THE PLURALITY OF MELODIC SIMILARITY

Alan Marsden
Lancaster Institute for the Contemporary Arts
Lancaster University, UK
A.Marsden@lancaster.ac.uk

ABSTRACT

Melodic similarity is a much-researched topic. While there are some common paradigms and methods, there is no single emerging model. The different means by which melodic similarity has been studied are briefly surveyed and contrasts drawn between them which lead to important differences in the light of the finding that similarity is dependent on context. Models of melodic similarity based on reduction are given particular scrutiny, and the existence of multiple possible reductions proposed as a natural basis for a lack of triangle inequality. It is finally proposed that, in some situations at least, similarity is deliberately sought by maximising the similarity of interpretations. Thus melodic similarity is found to be plural on two counts (differing contexts and multiple interpretations) and furthermore to be an essentially *creative* concept. There are therefore grounds for turning research on melodic similarity on its head and using the concept as a means for studying reduction and in musical creative contexts.

1. WHAT IS MELODIC SIMILARITY?

A common theme of music-computing research in the last couple of decades has been measurement of melodic similarity. Much of this research has been in the context of query systems, with the aim of finding a way of organising and searching a database of music so as to retrieve melodies similar to a given query. The idea has been used also as a basis for segmentation, for music analysis and for research on music cognition. This growing body of research, however, shows little agreement about what melodic similarity depends on, how to measure it, or even what it really is.

1.1 Seeking a similarity metric

The simple observation that some melodies are similar while others are different, and that the similarity can be closer or more distant, seems to have led many to believe that melodic similarity is a metric space. Formally, a metric is a function from two objects (here melodies) to a quantity (distance, difference) with the following properties [1, p.38]:

(a) non-negativity—the distance between two objects is never less than 0;
(b) self-identity—the distance is 0 if and only if the objects are the same;
(c) symmetry—the distance from *a* to *b* is the same as from *b* to *a*; and
(d) triangle inequality—the distance from *a* to *c* is never greater than the distance from *a* to *b* plus the distance from *b* to *c*.

The first two properties are rarely open to question in the case of differences between melodies, but it is not self-evident that (c) and (d) should be true. Symmetry is most obviously questioned when a short melody is compared to a long one. (A ring tone can be similar to a symphony, at least in the sense that it brings the symphony to mind when we hear it, but the symphony is unlikely to be considered similar to the ring tone.) This situation, however, rarely arises in the contexts considered below. Thus, while symmetry in melodic similarity is in need of thorough investigation, it will be assumed to apply in the remainder of this paper.

The property most commonly questioned is triangle inequality, and the common grounds for this are that melody *a* might be similar to melody *b* by virtue of property or component *x*, while melody *b* might be similar to melody *c* by virtue of a different property or component *y*. In such a situation there is no reason to expect the dissimilarity between *a* and *c* to be limited. Despite such easily imagined counter-examples, those who use systems of measurement with the property of triangle inequality have not reported failure to match human judgements of melodic similarity on the grounds that those judgements do not exhibit triangle inequality. Indeed it is not uncommon to adapt a measure precisely so that it has the property of triangle inequality (for example the development of Proportional Transportation Distance [2] from Earth Mover's Distance) with the objective of facilitating the organisation and searching of a database. (Meanwhile, others have taken the alternative path of investigating means for organising and searching databases without the need of triangle inequality [3].)

1.2 Contrasting empirical bases

Most studies have grounded their work on some kind of empirical basis, some raw 'truth' that certain melodies are similar and others are not. When we look at the detail, however, we find that very different paradigms have been used, firstly in the source of that 'truth' and secondly in the kind of relationship tested between melodies.

Many studies ask experimental subjects, often experts, to judge the similarity between pairs of melodies or extracts of melodies on a rating scale [4–8]. This has the advantage of directly generating measures of difference which will almost certainly have the first three properties of a metric. A rating of 0 or below is not an option; subjects are not asked to compare a melody to itself; and the set-up usually discourages asymmetric judgements. There is no guarantee, however, of triangle inequality. One objection to experimental procedures like this is that they are not realistic: musicians are rarely (if ever) in a situation when they have to match the similarity between melodies to a number. Such direct measurement was avoided in another study which also used expert judgement but subjects were asked to rank a set of melodies by their similarity to a reference melody rather than to simply compare pairs of melodies [9]. A measure of difference can be derived from the relative positions of melodies in the rankings. A potential disadvantage of the method, however, is that experts' judgment of the similarity between a pair of melodies is much more likely to be influenced by the context of the other melodies they are asked to rank simultaneously. This is avoided in approaches where subjects simply compare three melodies (identifying the pair which is most alike and the pair which is least alike) [10, 11]. Indeed, this approach is the one which places the least burden on experimental subjects, and it appears to have been successful for non-expert subjects, unlike the paradigms mentioned above. On the other hand, deriving metric data from these observations requires a method such as multi-dimensional scaling, and a large quantity of observations.

