Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Triennial Conference of the European Society for the Cognitive Sciences of Music, July 23-28, 2012, Thessaloniki, Greece
Cambouropoulos E., Tsougras C., Mavromatis P., Pastiadis K. (Editors)

Perceiving meaningful discourse structure in music and language

Jiaxi Liu

Faculty of Music, Cambridge University, United Kingdom
jl676@cam.ac.uk

ABSTRACT

Background

It may be assumed that the purpose of any form of communication is to transmit information. Philosophers and linguists typically analyse information in terms of truth or falsehood. Music, on the other hand, is taken to convey mostly emotional content; the existence of truth-conditional semantic meaning in music has remained an open debate between music researchers and linguists. There is good evidence, however, that both language and music consist of meaningful units that are structured hierarchically in similar ways. An understanding of the structural differences and commonalities between the two, as well as the way in which these meaningful structures are perceived and processed by humans, would be a valuable step towards an answer to the larger question of what kind of "meaning" is conveyed. If musical meaning is indeed functionally similar to linguistic meaning, then cognitive scientists, music therapists, advertising agencies, and even film score composers could all benefit from this new finding.

A large and growing body of research on language and music provides evidence for similar neural processing of music and language, suggesting that they share much in common (e.g. Steinbeis and Koelsch 2007). However, there remains considerable disagreement about whether the meaning of language is analogous to that of music. Many researchers reason that the lack of semantic meaning in music is what ultimately differentiates it from language (Pinker 1997 and Patel 2008). A comparative study on the structural makeup of these two domains would provide insight into how their respective meaningful elements are put together.

Fortunately, two comparable syntactic theories exist for music and text: Lerdahl and Jackendoff's Generative Theory of Tonal Music (GTTM) and Mann and Thompson's Rhetorical Structure Theory (RST). Both approaches construct their theories based on the recursive and hierarchical nature of musical and linguistic structure, respectively. Lerdahl and Jackendoff recognize that the musical relationships are based primarily on rhythm, while surface cues such as timbre, dynamics, and harmony are only secondary characteristics that ornament the underlying rhythmic patterns. Correspondingly for language, Mann and Thompson define the relationship between discourse segments in terms of nuclearity and coherence relations in linguistic literature. These nuclei outline the deep structure, e.g. functional clauses and trigger words like "but," "however," "therefore," etc. The roughly 26 logical relations that bind these nuclei are relationships like background, elaboration, and preparation among others. As a result, the theory produces a hierarchical structural analysis of discourse text.

Aims

Music studies have proposed that listeners do not pay much attention to the larger whole – the "global" structure – but do understand and are aware of "local," short-term structure (Tillman and Bigand 2004, Deliège et al 1997). However, linguistic data have yet to distinguish at which *level* meaningful structural perception occurs. This study undertakes a comparison between both media, additionally between musicians and nonmusicians.

Method

Two musical compositions were analysed for tree structure following the rules of GTTM: Beethoven's *Trio* from Piano Sonata, Op. 28 and the exposition of the *Andante cantabile* from Mozart's Piano Sonata, K. 333. Repeats were excluded from the recordings by Daniel Barenboim (1998) and Mitsuko Uchida (2003), respectively. Two texts with the same segment length and number of tree depths from a Wall Street Journal corpus annotated by Carlson et al 2001 were then paired with the musical stimuli (wsj0626, wsj1170, respectively). The branches of each tree were segmented at all tree depths, which were numbered from bottom up, starting at zero. This resulted in 22 total stimuli. The branches at each tree depth were cut and randomized as audio-visual music clips and visual text slides in iMovie projects (version 6.0.3).

One hundred students (50 musicians and 50 nonmusicians) at Northwestern University were paid $8 to participate in a 40-minute puzzle task. Subjects were defined as "musicians" if currently pursuing a degree in the Bienen School of Music and had completed at least one year of collegiate music theory and ear training. They were otherwise labeled "nonmusicians." Subjects were told that all the clips presented to them make up a real and complete work of music or text and were asked to drag slides onto a timeline to create what they considered the original composition. After a brief, observed training period during which subjects were familiarized with the iMovie interface, they performed the recreation task for selections from each of the four stimuli, which were randomized for the target depth and the ordering of the segments within those depths. A short subject information questionnaire followed.

Data were collected in the form of ordered segment strings. Edit distance analysis, a calculation that quantifies the minimal differences between two sequences, was applied on these data comparing the recreations to the originals within subjects and within stimuli. A second method of cluster analysis, which is a representation of internal cohesion, was applied to determine the subjects' local or global level of structural perception. Clustering assumes that similar observations group together and dissimilar observations do not; thus its output is a dendrogram (tree) of the groupings.

