Melodic Similarity - a Conceptual Framework

Ludger Hofmann-Engl

The Link
+44 (0)20 8771 0639
ludger.hofmann-engl@virgin.net

Abstract. Melodic similarity has been at the centre of research within the community of Music Information Retrieval (MIR) in recent years. Many different models have been proposed (such as transition matrices or dynamic programming). However, so it seems, all these models exhibit a number of shortcomings. The approach taken in this paper differs in several ways from previous attempts. The position is taken that no single model will be satisfactory in all contexts. Thus, rather than producing yet another (simple) model, the approach is taken to develop a conceptual framework whereupon a number of models can be based according to specific needs. The parameters which are to be inputted into a model are scrutinized leading to the definition of atomic beats (smallest time value of a melody), melota (substituting ambiguous definitions of pitch), dynama (partially equivalent to subjective loudness) and chronota (the cognitive correlate to time values). Describing melodies as chains based upon atomic beats allows for the mapping of any given melody onto any other melody via two specific reflections (where the first reflection is along the x-axis and the second reflection is mapping this image onto the second melody). The second reflection along a reflection curve (called similarity vector) displays two factors of similarity for each of the parameters (six factors altogether). The length of the similarity vector delivers information about how much two chains differ in average pitch, average loudness and average time division. The interval vector as derived from the similarity vector (subtracting the *ith* component from the (i+1)*th*) produces information of how closely two melodies are similar in shape. Finally, two specific models will be tested within an experimental setting.

1 Introduction

In recent years the interest in melodic similarity has been mushrooming. This interest has been driven by the developments within MIR (music information retrieval). A typical MIR situation can be described as the following: a user wishes to locate a musical piece within a database. The query can be by various means including musical notation, use of metadata (such as artists' names) or humming. Typically, a query by humming (QBH) requires the implementation of similarity algorithms, simply because the input by humming contains generally errors and excludes the possibility to search for perfect matches. A variety of similarity algorithm have been put forward, but they can be divided into four classes. These classes are (a) contrast models (e.g. Downie, 1999), (b) difference models (e.g. Maidin, 1998), (c) dynamic programming (e.g. Smith, McNab & Witten, 1998) and (d) Markov chains (e.g. Hoos, Renz & Görg, 2001). However, as pointed out by Hofmann-Engl (2002b), all these approaches suffer from a multiplicity of conceptual errors and an inability to understand melodic similarity as an issue which will require a sound conceptual framework, which then can serve as the basis for similarity models fashioned according to the specific needs. This is the main goal of this paper, to establish such a conceptual framework which cannot only be employed within the setting of MIR but for the creative process of composing (such as producing suitable variations to a theme) as well as for musical analysis. However, before we enter the discussion about such a conceptual framework, we will briefly scrutinize the parameters implemented in this framework and introduce the notation via atomic beats and the transformation of melodies. The main section investigating aspects of the framework will be followed by a short experimental evaluation of the approach taken here in this paper.

2 The relevant parameters

It seems sensible to ask before we develop a framework of melodic similarity, what we intend to input as the relevant parameters. Clearly, the three aspects pitch, duration and loudness will have to considered in one way or another. However, Hofmann-Engl (1989, 2001, 2002a, 2002b) expressed dissatisfaction with the terms pitch, duration, loudness and melody, as the terms are highly ambiguous (what one person considers to be a melody is not regarded as a melody by another person). Thus, we will consider a new terminology considering the relevant parameters from a cognitive point of view.

This is, instead of using objective or subjective values, we will define intersubjective values.

2.1 Meloton versus pitch

According to models of virtual pitch (e.g. Terhardt, 1979; Hofmann-Engl 1999) a tone does not only produce one distinct pitch, but a series of possible candidates which might serve as the pitch of the tone. Musical tones such as a piano 'a' produce often one strong candidate (e.g. 'a') and a series of very weak candidates (e.g. 'd' or 'f'), but other tones such as tones as produced by a drum instrument do not produce such a clear distinction between one strong and other weak candidates. This lead Hofmann-Engl (2001) to define the term meloton as such:

<u>Definition:</u>

The meloton is the psychological concept whereby a listener listens to a sound directing her/his attention to the sound with the intention to decide whether the sound is high or low.

True, this does not deliver a quantity we can represent, and hence we will have to define the value of a meloton somehow without using a physicalistic concept. In this context it seems most appropriate to consider an experimental setting as introduced by Schouten (1938). A group of listeners is asked to tune in a (sinusoidal) comparison tone with variable frequency to match a test tone best according to each listener's individual judgment. We expect to obtain a distribution of different responses. In case the majority of listeners tune into the same frequency under well defined conditions (compare Hofmann-Engl, 2002b), we will take the logarithm of this frequency as the melotonic value and call the meloton strong. In case there is no consensus amongst the listeners we will take the mean of the logarithmic frequencies to represent the melotonic value and call the meloton weak. The concept of meloton is by far superior to a concept of pitch, as we find that all tones fetch a melotonic value, and thus "drum-melodies" are not only conceivable but can also be captured by referring to the melotonic value.

