Understanding musical meaning: How does the listener parse music into phrases?

Susana M. Silva and São Luís Castro

Affiliation: University of Porto, Portugal
Address:
Susana M. Silva
Rua Duque de Palmela, 111, R/C, 4000-373 Porto
Email: zanasilva@clix.pt Phone: 96 436 67 25

São Luís Castro
Address:
FPCE-UP, Rua do Campo Alegre, 1021 4169-004 Porto
Phone: 22 607 9756 Fax: 22 607 9725
Email: slcastro@fpce.up.pt

Aims and Objectives

This study aims to contribute towards a better understanding of parsing in music and in prosody. Its goal is to compare how the same acoustic stimulus - modified speech - is segmented into major units depending on the instruction given to the listener ("it is music" vs. "it is speech"). The comparison is done for boundary marks and takes into account three types of parsing models (Gestalt, Parallelism and Intonation Groups).

Context

Generative frameworks state that *expressive strategies* in musical performance derive from *musical structure*. Expressive strategies are made up of expressive gestures that create changes in tempo, dynamics and articulation. The musical structure of a piece is regarded as immanent. It corresponds to the *group structure* of a piece, i.e., the way in which units - like sections or phrases - are delimited and organised (Clarke, 1988). According to this viewpoint, one can argue that a fundamental role in musical performance is to deliver some sort of a "parsing guide" to the listener. From the same viewpoint, the listener will have to decode that "parsing guide": s/he will have to identify phrase boundaries in order to communicate with the performer. Segmentation or parsing thus seems to be a nuclear component in the process of understanding music played by a performer. Identifying major units like phrases seems to be a more elemental task than, for example, decoding emotional (extra-musical) meaning.

Generative theories are mainly concerned with the performer's cognitive activity in delivering musical structure, namely as far as motor programming is concerned. But what about the listener? How is music segmentation processed? Does the listener rely mainly on domain-specific knowledge or, on the contrary, is music segmentation supported by a more general process?

One way of approaching this question is to compare segmentation processes in music with those that occur in another acoustic domain – speech. More specifically, music should be compared with prosody, since both deal with structured patterns of pitch, intensity and duration. While there is evidence regarding common resources in prosody and music processing (Patel, Peretz, Tramo & Labreque, 1998; Schoen, Magne, Shrooten & Besson, 2002; Thomson, Schellenberg & Husain, 2004), it is not known to what extent music segmentation relies on non-specific mechanisms such as those involved in speech parsing.

Segmentation models can be found in literature both for music and for speech: the importance of Gestalt (G) and Parallelism (P) models is discussed in the music domain (Bod, 2002; Bod, 2002a; Trehub, 2003; Drake & Bertrand, 2003; Patel, 2003; Shaeffer, Murre & Bod, 2004). Linguistic descriptions (Trask, 1996; Cruz-Ferreira, 1998, 1999; Lehrdahl, 2003) and psycholinguistic models (Levelt, 1993) refer to the Intonation Group (IG) as a significant prosodic unit related to clause.

In this study, all these models are tested in both domains (prosody and music), thus allowing comparison of behaviour determinants in parsing. Boundary cues that, according to the literature, may be important both to speech and music – pitch drop and final lengthening (Jusczyk & Krumhansl, 1993; Patel et al., 1998; Trainor & Adams, 2000; Lehrdahl, 2003) – are also tested. These cues were accommodated into the IG model, defining a subset of IGCC (Intonation Group with Common Cues) segments. The design of each type of model implied the definition of criteria for boundary setting. The stimulus was previously segmented according to such sets of criteria. Segmentation behaviours performed by participants were later confronted to the theoretical segmentation.

Methodology

Parsing behaviours for the same stimulus presented under two different instructions were compared. Half of the participants were told that the stimulus was modified speech, the other half that it was music (N = 20 x 2); within each group, half had musical training, the other half did not. The stimulus was an audio-to-MIDI transcription of an utterance. Their task was to break the stimulus into segments as they were listening to it, by pressing a key. For the purpose of control, participants also parsed a different version of prosodic speech derived from the original utterance (F0 synthesis).

