in concrete musical contexts usually show less perceptible independent aspects than the number of physical parameters suggest.

To sum up, MDS-algorithms designed for structures different from Euclidean vector spaces might be useful tools for establishing new timbre topologies and in exploring the relevance of cyclic structures for perception.

Static and dynamic spatiality

The above considerations suggest a reassessment of the static concept of timbre spaces. It seems plausible that spatial perception structures can arise locally in one given musical context, but not necessarily in another.
Timbral space perception might be favored by the listener's focused attention to minimal changes in phase and in amplitude, as long as the frequencies are held constant, but not at all in a two-voice polyphonic setting for flute and violin. And most probably, space perception in *minimal music* very soon tires and leaves room for dreaming.
On the other hand, there is evidence that listeners are able to distinguish more than three independent aspects in comparing vowels of equal pitch and loudness. Since there are two formants needed to distinguish different vowel types, they cannot be arranged on a straight line with perceptual evidence. Furthermore, the same vowel can be more or less sharp, and it can be sung with more or less vibrato [Lehrdahl 1987]. All these aspects can be varied independently, which indicates a structure of at least four dimensions.
The fact that the German vowels *a* – *u* – *i* cannot be arranged with psychometric evidence on a straight line may be completely irrelevant, when a sharp *a* with no vibrato, a still sharper *a* also with no vibrato, a dark *u* with strong vibrato and a sharp *i* with little vibrato are compared. The focus of interest could be the nuances in sharpness and vibrato and, although the difference of the vowel types is recognized, their geometrical constellation is insignificant. And everything is totally different, if the same vowel qualities are presented in a melody with a text that makes use of expressive dynamics.

When we compare optical colours, there is no need to know that the colours of equal intensity are better represented on a kind of shoe sole than on a circle or on a triangle, not even whether a two-dimensional representation, which fits with dissimilarity estimations of average persons not being colour-blind, is possible or not. We do not need to know that under normal circumstances all three types of colour-receptor cells are always excited simultaneously – so that according to Helmholtz a distinction between simple and compound colors makes no sense – in order to be able to recognize similarities and contrasts of colors.

It is a challenge to find new ways of analyzing music, which are adaptive in dimensionality, time and context, and modes of visualizing the analysis interactively and simultaneously.

Analysis of spectra with overlapping band-pass filters

In some psychometric interpretations of the axes in a configuration of timbre points the spectral properties of sounds are mapped onto a single psychological dimension, their sharpness. Benedini, however, gives a class of stationary sounds, in which this kind of reduction causes a loss of perceptible information [Benedini 1978].
By use of overlapping filters, the spectral information of sound can be analyzed similarly to the way the incoming light gets analyzed by the color cells of our visual system. Using B-spline interpolation techniques, this kind of analysis can be done efficiently. By varying the number of resonance filters the dimension can also be varied dynamically. Three filters correspond with optical color classification, from two filters there can be still deduced a sharpness measure. [Muzzulini 2000, pp. 267–268; Muzzulini 2004, pp. 328–331]

Keeping the frequency range of the filters constant over time – independent of actual pitches – is in accordance with a formant theory of timbre, in contrast to the concept of Hauer as outlined above.

Prototypes in timbre spaces

Bearing in mind Schoenberg's [1912] concept of 'Klangfarbenmelodie', Lehrdahl proposes a procedure for defining prototypes in the universe of timbre. According to him, it is a prerequisite for the development of a musical language of certain complexity to operate on a limited set of reference tokens, and without its restriction to a few pitch prototypes, the highly developed western musical semantics would be unthinkable [Lehrdahl 1987].
His basic idea is the translation of generic musical concepts as scales and consonance/dissonance, the latter psychologically related to tension/relaxation, into timbres, in order to open them up for structural musical composition.
For this purpose, a set of timbre prototypes is to be distributed well over a range of timbres. In a cubic timbre space for instance, the prototypes can be positioned on a three-dimensional grid, so that few timbres cover a large range of timbres. (A similar procedure was the selection of 8 and later of 256 color values in older graphic formats for color representation on computer screens.)

If a given set of timbre prototypes were consistently recognizable to the audience, it would be possible to use them to define musical rule systems. Lehrdahl gives such an example, which defines a 'diatonic scale' within a two-dimensional system of vowel timbres, but he points to the topological lacking of octaves in timbre.
The mapping would be probably much more convincing, if the set of prototypes was taken not from a plane but from a Moebius strip, since the Moebius strip is cyclic in its very nature and

can express both, the closed sequence of thirds f-a-c-e-g-b-d-f and the sequence of fifth f-c-g-d-a-e-b-f [cf. Mazzola 1990, pp.176-199], better than the selection of seven points out of a 3×3-grid.
Possibly, there are psycho-acoustical properties (not stemming from bi-polarity), which map perfectly onto the border of a Moebius strip.
The application of non-classical MDS-algorithms could give some insight in this respect. The algorithm decides if a given set of prototypes matches with the distance estimations taken on a Moebius strip, on a torus, or on a cylinder better than on a Euclidean plane.
If non-Euclidean timbre topologies were learnable by educated listeners and if there were no three-dimensional timbre space around a given timbral Moebius strip, the introduction of the diatonic scale into timbres would be justified.