Clearly, we expect that the concept of pitch and meloton will in many cases coincide, but the fundamental difference remains that pitch is a cognitive predictor while the meloton is a cognitive measurement.

2.2 Dynamon versus loudness

In analogy to the term meloton, we define the term dynamon:

Definition:

The dynamon is the psychological concept whereby a listener listens to a sound directing her/his attention to the sound with the intention to decide whether the sound is loud or soft.

The experimental measurement of the dynamic value follows the same idea as did the measurement of the melotonic value. A group of listeners is asked to tune in a (sinusoidal) comparison tone with variable loudness so as to match a test tone best according to each listener's individual judgment. We expect to obtain a distribution of different responses. In case the majority of listeners tune into the same loudness under well defined conditions, we will take the logarithm of this loudness as the dynamic value and call the dynamon strong. In case there is no consensus amongst the listeners we will take the mean of the logarithmic loudness to represent the dynamic value and call the dynamon weak.

Note that relative dynamic values seem to be of greater importance than absolute values (otherwise a piece played back at various loudness levels would alter the quality of the piece substantially). We also face the situation that loudness perception depends on the room characteristics and the characteristics of the individual listener far more than does the perception of pitch. The issue is discussed in detail by Hofmann-Engl (2002b).

2.3 Chronoton versus duration

In order to ensure an equal treatment of the parameters we expect to be of importance in the context of melodies, we introduce the following definition:

Definition:

The chronoton is the psychological concept whereby a listener listens to a sound directing her/his attention to the sound with the intention to decide whether the sound is short or long.

The measurement of the chronotonic value will have to be conducted this time in a different way. A group of listeners will be presented with the test tone for which we intend to obtain the chronotonic value. After the test tone is heard the listener will be asked to adjust a control tone (by switching it on and off) so as to match the duration of the test tone best according to each listener's individual judgment. However, the term strong and weak chronoton will bear different meaning this time. Generally, listeners will be asked to segment an audio stream into segments (tones). Segmentation will occur due to sudden dynamic or frequency changes. Should the majority of listeners segment the audio stream in the same fashion we will accept their segmentation and talk about strong chronota. Should there be no consensus, we will incorporate the peak segmentations and talk about weak chronota.

We would expect that chronota are strong in general. However, where tones display time-dependent melota (such as glissandi) or time-dependent dynama (such as crescendi), we might expect higher variations of segmentations.

3.4. Melodic chains versus melodies

We are now in the position to define what we will call melodic chains.

<u>Definition:</u>

A chain is a sequence of a finite amount of tones.

True, this definition does not deviate from the more conventional concept of what we consider melodies except that the term melody is highly ambivalent and implies some form of musical judgment. We denote a chain in the form:

$$ch = [t_1, t_2, ..., t_n]$$

If we are interested in the melotonic contents of a chain, we write:

$$M(ch) = [m_1, m_2, ..., m_n]$$

If we are interested in the dynamic contents of a chain, we write:

$$D(ch) = [d_1, d_2, ..., d_n]$$

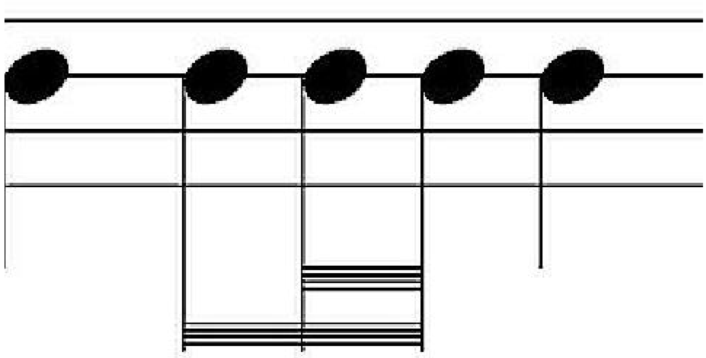
and if we want to depict the chronotonic information of a chain, we write:

$$C(ch) = [c_1, c_2, ..., c_n]$$

Note, that the introduction of chains, melota, dynama and chronota does not mean a radical change or break from traditional approaches. However, it has been a shift towards cognitive measurement and an equal treatment of the parameters, which we considered to be relevant.