Other studies have avoided direct judgment of similarity, whether by experts or naive listeners. Some have depended on categorisation of melodies either from existing musicological studies [12, 7] or on the basis of geographical origin [13]. In these cases a useful metric cannot be derived from the empirical data, since distances between melodies are all either 0 or 1 according to whether or not the melodies belong to the same category. However, the data can still be used to verify a computational metric on the grounds that the computed distance for melodies within a category should be less than the distance between melodies from different categories.

Yet other studies have attempted to judge similarity on the basis of some real musical activity. Studies aimed at producing metrics for use in query-by-humming systems have been based on asking subjects to sing a known melody [14, 15]. The subjects make mistakes, so the resulting melody is not the same as the original, but it is assumed to be more similar to the original than to other melodies. Subjects can also be asked to deliberately vary a melody [16], and once again the variations are assumed to be more similar to the original than to other melodies. (Others used a related approach of introducing artificial variations into melodies, but this was usually to generate test materials which were then subject to expert judgement of similarity.)

1.3 Similarity and cognition

Do all these paradigms study the same thing? Certainly there are other musical phenomena whose underlying models are robust under different experimental paradigms (models of tonal perception via pitch-frequency profiles are one example), and these suggest stable underlying cognitive functions. The data on melodic similarity has been shown to be relatively consistent from one expert to another and from one occasion to another under the same paradigm, but I am not aware of evidence of consistency between different paradigms. Indeed, there is clear evidence for what one might expect from other aspects of human behaviour: that judgements of melodic similarity are dependent on context. Müllensiefen and Frieler have demonstrated that a different model is required to account for similarity judgements which use the same paradigm but in which the set of melodies to be compared is different [7].

In fact, the contexts in these various experiments have been very different. The nature of melodic materials has varied widely, and crucially the instructions and information given to the subjects have also varied. Sometimes subjects have been given no further instruction than to rate the similarity between two melodies. On other occasions they have been given guidance such as to imagine that the comparison melody is a student's attempt to reproduce a teacher's melody and to think of the similarity rating as a mark [7]. (Note that in this case the similarity judgement can no longer be assumed to be symmetric.) Sometimes subjects' attention has been drawn to particular aspects of the melody, for example by being told in advance that the experiment was concerned with contour [8].

The differences in paradigm also introduce significant issues. If data is derived from real musical behaviours which do not involve explicit similarity judgements, we can only assume that similarity is a governing factor; if data is not derived from real musical behaviours we cannot be certain that it has any real musical relevance. Even in the cases based on explicit expert judgements of similarity, there are important differences. As stated above, we cannot be certain that judgement of melodic similarity has the property of triangle inequality. Even if it does not, subjects can give answers with confidence when asked to rate the similarity between two melodies, or even to judge the most similar and least similar pairs in a triple. However, in a ranking task such as used in [9] the subjects might be in a position of having to balance competing similarity judgements, depending on how they interpret the instructions. If they consider their task to be simply to ensure that the melody ranked x is no less similar to the reference than the melody ranked $x + 1$, no competing rankings can arise. If, however, they also believe that a ranking implies that the melody ranked $x + 2$ is less similar to the one ranked x than the one ranked $x + 1$, then in the absence of triangle inequality, a subject might find it impossible to find a ranking which meets both criteria: melodies *a*, *b* and *c* might have decreasing similarity to the reference, and so be ranked x, $x + 1$ and $x + 2$, but *c* might be more similar to *a* than *b*, implying instead the ranking x, $x + 2$ and $x + 1$.

It is not safe, therefore, to assume that these studies investigate the same phenomenon of melodic similarity. Until there is evidence that data produced under these various paradigms is compatible, and in particular evidence that melodic similarity does exhibit triangle inequality, it is probably better to consider melodic similarity to be a family of possibly related phenomena.