Results

A considerable proportion of segments marked under the music instruction are common to segments marked in the speech instruction. Speech segmentation models (Intonation Groups) are found in speech, and music segmentation (Gestalt and Parallelism) models are found in music, but a music model (Parallelism) is predominant in both conditions. Speech segmentation models are less used in the music condition than music segmentation models are in the speech condition. IG and IGCC (speech models) segments are only found in speech. Besides, the average number of marked segments is significantly higher in speech (M=15) than in music (M=10), F (1, 36) = 9, 858, p=0.003. The designed models account for 53% of obtained segments. An emergent category, named Concatenated Parallels (IV), accounts for 30% of all marked segments, and for 37% of common segments. Concatenated parallels are more used in music (48% of marked units) than in speech parsing (27%), F(1,36)=8,017, p=0.03). Differences between musicians and non-musicians were limited to consistency in parsing the experimental stimulus, musicians showing more consistent behaviours.

Key Contribution

Our findings suggest that resources for music parsing (parallelism principles) have been largely used in speech parsing independently of musical expertise. There seem to be, however, resources that are exclusive to dealing with speech prosody, shorter, IG-typed units being accepted by listeners in speech processing. Whatever the explanation for the use of parallelism in speech, listening to music seems to be a processing activity involving mechanisms also required by speech processing. Regarding the "mission" of the music performer, highlighting parallel traits (motives) in music seems to be of major importance in helping the listener to rebuild the meaning of a piece. The same can also apply to improvisational performance, where parallel traits can be judged to be more powerful than other kinds of markers in structuring the listening act. Further research is needed.

References

Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research*, 17, 289-308.

Bod, R. (2002a). Memory-based models of melodic analysis: Challenging the Gestalt principles. *Journal of New Music Research*, 31, 27-37.

Clarke, E. (1988). Generative principles in music performance. In J. Sloboda (Ed.). *Generative Processes in Music* (pp. 1-26). Oxford: Clarendon Press.

Cruz-Ferreira, M. (1998). Intonation in european portuguese. In D. Hirst & A. di Cristo (Eds.). *Intonation systems. A survey of twenty languages* (pp. 167-178). Cambridge: Cambridge University Press.

Cruz-Ferreira, M. (1999). Portuguese (European). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet* (pp. 126-130). Cambridge: Cambridge University Press.

Drake, C. & Bertrand, D. (2003). The quest for universals in temporal processing in music. In I. Peretz & R. Zatorre (Eds.). *The cognitive neuroscience of music* (pp. 21-31). Oxford: Oxford University Press.

Jusczyk, P. & Krumhansl, C. (1993). Pitch and rhythm patterns affecting infants' sensitivity to musical phrase structure. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 627–640.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.

Lehrdahl, F. (2003). The sounds of poetry viewed as music. In I. Peretz & R. Zatorre (Eds.). *The cognitive neuroscience of music* (pp. 413-429). Oxford: Oxford University Press.

Levelt, W. (1993). The architecture of normal spoken language use. In G. Blanken, H. Grimm, J. Dittmann (Eds.), *Linguistic disorders and pathologies: An international handbook* (pp. 1-15). Walter De Gruyter Inc: New York.

Patel, A. (2003). A new approach to the cognitive neuroscience of melody. In I.Peretz & R. Zatorre (Eds.). *The cognitive neuroscience of music* (pp. 325-345). Oxford: Oxford University Press.

Patel, A., Peretz, I., Tramo, M. & Labreque, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain and Language*, 61**,** 123–144.

Schaffer, R., Murre, J. & Bod, R. (2004). Limits to universality in segmentation of simple melodies. *Proceedings of the 8th international conference on music perception and cognition* , 1-4, Evanston IL, 2004 ICMPC8 S.D. Lipscomb, R Ashley , R. O. Gierdingen & P. Webster (Ed). Adelaide, Australia: causal productions. Retrieved August 18, 2004, from http://staff.science.uva.nl/~rens/icmpc04.pdf.

Schoen, D., Magne, C., Schrooten, M., Besson, M. (2002, April 11-13). The music of speech: electrophysiological approach. *Proceedings from Speech Prosody [SP] Conference,* 635-638, Aix-en-Provence, France. Paper retrieved August 20, 2004, from http://www.isca-speech.org/archive/sp2002/sp02_635.pdf.

Thomson, W., Schellenberg, E. & Husain, G. (2004). Decoding speech prosody: do music lessons help? *Emotion*, 4(1) 46-64.

Trainor, L. & Adams, B. (2000). Infants' and adults' use of duration and intensity cues in the segmentation of tone patterns. *Perception & Psychophysics*, 62, 333–340.

Trask, R. (1996). *A dictionary of phonetics and phonology*. London and New York: Routledge.