3 Atomic beats and atomic notation

It is interesting to observe that conventional musical representation does little justice to an adequate representation of time. This is, the x-axis of a typical score does not correlate to the actual time flow. The author argues that some of the cognitive misconceptions about melodic similarity are due to this form of misrepresentation of time. Hopefully, this will become more obvious as the text develops. We will now introduce the concept of atomic notation by means of an example. Let us consider the following c-chain:



In a first instance we can write:

$$C(ch) = \{1/4, 1/8, 1/16, 1/16, 1/4\}$$

However, if we consider that the 1/16th notes are the smallest chronota (=durations) and that all other chronota are multiples of the 1/16th value, we can denote the same c-chain in the following manner:

$$C(ch) = [4, 4, 4, 4; 2, 2; 1; 1; 4, 4, 4, 4](1/16)$$

This is, the c-chain consists of 12 atomic beats taking 1/16th to be the atomic beat. The first value (a quarter note) stretches over four atomic beats, the 1/8th over two atomic beats and so on. The quarter note is four times longer than the atomic beat and hence the c-chain fetches for the first four atomic beats the value 4, then the value 2 and so on. As we will see, this form of representation will enable us not only to define melodic transformations but also to establish a conceptual framework of melodic similarity. However, before we will do so, we present the general form of a c-chain, given as:

$$C(ch) = [c_{11}, c_{12}, \ldots, c_{1m_1}; c_{21}, c_{22}, \ldots, c_{2m_2}; \ldots; c_{n1}, c_{n2}, \ldots, c_{nm_n}](1/a)$$

where *C*(*ch*) is a chronotonic chain in atomic notation, $c_{i1}, c_{i2}, \ldots c_{im_i}$ ***the i*th chronoton in atomic notation,** m_i **the number of atomic beats covered by the *i*th chronoton, *n* the length of the the c-chain and 1/*a* the atomic beat.**

Dealing with melodic chains in general, we obtain:

$$ch = [t_{11}, t_{12}, \ldots, t_{1m_1}; t_{21}, t_{22}, \ldots, t_{2m_2}; \ldots; t_{n1}, t_{n2}, \ldots t_{nm_n}](1/a)$$

***where ch* is a melodic chain in atomic notation,** $t_{i1}, t_{i2}, \ldots, t_{im_i}$ **atomic notation,** m_i **the number of atomic beats covered by the *i*th tone, *n* the length of the chain and 1/*a* the atomic beat. Each tone consists of a meloton, dynamon and chronoton**.

We will give an example. The opening of the theme from Mozart's a-major sonata (Köchel Nr. 331) is:

The chain *ch* is given as:

ch = [(c#, 3/16, p), (d, 1/16, p), (c#, 1/8, p), (e, 1/4, p), (e, 1/8, p), (b, 3/16, p), (c#, 1/16, p), (b, 1/8, p), (d, 1/4, p)]

Note, we are dealing here with a score rather than a transcript of a sound source. Hence, we are dealing with predicted values rather than measured values. Still, for the purpose of illustration, we will assume that the score does represent approximated measured values. However, this is not true for the dynama. Clearly, Mozart did not intend the piece to be performed without any dynamic variation, but following 18th century notational practice, Mozart did not feel the need to be more specific about the dynamics of the piece. Rating dynama on a scale from 1 to 9 (1 = soft, 9 = loud), we might assume that the following chain might be an appropriate interpretation of the score:

ch = [(c#, 3/16, 4), (d, 1/16, 1), (c#, 1/8, 2), (e, 1/4, 3), (e, 1/8, 2), (b, 3/16, 4), (c#, 1/16, 1), (b, 1/8, 2), (d, 1/4, 3)]

The atomic beat of the excerpt is 1/16. Thus the excerpt falls into 22 atomic beats. We obtain the three chains:

$M(ch)$ = [c#, c#, c#; d; c#, c#; e, e, e, e; e, e; b, b, b; c#; b, b; d, d, d, d](1/16)

$C(ch)$ = [3, 3, 3; 1; 2, 2; 4, 4, 4, 4; 2, 2; 3, 3, 3; 1; 2, 2; 4, 4, 4, 4](1/16)

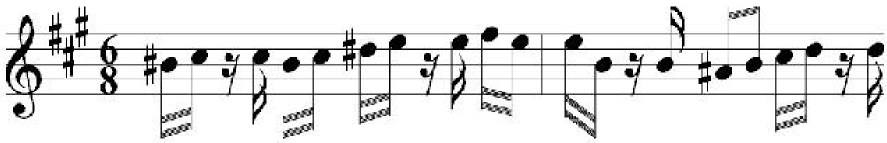
$D(ch)$ =[4, 4, 4; 1; 2, 2; 3, 3, 3, 3; 2, 2; 4, 4, 4; 1; 2, 2; 3, 3, 3, 3](1/16)

This form of notation preserves the information of all three parameters equally well relating them to quantasized time events. The main advantage of this method might only become fully apparent in the context of melodic similarity, however the transformation theory as presented next will enable us to see some useful aspects of this notation.

4 MELODIC TRANSFORMATIONS

There are several ways of introducing melodic transformations. However, the author decided it would be most appropriate to refer to an example, and then to explain the underlying concept in more detail.

Let us consider the beginning of the 1st variation of the Mozart piece we were talking about. This variation opens as:



As before, we have no sufficient dynamic information. We assume that the following d-chain represents an acceptable interpretation of the piece: $D(ch_v)$ = [4; 3; 0; 2; 4; 2; 3; 2; 0; 2; 3; 1; 4; 3; 0; 2; 4; 2; 3; 2; 0; 2]**.** It is interesting to note that in this case standard and atomic notation produce the same chains. This is, because now all chronota are 1/16 notes. We further obtain the m-chain $M(ch_v)$ = [b#; c#; -; c#; b#; c#; d#; e; -; e; f#; e; e; b; -; b; a#; b; c#; d; -; d] and the c-chain $C(ch_v)$ = [1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1].