2. MEASURING SIMILARITY

As mentioned above, different approaches to measuring melodic similarity have arisen from different objectives. A common one has been the retrieval of melodies from a database, but there are others also. Some seek to use measures of similarity as an aid in ethnomusicological studies, for example to find variants of a folk song, or to trace the provenance of a song. Others aim to use it as a tool in music analysis. In each case, the kinds of differences one is likely to find in melodies are likely to vary, and an approach founded on behaviours should take these into account. For example, in a query-by-humming system, a similarity metric should ideally be based on the kinds of errors which singers make when trying to recall and reproduce a melody. Similarly, similarity in folk songs should take into account the kinds of changes commonly introduced in oral traditions (either accidentally or deliberately), which might vary from one culture to another. In music analysis, one is generally concerned not with mistakes or accidental changes, but with deliberate and crafted variations of musical materials. In the remainder of this paper, I will concentrate on similarity in this context.

2.1 Similarity based on reduction

It is common to regard melodies as having an underlying structure, and to consider melodies sharing the same structure to be similar (at least in one sense) even if their surface details are quite different. To account for this kind of similarity, studies have been based on comparing melodies not note-by-note, but on the basis of a reduction of the melodies (generally in a tree structure) which progressively removes decorative notes until only the main outline of the melody is left [16–19].

Rizo and colleagues [16, 17] derive the reduction of a melody by selecting one of the notes occurring in each span based on a small number of rules. The spans are determined by the metre, so that, in 4/4 for example, there is a span for each bar, at the next level down two spans for the minims (half notes), then four spans for the crotchets (quarter notes), etc., halving each span at the level above. There are also higher-level spans which group bars into pairs, etc. The result is a tree structure in which each node corresponds to a specific time span, and the rhythm of the melody is completely defined by the tree structure. The reduction is built bottom-up by

(a) always selecting a note in preference to a rest,
(b) selecting a harmonic note in preference to a non-harmonic one, and
(c) selecting the note at the head of the span if both are harmonic.

A harmonic analysis of the melody must be generated before reduction, and this is currently done by hand. A measure of similarity based on the tree edit distance between the reductions of melodies was compared with edit distance on the melodic surfaces alone. The reduction-based similarity measure proved to perform better at distinguishing variations of a melody from unrelated melodies [16].

The approach of Orio & Rodà [18] is similar, in that it generates a tree based on the metrical structure, and notes are selected within each span partly on the basis of a harmonic analysis. The selection, however, is based on a more complex set of weights using the relation of the note to the underlying harmony (fifth, third or root), the function of that harmony, and the position in the metre. Furthermore, similarity between melodies is not based on the edit distance between trees. Melodies are segmented (using pre-existing segmentation schemes) and the segmentation propagated to higher levels of the tree. The resulting melodic segments, expressed as interval patterns, are placed in a directed acyclic graph (DAG) in which parent-child relations between segments copy those relations in the reductions. The difference between two segments is then measured by the minimum path length between the segments in the DAG, and the difference between two melodies is the average difference between their component segments. This method was not tested against other measures of melodic similarity.

The reductions produced by my own system [19, 20] are intended to more closely mimic the reductions of Schenkerian analysis. Furthermore, they are based not just on melodies but on a full musical texture (generally extracts from piano pieces). The reduction process is therefore considerably more complex than those outlined above. In particular, the reduction tree does not necessarily follow the metrical structure (as indeed it does not in many Schenkerian analyses), and no prior harmonic analysis is necessary (though specification of the key and metre is). While early results matched actual analyses to a promising degree [20], an attempt to use the same system of reduction for demonstrating the similarity underlying themes and variations produced less promising results [19]. Matching themes and variations via reductions proved no better than matching on the basis of the surfaces alone.

