Hooked on Music Information Retrieval

W. BAS DE HAAS [1]
Utrecht University

FRANS WIERING
Utrecht University

ABSTRACT: This article provides a reply to 'Lure(d) into listening: The potential of cognition-based music information retrieval,' in which Henkjan Honing discusses the potential impact of his proposed *Listen, Lure & Locate* project on Music Information Retrieval (MIR). Honing presents some critical remarks on data-oriented approaches in MIR, which we endorse. To place these remarks in context, we first give a brief overview of the state of the art of MIR research. Then we present a series of arguments that show why purely data-oriented approaches are unlikely to take MIR research and applications to a more advanced level. Next, we propose our view on MIR research, in which the modelling of musical knowledge has a central role. Finally, we elaborate on the ideas in Honing's paper from a MIR perspective in this paper and propose some additions to the *Listen, Lure & Locate* project.

INTRODUCTION

FROM its shady beginnings in the 1960s, Music Information Retrieval (MIR) has evolved into a flourishing "multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all" (Downie, 2004, p. 12). This store is indeed vast, especially in the digital domain. Reportedly, iTunes has over 13,000,000 songs for sale [2]. Apple claims that the iPod Classic may contain 40,000 music tracks. Based on Apple's figures [3], it would take around 110 days of continuous listening to play through all these, and nearly 100 years for all iTunes. All of this music is accessible in the sense that it can be reached within a few mouse clicks, if you know where to click. This is where the trouble begins, for under many circumstances listeners do not have the necessary information to be able to do so. Consider the following situation, provided by the first author.

> As a guitar player, I am quite fond of Robben Ford, a guitar player who, in my view, takes blues improvisation to the next level by combining it with various elements of jazz. I find his albums from the late eighties and early nineties especially delectable. Even though some pieces remain appealing even after hundreds of listening experiences, I cannot suppress the desire for something new yet similar in certain aspects, e.g. instrument, groove, ensemble, emotional intensity, etc. Unfortunately, Robben Ford will not be able to fulfil this need at short notice: it is unclear when he will release new material. Also, in my view, his recent work does not exhibit the same vitality that his older work does. However, since the world provides for numerous excellent, well-schooled, creative guitar players, other artists might be capable of satisfying my aching desire. Nowadays, their music will very probably be readily available online, e.g. via iTunes, mySpace or the like. The only problem is actually finding it, especially when the artists are not among the ones best known to the general public. In other words, I do not know the name of the artist I am looking for, I do not know where (s)he comes from, let alone the title of a piece. Hence, submitting a textual query to a search engine like Google

approximates random search. Clearly, search methods are needed that are specifically designed for musical content.

This is where MIR aims to provide a helping hand. However, delivering the right music–or, in the terminology of Information Retrieval, *relevant* music–by means of computational methods is not a solved problem at all. What music is relevant to such common but complex user needs as the one described above, depends on factors such as the user's taste, expertise, emotional state, activity, and cultural, social, physical and musical environment. Generally, MIR research abstracts from these problems by assuming that unknown relevant music is in some way similar to music known to be relevant for a given user or user group. As a consequence, the notion of music similarity has become a central concept in MIR research. It is not an unproblematic concept, though.

Similarity has generally been interpreted in MIR to be a numerical value for the resemblance between two musical items according to one or more observable features. For audio items, such features could be tempo or various properties of the frequency spectrum; for encoded notation, pitch, duration, key and metre are example features. The first problem to solve is to extract such features from the musical data using appropriate computational methods. The next problem is to create effective methods for calculating similarity values between items. Based on these values, items can then be ranked or clustered, for which again various methods can be devised. Finally, these methods need to be evaluated individually and in combination. This is usually done by comparing their result to the ideal solution, the *ground-truth* produced by humans manually (or aurally) performing the same task. Usually, the computational result is far from perfect. Indeed, there seems to be a *glass ceiling* for many MIR tasks that lies (for audio tasks) at around 70% accuracy (Aucouturier & Pachet, 2004; Wiggins, Mullensiefen, & Pearce, 2010, p. 234). Many MIR researchers have concluded from this that not all information is in the data, and that *domain knowledge* about how humans process music needs to be taken into account.