Now transforming *ch* into ch_v, will require some form of transformation process. We follow the mechanism as proposed by Hofmann-Engl (2001, 2002b) whereby *ch* will be mapped onto ch_v via a chain of reflection points. In order to illustrate this, we will consider the m-chain of the theme compared to the m-chain of the first variation. We find:

$M(ch)$ = [c#, c#, c#; d; c#, c#; e, e, e, e; e, e; b, b, b; c#; b, b; d, d, d, d](1/16)
and$M(ch_v)$ = [b#; c#; -; c#; b#; c#; d#; e; -; e; f#; e; e; b; -; b; a#; b; c#; d; -; d](1/16)

We will depict the chains in log frequencies (where a rest will be mapped onto a 0 value) Thus we obtain:

$M(ch)$ = [2.75, 2.75, 2.75; 2.77; 2.75; 2.75; 2.81, 2.81, 2.81, 2.81; 2.81, 2.81; 2.71, 2.71, 2.71; 2.75; 2.71, 2.71; 2.77, 2.77, 2.77, 2.77]

and

$M(ch_v)$ = [2.73; 2.75; 0 ; 2.75; 2.73; 2.75; 2.79; 2.81; 0 ; 2.81; 2.85; 2.81; 2.81; 2.71; 0 ; 2.71; 2.69; 2.71; 2.75; 2.77; 0 ; 2.77]

Mapping the values of each atomic beat of ch onto $C(ch_v)$, we obtain:

$R_m(ch)$ = [2.74, 2.75, 1.38, 2.76, 2.74, 2.75, 2.80, 2.81, 1.41, 2.81, 2.83, 2.81, 2.76, 2.71, 1.34, 2.73, 2.70, 2.71, 2.76, 2.77, 1.39, 2.77]

This reflection chain as such makes it possible to map two m-chains onto each other. Moreover it delivers some information on how closely the two m-chains in question are inversions to each other (the more straight the reflection line the closer the transformation to an inversion). Clearly, in our case we can see that Mozart did not have the concept of an inversion in mind when he devised the first variation. In fact we will later see, that his intention must have been very different.

We now consider the c-chains. Rewriting them in $\log_2$ notation, we get:

$C(ch)$ = [1.6, 1.6, 1.6; 0; 1, 1; 2, 2, 2, 2; 1, 1; 1.6, 1.6, 1.6; 0; 1, 1; 2, 2, 2, 2] and
$C(ch_v)$ = [0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0; 0]

Mapping both c-chains onto each other will require the reflection chain:

$R_c(ch)$ = [0.8, 0.8. 0.8, 0, 0.5, 0.5, 1,1 ,1, 1, 0.5, 0.5, 0.8, 0.8, 0.8, 0, 0.5, 0.5, 1, 1, 1, 1]

We saw that the more straight the reflection chain in the context of m-chains, the closer is the transformation related to an inversion. The same is true in the context of c-chains, but what exactly is an inversion of a c-chain?

We consider the c-chain $C(ch_1)$ = [4, 4, 4, 4, 2, 2, 4, 4, 4, 4](1/16) and the c-chain $C(ch_2)$ = [1, 1, 1, 1, 2, 2, 1, 1, 1, 1](1/16). We get in $\log_2$: $C(ch_1)$ = [2, 2, 2, 2, 1, 1, 2, 2, 2, 2] and $C(ch_2)$ = [0, 0, 0, 0, 1, 1, 0, 0, 0, 0]. We obtain the reflection chain: $R_c(ch)$ = [1, 1, 1, 1, 1, 1, 1, 1, 1, 1]. This is a straight line, and hence $C(ch_2)$ is the inversion to $C(ch_1)$. This is, chronota which are longer than the values of the reflection line, will be split into smaller chronota, and chronota which are shorter than the values of the reflection line, will be fused into longer chronota. As far as the author is aware, inverting c-chains is not a commonly known or used transformation.

As we can observe, Mozart's variation and theme are not close chronotonic inversions.

We finally map the dynamic chains onto each other. We found earlier the possible interpretations for the theme and the first variation to be:

$D(ch)$ = [4, 4, 4; 1; 2, 2; 3, 3, 3, 3; 2, 2; 4, 4, 4; 1; 2, 2; 3, 3, 3, 3]
and
$D(ch_v)$ = [4; 3; 0; 2; 4; 2; 3; 2; 0; 2; 3; 1; 4; 3; 0; 2; 4; 2; 3; 2; 0; 2]

Assuming that our dynamic values are based on a logarithmic scale, we obtain the reflection chain:

$R_d(ch)$ = [4, 3.5, 2, 1.5, 3, 2, 3, 2.5, 1.5, 2.5, 2.5, 1.5, 4, 3.5, 2, 1.5, 3, 2, 3, 2.5, 1.5, 2.5]

As we can see, the d-chains are also not close inversions to each other either. However, generally we calculate the reflection points as:

$$p_{ri} = \frac{t_{1i} + t_{2i}}{2}$$

where p_{ri} is the reflection point at the place *i*, t_{1i} the value of the first chain at the place *i* and t_{2i} the value of the second chain at the place *i*.

