Chameleongroup – online publication December 2004

Melodic similarity – providing a cognitive groundwork

Ludger Hofmann-Engl
Essex University
Department of Psychology
The Link Schools London
hofmann-engl@chameleongroup.org.uk

Ludger Hofmann-Engl
18 Howberry Road
London CR7 8HY
UK

Abstract

This article deals with the fundamental question of how to evaluate melodic similarity as a cognitive concept. As will be shown, existing approaches to melodic similarity do by large not consider the underlying cognitive principles behind the concept of melodic similarity. Thus, a great deal of similarity models have been developed over the last years making the question of how to rate these models more and more urgent. However, as this article endeavors to demonstrate, much of the issue at hand can be resolved when considering empirical data, musical experience and cognitive principles. Still, the emphasis of this article will not be so much the investigation on how much models comply with the cognitive constrains, but to produce a path based upon these constrains, which allows us to sketch a model. As we will see, this article cannot give final answers, but can demonstrate that enough data are available in order to give some direction to the further development of an understanding of melodic similarity.

1. Introduction

Melodic similarity has received a steady increase of interest within the last decade. While it appears that earlier works were predominantly influenced by algorithms which were original designed for the comparison of strings such as DNA strings or letter strings (compare Smith, McNab & Witten, 1998), we now find that many more algorithms have been developed such as the geometric measure (Ó Maidín, 1998), transportation distances (Typke, Giannopoulos, Veltkamp, Wiering & Oostrum van 2003), musical artist similarity (Ellis, Whitman, Berenzweig & Lawrance, 2002), probalistic similarity (Hu, Dannenberg, & Lewis, 2002) statistical similarity measures (Engelbrecht, 2002), transformational models (Hofmann-Engl, 2001, 2002a, 2002b, 2003a, 2003b) and transition matrices (Hoos, Renz & Görg, 2001). Additionally, given the fact that we are faced with a number of similarity measures the need for comparison has been recognized (e.g. Toiviainen & Eerola, 2002; Grachten, Arcos & Mántaras 2002; Müllensiefen & Frieler, 2003, 2004). However, as much as a comparison of the reliability and validity of these models is required, the method of how such a comparison ought to be conducted is more than unclear. In fact, it appears to the author that there are two methods which could be employed. The first method would look into the explicit and implicit underlying concepts of a model. Should it turn out that such underlying concepts are contradictory, that they are musically meaningless or in conflict with the cognitive science, then such a model will have to rejected. Models which pass the criteria as acceptable will inevitably be of a similar status. The second method consists of inputting the same set of data into different models and to see which model produces the best match. However, this second method, which for instance has been employed by Müllensiefen & Frieler (2003, 2004) and Schmuckler (1999), is valuable but comes with some problems of its own: This is, assuming that a model M_1 proves to be the best match to the data set S_1, there is no guaranty that the model M_2 will be a better match to the data set S_2 and another model M_3 the best match for the superset of S_1 and S_2. In fact this is exactly what happened to the study by Müllensiefen & Frieler (2003, 2004). These researchers found that the order of the models they compared was consistent to some extend only. Unless there is some theoretical foundation there is no way of deciding which of the models fulfills the minimum requirements. This is the goal of this article; to determine the psychological constrains and to analyze the parameters and how these parameters will have to be inputted into a similarity model. In the first instance we will examine the psychological approach to similarity modeling. We then consider the psychological values or parameters and which features are of importance in the context of melodic similarity. Finally, we will investigate how these features can be implemented into a similarity model.

2. Similarity modeling within the cognitive science

There have been extensive discussions over whether Tversky's approach (1977) is the "correct" approach or whether Shepard's approach (1987) is more suitable (e.g.

Goldstone, 1994; Hofmann-Engl, 2003b). However, the author believes that this apparent dichotomy between the two models does not exist. In fact both models complement each other if some generalizations are undertaken.

Tversy's model appears in the following form:

$$S = \theta c + \alpha a + \beta b \qquad (1)$$

> where S is the similarity θ, α, β are empirical constance, c the count of common features, a the count of features present in object A but not in B, and b the count of features present in B but not in A.

This model might seem suitable at first, but a simple mind experiment can demonstrate that it cannot be the final answer. Let us assume we have three triangles *A*, *B* and *C* of the same color and equal in every other respect except size. This is, *A* shall be the smallest and *C* the largest triangle. Clearly, *A* ought to be more similar to *B* than to *C*. Tversky's model is unable to handle this situation as it will regard all three triangles as being of the same similarity status.

In contrast to Tversky's model, Shepard's model can handle this example. The Shepard model is written in the following form:

$$d(x, y) = \left(\sum_{k=1}^{D} |x_i - y_i|^p\right)^{(1/p)} \qquad (2)$$

> where $d(x,y)$ is the distance between the objects x and y, D is the dimension of the objects, x_i is the ith feature of the object x, y_i is the ith feature of the object y and p is an empirical constance.