In the target article (Honing, 2010), Henkjan Honing takes a critical stance towards data-oriented approaches to MIR and argues for a cognition-based approach to MIR. Specifically, he proposes to focus on the notion of the 'hook' as a key feature in the human memorization, recall and appreciation of music (Burns, 1987). We share his critical stance to a large extent and welcome his suggestion for researching the hook (about which more later), but first we present a brief outline of MIR and elaborate some of the issues in data-oriented MIR in more depth. Even though the primary goal of MIR is not to develop theories and models of music that contribute to a better understanding of music, we will argue that it is not possible to develop effective MIR systems without importing knowledge from music theory and especially music cognition into these systems. Moreover, we will discuss how quantitatively testing knowledge-based MIR systems can provide evidence for the validity of the musical models used in the system and thereby contribute to the understanding of music.

A BRIEF OUTLINE OF MUSIC INFORMATION RETRIEVAL

The MIR research community is shaped to a large extent by the International Society for Music Information Retrieval (ISMIR [4]; Downie, Byrd, & Crawford, 2009) and its yearly conference. Even a quick glance through the open access ISMIR proceedings [5] shows that MIR is a very diverse research area. Here we can only present a condensed overview: for more comprehensive overviews we refer to Orio (2006) and Casey et al. (2008).

Within the community at least three views on musical data coexist; that is, music as represented by metadata (originating from the library science subcommunity), by encoded notation (from musicology) and by digital audio (from digital signal processing). Computational search and classification methods have been designed independently from each of these viewpoints, although occasionally viewpoints have been combined. In addition, much research goes into providing infrastructural services for MIR methods and systems, for example research in feature extraction, automatic transcription and optical music recognition. Automatic analysis of music tends to be subsumed under MIR as well, for example the creation of analytical tools, performance analysis and quantitative stylistics. Much research is directed towards visualisation, modelling mood and emotion, interfaces for engaging with music (e.g. playing along, Karaoke), playlist generation, collaborative tagging and industrial opportunities. Despite this diversity, there seems to be a strong awareness of coherence within the MIR community, which can largely be explained by a shared ideal, very similar to Downie's quoted above, of universal accessibility of music and by a common commitment to empirical observation, modelling and quantitative evaluation.

Today's achievements and advances in MIR are probably best illustrated by the Music Information Retrieval EXchange (MIREX [6]; Downie, 2008). MIREX is a community-based initiative that provides a framework for objectively evaluating MIR related tasks. Each year during the ISMIR conference, the results of the evaluation of around 15-20 different tasks are presented, each with on average 8-9 submissions. Example tasks include Audio Beat Tracking, Audio Key Detection, Structural Segmentation, Query by Singing/Humming and Symbolic Melodic Similarity. Submissions that have been evaluated for the last task over the years include geometric, sequence alignment, graph-based, n-gram and implication-realisation based approaches. Considerable progress has been made in most tasks since 2005. Especially the transcription of low-level audio features into higher-level musical symbols such as beats, notes, chords and song structure has seen substantial improvement.

Nevertheless, moving from research prototypes to industry-strength applications is difficult and the number of functioning systems that are actually capable of solving issues like the one raised in the previous section is very limited. To get a somewhat sensible answer one's best bet is still to use services such as Last.fm [7] and Pandora [8], which are based on very rich metadata (including social tagging). Although the retrieval performance of these services is among the best currently available on the web, we are convinced that integrating content-based search methods into these will improve their performance. But, as Honing has observed, to realise this potential, a major step beyond the dominant data-oriented approach needs to be taken.