A reflection chain will have the form:

$$R(ch) = [p_{r_1}, p_{r_2}, \ldots, p_{r_n}]$$

We will now consider a vector notation of the reflection chains.

5 Melodic vectors and reflection matrices

Instead of depicting melodic chains as a sequence of tones in atomic notation, there lies a great advantage in representing them in form of (n+1)-dimensional vectors, simply because reflections via reflection matrices is mathematically well defined. Thus, a melodic chain consisting of n atomic beats, will be represented in the form:

$$\vec{M} = \begin{pmatrix} t_1 \\ t_2 \\ . \\ . \\ t_n \\ 1 \end{pmatrix} \cdot 1/a$$

where $\vec{M}$ is the melodic vector, t_1 the first tone in atomic notation, t_2 the second tone, t_n the *n*th tone and 1/a the atomic beat.

Reflecting the melodic vector $\vec{M}$ of the dimension n+1 onto the melodic vector $\vec{M}'$ of the dimension n+1 will require the following reflection matrix:

$$R = \begin{pmatrix} -1 & 0 & . & . & 0 & 2p_1 \\ 0 & -1 & & & & 2p_2 \\ . & & . & & & . \\ . & & & . & & . \\ 0 & 0 & & & -1 & 2p_n \\ 0 & 0 & & & 0 & 1 \end{pmatrix}$$

where R is the reflection matrix and $p_1, p_2 .. p_n$ the reflection points of the reflection chain as defined previously.

This is, we find:

$$R \cdot \vec{M} = \begin{pmatrix} 2p_1 - t_1 \\ 2p_2 - t_2 \\ . \\ . \\ 2p_n - t_n \\ 1 \end{pmatrix} = \vec{M}'$$

where $t_1, t_2 .. t_n$ are the tone components of the melodic vector $\vec{M}$.

For a more detailed description of the algebra underlying these reflection matrices compare Hofmann-Engl (2002b).

6 A concept of melodic similarity

Just as we introduced melodic reflections above, so will we now introduce a concept of melodic similarity by referring to the Mozart example.

The m-chains of the theme and the variations were:

$M(ch)$ = [2.75, 2.75, 2.75; 2.77; 2.75; 2.75; 2.81, 2.81, 2.81, 2.81; 2.81, 2.81; 2.71, 2.71, 2.71; 2.75; 2.71, 2.71; 2.77, 2.77, 2.77, 2.77]

and

$M(ch_v)$ = [2.73; 2.75; 0 ; 2.75; 2.73; 2.75; 2.79; 2.81; 0 ; 2.81; 2.85; 2.81; 2.81; 2.71; 0 ; 2.71; 2.69; 2.71; 2.75; 2.77; 0 ; 2.77]

Reflecting (=inverting) the theme along the x-axis, we obtain:

$-M(ch)$ = [-2.75, -2.75,- 2.75;- 2.77;- 2.75;- 2.75;- 2.81,- 2.81,- 2.81,- 2.81;- 2.81,- 2.81;- 2.71,- 2.71,- 2.71; -2.75;- 2.71,- 2.71;- 2.77,- 2.77,- 2.77,- 2.77]

Reflecting $-M(ch)$ onto $M(ch_v)$, we obtain the similarity chain $S_m(ch, ch_v)$:

$S_m(ch, ch_v)$ = [0.02, 0, 2.75, 0.02, 0.02, 0, 0.02, 0, 2.81, 0, 0.04, 0, 0.1, 0, 2.71, 0.04, 0.02, 0, 0.02, 0, 2.77, 0]

We illustrate this in figure 1.

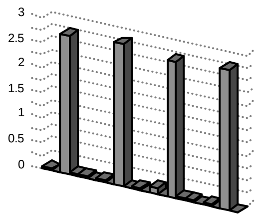


Fig. 1. The x-axis represent the atomic beat and the y-axis the corresponding value of the melotonic similarity chain, where Mozart's theme and variation are compared

As we can see, the lines reach a peak each time the first variation of the Mozart theme is on a rest. All other times, we find small deviations between the theme and the variation. Variation and theme even coincide in 9 atomic beats. Clearly, theme and variation fetch high melotonic similarity.

If we form the similarity chain comparing the c-chains, we obtain:

$S_c(ch)$ = [1.6, 1.6, 1.6; 0; 1, 1; 2, 2, 2, 2; 1, 1; 1.6, 1.6, 1.6; 0; 1, 1; 2, 2, 2, 2]

As we can see $S_c(ch, ch_v)$ and $C(ch_v)$ are identical. This is, because all chronotonic values of the variation fetch the value 0.