Applying this model to our mind experiment above, setting $x_i - y_i$ to be the difference in size, we obtain: $d(A,B) < d(A, B)$, just as we would expect it. This model works also in a musical context. As found by Egmond, Povel and Maris (1996) and confirmed by Hofmann-Engl & Parncutt (1998), melodic similarity is correlated to the transposition interval. This is, the further a melody is transposed, the smaller are the measured similarity ratings. If for instance the melody $M_1 = [m_1, m_2, ..., m_n]$ then follows that $M_2 = [m_1 + c, m_2 + c, ..., m_n + c]$. Regarding each pitch as an attribute, we obtain:

$$d(M_1, M_2) = (|(m_1 + m_2 + ... m_n) - (m_1 + c + m_2 + c + ... + m_n + c)|^p)^{(1/p)} = nc$$

The fact that Shepard's model yields a distance where c is a factor, is in accordance with the above mentioned findings. However, the factor n is in contrast to experimental data: longer melodies do not produce increasingly smaller similarity ratings (Hofmann-Engl & Parncutt, 1998). Dividing Shepard's model by n (as proposed by Kluge, 1996) is a step into the right direction.

However, Shepard's model gets into more serious trouble when we consider a second mind experiment. This time, let us compare three cars which are all equal except in speed and weight. According to Shepard's model, we are supposed to work out the difference in speed Δs, the difference in weight Δw and then add this differences. This is: $d(x, y) = \Delta s + \Delta w$. However, adding weight measure (for instance kg) to a speed measure (for instance km/s) is entirely meaningless. Clearly, the only way around this problem is to work out the speed and weight similarity (not fetching any units) separately first, and then to add these similarity ratings according to the following principle:

$$S(x,y)=c_1 S(D_1)+c_2 S(D_2)+...+c_n S(D_n) \quad 3$$

with: $S(D_i)=f(d_i(x,y))$, $0 \leqslant S(D_i) \leqslant 1$ and $c_1+c_2+...+c_n=1$

where $S(x,y)$ is the similarity between the objects x and y, c_1, c_2, ..., c_n are empirical constances , $S(D_i)$ is the similarity across the ith dimension, f is a function and $d_i(x,y)$ is the distance between x and y along the ith dimension.

Setting the similarity of the ith dimension to be greater equal 0 and smaller equal 1 and setting that the constances c_1, c_2, ..., c_n add up to 1, ensures that the similarity rating fetches a number between 0 and 1. This is, two objects can at best be equal (100% similarity) or totally different (0% similarity).

Interestingly, this model is a generalization of Tversky's model incorporating Shepard's distance model. This is, if two objects have a feature in common, which means they are identical along one specific dimension D_i, then $S(D_i) = 1$, if they are totally different along another dimension D_j, which means they have nothing in common along this dimension, then the similarity rating $S(D_j) = 0$, and hence the overall similarity rating is reduced.

If we want to set the range for the similarity functions to be between 0 and 1, it appears to be appropriate to set either:

$$S(x,y)=f_1 e^{-f_2(d(x,y))} \quad 4$$

or

.

$$S(x,y)=f_1 \sum_{j=1}^{n} e^{-f_2(d(x_i-y_i))} \quad 5$$

where $S(x,y)$ is the similarity between the objects x and y, f_1 and f_2 are functions, $d(x, y)$ is the distance between x and y and $d(x_i-y_i)$ the ith distance between x and y and n is the number of distances between x and y.

As we will discuss further on, model 5 seems to be more appropriate than model 4.

Another psychological issue, which is of importance, is the question whether similarity judgments are symmetrical or not. While Tversky (1977) found asymetries under specific condition (when comparing New York with Tel Aviv), Hofmann-Engl & Parncutt (1998) did not find any asymmetries in the context of melodic similarity judgments. However, the stimuli used by these researchers were highly artificial (not even using equal temperament) and very short. Hence, the question whether such asymmetries will be found when more complex melodies, which are of different salience, are compared is an open question.

There are a great deal of other psychological issues which have been shown to be of importance (compare Goldstone, 1994; Deliege, 2000 and Hofmann-Engl, 2003b) including data which indicate that similarity judgments dependent on age, expertise, language/culture, context, the extraction of the relevant dimensions and categorization. However, Hofmann-Engl (2003b) argued that to incorporate these aspects into a formal melodic similarity model would be overambitious at the present time.

However, we can conclude that a cognitive model of melodic similarity will be of the form as written in formula (3). We will now investigate the parameters which appear to be of cognitive relevance in the context of melodic similarity.

3. Melodic similarity parameters

Generally, when melodic similarity is investigated, there are mostly two parameters considered. The two dimensions are: Pitch and Duration (e.g. Ó Maidín, 1998; Smith, McNab & Witten, 1998; Typke, Giannopoulos, Veltkamp, Wiering & Oostrum van, 2003). However, to the knowledge of the author there has been only one attempt to consider dynamic values (Hofmann-Engl, 2003b) as an independent dimension. Still, maybe more far reaching than deciding whether to implement dynamic values or not, is a point made by Shepard (1987): it is common psychological knowledge that physical values such as measured time, frequency and sound pressure are somewhat correlated to our perceptional and cognitive makeup, but they are not synonym. Therefore, Shepard suggested that a cognitive model of similarity cannot be based upon physical dimensions but has to be based upon psychological dimensions. This means, in the context of melodic similarity, that inputting the fundamental frequencies of a series of tones is not satisfying from a psychological point of view. Taking this critique into account, Hofmann-Engl (2001, 2002a, 2003b) attempted to construct such psychological dimensions coining the terms *meloton*, *chronoton* and *dynamon*. We will briefly outline these concepts below.