THE LIMITATIONS OF DATA-ORIENTED MIR

Before we present our critical notes on the machine learning paradigm, we would like to stress that machine learning is an important area of research with many useful, practical and theoretical applications in MIR. However, we will argue here that a purely data-oriented approach is not sufficient for meaningful music retrieval, i.e. music retrieval that steps beyond mere similarity of content features in order to deliver music that actually makes sense to the user 'here and now,' and contributes to his/her experience, enjoyment and/or understanding of music (Wiering, 2009). We distinguish the following six limitations to data-oriented MIR.

Ground-Truth and Data. Probably the greatest weakness of data-oriented approaches is the data itself, or more precisely, the lack of it. To be able to train a supervised learning algorithm a very substantial amount of data has to be annotated–in the case of MIR often by musical experts–and it is known that the larger the amount of data is, the better the algorithm performs (Harris-Jones & Haines, 1998). Obtaining such ground-truth data is a costly and time-consuming enterprise, which moreover is often frustrated by copyright issues. In practice, therefore, such sets tend to be small. In addition, it is hard to generalise annotations and similarity judgments of a small number of experts collected in an experimental setting to the much larger population of possible end-users studying or enjoying music in 'ecological' circumstances.

Danger of Overfitting. Supervised learning algorithms are all based on optimizing the parameters of a model in such way that the difference (error) between the predictions of this model and the expert annotations is minimised. Obviously, the more flexible a model is the better it can fit the data. As a consequence, a flexible model will often have a larger prediction error on other data sets than a less flexible model because it was trained to explain the noise in the training set as well. This process is often referred to as *overfitting* (e.g. Bishop, 1995, p. 332; Pitt & Myung, 2002). In MIR, the specific issue is that there are only a few annotated data sets that can be used for testing trained systems. It is therefore hard to assess the claimed retrieval results: it is often unclear if these systems present an improvement that can be generalised to other data sets, or if they are merely overfitting the currently available data sets.

Curse of Dimensionality. Music is a complex phenomenon; therefore a considerable number of musical features need to be taken into account at any given point in time. For example, a MIR system may need information about simultaneously sounding notes, their timbre, intonation, intensity, harmonic function, and so on. As a result the input vector, i.e. the list of numerical values representing these features, is often high-dimensional. This introduces the so-called curse of dimensionality (Bishop, 1995, p. 7): the problem of the exponential increase in volume of the search space caused by adding extra dimensions to the input data, whereas the data itself becomes very sparse in this space. The amount of training data also needs to increase exponentially in order to attain a model with the same precision as a corresponding low-dimensional one.

Neglecting Time. One of the most distinctive features of music is that it evolves over time: there is no such thing as timeless music. The fundamental role of time is illustrated by the fact that the perception

of a musical event is largely determined by the musical context in which it occurs, i.e. what the listener has heard before (e.g. Krumhansl, 2001; Schellenberg, 1996). A significant number of data-oriented approaches completely disregard this fact. For example, when dealing with audio data, a common paradigm is to split an audio file up into small (overlapping) windows. Subsequently, a feature vector is created for each window, which contains characteristics of the signal (for example chroma features or Mel Frequency Cepstral Coefficients; Logan, 2000). These feature vectors are inputted into a classifier for training and the temporal order of the feature vectors, and thus the notion of musical time, is lost in the process. In a sense this resembles analyzing the story in a movie while randomly mixing all the individual frames.

Nothing to Learn? Another drawback of most data-oriented approaches is that it is hard to grasp what a system has actually learned. For instance, it is quite hard to interpret the parameters of a neural network or a hidden Markov model. This makes it difficult to infer how a system will respond to new unseen data. After all, how would one know whether the machine learning process did not overfit the system to the data? Moreover, for music researchers it is impossible to learn anything from the predictions of the model, because the model itself is difficult to interpret in humanly understandable, let alone musical terms.

Not All Information Is in the Data. Last, but not least, music only becomes music in the mind of the listener. Hence, only part of the information needed for sound judgment about music can be found in the musical data. An important piece of information that is lacking in the data is the information about which bits of data are relevant to the musical (search) question and which bits are not, because this is often not clear from the statistical distributions in the data. For instance, in a chord sequence not every chord is equally important (for example passing chords or secondary dominants) and a harmonic analysis of the piece is needed to identify the important chords. Similarly, most musically salient events occur at strong metrical positions, and a model is needed to determine where these positions are.