As we can see in the figure below, we obtain a pattern, whereby the chronotonic similarity is highest on the 4th atomic beat and lowest on the 7th to 10th atomic beat. Clearly, the theme and the variation seem to be little similar in terms of chronotonic similarity.

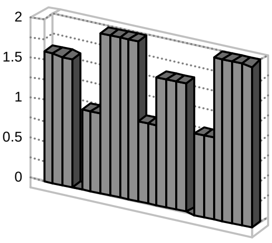


Fig. 2. Here, the x-axis represents the atomic beats and the y-axis the corresponding values of the chronotonic similarity chain, where Mozart's theme and variation are compared.

Before we will discuss more general features of the similarity chains, we will also produce the dynamic similarity chain. However, it must be stressed that this is somewhat problematic as the dynamic chains are the result of interpretation by the author rather than based on Mozart's score. Even more, going by Mozart's instructions, both the theme and the variation are to be performed piano. Accordingly, all values of the similarity chain would fetch the value 0. However, if we considered our earlier interpretation as acceptable, we obtain the following similarity chain:

$S_d(ch)$ = [0, 1, 4, 1, 2, 0, 0, 1, 3, 1, 1, 1, 0, 1, 4, 1, 2, 0, 0, 1, 3, 1]

As shown in figure 3, the similarity between theme and variation is smallest on the atomic beats 3 and 15. This is a rest in the variation (dynamon value 0) is compared to a tone on an down beat. However, we find, unlike in the context of chronotonic similarity that the dynamic similarity chain still fetches a 0 value on 6 atomic beats. Hence, we assume that the dynamic similarity is higher than the chronotonic similarity. This does not surprise: Mozart clearly overrides the original rhythm in his first variation by implementing a monotonous c-chain, while the dynamics would remain somewhere similar, because the m-chains are similar and because the accents will be somewhat the same as both pieces are written in a 6/8 time.

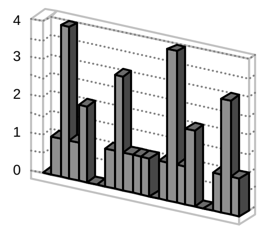


Fig. 3. Again, the x-axis represents the atomic beats while the y-axis represents the corresponding values of the dynamic similarity chain, where Mozart's theme and variation are compared

We finally point out two specific features of similarity chains: (A) A straight line such as $S(ch)$ = [1, 1, ..., 1] indicates in the context of melotonic similarity a transposition, in the context of dynamic similarity a volume change and in the context of chronotonic similarity a split or fusion of chronota. (B) The curvier the line (e.g. $S(ch)$ = [1, 4, 2, 9 ....], the more the two compared chains differ in shape. As shown by Hofmann-Engl & Parncutt (1998), melotonic similarity can be predicated by referring to the transposition interval and the interval difference (shape). Further (Hofmann-Engl 2002a) produced data which indicate that the distance of the chrontonic similarity chain is the sole predictor for chronotonic similarity.

The points of a similarity chain are given as:

$$p_{s_i} = \left| p_{1_i} - p_{2_i} \right|$$

where p_{si} is the ith point of the similarity chain, p_{1i} is the ith point of the first chain and p_{2i} is the ith point of the second chain

7 Similarity and interval vector

Mathematically, the similarity chain is the results from the composition of two reflections. Given the two m-vectors $\vec{M}_1$ and $\vec{M}_2$. We map $\vec{M}_1$ onto $\vec{M}_2$ via the following reflections:

$$\vec{M}_2 = \begin{pmatrix} -1 & 0 & . & . & 0 & -t_{11} + t_{21} \\ 0 & -1 & & & 0 & -t_{12} + t_{22} \\ . & & . & & & . \\ . & & & . & & . \\ 0 & 0 & & & -1 & -t_{1n} + t_{2n} \\ 0 & 0 & . & . & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & . & . & 0 & 0 \\ 0 & -1 & & & 0 & 0 \\ . & & . & & & . \\ . & & & . & & . \\ 0 & 0 & & & -1 & 0 \\ 0 & 0 & . & . & 0 & 1 \end{pmatrix} \begin{pmatrix} t_{11} \\ t_{12} \\ . \\ . \\ t_{1n} \\ 1 \end{pmatrix} = \vec{M}_2$$

Isolating the last column of the first matrix in the subspace n, we obtain the similarity vector $\vec{S}$:

$$\vec{S} = \begin{pmatrix} -t_{11} + t_{21} \\ -t_{12} + t_{22} \\ . \\ . \\ -t_{1n} + t_{2n} \end{pmatrix}$$

The length of the similarity vector serves as a predictor for the following similarity features:

- **Melotonic** (pitch) **similarity**: The longer the similarity vector, the more the two chains differ in average pitch. In case all components of the similarity vector fetch the same value, we are dealing with a transposition.
- **Dynamic** (loudness) **similarity**: The longer the similarity vector, the more the two chains differ in average loudness. In case all components of the similarity vector fetch the same value, we are dealing with a volume change.
- **Chrontonic** (rhythmic) **similarity**: The longer the similarity vector, the more the two chains differ in density (while one chain may consist of long durations, the other chain may consist of short durations). In case all components of the similarity vector fetch the same value, all chronota are split into chronota of the same ratios (e.g. quarter into two eighths, an eighth into two sixteenths etc.).