3.1The Meloton

Definition

The *meloton* is the cognition resulting from the intentional listening out for the quality of a sound which is signified by the statement that the sound is high or low.

This definition is vague and does not deliver us any values which we can input into a similarity model. Hence, we have to determine values which will be called *melotonic* values. There are two principle ways of obtaining such values: we either predict such values through some pitch extracting model or we measure such values. The measurement of these values could be obtained by employing a experimental setup as introduced by Schouten (1938), where participants were asked to adjust a variable sinusoidal tone so as to match the tone for which we are seeking to determine the melotonic value. In a procedure described in detail in Hofmann-Engl (2003b), the melotonic value is obtained by taking the frequency of the sinusoidal tone where the majority of participants have tuned in (measured in log frequency) or in case there is no majority to accept the mean frequency of all measured frequencies as the melotonic value. The prediction of melotonic values through pitch extracting models, such as the model by Meddis & Hewitt (1991) or the model by Terhardt, Stoll & Seemann (1982) and Hofmann-Engl' smodel (1990) have not as yet been sufficiently tested against measured data, and hence the reliability of these models cannot be established as yet (although this work is in progress).

3.2The Dynamon

Definition:

The dynamon is the cognition resulting from the intentional listening out for the quality of a sound which is signified by the statement that a sound is loud or soft.

Again, this definition is vague and does not deliver us any values which we can input into a similarity model. Hence we have to determine values which will be called *dynamic* values. Hofmann-Engl (2003b) proposed to measure the *dynamic* value of tone by asking participants to adjust a sinusoidal tone (1 kHz) or narrow noise band around 1 kHz until it matches the loudness of the test tone. Again, should the majority of participants tune in on a peak (measured in dB), this peak will represent the dynamic value of the tone. If there is no peak then the mean value of the measured loudnesses will serve as the dynamic value. A detailed description can be found in Hofmann-Engl (2003b). A predictor for the dynamic value is the perceptual center (compare Stecker, 1996).

3.3.The Chronoton

Defintion:

The chronoton is the cognition resulting from the intentional listening out for the quality of a sound which is signified by the statement that the sound has a shorter or longer duration.

As observed by Allan (1979) subjective time is generally directly proportional to physical time. However, we face another issue in the context of chronota. True, that we might be able to determine the duration of one single chronoton, but to determine where a tone starts and where it ends within a melody, can be a rather difficult question. For instance, a fast run performed on an instrument might cross the boundaries of the fusion threshold and be perceived as a glissando rather than a series of single notes. On the other hand, there are situation where one continuous sound will be segmented into two due to the Doppler effect. How to deal with these issues is described in detail by Hofmann-Engl (2003b).

We are now in the position to discuss how these parameters can effect similarity judgments in the context of melodic similarity.

4. Factors affecting melodic similarity

The question which factors are of cognitive relevance is quite possibly the most important one. Strangely, so it appears, this question has never been asked in a systematic fashion except by Gómez, Klapuri & Meudic (2003) and Hofmann-Engl (2003b).

In order to determine these factors, there are two methods at hand. Firstly, we can refer to experimental data and secondly, where there are no or limited data, we can perform thought experiments based upon musical experience and common sense. The author will open the discussion with a thought experiment which, he believes, can demonstrate that contour is not a melodic similarity predictor.

Let S be the set of all melodies possessing the contour: up – up. As a subset of S we can construct the subset $S_{equal} = \{ (m_1, m_1 + 1, m_1 + 2), (m_1, m_1 + 2, m_1 + 4), ..., (m_1, m_1 + n, m_1 + 2n)\}$. Thus, the melodies: $M_1 = [c, d, e]$, $M_2 = [c, d\#, f\#]$, $M_3 = [c, e, g\#]$ and $M_4 = [c, f, b^b]$ or even $M_5 = [c, f\#\ c_{octave}]$ are elements of S_{equal}. Not only, that a contour model will fail to differentiate between different degrees of similarity (there is no musical reason to assume that the distances $d(M_1, M_2)$ and $d(M_1, M_5)$ are equal), but it identifies all elements of S_{equal} to be the same. This would imply that no difference was to be heard, and this is most certainly not the case. Even more complex contour measures such as Friedman's (1987) or Marvin & Laprade's (1987) will fail in this context as will Steinbeck's (1982) and Müllenseifen and Frieler's (2003).

There are a number of experimental studies which investigate contour similarity (e.g. Hofmann-Engl & Parncutt, 1998; Schmuckler, 1999; Müllenseifen and Frieler's, 2003) where it has been shown that contour appears to be a reasonable similarity predictor. Therefore the question: If our thought experiment demonstrated that contour fails at least in the above example, how then can contour be a reasonable predictor in other cases? In order to answer this question we have to consider the experiments by Hofmann-Engl & Parncutt (1998), which is described in more detail in Hofmann-Engl (2003b). There, contour differences accounted for 61% of the variance ($p < 0.008$). However, computing the correlation between the measured data and the exact interval differences accounted for 72% ($p < 0.001$) of the variance. Moreover, multiple correlation including exact interval difference and contour as similarity predictors accounted for 75% of the variance with contour showing up to be insignificant ($p > 0.2$). It seems there is only one explanation: Each time two melodies differ in contour, they also differ in their exact intervals. However, at times there might be an interval difference (one melody going up two steps while the other melody going up one step only) but no contour change. Interestingly, the studies by Müllenseifen and Frieler's (2003, 2004) confirm that contour is an embedded factor. Hence, contour captures partially interval difference but is embedded in the latter. Thus, we conclude that interval difference is a similarity predictor and contour is not.