Furthermore, the perception of musical events strongly depends on the context in which the event occurs. This context depends on cultural, geographical and social factors, and specific user taste. One needs only to imagine playing a piece that is appropriate in a church at a dance party (or vice versa) to realise this. It is known that musically schooled as well as unschooled listeners have extensive knowledge about music (Bigand & Poulin-Charronnat, 2006; Honing & Ladinig, 2009). Herein, exposure to music plays a fundamental role. In other words, humans acquire a significant amount of musical knowledge by listening to music, and as for each human this exposure has been different, the outcome is bound to be different as well.

An often-heard argument in favour of machine learning is that, if humans can acquire this knowledge, machines must be capable of doing it as well. However, we argue that, even in theory and under perfect circumstances, a data-oriented approach to MIR is not sufficient. Given a very complex model with enough degrees of freedom, similar to the human brain; thousands of hours of music; and, most importantly, the required relevance feedback, one still needs a model that captures the (music) cognitive abilities similar to those of a newborn, which enables the acquisition of musical skills. The reality is that models of such complexity do not exist, nor is there any certainty that they will come into existence soon, let alone that they can be learned from a few songs. Therefore, in practice purely data-oriented approaches have considerable limitations when dealing with musical information.

A MODEL/KNOWLEDGE-BASED ALTERNATIVE

Music is sound, but sound is most certainly not necessarily all there is about music, since music can also be said to exist with no sound present–the phenomenon of the 'earworm,' i.e. a melody that spontaneously appears and consequently sticks in one's mind (Honing, 2010), is sufficient proof of this. This then raises the question what the role of sound is in music. Herein, we adopt the view of Wiggins (2009, 2010), which is in turn based on Milton Babbitt's work (1965). According to Wiggins, music can reside in three different domains (see Figure 1): the acoustic (or physical), the auditory (or perceived/internalised) or the graphemic (or stored/notated). Music as sound belongs to the acoustic domain, which embraces all the physical properties of sound. The graphemic domain can be viewed as an unlimited collective memory that is able to store musical information: this can be a musical score, but also a digital representation such as a CD or MP3 file. The auditory domain accounts for all musical phenomena that occur inside the human mind. Each of these domains captures a different *aspect* of music. All together these domains describe everything that we consider music, while music itself is something abstract and intangible that is constantly redefined by composers, musicians and listeners.
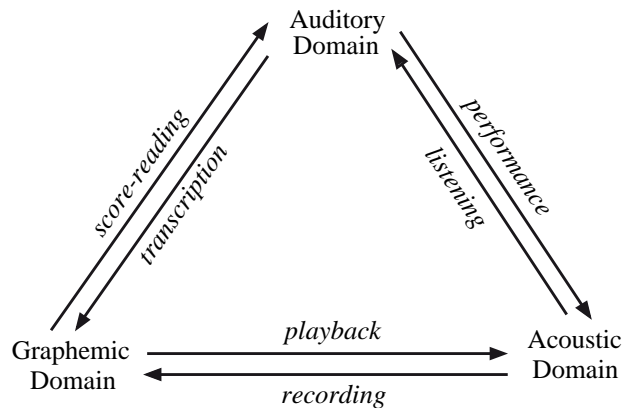
Fig. 1. Babbitt's trinity of domains, with Wiggins' addition of transformations between them, quoted from Wiggins (2009).

Categorizing musical phenomena in three different domains does not imply that all three domains are equally important. Music can exist without sound being present, since people can imagine music without listening to it or even create it, like Beethoven when he was deaf. On the other hand, improvised music is often performed with little or no graphical information, and should ideally be experienced in a live setting rather than a recording. However, without human intervention there is no music. The fundamental source, but also the purpose of music can only be found in the human mind, without which music cannot exist. Therefore, a deeper understanding of what music is can only be gained by investigating the human mind.