In our example above we did not only consider the overall distance of the similarity chains but also the "curviness" of the chains. At first, we might think that this curviness could be measured by the angle between the diagonal and the similarity vector. However, this is not the case, as the following example will illustrate. Given the two m-vectors:

$$\vec{m}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \vec{m}_2 = \begin{pmatrix} 1 \\ x \end{pmatrix}$$

where x is a variable, we find that the similarity vector fetches the value:

$$\vec{S} = \begin{pmatrix} 0 \\ x - 1 \end{pmatrix}$$

Now, we find that for $x > 1$, that the angle between the similarity vector and the diagonal $D = (1, 1)$ is:

$$\text{angle}(\vec{S},\vec{D}) = \cos^{-1}\frac{(1\cdot 0)+(x-1)}{\sqrt{1+1}\cdot\sqrt{(x-1)^2}} = \cos^{-1}\frac{x-1}{\sqrt{2}\cdot(x-1)} = \cos^{-1}\frac{1}{\sqrt{2}} = 45^\circ$$

This is true for any value of x>1. However, the curviness of the similarity vector will be affected by the value of x and hence, the angel is not a suitable measurement. Instead of the angle, we will introduce the interval vector in the subspace n-1 as:

$$\vec{I} = \begin{pmatrix} s_2 - s_1 \\ s_3 - s_2 \\ . \\ . \\ s_n - s_{n-1} \end{pmatrix}$$

where s_1, s_2, ..., s_n are the components of the similarity vector.

The length of the interval vector serves as a predictor for the following similarity features:

- **Melotonic** (pitch) **similarity**: The longer the interval vector, the more the two m-chains differ in their interval sequences. Note, interval difference captures by how many cents two intervals differ, while contour captures directions only. As we will see, contour is not a similarity predictor, while interval difference is.
- **Dynamic** (loudness) **similarity**: The longer the interval vector, the more two d-chains differ in their dynamic interval sequences.
- **Chrontonic** (rhythmical) **similarity**. As we will see in the next section, experimental evidence indicates that the chrontonic interval vector is not a similarity predictor.

Note, chrontonic and rhythmical similarity are not identical. This is, rhythmical similarity incorporates both durations and accents. Chrontonic similarity incorporates durations only. Accents are to be seen a dynamic aspect and hence they are covered by dynamic similarity.

Finally, we have not considered chains of different length. Hofmann-Engl (2002) proposed the following approach. Let *L* be the length of the chain *ch* and *L*' the length of the chain *ch*', we can write:

$$L' = a \cdot L$$

Basing a similarity predictor upon the factor a, we can correlate similarity in the following fashion:

$$S \propto \ln^2 a$$

Note, the lengths of the similarity and interval vectors are correlated to similarity. However, similarity models might be based upon the lengths but will not necessarily be identical with the lengths. We will now consider some experimental findings.

7 Experimental findings

Hofmann-Engl & Parncutt (1998) conducted two experiments testing melotonic similarity and Hofmann-Engl (2002) undertook one experiment testing chrontonic similarity. Dynamic similarity remains untested. The results will be briefly summarised in the following subsections.

7.1 Two experiments on melotonic similarity

Both experiments were conducted over a mixed sample of 17 and 20 people. The stimuli were melodic fragments consisting of 1 to 5 tones of equal length (isochronous). A fragment *a* was played followed by a fragment *b* and the participants were asked to rate the similarity on a scale of 1 to 9. The stimuli included (a) transpositions, (b) contour changes, (c) interval changes, (d) tempo changes and (e) inversions.

The following results emerged:

- No order effect was observed (significance level 95%). This is, it did not matter whether the fragments were played in order a - b or in order b - a.
- Tempo change was ignored by the participants (even when the tempo was changed by a factor 6). Clearly, participants recognized tempo changes, but decided to ignore them.
- In the first experiment a correlation between transposition and similarity of $r^2 = 0.72$ ($p < 0.005$) was established. In the second experiment length of fragments and transposition interval were manipulated simultaneously. Multiple regression revealed a correlation of $r^2 = 0.79$ (with p(interval) < 0.001 and p(length) < 0.01). Inputting the data into a model based upon the similarity vector produced a correlation of $r^2 = 0.92$ ($p < 0.003$).
- Inverting fragments did not produce significantly different similarity ratings (significance level 95%).
- Inputting interval difference and contour difference revealed that contour difference is not a significant predictor ($p > 0.2$).