In case the reader is still in doubt as to whether to accept that contour is no predictor or not, the author proposes one more thought experiment: Suppose a melody progresses form *c* to *d*. Now we substitute the *d* by a *b* (semitone down form *c*). In a second instance we substitute the *d* by a *f#* two and a half octaves above *c*. Well, all musical experience will clearly see (hear) that the second change is far more drastic than is the first one, although the contour is changed in the first but not in the second case. Hence, contour cannot be a similarity predictor, and interval difference appears to be a similarity predictor.

For the purpose of clarity we define the melotonic interval difference, which we will call melotonic interval distance from now on. Let two m-chains $m\text{-}ch_1$ and $m\text{-}ch_2$ be: $m-ch_1=[m_{11,} m_{12,} m_{13,} \ldots, m_{1n}]$ and $m-ch_2=[m_{21,} m_{22,} m_{32,} \ldots, m_{2n}]$. The melotonic intervals of $m\text{-}ch_1$ are $m_{1i}-m_{1(i+1)}$ and the intervals of $m\text{-}ch_2$ are $m_{2i}-m_{2(i+1)}$. The melotonic interval distance is:

$$d_i(m-ch_{1,} m-ch_2)=[(m_{1i}-m_{1(i+1)})-(m_{2i}-m_{2(i+1)})]$$

where $d_i(m-ch_{1,} m-ch_2)$ is the *i*th melotonic between the melotonic chains $m\text{-}ch_1$ and $m\text{-} ch_2$, m_{1i} is the *i*th meloton of $m\text{-}ch_1$ and m_{2i} is the *i*th meloton of $m\text{-}ch_2$.

A second and at first surprising factor was discovered by Egmond, Povel & Maris (1996) and confirmed by Hofmann-Engl & Parncutt (1998). The factor in question is *transposition interval*. This is, both studies show that the larger the transposition interval is, the smaller is the measured similarity. These findings are supported by a study conducted by Levitin (1994) who demonstrated that some form of absolute hearing is more common than previously thought. Levitin's study is interesting as it showed that participants tend to sing their favorite pop song in the right key but don't show such consistency when recalling a nursery rhyme. Generally, we hear pop songs just in one key and nursery rhymes in many different keys. This might be an important fact when devising a retrieval algorithm and adds support to the argument above.

The fact that similarity ratings decrease with increasing transposition interval might appear at odds with our musical intuition, and indeed most similarity models (e.g. Ó Maidín, 1998; Engelbrecht, 2002; Müllensiefen & Frieler, 2004) are transposition invariant. However, conducting yet another thought experiment, we might understand the issue better.

Let us assume, we hear the nursery rhyme *Mary had a little lamb* played in C major: *e, d, c, d, e, e, e,* .... and then repeated in D major: *f#, e*, *d*, *e*, *f#*, *f#*, *f#* ..., the change might be drastic enough for us to be unable to recognize the melody for the first four or five notes altogether. Hence, if we do not recognize the melody, there must have been a similarity change. The reader might argue that this change is due to a change of key, but as the study by Egmond, Povel and Maris (1996) demonstrated, key relationships are a minor or no significant similarity predictor altogether. Additionally, the concept of modulation or the Neapolitan Sixth chord would make little sense. Why, for instance, modulate and repeat a theme in transposition if there is no perceptual difference between transposed material? We conclude, transposition is a similarity factor.

A third factor which influences similarity judgments was observed by Gabriellson (1973) and Hofmann-Engl (2002): this is *tempo*. In an experiment, which investigated rhythmic similarity, it was found that similarity ratings decrease when rhythmical sequences are played at increasingly different tempo. This effect, however was not observed when isochronous melodies underwent tempo variations (Hofmann-Engl & Parncutt, 1998). Hence, we consider tempo variations a factor which determines chronotonic (rhythmic) similarity and not melotonic similarity. Additionally, the fact that tempo changes have an effect on melody recognition (which is interlinked with similarity) has been observed by Andrews, Dowling & Bartlett (1998). Hence, we conclude that tempo is a similarity factor.

A second chronotonic similarity factor was observed by Hofmann-Engl (2002a) based upon what has been coined atomic notation. The method whereby rhythms are presented as a series of values which themselves are multiples of a set value has been used widely (e.g. Nettheim, 1992; Lemström & Laine, 1998). Now, atomic notation is a modification of this representation and was introduced by Hofmann-Engl (2002a, 2002b, 2003a, 2003b). We give an example (figure 1):

Figure 1: The opening of the piano sonata in A major be W. A. Mozart KV 331.