From the point of view of a MIR researcher, the graphemic domain is particularly interesting. Analogous to written language, the printing press, or photography, the graphemic domain emerged to compensate for one of our brains' major deficits: its lack of ability to precisely recall and/or reproduce something that happened in the past. This brings us back to the problem sketched at the beginning of this paper of the immense amount of valuable, but often unorganised musical material that we want to unlock for a general audience. We believe that this can only be done in an automated fashion if the machine has a reasonable understanding of how this data is processed by the human mind. In turn, scientifically studying and formalizing the cognition of music can only achieve this.

We are certainly not the first to call for a more music-cognitive inspired approach to MIR. Already at the first ISMIR in 2000, David Huron presented a paper entitled 'Perceptual and Cognitive Applications in Music Information Retrieval' (Huron, 2000). Other MIR researches, like Aucouturier and Pachet (2004) and Celma (2006), recognised that data-only MIR methods suffered from a *glass ceiling* or a *semantic gap*, whereupon scholars like Honing (2010) and Wiggins (2009, 2010) once more stressed the importance of music-cognitively oriented alternatives. We share this stance and believe that complementing low-level bottom-up with top-down approaches that start from what knowledge we already have about music, can have a positive effect on retrieval performance since they sidestep the issue of automatically assembling that knowledge in the first place.

Nonetheless, it is certainly not all doom and gloom in the field of MIR. There exist some successful MIR approaches that are based on musical knowledge. Some examples include: using perceptual models (Krumhansl, 2001) to search for the musical key (Temperley, 2001, Ch. 7); improving harmonic similarity by performing automatic harmonic analyses (de Haas, Rohrmeier, Veltkamp, & Wiering, 2009) or by consulting Lerdahl's (2001) Tonal Pitch Space (de Haas, Veltkamp, & Wiering, 2008); making F0 estimation easier with a filter model based on human pitch perception (Klapuri, 2005); using Gestalt principles for grouping musical units (Wiering, de Nooijer, Volk, & Tabachneck-Schijf, 2009); or retrieving melodies on the basis of the Implication/Realization model (Grachten, Arcos, & Lopez de Mantaras, 2004; Narmour, 1990), to name a few.

Although we are convinced that MIR approaches grounded in musical knowledge have an advantage over mere data-oriented approaches, still some of the arguments posed in the previous section

also hold for non-data-oriented approaches. However, we will argue below that these arguments do not have such severe consequences when MIR research does not rely on the musical data alone.

Ground-Truth and Data. Ground-truth data is essential to the *evaluation* of advances in MIR and the lack of it will restrict MIR in progressing further. However, knowledge- or model-based approaches to MIR do not require ground-truth data for *training*, thus reducing the overall data needed .

Danger of overfitting. Knowledge-based approaches are vulnerable to overfitting as well. After all, a specific parameter setting of a musical model that is optimal for a certain data set does not necessarily have to be optimal for other data sets. However, overfitting is interpretable and easier to control, because the parameters have a (musical) meaning and the MIR researcher can predict how the model will respond to new data.

Curse of Dimensionality. Prior knowledge about music can help a MIR researcher to reduce the dimensionality of the input space by transforming it into a lower dimensional feature set with more expressive power. To give a simple example, instead of representing a chord with notes in eight octaves a MIR researcher could assuming octave equivalence and choose to represent it with only twelve pitch classes. This reduces the dimensionality of the input vector and reduces the space of possible chords considerably.

Neglecting Time. Music only exists in time. If a musical model disregards the temporal and sequential aspect of music, it fails to capture an essential part of the musical structure. Hence, it might be wise to reconsider the choice of model. Besides, there are plenty of musical models that do incorporate the notion of musical time, e.g. models for segmenting music (Wiering et al., 2009), musical expectation (Pearce & Wiggins, 2006), or (melodic) sequence alignment (van Kranenburg, Volk, Wiering, & Veltkamp, 2009), to name a few.

