PREDICTING DEVELOPMENT OF RESEARCH IN MUSIC BASED ON PARALLELS WITH NATURAL LANGUAGE PROCESSING

Jacek Wołkowicz
Dalhousie University
Faculty of Computer Science
`jacek@cs.dal.ca`

Vlado Kešelj
Dalhousie University
Faculty of Computer Science
`vlado@cs.dal.ca`

ABSTRACT

The hypothesis of the paper is that the domain of Natural Languages Processing (NLP) resembles current research in music so one could benefit from this by employing NLP techniques to music. In this paper the similarity between both domains is described. The levels of NLP are listed with pointers to respective tasks within the research of computational music. A brief introduction to history of NLP enables locating music research in this history. Possible directions of research in music, assuming its affinity to NLP, are introduced. Current research in generational and statistical music modeling is compared to similar NLP theories. The paper is concluded with guidelines for music research and information retrieval.

1. INTRODUCTION

Along with the information revolution triggered by the introduction of computers, new opportunities have emerged for music artists and researchers. When some began executing computational problems on large mainframe computers others used them to generate early computer music. At this point of time one started to think how computers might be used to process and analyze music matter. Music, similarly to human speech, accompanied human evolution from its beginning, so deep understanding of music can allow better understanding of better human cognition. However, music data is in most cases still treated as unstructured binary data left on the same shelf with images, movies, computer programs; opposite to textual data, which are easy to process, search, index, driven by a large number of available computer aided techniques provided by natural language processing, information retrieval or text data mining like classification, analysis, generation, summarization, indexing, searching, translation and much more. However, music can be treated as a natural language and could be processed in the similar way as text. Although there are substantial differences between written text and music, they have many features in common.

NLP level	Music research areas
phonetics	Waveform analysis, audio signals
phonology	Sound events identification
morphology	Score symbols, symbolization
syntax	N-grams, shallow reduction and parsing
semantics	Harmonics, phrase level, parsing
pragmatics	Phrases, voice leading
discourse	Interpretations, context of a piece

Table 1. NLP Levels with respective music research tasks.

2. MUSIC AS A NATURAL LANGUAGE

By definition, a natural language is any language which arises in an unpremeditated fashion as the result of the innate facility for language possessed by the human intellect. Not everybody agrees that music fits this definition, but music researchers, who know the rules of music, are usually more prone to agree with it. Music, as well as text, has the symbolic representation that has its origins dated back in ancient times. Music and language are the only old human creative activities where symbolic representation is commonly used. Others, like painting, sculpture, dance did not have such common symbolic notation. Music notation cannot be directly ported into computers like text is, but this is only a representation issue that could be easily overcome. For instance, the argument that text can easily be split into words - the basic features for Natural Language Processing (NLP) and Information Retrieval (IR), which is not the case for music, can be countered if one mentions that there are natural languages that do not use anything to separate words, like Thai.

3. MUSIC RESEARCH PARALLEL TO NLP

In order to treat music as a natural language, one has to show that music processing works on the same classes of problems as NLP does. One distinguishes certain levels of a text processing, listed in the Table 1. NLP tries to convey the research through all those levels, from recording (a voice, speech) to understanding (the meaning of a speech).

These levels also exist for music. Similarly to a natural language, music can be recorded and presented primarily as a waveform. On the 'phonetics' level one tries to investigate the structure of a sound, separate and distinguish

between notes or instruments. However, music is much more complex in this area and sound recognition tasks are still facing basic problems.

The second very important similarity results from the fact both domains use symbolic notations. Music score also consists of characters which are called notes. Similarly to NLP's morphology and syntax — music has hidden, grammar-like structure, hidden rules. Part of it is the harmony. It determines how to put words (notes) together, how to build well-formed phrases using them. It also manages the musical meaning of a piece of which the basic exemplification is a progression of chords and notes. In the case of notes and their dependencies — we may talk about the syntax of the music while in the case of chords or harmonic progressions — about the semantics of the certain phrase or given the phrasing — the pragmatics of the excerpt. This is very similar in its form to one of the main areas of NLP, which is grammatical analysis. The highest level of NLP (discourse) is also common in music in a form of ideas, desires or aspirations (romantic music) of a composer as well as pictures and actions behind it (program music). Dukas's "The Sorcerer's Apprentice" or Smetana's "die Moldau" are the very good example of such music.