The chronotonic chain (rhythm) of this incipit is depicted by extracting the atomic beat (smallest chronotonic value). In the example above the smallest chronotonic value is 1/16 note. The *c#* at the beginning is three times longer than the atomic beat, the second note is one time as long, the third note twice as long and so on. Additionally, we quantize the time line into 1/16 values. Each time interval finally fetches the value which the corresponding chronoton fetches in terms of multiples of the atomic beat (for more detail compare Hofmann-Engl, 2003b). For the incipit above, we have 24 time intervals and hence, we obtain the following chronotonic chain c-chain:

C-Chain(theme) = [3, 3, 3; 1; 2, 2; 4, 4, 4, 4; 2, 2; 3, 3, 3; 1; 2, 2; 4, 4, 4, 4; 2, 2](1/16)

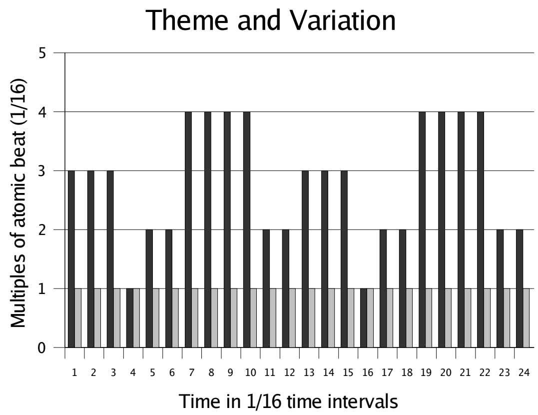
The corresponding two bars of the first variation of Mozart's theme are the following (figure 2):

Figure 2: The incipit of Mozart's first variation of the piano sonata in A major (KV 331).

In this case now we obtain the c-chain:

C-chain(variation) = [1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1; 1](1/16)

Plotting both chains in a box diagram, we obtain graph 1:



Graph 1: The dark gray columns represent the chronotonic values of the theme and the light gray columns the chronotonic values of the variation. While the variation produces a series of columns with identical height (as the variation is based upon a series of 16ths notes), we find that the chronotonic values of the theme (as the theme does not possess an isochronous rhythm) differ.

We find that the dotted eighths note (*c#*) of the theme produces the first three dark gray columns. The fourth column fetches the value 1 only because the *d* is a 16ths note and thus is identical with the atomic beat and so on. All light gray columns fetch the value 1 as the variation of Mozart theme is based upon 1/16 note values.

As shown by Hofmann-Engl (2002a, 2003b), the distance between the multiples of each time unit is a powerful predictor of chronotonic (rhythmic) similarity (with a correlation of over 88%). We will define the chronotonic similarity now:

Let M_1 and M_2 be two melodies of the same length. Further, Let $C\text{-}Ch_1$ and $C\text{-}Ch_2$ be the chronotonic chains of M_1 and M_2 respectively. With a as the atomic beat for both chains, we can write:

$C-ch_1=[c_{11,}c_{12,}c_{13,}..,c_{1n}](a)$ and $C-ch_2=[c_{21,}c_{22,}c_{23,}...,c_{2n}](a)$

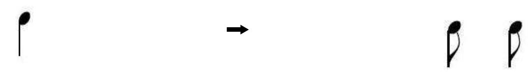
Now, as mentioned above, chronotonic similarity is correlated to the distances c_{1i} and c_{2i} . Hence, we can, in accordance with formula 5, write:

$$S(C-ch_1,C-ch_2)=\sum_{i=1}^{n} e^{-f(d(c_{1i}-c_{2i}))} \qquad 6$$

where $S(C\text{-}ch_1,\ C\text{-}ch_2)$ is the chronotonic (rhythmic) similarity between the two chronotonic chains $C\text{-}ch_1$ and $C\text{-}ch_2$, n the length of the chronotonic chains and $d(c_{1i}-c_{2i})$ the distance between the *i*th chronoton of the chains $C\text{-}ch_1$ and $C\text{-}ch_2$.

Inputting the data of an experiment into a specific form of this model produced a correlation of 89% with $p < 0.001$ (compare Hofmann-Engl, 2002a, 2003b). Hence, we accept that the principles behind this model are valid at least to some extend. However, the question is, what principles lie behind this chronotonic distance measure.

The most important principle is concerned with the "splitting" of chronota (durations). For instance, splitting a quarter into two eighths and in a second instance into one dotted eighth and one sixteenth, we get:



and:

In the first case, we obtain a split ratio 1:1 and in the second case 3:1. As shown in the experiment by Hofmann-Engl (2003b), the smaller the split ratio is, the larger is the measured similarity. This factor is captured by the chronotonic similarity measure as described above (formula 6).

The second factor is the following: Comparing one quarter note with two eighths notes produces higher similarity ratings than a quarter note compared to four sixteenths notes (Hofmann-Engl, 2003b). This factor is captured by the chronotonic similarity measure too. Hence, we accept the chronotonic distance as a similarity factor.

Dynamic similarity, to the knowledge of the author, has not been described in the literature except by Hofmann-Engl (2003b). Additionally, there are, so it appears, no experimental data available. Hence, we will conduct a thought experiment.