Results

When compared to a generated baseline of random edit distance values, subjects performed either much higher or much lower than chance, indicating that they were paying attention to the task. As expected, musicians were generally better than nonmusicians for the music stimuli, but only at intermediate levels ($p=0.03$ from a two-sample, equal-variance t-test). For text stimuli, there was no significant difference between subjects regardless of tree depth. This result was also not unexpected since musicians and nonmusicians had comparable linguistic knowledge.

Within the subjects, music recreations were only significantly better than text at intermediate levels. Interestingly, most nonmusicians have had lengthy musical training (average training of 5.9 years), yet did not perform on par with the musicians (average training of 12.1 years) for the music stimuli until at the upper levels. These observations may be explained by the influence of music theory knowledge in a "musician's" perception of underlying discourse structure.

For nonmusicians, the comparable edit distance similarities in music and text are unsurprising, since the tree structures of the two media are highly similar. These results suggest that nonmusicians did not perform poorly in recreating the music stimuli, only not as well as the musicians did.

Cluster analyses indicated that for music, musicians differed from nonmusicians in that the first attended to both structural (global) and surface (local) cues in their recreation process while the latter relied on surface (local) cues. As seen in the dendrograms, the segments attached one at a time rather than forming sub-trees that joined together. As a result, the musicians' dendrograms included more peripheral segments, i.e. external nodes of the tree. From self-reports of their recreation methods, nonmusicians described listening for continuity between segments in cues such as dynamics, tempo, and changes in key.

This was not the case for text, where all subjects were aware of global structure but depended primarily on segment adjacency (local cues) for the puzzle task. Nonetheless, musicians reported first creating the beginnings and ends of chunks before deciding what fits in between. Consequently, their recreations resulted in often balanced cluster dendrograms and nonmusicians in unbalanced ones. This trend indicates the musicians' willingness to approach and frame their recreations within a structural perspective regardless of medium.

Conclusions

These results confirm and add to previous findings on the perception of linguistic and musical discourse structure. Although musicians and nonmusicians may neurologically process music in a similar manner, they attend to different musical cues: musicians used both global and local ones while nonmusicians favoured local cues. Additionally, for text, all subjects performed comparably at the puzzle task and were aware of global structure but preferred to use local cues for recreation. These findings partially resonate with the work of Deliège et al 1997 and Tillman and Bigand 2004. However, this study also refines their claims and provides quantitative evidence for differing attention paid to structural elements in music *and* text by musicians versus nonmusicians.

Furthermore, the musical recreations of nonmusicians were not considerably worse than their corresponding texts, only worse than the recreations of their musician counterparts. This indicates that nonmusicians are knowledgeable about musical structure, as much as they are about linguistic structure. Such evidence that all subjects were able to construct meaningful recreations regardless of the compositional medium suggests that musical and linguistic meaning may share a similar degree of denotative quality.

Keywords

Discourse structure, meaning, hierarchical recursion, perception of music and text

REFERENCES

Barenboim, D. (1998). Piano Sonata in D major, Op. 28, *Trio* by Ludwig van Beethoven [CD]. EMI Classics, ASIN: B00000C2KP.

Carlson, L., Marcu, D. and Okurowski, M.E. (2001). Building a discourse-tagged corpus in the framework of rhetorical structure theory. Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue, Eurospeech 2001, Denmark, September 2001.

Deliège, I., Mélen, M., Stammers, D. and Cross, I. (1997). Musical Schemata in Real-Time Listening to a Piece of Music. *Music Perception*. 14(2): 117-160.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press.

Mann, W.C., and Thompson, S. A. (1987). Rhetorical structure theory: A theory of text organization. USC/Information Sciences Institute Technical Report Number RS-87-190, Marina del Rey, CA.

Patel, A. D. (2008). *Music, Language, and the Brain.* Oxford University Press, USA.

Pinker, S. (1997). *How the Mind Works*. Norton, New York.

Steinbeis, N. and Koelsch, S. (2007). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*.

Tillman, B., and Bigand, E. (2004). The Relative Importance of local and global structures in music perception. *Journal of Aesthetics and Art Criticism*, 62: 211-222.

Uchida, M. (2003). Piano Sonata in C major, K. 330, exposition of *Andante cantabile* by Wolfgang Amadeus Mozart [CD]. Philips, ASIN: B00005QDYG.