Nothing to Learn? Using a MIR approach based on cognitive models might not only be beneficial for retrieval performance; it can also be used to evaluate the model at hand. When such a MIR system is empirically evaluated, the model is also evaluated by the experiment–albeit implicitly. The evaluation can provide new insights into the musical domain used and thereby contribute to the improvement of the musical model.

Not All Information Is in the Data. This point has been extensively made above and needs no further explanation.

TAGGING IN MIR

Tagging, the creation of (generally short) textual annotations to web resources by end users, has recently become a popular topic in MIR research. Nowadays, Last.fm [7] tags are widely used: they are studied in 24 out of 110 papers in the ISMIR 2010 proceedings, mostly in genre and emotion classification. Nine papers from the same collection mention online tagging 'games with a purpose,' such as TagATune and MoodSwings. Tagging and games are often seen as an answer to the problem of data scarcity. By mobilizing the collective effort of countless web users, annotations can be created in numbers that cannot possibly reached by means of local experiments with domain experts. Also, annotations often convey the listeners' emotional response to music. Therefore, annotations may help to solve in the problems inherent in data-oriented MIR. Unfortunately social music sites only allow tagging at the song level. Online games usually work with short clips (10-30 seconds). In MoodSwings, five clips are used for each song, allowing some study of 'mood distribution' over time (Schmidt & Kim, 2010). Mandel, Eck, and Bengio (2010) used Mechanical Turk [9] to create annotations for multiple clips from one song to study if better overall song features could be derived from these.

Honing's (2010) proposal shares several features with the work just discussed in that it aims to mobilise the work power of end users within a Web 2.0 community sharing their annotations. It goes beyond these approaches in several respects. First, listeners will be able to mark an arbitrary segment from the entire track for annotation, rather than a given track or clip. Second, the object of the annotation is locating the 'hook,' the special, captivating, sticky passage that characterises the song. Finally, it aims beyond ground-truth creation at understanding the important musical phenomenon of the hook, about which, as Honing observes, very little is known.

From the perspective of MIR, we are convinced that a large annotated corpus of hook locations would be a very powerful new resource. The corpus could itself be much more efficiently searched than a collection of complete pieces: there is less data to search, and we know that the available data is highly distinctive. Also, the hooks and annotations can be used as a ground-truth for designing computational

models for musical hooks, or more generally musical salience, and on top of these models, similarity measures could be designed that further abstract from the low-level features used by most of today's similarity measures. Ultimately, this might lead to effective, even commercially viable, methods for automatically creating and processing thumbnails, that is, short, representative fragments of an audio track.

Concerning the tags attached to the hooks, we expect the same limitations as with tagging in general. People have well-known problems in describing music. They tend to tag what they like and ignore what they dislike. Furthermore, they may respond to social pressure, reproduce judgements they have heard before, etc. rather than directly record their experience. Finally, we expect the tags will not be too different from the ones for complete tracks (if available): it seems likely that tags for these are first of all based on the most salient part of the music. It might be worthwhile to stimulate people to listen to music they do not know or do not particularly appreciate for two reasons: to get more diverse and rich tags for each item and to enable the emergence of a new kind of music criticism. Hopefully the Lure part of the proposed *Listen, Lure & Locate* environment will be effective in this respect: there is a serious design challenge here, both technically and socially.

It seems the tagging idea could be taken a couple of steps further. Listen, Lure and Locate focuses on remembering music, but could forgetting, which is an equally important aspect of our musical memory, somehow be accommodated? An initial suggestion would be to annotate 'anti-hooks,' such as disposable building blocks that surround hooks, music that is easy to forget and music that one would like to forget but is unable to. No doubt the features that contribute to forgetting (or wanting to forget) are also informative for understanding musical salience. Listeners may need an extra stimulus to annotate anti-hooks, but no doubt a sufficiently rewarding online game could be designed for this.

Another idea is to enable linking between musical segments. In addition to tagging a section of a piece that they find special, intriguing or moving, users could link that section to another section in an different (or the same) piece that creates a very similar experience. By doing so, one abstracts from the subjective semantics that are associated with tags. Obviously, these links could also be tagged by the user to describe the meaning of the link. Such links and tags are particularly interesting for MIR research because they capture the kind of similarity that is salient and relevant for the user.