Firstly, if we imagine to listen to a piece of music through a stereo system, we might want to change the volume according to our liking. If however, the volume change made no difference to the way we perceive the music, there would be no point having a volume control on our stereo system. Hence, volume change is of perceptual relevance. If we can distinguish between two play back situations, whereby a volume change occurred, we can rate the similarity between these two situations. Let us now consider three payback situations P_1, P_2 and P_3. Clearly, we can order the similarity ratings now. Let us further assume that P_1 is the loudest and P_3 is the softest playback situation, a change along one dimension occurred, similarity judgments can be made, and hence P_1 will be closer to P_2 than to P_3. Admitted, volume change might be a small similarity factor, but there is good reason to assume that it will be.

Following a similar thought experiment, we can reason that not only volume change will affect similarity but the distance between dynamic intervals as well. This is, for two dynamic chains *d-ch*$_1$ and *d-ch*$_2$, we write: $d-ch_1=[d_{11,}d_{12,}d_{13,}\ldots,d_{1n}]$ and $d-ch_2=[d_{21,}d_{22,}d_{23,}\ldots,d_{2n}]$. The dynamic intervals are defined as $d_{1i}-d_{1(i+1)}$ and

$d_{2i} - d_{2(i+1)}$. Finally, we obtain the dynamic interval distance:

$$d_i(d-ch_{1,} d-ch_2) = [(d_{1i} - d_{1(i+1)}) - (d_{2i} - d_{2(i+1)})]$$

where $d_i(d-ch_{1,} d-ch_2)$ is the ith dynamic distance between the dynamic chains $d\text{-}ch_1$ and $d\text{-}ch_2$, d_{1i} is the ith dynamon of $d\text{-}ch_1$ and d_{2i} is the ith dynamon of $d\text{-}ch_2$.

Summarizing the factors which we consider to be of importance in the context of melodic similarity, we find the following:

- melotonic distance (difference between the pitch values)
- melotonic interval distance (distance between pitch intervals)
- chrontonic distance (distance between the durations under atomic notation)
- tempo distance
- dynamic distance (difference between dynamic values)
- dynamic interval distance (distance between relative dynamic values)

Note, in order to compute melotonic and dynamic values, we need to represent those in atomic notation.

We are now in the position to sketch a melodic similarity model which is thought to be of cognitive relevance.

5. Outline of a cognitive melodic similarity model

The first question we might ask, is, which are the independent and which are the dependent dimensions. We will open the discussion with the tempo factor.

As shown in the experiments by Hofmann-Engl & Parncutt (1998), tempo variations on isochronous melodies did not affect the measured similarity ratings significantly. However, tempo variations, in the context of rhythms with monotonous pitch, were a strong and significant similarity factor (Hofmann-Engl, 2002a). Thus, we conclude that the tempo dimension is independent of the melotonic dimension but not of the chronotonic dimension. If however, tempo and chronotonic similarity are not independent, then we treat both as part of the chronotonic dimension by writing:

$$S_c = S_{tempo} * S_{c\,distance} \qquad 7$$

where S_c is the chronotonic similarity, S_{tempo} the tempo similarity and $S_{c\,distance}$ the similarity based upon the chronotonic distance

Now, the similarity based upon melotonic distance and melotonic interval distance are dependent upon each other. This is, changing the melotonic position of a tone, will alter both the melotonic distance and the melotonic interval distance. Hence the melotonic similarity is:

$$S_m = S_{m\,distance} * S_{m\,interval} \qquad 8$$

where S_m is the melotonic similarity, $S_{m\ distance}$ the melotonic distance similarity and $S_{m\ iinterval}$ the similarity based upon the melotonic interval similarity.

In a similar fashion, we obtain for the dynamic similarity:

$$S_d = S_{d\,distance} * S_{d\,interval} \qquad 9$$

where S_d is the dynamic similarity, $S_{d\ distance}$ the dynamic distance similarity and $S_{d\ iinterval}$ the similarity based upon the dynamic interval similarity.

The overall melodic similarity then will be:

$$S_{mel} = \alpha S_m + \beta S_d + \gamma S_c \qquad 10$$

where S_{mel} is the melodic similarity, S_m the melotonic similarity, S_d the dynamic similarity, S_c the chronotonic similarity, α, β,γ are empirical constances and $\alpha+\beta+\gamma=1$

We now return to the question, which had been raised above, of whether formula 4 or formula 5 deliver a more appropriate similarity model. Formula 4 implies that an overall distance will be computed and then the similarity. Formula 5 implies that one distance at a time will be computed and then its corresponding similarity and then an overall similarity in the end. The author argues that it appears more likely that the latter process corresponds to the cognitive process involved. This is, single similarity ratings are made which are in the end put together in order to form an overall picture. However, there is no empirical evidence, to the knowledge of the author, that this is the case. Still, from an introspective point of view, formula 5 seems more plausible than formula 4, and hence, we will operate with formula 5.