2.2 Multiple reductions

One possible reason for the disappointing results in [19] might have been poor reductions. I did not check each reduction for accuracy (after all, in the absence of prior

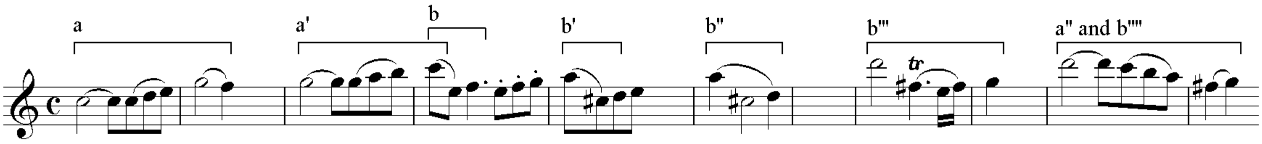

Figure 1. Extracts from Mozart's string quartet in C major, K. 465, first movement.

analyses, there is no test of accuracy other than expert analytical judgement), but I did note a number of cases where the reduction of a theme seemed incorrect. One important finding from the research on computational Schenkerian analysis is that a very large number of reductions is possible on the basis of the 'rules' inferred from writings on Schenkerian analysis alone [20]. Indeed, music analysts commonly recognise that alternative analyses of the same piece of music are possible and valid. If multiple reductions are possible, how should a similarity-measurement procedure based on reduction select which reduction to use?

It is instructive in this context to compare reduction-based similarity with edit distance, or more specifically Levenshtein distance. This measures the difference between two sequences in terms of the number of deletions, insertions and substitutions required to transform one sequence into another. Since reduction depends on selecting one of the notes of a pair (in most cases; Rizo et al. allow selection from a triple if warranted by the metre) each reduction step can be considered as equivalent to a deletion. Note that an insertion in one sequence is equivalent to a deletion in the other, and a substitution is equivalent to a deletion in both sequences. Thus if the difference between two melodies is measured as the minimum of the sum of reduction steps necessary to arrive at the same reduction for both melodies minus the number of reductions which take place in equivalent places (to account for substitutions), the difference is equivalent to the Levenshtein distance between the two melodies. (There is thus a strong correspondence between Levenshtein distance and the metric used by Orio & Rodà.) However, this assumes that the two melodies can be freely aligned in the way which allows for the minimum number of deletions, insertions and substitutions. Reduction, on the other hand, in all of the cases examined, is constrained by the rhythm, metre and other characteristics of the melody. Thus computing difference by reduction can be seen as similar to computing the Levenshtein distance with constraints on how the melodies may be aligned. I say 'similar' because it is difficult to see how the constraints of harmony and melody could be applied without actually performing the reduction. However, it might be a useful approach (especially in the light of the considerable computational complexity of reduction as performed in [20]) to use alignment constrained by metre as a means of guiding reduction.

3. SIMILARITY AND CREATIVITY

3.1 Finding similarity

Reduction is not the only approach to similarity which depends on a step which is potentially subject to multiple interpretations. A number of similarity-measurement systems depend on segmentation, which also is not an unequivocally definite process. If this is the case, we should expect similarity judgements to vary according to the degree of freedom (or inclination) that subjects have to interpret the melodies in multiple ways.

At one extreme are probably the situations when someone compares themes and variations or when a teacher assesses a student's performance. In both cases, there is a presumption that the melodies should be similar, and so listeners are likely to *seek* the interpretations which allow maximum similarity. At the other extreme are situations when listeners have to make snap judgements or when they are asked to rank melodies for similarity. In the first case there will not be time for multiple interpretations; in the second there is an inclination to find difference as much as to find similarity.

If, in some situations at least, similarity is judged on the basis of maximising the similarity between interpretations of two melodies, we should *expect* triangle inequality to be violated: that melody *b* can be interpreted in different ways to be similar to both *a* and *c* does not imply that there is any way to interpret *a* to be similar to *c* (at least not in general; this conjecture would have to be tested with respect to specific methods of interpretation, such as reduction methods).

3.2 An example

There is no direct evidence for such multiple interpretation in similarity studies I know of, but I can retrieve a candidate case from a music analysis I made some years ago [21, 22]. Figure 1 shows extracts from the first violin part of Mozart's string quartet in C major, K. 465 ("Dissonance"). The allegro begins with the theme shown as **a**. This is immediately repeated a tone higher (not shown) and then, with a slight modification, as **a'**. The last note of **a'** begins a new motive **b** which appears to contrast with **a** (descending instead of rising; made up largely of shorter notes; containing a large leap instead of mostly steps). This is repeated at **b'** (reinforcing the identity of the motive) and then in rhythmic transformation some bars later at **b''** (where the recognition of similarity is aided by using exactly the same pitches). Several bars later the figure identified as **b'''** is heard, whose similarity to **b''** is aided by the equivalent durations of the second note (though in the case of **b'''** it is decorated with a trill). Finally, beginning on the same pitch as **b'''** and ending with the same pair of pitches, a figure is heard which is also clearly similar to **a** by inversion. (Indeed, to help make this clear, the intervening music has presented several other versions of **a** without inversion.) This figure is easily recognised as similar to *both* **a** and (with the aid of the intermediate transformations) **b**.