CONCLUSION

We have presented a brief overview of the field of MIR and argued that knowledge about music is necessary for improving the current state-of-the-art in MIR. We believe that the assumption that all information is in the data will hamper the development of high performance, usable and effective MIR systems and that in particular knowledge about music cognition can aid in overcoming the limitations that MIR is facing today.

We have reviewed Honing's article specifically from the viewpoint of MIR. There are other perspectives as well that we decided not to explore, such as Internet culture, music criticism and music cognition itself. From the MIR perspective, the most salient characteristic of the *Listen, Lure & Locate* proposal is that it extends crowdsourcing by enabling more focused data to be created about a very high-level cognitive phenomenon in music: the hook. It might also be possible that the tags themselves will be more informative, but only insofar as the system will be able to alter the listening and annotation behaviour of the participants.

It is easy to imagine how novel methods could be created to mine such data. To develop such methods into a service to end-users sounds tempting, but in reality gathering annotations for millions of songs in competition with already existing social music sites presents a great challenge. The real value lies in the potential of the data to create new cognitive models of music, which in turn may be implemented as part of MIR systems. We believe that such systems have great potential and reflect the direction in which MIR research ought to move.

NOTES

All hyperlinks were accessed November 2010.

[1] Utrecht University, Department of Information and Computing Sciences, PO Box 80.089, 3508 TB Utrecht, The Netherlands {bas.dehaas, frans.wiering}@cs.uu.nl
[2] http://en.wikipedia.org/wiki/ITunes_Store
[3] http://www.apple.com/ipodclassic/features.html
[4] http://www.ismir.net
[5] http://www.ismir.net/proceedings/
[6]http://www.music-ir.org/mirex/wiki/MIREX_HOME
[7] http://www.last.fm/
[8] http://www.pandora.com
[9]http://www.mturk.com

REFERENCES

Aucouturier, J. & Pachet, F. (2004). Improving timbre similarity: How high is the sky. *Journal of Negative Results in Speech and Audio Science*, Vol. 1, No. 1, pp. 1–13.

Babbitt, M. (1965). The use of computers in musicological Research. *Perspectives of New Music*, Vol. 3 No. 2, pp. 74–83.

Bigand, E. & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, Vol. 100 No. 1, pp. 100–130.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition.* Oxford University Press, USA.

Burns, G. (1987). A typology of 'hooks' in popular records. *Popular Music*, Vol. 6, No. 1, pp. 1–20.

Casey M., Veltkamp R.C., Goto M., Leman M., Rhodes C., & Slaney M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, Vol. 96, No. 4, pp. 668–696.

Celma, O. (2006). Foafing the music: Bridging the semantic gap in music recommendation. *Web Semantics: Science, Services and Agents on the World Wide Web,* Vol. 6, pp. 250–256.

Downie, J.S. (2004). The scientific evaluation of music information retrieval systems: Foundations and future. *Computer Music Journal*, Vol 28, No. 2, pp. 12–23.

Downie, J.S. (2008). The music information retrieval evaluation exchange (2005-2007): A window into music information retrieval research. *Acoustical Science and Technology*, Vol. 29, No. 4, pp. 247–255.

Downie, J.S., Byrd, D., & Crawford, T. (2009). Ten years of ISMIR: Reflections on challenges and opportunities. *Proceedings of the 10th Society for Music Information Retrieval Conference (ISMIR)*, pp. 13–18.

Grachten, M., Arcos, J.L., & Lopez de Mantaras, R. (2004). Melodic similarity: Looking for a good abstraction level. *Proceedings of the 5th Society for Music Information Retrieval Conference (ISMIR)*, pp. 210–215.

de Haas, W.B.,Veltkamp, R.C., & Wiering, F. (2008). Tonal pitch step distance: A similarity measure for chord progressions. *Proceedings of the 9th Society for Music Information Retrieval Conference (ISMIR),* pp. 51–56.

de Haas, W.B., Rohrmeier, M., Veltkamp, R.C., & Wiering, F. (2009). Modeling harmonic similarity using a generative grammar of tonal harmony. *Proceedings of the 10th Society for Music Information Retrieval Conference (ISMIR)*, pp. 549–554.