The next step is crucial. True, that now there exists an infinite amount of possible formulae which could be used to be implemented into a similarity model, but this of no benefit to us. Hence, we need to introduce two further constrains. The first constrain arises from the fact that a model on melodic similarity will at best always be of an approximative nature. Therefore, it does not make sense to construct formulae which are overtly complex. The second constraint is that, ideally, a similarity model can be constructed within a systematic and mathematical framework. Such a mathematical framework (the only one known to the author) based upon melodic transformation, has been proposed by Hofmann-Engl (2001, 2002a, 2002b, 2003a, 2003b). This transformation theory however exceeds the limits of this article but can be found in great detail in the thesis by Hofmann-Engl (2003b). We will simply list the models here:

$$S_{m\,distance} = \sqrt{\frac{\sum_{i=1}^{n} (e^{(-c_1 D_{mi}^2)})^2}{n}} \quad 11$$

$$S_{m\,interval} = \sqrt{\frac{\sum_{i}^{n-1} (e^{(-c_2 I_{mi})^2})^2}{n-1}} \quad 12$$

$$S_{d\,distance} = \sqrt{\frac{\sum_{i=1}^{n} (e^{(-c_3 D_{di})^2})^2}{n}} \quad 13$$

$$S_{d\,interval} = \sqrt{\frac{\sum_{i=1}^{n-1} (e^{(-c_4 I_{di})^2})^2}{n-1}} \quad 14$$

$$S_{c\,distance} = \sqrt{\frac{\sum_{i}^{n} (e^{(-c_5 D_{ci})^2})^2}{n}} \quad 15$$

and

$$S_{tempo} = e^{-c_6 \ln^2 a} \quad 16$$

where $S_{m\,distance}, S_{m\,interval}, S_{d\,distance}, S_{d\,interval}, S_{c\,distance}, S_{tempo}$ are the relevant similarities, $c_{1,} c_{2,} c_{3,} c_{4,} c_{5,} c_6$ are empirical constances, D_{mi}, D_{di}, D_{ci} the *i*th melotonic, dynamic and chronotonic distances respectively, I_{mi}, I_{di} the *i*th melotonic and dynamic interval distances, *a* the tempo factor and *n* the length of the atomic chains.

While formulae 11 to 15 are more less self explanatory, formula 16 is not. However, it will become clear, if we follow the process of how similarity is to be computed.

Let as assume that we are about to compare two melodies M_1 and M_2. Let us further assume that M_2 lasts longer than M_2 by the factor *a* (tempo distance). We now can either stretch M_2 by multiplying it by the factor a or compress M_2 by the factor $1/a$. Inputing either a or $1/a$ into formula (16) produces the same similarity rating. This is the reason why $\ln^2$ is part of formula 16. The next step is to find the common atomic beat for both melodies and represent the melotonic, dynamic and chronotonic values in atomic notation (compare Hofmann-Engl, 2003a). Subsequently, the data are inputed into formulae 11 to 16 and the overall similarity is computed according to formula 10.

Although Hofmann-Engl (2003b) produced some experimental data which delivered values for the constances c_1, c_2, c_5 and c_6, the data are too few to be considered reliable. A great deal of experimental investigation is needed in order to produce reliable data for the constances. However, it can be speculated that these constances might be dependent apon age, expertise and culture. Still, it seems that working in cents that c_1 and c_2 are in the region of 10^{-6} to 10^{-7}, c_5 in the region of 10^{0} (where one atomic beat lasts 0.1 sec.) and c_6 in the region of 10^{-1}.

There are a great deal of other factors which might be of importance in the context of melodic similarity, such as melodic symmetries, tone repetitions or changing tones (compare Steinbeck, 1982). However, so the author argues, unless a stable model is established based upon the factors mentioned above, it appears on overambitious undertaking trying to understand let alone implement more complex factors.

6. Conclusion

Considering that a great variety of melodic similarity models exists, the author set out to investigate the issue at hand from a strict psychological, or more precisely, cognitive point of view. Such an approach is not only appropriate and long overdue but necessary considering that we are not interested in some abstract formulae or algorithms but human cognition. Commencing with some general psychological remarks on the cognition of similarity, the author proceeded to define the subjective dimension which are to determine the cognition of melodic similarity. We then investigated rigorously empirical data as well as referred to musical experience in order to determine the factors of the cognition of melodic similarity. As it appeared that none of the existing models take all the determining factors into account in a systematic fashion except Hofmann-Engl (2002b, 2003a, 2003b), we presented some features of this model. It was the purpose of this article to demonstrate that on the one hand, there is sufficient psychological knowledge to path the way towards melodic similarity and yet to demonstrate that a great deal of experimental data will be needed before more definite answers can be given.

References

Allan, L. G. (1979). The perception of time. In: Perception & Psychophysics, 26, 340-354

Andrews, M. W., Dowling, W. J. & Bartlett, J. C. (1998). Identification of Speeded and Slowed Familiar Melodies by Younger, Middle-Aged, and Older Musicians and Non-Musicians. Psychology of Aging 13.3, 462-471

Deliége, I. (2000). Perception of similarity and related Theories. In: Proceedings of ICMPC 6, Keele, UK

Egmond van, R. & Povel, D-J. & Maris, E. (1996). The influence 0f height and key on the perceptual similarity of transposed melodies. Perception & Psychophysics, vol. 58, 1252-1259

Ellis, D. P. W, Whitman, B., Berenzweig, A. & Lawrence, S. (2002). The Quest for Ground Truth in Musical Artist Similarity. In: Proceedings of ISMIR 2002, IRCAM, Paris

Engelbrecht, R. (2002). Statistical Comparison Measure for Searching in Melody Databases. Online publication at:
www-mmdb.iai.uni-bonn.de/download/ publications/melodic-comparison.pdf

Friedmann, M. L. (1987). My contour, their contour. J. of Music Theory 31, 268-274