Is it true, then, that **a** is similar to **b**, despite the fact that at first the motives seemed to be contrasted? If it is, then we must reduce **a** in *different* ways to find maximum similarity in each case. To find maximum similarity between **a** and **a'**, we must reduce **a** by removing the appoggiatura on the last note, which implies that the remaining notes are passing notes from C to F. To find maximum similarity between **a** and **b**, on the other hand, the first step must be to reduce out the quavers in **a** and regard the appoggiatura (neighbour note) as prior. It was my contention in the original analysis [22] that Mozart intended this play with our sense of the difference and similarity between these motives as a way of capturing the listener's interest.

3.3 Exploring similarity through creativity

Listening to music, or indeed any human process with music, involves interpretation, and interpretation is always a *creative* act. When musicians say two melodies are similar, the arguments above suggest that the musicians have *created* that similarity as much as recognising it. While it is now not uncommon for researchers to claim that a single measure of melodic similarity for all situations is an impossibility (e.g., [7]), this argument suggests that it is an impossibility in *any* situation. The best one can hope for is a measure which will usefully approximate human judgements of similarity in such situations.

The distinction is perhaps technical, since no researchers have claimed to derive a perfect measure of melodic similarity, but it does imply a radically different research perspective. In particular, it suggests that melodic similarity might profitably be explored in explicitly creative situations. For example, a system which aimed to allow users to compose music on the basis of arranging similar and contrasting melodic fragments might be based on competing models of similarity. Then by observing users' interaction with the software (probably silently through background monitoring), data could be gathered about which model was most useful for achieving the users' artistic goals.

3.4 Using similarity to explore reduction

Another possible research direction which turns previous research on its head is to use similarity as a means for investigating reduction rather than the other way around. As mentioned above, the bases for making Schenkerian reductions are not well understood, and there are not pre-existing paradigms for their discovery. If my hypothesis that similarity, at least in some situations, is based on finding the maximally similar reductions of two melodies, then melodies which are known to be similar could be used as ground truths for guiding reduction. This has the advantage over the approach taken in [20] that, instead of being based on the activities of experts directed towards either pedagogy or analytical debate, it is based on the practice of real composers and listeners. Sets of variations, in particular, provide a promising ground for such investigations.

3.5 Creativity of music information retrieval

Researchers who develop systems for measurement of musical similarity generally take a scientific approach, judging their success or failure by the degree to which results match observations. Yet they too are creative, or at least have a creative influence, not only in the general sense of making something new, but also in a musical sense. They might not make new pieces of music, but their work will certainly lead to new kinds of musical experience.

The recent past provides numerous examples of similar creative impact of scientific advances. The invention of MP3 encoding, for example, in conjunction with the internet, has created an entirely new environment in which to discover, obtain, experience and even create music, crucially creating new kinds of musical community [23]. The iPod and similar personal music devices (also dependent on the technology of MP3 and related encodings) has also radically affected common experiences of music. In contrast to previous centuries when the only way to experience music on one's own was to play it oneself, listening to music has become commonly an isolated and personal experience. Indeed, listeners commonly report using a mobile music device in order to create a 'personal space' [24], quite the opposite of the traditional necessary association of music with a social or communal space.

If the work of those who research melodic similarity leads to ubiquitous software which allows music to be rapidly retrieved on the basis of its similarity to a given model, what will be the impact on our musical culture, and on the nature of music which is created? And, since judgements of similarity are context-dependent (as discussed above), what will be the consequent effect on people's concepts of melodic similarity? Musical scientists too do not escape the uncertainty principle: in investigating melodic similarity they affect the very culture which generates the concept of melodic similarity itself.

4. CONCLUSION

Melodic similarity seems not to be a single relationship, but to be plural on at least two counts. Firstly, it differs from one context to another. Secondly, it depends of differing interpretations. The second of these is undoubtedly a creative act (though listeners do not generally regard themselves as creative). In enabling new ways of experiencing and encountering music, researchers of melodic similarity also have a creative impact on musical culture.

5. REFERENCES

[1] M.A. Armstrong, Basic Topology, Springer-Verlag, 1983.