- A model (compare Hofmann-Engl, 2001) based upon the similarity and interval vector produced a correlation of $r^2 = 0.74$ within the first experiment The second experiment was designed to test specific cases (such as multiple correlation between transposition, length and similarity, hence the data were not inputted into the model.

7.2 Experiment on Chrontonic Similarity

This experiment was conducted over a mixed sample of 18 people. The stimuli were rhythmical fragments consisting of 1 to 9 durations (e.g. 6 eighths notes compared to 3 quarter notes). All tones had the same loudness and frequency. A fragment *a* was played followed by a fragment *b* and the participants were asked to rate the similarity on a scale of 1 to 9. The stimuli included (a) split ratio (e.g. splitting a quarter into two equal parts (eighth notes) and during another trial splitting a quarter into a ratio 7:3 which is approximately a dotted eighth and a sixteenth note), (b) reversal (i.e. playing a rhythmical pattern backwards), (c) complexity (i.e. comparing a simple rhythmical sequence with a complex sequence), and (d) tempo changes.

The following results were obtained.

- One of the trials produced an order effect (significance level 95%). However, the order effect disappears when setting the significance level at 96%.
- Split ratio: It was found that the more equal the split ratio of a duration is the smaller are the similarity ratings. This is, considering the split ratio 1:x, we obtain minimal similarity for x=1 and increasing similarity for increasing values for x. It also was found that the length of the fragments is a second predictor (the more durations are compared the higher the similarity). Multiple correlations produced $r^2 = 0.77$ with p(split ratio) < 0.001 and p(length) < 0.02.
- Reversing rhythmical sequences did not produce any effect (t-test, $p < 0.05$).
- Complexity: It was shown that trials comparing simple patterns with complex patterns produced a significantly lower similarity rating (t-test, $p < 0.004$).
- Tempo change: It was found that changing tempo in the context of chrontonic (rhythmical) similarity affects the similarity ratings. This is, the larger tempo change the smaller the similarity ($r^2 = 0.77$, $p < 0.001$).
- Implementing the data into a model (compare Hofmann-Engl, 2002a), produces a correlation of $r^2 = 0.79$, $p < 0.001$. However, the interval vector appeared to have negative influence and hence it was omitted.

8 Conclusion

In this paper we argued that existing approaches to the phenomenon of melodic similarity are insufficient. Instead of replacing these models by yet another model, we presented a novel musical representation in form of atomic chains. These chains enabled us to transform any given chain into any other chain. We then introduced the concept of the similarity and interval vector which can be considered as the theoretical framework for melodic similarity. Finally, we presented some experimental data which are in support of this approach. In terms of creativity we argue that no specific model will be needed as long as the composer is aware of the factors which determine melodic similarity. However, in order to produce a more comprehensive knowledge of melodic similarity, much experimentation (such as testing dynamic similarity) will be needed.

References

1. Downie, J.S. (1999). Evaluating a simple approach to music information retrieval: Conceiving melodic N-grams as text. Graduate Programme in Library and Information science. London, Ontario, University of Western Otario: 179.
2. Hofmann-Engl, L. (1989). Beiträge zur theoretischen Musikwissenschaft. M 65+1, TU Berlin
3. Hofmann-Engl, L. & Parncutt, R. (1998). Computational modelling of melodic similarity judgments - two experiments on isochronous melodic fragments. Online: http://www.chameleongroup.org.uk/research/sim.html
4. Hofmann-Engl, L. (1999). Virtual pitch and pitch salience in contemporary composing. In Proceedings of the VI Brazilian Symposium on Computer Music, Rio de Janeiro
5. Hofmann-Engl, L. (2001). Towards a cognitive model of melodic similarity. In: Proceedings of the 2nd annual ISMIR, Bloomington
6. Hofmann-Engl, L. (2002a). Rhythmic Similarity: A theoretical and empirical approach. In Proceedings of ICMPC 7, Sydney
7. Hofmann-Engl, L. (2002b). Melodic transformations and similarity: a theoretical and empirical approach. PhD thesis, Keele Univerisity
8. Hoos, H., Renz K., Görg, M.(2001). Guido/Mir - an experimental Musical Information Retrieval System based on GUIDO Music Notation. In Proceedings of ISMIR 2001, Bloomington, Indiana.
9. Maidin O, D. (1999). A Geometrical Algorithm for Melodic Differences. In: Melodic Similarity - Concepts, Procedures, and Applications (ed. Hewlett & Selfridge-Field). Computing in Musicology, vol 11.

10. Schouten, J. F. The perception of subjective tones. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen, 41, 1938, 1418-1424
11. Smith, L.A.,McNab, R. J. & Witten, I. H. (1999). Sequence-based melodic comparison: A dynamic-programming approach. In: Melodic Similarity - Concepts, Procedures, and Applications (ed. Hewlett & Selfridge-Field). Computing in Musicology, vol 11.

12. Terhardt, E. (1979). Die psychoakustischen Grundlagen der musikalischen Akkordgrundtöne und deren algorithmische Bestimmung. In Tiefenstruktur der Musik, TU Berlin, 1979