Without testing different space structures on the same set of distance estimations, it cannot be decided whether the above-mentioned twelve distinguishable phase steps within a two-tone complex better correspond with a Moebius strip, on which geometrically isomorphic but not identical signals (retrogrades) have exactly the same position, but on opposite sides of the strip, than with a simple circle. Perhaps topologies of this kind, which consider amplitude and phase from a unifying geometrical viewpoint in the way described, serve as a natural example of the curved nature of the universe of timbres.

Mc Adams et al. [1992] examine a very simple but essential musical structure relation, namely the direct sequence of a two-tone pair in its transformation into a timbre space, and they estimate the concept of timbral analogies to be learnable (see Figure 8).

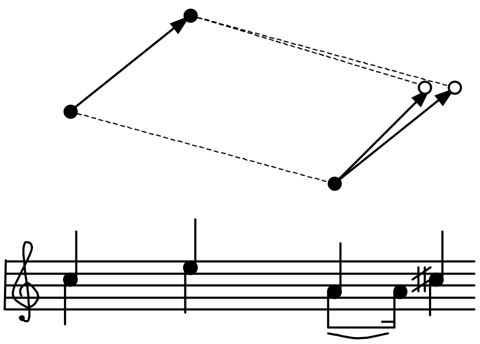

Figure 8. Timbral analogies according to McAdams et al. [1992]. The listeners of the experiment have to decide which of the two pairs of timbres form a true parallelogram. This task is similar to finding the correct direct sequence in the two rhythmical variants of a melody as shown below.

The mastering of the more sophisticated geometry of the Moebius strip would be equivalent to the learning of the triadic interpretation of a seven-tone scale in the sense of Mazzola [1990].
If remains an open question, if the compositional exploitation of structural relationships of this kind in timbre spaces is understandable or if it gives at least occasion for satisfying aural experiences.

Acknowledgements

Some of the ideas presented here are already sketched out in Muzzulini [2001; 2004], I am very grateful for the help of me English teacher June Graff.

Bibliography

[Benedini 1978] Kurt Benedini: Psychoakustische Messung der Klangfarben-Ähnlichkeit harmonischer Klänge und Beschreibung der Zusammenhänge zwischen Amplitudenspektrum und Klangfarbe durch ein Modell. Diss., TU München 1978

[Craig et al. 1962] James H. Craig, Lloyd A. Jeffress, Effect of Phase on The Quality of a Two-Component Tone, JASA 34, 1962, 1752–1760

[Fleischer 1976] Helmut Fleischer: Untersuchung zur Hörbarkeit von Phasenänderungen. TU München 1976

[Gowers 2002] Timothy Gowers, Mathematics, A very short introduction, Oxford University Press, New York 2002

[Grey 1975] John M. Grey, An exploration of musical timbre using computer-based techniques for analysis, synthesis and perceptual scaling, PhD Diss. Stanford University, Center for Computer Research in Music and Acoustics, Report No. STAN-M-2, 1975 (UMI 1987)

[Handschin 1948] Jacques Handschin, Der Toncharakter, Eine Einführung in die Tonpsychologie, Atlantis, Zürich 1948

[Hauer 1918] Josef Hauer, op. 13 Über die Klangfarbe

[Helmholtz 1859] Hermann von Helmholtz, Ueber die Klangfarbe der Vocale, Annalen der Physik und Chemie, Band CVIII, Leipzig 1859, 463–466

[Helmholtz 1867] Hermann von Helmholtz, Handbuch der physiologischen Optik. Allgemeine Encyklopädie der Physik Bd. 9, Voss, Leipzig 1867

[Jewanski 1999] Jörg Jewanski, Ist C = Rot? Eine Kultur- und Wissenschaftsgeschichte zum Problem der wechselseitigen Beziehung zwischen Ton und Farbe, Von Aristoteles bis Goethe, Berliner Musik Studien 17, Studio, Verlag Schewe, Sinzig 1999

[Lehrdahl 1987] Fred Lehrdahl, Timbral hierarchies, Contemporary Music Review, 1987, Vol. 2, 135–160

[Licklider 1951] J. C. R. Licklider, A Duplex Theory of Pitch Perception, Experientia 7, 1951, 128–134

[Mazzola 1990] Guerino Mazzola: Geometrie der Töne, Elemente der Mathematischen Musiktheorie, Unter Mitarbeit von Daniel Muzzulini und einem Beitrag von Georg Rainer Hoffmann, Birkhäuser, Basel 1990

[Mc Adams et al. 1992] Stephen McAdams, Jean-Christophe Cunibile: Perception of timbral analogies, Philosophical Transactions of the Royal Society (vol 336), London Series B, 1992

[Muzzulini 2000] Daniel Muzzulini, Klänge und Farben – Spektren in der Akustik und Optik, in: Antonio Baldassarre et al (ed), Musik denken, Ernst Lichtenhahn zur Emeritierung, Peter Lang, Bern 2000, 255–269

[Muzzulini 2004] Daniel Muzzulini, Genealogie der Klangfarbe, Diss. Universität Zürich, http://www.dissertationen.unizh.ch/2004/muzzulini/timbre.pdf

[Schönberg 1911] Arnold Schönberg, Harmonielehre, Universal-Edition, Leipzig-Wien 1911