[2] P. Giannopoulos & R.C. Veltkamp, "A pseudo-metric for weighted point sets," in Proceedings European Conference on Computer Vision (ECCV 2002), LNCS 2352, Springer, 2002, pp. 715–730.

[3] R. Typke & A.C. Walczak-Typke, "A tunnelling-vantage indexing method for non-metrics," in Proceedings International Conference on Music Information Retrieval (ISMIR), Philadephia, 2008, pp. 683–688.

[4] T. Eeerola, T. Järvinen, J. Louhivuori & P. Toiviainen, "Statistical features and perceived similarity of folk melodies," in Music Perception, 2001, pp. 275–296.

[5] T. Eerola & M. Bregman, "Melodic and contextual similarity of folk song phrases," in Musicae Scientiae, 2007, pp. 211–233.

[6] D. Müllensiefen & K. Frieler, "Cognitive adequacy in the measurement of melodic similarity: algorithmic vs. human judgments," in Computing in Musicology, 2004, 147–176.

[7] D. Müllensiefen & K. Frieler, "Modelling experts' notions of melodic similarity," in Musicae Scientiae, 2007, pp. 183–210.

[8] M.A. Schmuckler, "Melodic contour similarity using folk melodies," in Music Perception, 2010, pp. 169–193.

[9] R. Typke, R. Wiering & R.C. Veltkamp, "Transportation distances and human perception of melodic similarity," in Musicae Scientiae, 2007, pp. 153–181.

[10] H. Allan, D. Müllensiefen & G. Wiggins, "Methodological considerations in studies of musical similarity," in Proceedings International Conference on Music Information Retrieval (ISMIR), Vienna, 2007, pp. 473–478.

[11] A. Novello, M.M.F. McKinney & A. Kohlrausch, "Perceptual evaluation of inter-song similarity in western popular music," in Journal of New Music Research, 2011, pp. 1–26.

[12] A. Volk, P. van Kranenburg, J. Garbers, F. Wiering, R.C. Veltkamp & L.P. Grijp, "A manual annotation method for melodic similarity and the study of melody feature sets," in Proceedings International Conference on Music Information Retrieval (ISMIR), Philadephia, 2008, pp. 101–106.

[13] Z. Juhasz, "A systematic comparison of different European folk music traditions using self-organising maps," in Journal of New Music Research, 2006, pp. 95–112.

[14] N. Hu, R. Dannenberg, A.L. Lewis, "A probabilistic model of melodic similarity," in Proceedings of the International Computer Music Conference (ICMC), San Francisco, 2002, pp. 509–515.

[15] B. Pardo & J. Shifrin, "Name that tune: a pilot study in finding a melody from a sung query," in Journal of the American Society for Information Science and Technology, 2004, pp. 283–300.

[16] J.F. Bernabeu, J. Calera-Rubio, J.M. Iñesta & D. Rizo, "A probabilistic approach to melodic similarity," in Proceedings of Music and Machine Learning, 2009, pp. 48–53.

[17] D. Rizo, Symbolic music comparison with tree data structures, PhD thesis, University of Alicante, 2010.

[18] N. Orio & A. Rodà, "A measure of melodic similarity based on a graph representation of the music structure", in Proceedings International Conference on Music Information Retrieval (ISMIR), Kobe, 2009, pp. 543–548.

[19] A. Marsden, "Recognition of variations using automatic Schenkerian reduction," in Proceedings International Conference on Music Information Retrieval (ISMIR), Utrecht, 2010, pp. 501–506.

[20] A. Marsden, "Schenkerian analysis by computer: a proof of concept," in Journal of New Music Research, 2010, 269–289.

[21] A. Marsden, "Listening as discovery learning," in Contemporary Music Review, 1989, pp. 327–340.

[22] A. Marsden, Analysing music as listeners' cognitive activity, a study with reference to Mozart, PhD thesis, Cambridge University, 1987.

[23] A. Tanaka, "Interation, experience and the future of music", in K. O'Hara & B. Brown (eds.), Consuming Music Together: Social and Collaborative Aspects of Music Consumption Technologies, Springer, 2006, 271–292.

[24] E. Nettamo, M. Nirhamo & J. Häkkilä, "A cross-cultural study of mobile music – retrieval, management and consumption", in Proceedings 18th Australia conference on Computer-Human Interaction (OZCHI '06), 2006, 87–94.