4. HISTORY OF NLP AND MUSIC RESEARCH

The history of NLP reaches beginnings of the history of computation since it is believed that the ability of computers to process natural language as easily as we do will signal creation of real intelligent machines. It comes from the assumption that the use of language is an inherent part of human cognitive abilities. Based on that a test, known as the Turing test, has been proposed to determine if a machine is intelligent. The main objective of this test is that a truly intelligent machine could carry on a discussion with a human so that the latter could not recognize if he is talking to a human or to a machine. This definition of intelligence triggered the research in NLP with the ultimate goal to create such a conversational program. If it is a feasible goal one has to actually and physically implement the way people think, reason and formulate thoughts.

At this point we can see a resemblance between human speech and music. Assuming music is a natural language, the Turing test would consist in generating musical pieces so that an expert could not recognize if an author is a machine or not. Why not a layman? Because it would be equivalent to a Turing test where an interlocutor does not know the language of the talk. We would call it a *Soft Turing Test* which, unlike for regular human natural languages, makes more sense for music.

4.1 Introduction of Grammars

There has been some pioneering work in both areas, NLP and Music Research. Some natural language analysis in a form of linguistics theories were made before the introduction of computing machines. As an early, pre-computer music research one can point work of Heinrich Schenker

with his Reduction theory in the beginnings of 20th century.

These things changed rapidly for NLP just after computers were invented. Interface to computers is textual and this is a natural format of computer files. Moreover, there was a very strong need for developing automatic machine translation. The beginning of this - the Noam Chomsky's theory of context free grammars for natural texts - dates back to 1956. Representing and processing music was not the top priority of that times. As a similar work in the field of music one can point the book "A Generative Theory of Tonal Music" by Lerdahl and Jackendoff published in 1983. Both of this approaches deal with the respective areas in the same way - by introducing a formal grammar that may generate instances in the given areas.

4.2 The period of 'Look ma, no hands'

The early NLP researchers were very optimistic. The research was driven by the goal of developing automatic machine translation. Various systems were created but, although they worked perfectly on several, very limited examples, they were failing in the real-world applications. The research came to the point where nothing more could be achieved, and yet they haven't created any robust system that will work on real data. This created a crisis in the whole field. It looked that despite their complicated system, it is not possible to mimic human cognition in the area of natural languages.

It is likely the time where the music research has just come. It is not that crucial as for natural languages, since one can still try to trick unexperienced listeners and thus pass the *Soft Turing Test* with the system that does not demonstrate the full understanding of the matter it deals with. Another sign that the field of music research might be in this kind of situation is the introduction of this kind of tracks, where one asks about the future of the field(e.g. fMIR).

4.3 Present NLP

The old approach - to create a model that will solve all our problems in the area of cognition of human speech — seemed to be wrong. Current research of NLP lies on one hand on creating more precise generative models that includes probability or other aspects of other features of languages (for example using attribute value matrices — AVMs). On the other hand, stochastic NLP has gone towards classical Data Mining (DM) where with the use of shallow parsing with mining on textual data gives much better results than simple DM. NLP techniques are also being injected to Information Retrieval increasing performance of this systems. Finally, the great improvement has been done in machine translation, the first goal of NLP researchers, where Google Translate or Wave's Rosy are the examples. This is the direction which current music research may follow.

5. RELATED WORK

Current research in music concentrates around Music Information Retrieval, both for the signal and symbolic music representations. In most cases it deals on basic issues how computers should deal with music data in general. The level of music interpretation does not go into semantics, probably because it is vague what the meaning in music is. However, one should notice that current text Information Retrieval benefits from the semantic layer of text (text classification, ontologies and relations between terms, dependencies between documents, linguistic layer of text).