Harris-Jones, C., & Haines, T.L. (1998). Sample size and misclassification: Is more always better? *Proceedings of the Second International Conference on the Practical Application of Knowledge Discovery and Data Mining*.

Honing, H., & Ladinig, O. (2009). Exposure influences expressive timing judgments in music. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 35, No. 1, pp. 281–288.

Honing, H. (2010). Lure(d) into listening: The potential of cognition-based music information retrieval. *Empirical Musicology Review,* Vol. 5, No. 4, pp. 146–151.

Huron, D. (2000). Perceptual and cognitive applications in music information retrieval. *Proceedings of the 1st Society for Music Information Retrieval Conference (ISMIR),* pp. 83–92.

Klapuri, A. (2005). A perceptually motivated multiple-F0 estimation. *IEEE Workshop on Method Applications of Signal Processing to Audio and Acoustics,* pp. 291–294.

van Kranenburg, P., Volk, A., Wiering, F., & Veltkamp, R.C. (2009). Musical models for folk-song melody alignment. *10th International Society for Music Information Retrieval Conference (ISMIR)*, pp. 507−512.

Krumhansl, C.L. (2001). *Cognitive Foundations of Musical Pitch.* Oxford University Press, USA.

Lerdahl, F. (2001). *Tonal Pitch Space.* Oxford University Press.

Logan, B. (2000). Mel Frequency Cepstral Coefficients for Music Modeling. *Proceedings of the 1th Society for Music Information Retrieval Conference (ISMIR).*

Mandel, M.I., Eck, D., & Bengio, Y. (2010). Learning tags that vary within a song. *Proceedings of the 11th Society for Music Information Retrieval Conference (ISMIR)*, pp. 399–404.

Narmour. E. (1990). *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. University of Chicago Press.

Orio, N. (2006). Music retrieval: A tutorial and review. *Foundations and Trends in Information Retrieval*, Vol. 1, No. 1, pp. 1–96.

Pearce, M.T., & Wiggins, G.A. (2006). Expectation in melody: The influence of context and learning. In *Music Perception,* Vol. 23, No. 5, pp. 377–405.

Pitt, M.A., & Myung, I.J. (2002). When a good fit can be bad, *Trends in Cognitive Sciences*, Vol. 6, No. 10, pp. 421–425.

Schellenberg, E.G. (1996). Expectancy in melody: Tests of the implication-realization model. *Cognition,* Vol. 58, No.1, pp. 75–125.

Schmidt, E.M., & Kim, Y.E. (2010). Prediction of time-varying musical mood distributions from audio. *Proceedings of the 11th Society for Music Information Retrieval Conference (ISMIR)*, pp. 465–70.

Temperley, D. (2001). *The Cognition of Basic Musical Structures.* Cambridge, MA: MIT Press.

Wiering, F. (2009) Meaningful music retrieval. *1st Workshop on the Future of MIR*, pp. 1-3.
http://www.columbia.edu/~tb2332/fmir/Papers/Wiering-fmir.pdf

Wiering, F., de Nooijer, J., Volk, A., & Tabachneck-Schijf, H.J.M. (2009). Cognition-based segmentation
for music information retrieval systems. *Journal of New Music Research*, Vol. 38, No. 2, pp. 139–154.

Wiggins, G.A. (2009). Semantic gap?? Schemantic schmap!! Methodological considerations in the
scientific study of music. *11th IEEE International Symposium on Multimedia*, pp. 477–482.

Wiggins, G.A. Mullensiefen, D., & Pearce, M.T. (2010). On the non-existence of music: Why music theory
is a figment of the imagination. *Musicæ Scientiæ*, pp. 231–255.