Gabrielsson, A. (1973). Similarity ratings and dimension analysis of auditory rhythm patterns: I & II. Scandinavian Journal of Psychology 14, 138-176

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. Cognition 52: 125–157

Gómez, E. Klapuri, A. & Meudic, B. (2003). Melody Description and Extraction in the Context of Music Content Processing. Journal of New Music Research Vol.32 .1

Grachten M., Arcos J. L. & Mántaras de, R. L. (2002). A comparison of two approaches to melodic similarity. In: Proceedings of ICMAI' 02, Edinburgh

Hofmann-Engl, L. J. (2003b). Melodic similarity - a theoretical and empirical approach. PhD thesis, Keele University, UK
(available online at: www.chameleongroup.org.uk/research/thesis.html)

Hofmann-Engl, L. (2002a). Rhythmic similarity - a theoretical and empirical approach. In Proceedings: ICMPC, Sydney
(available online: www.chameleongroup.org.uk/research/rhythm.html)

Hofmann-Engl, L. J. (2002b). Melodic Similarity - a conceptual framework. In Proceedings of Understanding and creating Music, 2nd International Conference, Naples 2002

(available at: www.chameleongroup.org.uk/research/A36.html)

Hofmann-Engl, L. J. (2001). Towards a model of melodic similarity. In Proceedings of ISMIR 2001, Bloomington, Indiana
(available at: ismir2001.ismir.net/pdf/hofmann-engl.pdf)

Hofmann-Engl, L. & Parncutt, R. (1998). Computational modeling of melodic similarity judgments - Two Experiments on Isochronous Melodic Fragments.
Online Publication: www.chameleongroup.org.uk/research/sim.html

Hofmann-Engl, L. (1990). Sonanz und Virtualität. MA thesis, Berlin (unpublished)

Hoos, H., Renz, K. & Grög, M. (2001). Duido/Mir – an experimental Musical Information Retrieval System based on GUIDO Music Notation: In Proceedings of ISMIR 2001, Bloomington, Indiana

Hu, N., Dannenberg, R. B. & Lewis, A. L. (2002). A Probabilistic Model of Melodic Similarity. In: Proceedings of the 2002 International Computer Music Conference. San Francisco, pp. 471-4.

Kluge, R. (1996). Ähnlichneitskriterien für Melodieanalyse. In: Systematische Musikwissenschaft, 4.1-2, 91-99

Lemström, K. & Laine, P. (1998). Musical information retrieval using musical parameters. In: Proceedings of the International Computer Conference 1998

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. In: Perception & Psychophysics 58, 927-935

Ó Maidín, D. (1998). A Geometrical Algorithm for Melodic Difference. n: Melodic Similarity – Concepts, Procedures, and Applications (ed. Hewlett & Selfridge-Field), MIT Press

Marvin, E. W. & Laprade, P. A. (1987). Relating musical contours: Extensions of a theory for contour. J. of Music Theory 31, 225-267

Meddis, R. & Hewlitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch Identification. In: J. Acoust. Soc. Am. 89.6, 2866-2882

Müllenseifen, D. & Frieler, K. (2004). Measuring melodic similarity: Human vs. algorithmic judgments. In: Proceedings of CIM04, Graz

Müllenseifen, D. & Frieler, K. (2003). Cognitive Adequacy in the Measurement of

Melodic Similarity: Algorithmic vs. Human Judgments. In: Computing in Musicology 13 (Editor: Hewlett & Selfridge-Field), MIT Press

Nettheim, N. (1992). On the spectral analysis of melody. In: Journal of New Music Research 21, 135-148

Schouten, J. F. (1938). The perception of subjective tones. In Proceedings of the Koninklijkke Nederlandse Akademie van Wetenschappen 41, 1418-1424

Shepard, R. N. (1987). Toward a universal law of generalization of psychological science. Science, vol 237, 1317-1323

Schmuckler, M. A. (1999). Testing Models of Melodic Contour Similarity. In: Music Perception 16.3, pp 109-150

Smith, L. A., McNab R. J. & Witten, I. H. (1998). Sequence-Based Melodic Comparison: A Dynamic Programming Approach. In: Melodic Similarity – Concepts, Procedures, and Applications (ed. Hewlett & Selfridge-Field), MIT Press

Stecker, G. C. (1996). The role of temporal structure in loudness perception. In: Proceedings of the 18th Annual Stanford-Berkeley Cognitive Psychology Symposium, Stanford, CA

Steinbeck, W. (1982). Struktur und Ähnlichkeit: Methoden automatisierter Melodienanalyse. (Kieler Schriften zur Musikwissenschaft, XXV). Kassel, Basel, London. Bärenreiter

Terhardt, E., Stoll, G. & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. J. Acoust. Soc. Am. 71, 679-688

Toiviainen, P. & Erola, T. (2002). A computational model of melodic similarity based on multiple representations and self-organizing maps. In: Proceedings of ICMPC 7, Sydney

Tversky, A. (1977). Features of Similarity. Psychological Review, vol 84, 327-352

Typke R., Giannopoulos P., Veltkamp R. C., Wiering F., Oostrum van R. (2003). Using Transportation Distances for Measuring Melodic Similarity. In Proceedings of ISMIR 2003, Washington, D.C.