We would like to emphasize the work of Lerdahl and Jackendoff [5], who first describe a generative approach that one can use toward the music. They describe it in a computational linguistics manner, using preference rules approach, mentioning that it could be possible to implement their rules in a working system. For the implementation of their system we had to wait for a long time, since they did several elisions of some tough to define, important basic notions, understandable by humans, but hard to implement on a machine. A recent try, ATTA [3], deals with all the implementation issues by introducing several important limitations to the system, which does not go beyond syntactic level, leaving behind harmony issues.

Another preference rules generative approach, that introduces very important component of modern NLP - probability, is described by Temperley [7]. Probabilities and corpus based statistics is an inherent part of all modern NLP theories hence probabilities can model the meaning of text by inferring dependencies within it [6]. This work reminds of the idea of probabilistic grammars introduced earlier for text and proposed for music by Bod [1].

Statistical analysis is a very important component of NLP models and it has played (Cope [2]) and will play a major role in music research. In many cases, solutions to many problems that gave good results for texts, could give comparable results in music area. As an example, the n-gram method of authorship attribution developed for natural language texts [4] gave good results for composer recognition of musical pieces [8].

6. FUTURE OF MUSIC RESEARCH

If the hypothesis that NLP and current computational music research are in a sense the same but operate in two similar but not identical fields, both fields could benefit from this legacy. For instance, applications that span large number of levels of NLP (e.g. try to draw some high level conclusions based on low level music representations) would work better, if they focus on a few levels only. As we have pointed out the layers of NLP, some of them are not that well covered for the music matter. Lots have been done in the areas of music 'phonetics', 'phonology' and 'morphology'. We notice some recent work in the area of 'semantics' but there is no models in higher, much more interesting but complicated levels: 'semantics' 'pragmatics' and 'discourse'. Those areas define the meaning of the data we deal with, the understanding of undergoing structure and

the flow of composers ideas within a piece. In general, one can stack different applications given the structure of NLP i.e. the output of a model that operates on syntactic level could be an input of a model operating on semantic level.

A few tasks that are relevant for music research and are well developed within NLP are sentiment analysis, genre classification, automatic summarization or idiom extraction. Other approach would be to enhance MIR with some semantic aspects of music matter - music ontologies with an application of shallow parsing (or alternatively, local reductions) to reach the level of current state-of-the-art of textual Information Retrieval. However, it is still not clear how to represent meaning of music in computational tasks but in this case statistical approach and data mining techniques may be relevant tool to describe this phenomenon.

7. CONCLUSIONS

The domain of music research resembles research in NLP. Both fields operate on the similar types of data that share common features. Both domains deal with data that are easily perceivable by humans but pose a lot of problems to make them fully understandable by computers. Research in both areas uses similar techniques and should be able to take from each other in the areas that are more developed in one of them. Music researchers could share their insight in tasks like voices separation or boundaries detection while benefit from NLP's statistical methods, automatic approaches to semantics or aiding information retrieval and data mining with natural language understanding. It is not necessary that all those inherited techniques and approaches will work but definitely, it is worth trying.

8. REFERENCES

[1] Rens Bod. A unified model of structural organization in language and music. *JAIR*, 17(1):289–308, 2002.

[2] David Cope. Computer modeling of musical intelligence in emi. *Comp. Music Journ.*, 16(2):69–83, 1992.

[3] Masatoshi Hamanaka, Keiji Hirata, and Satoshi Tojo. Implementing a generative theory of tonal music. *Journal of New Music Research*, 35:249–277, 2006.

[4] Vlado Keselj, Fuchun Peng, Nick Cercone, and Calvin Thomas. N-gram-based author profiles for authorship attribution. In *Proc. of the PACLING03 Conf.*, 2003.

[5] Fred Lerdahl and Ray Jackendoff. *A Generative Theory of Tonal Music*. The MIT Press, 1983.

[6] Christopher D. Manning and Hinrich Schütze. *Foundations of Statistical Natural Language Processing*. The MIT Press, June 1999.

[7] David Temperley. *Music and Probability*. The MIT Press, 2007.

[8] Jacek Wołkowicz, Zbigniew Kulka, and Vlado Keselj. N-gram-based approach to composer recognition. *Archives of Acoustics*, 33(2008)(1):43–55, Jan 2008.