Performa '09 – Conference on Performance Studies
University of Aveiro, May 2009

Phrasing in music and in language: An open question in the cognitive sciences

Susana Silva and São Luís Castro, Universidade do Porto,
Laboratório de Fala da Faculdade de Psicologia e de Ciências da Educação
labfala@fpce.up.pt

Aims

In this paper, we aim at understanding how continuous musical streams are segmented, that is, phrased by listeners. Specifically, we discuss how phrasing processes might be similar in music and in language. We focus on the extent to which local cues like pauses between phrases, and relational cues taking segments as coherent configurations, would be involved in both domains.

Context

In order to manage time-dependent objects such as music or speech, humans perform segmentation of these objects into smaller units. Some of these units are phrases, and dealing with mental representations of such segments is broadly referred to as phrasing. An intuitive analogy between the musical phrase and the speech phrase (part-of-sentence) is commonly taken for granted. However, what kind of evidence on this issue do we have from cognitive neuroscience?

Recent neurocognitive studies have revealed a pattern of electrophysiological brain responses that is associated with the processing of phrase boundaries in speech. This pattern is a positive peak in event-related potentials occurring after a phrase boundary, which became known as the closure positive shift or CPS (Steinhauer, Alter & Friederici, 1999). Interestingly, a CPS also occurs for speech without lexical content, that is, pure prosody (Pannekamp et al., 2005), and it is also present when musicians listen to phrased musical excerpts (Knoesche et al., 2005; Neuhaus, Knoesche & Friederici, 2006; Nan, Knoesche & Friederici, 2006). Does this mean that the human mind/brain deals with music as it deals with speech prosody?

In order to tackle that question, it is best to clarify what is the relevant level of comparison, or analogy, between both domains. Is it that the acoustical properties of a phrase are similar in music and in speech, leading the brain to attend to similar acoustical events and to process them with the same resources? Is it that the brain responds in a similar fashion to both kinds of auditory streams because they unfold in time and require online processing? Concerning this issue an interesting proposal was made by Patel (2008), who, in order to reconcile contradictions emerging in the study of music-language relations, invoked the distinction between representations and functions. *Representations* are instances of domain-specific knowledge that are fundamentally different in music and in language. *Functions* are neural resources that operate upon knowledge and that can be shared across domains. This distinction was proposed to be relevant for syntax in music and in language. Here, and speculating for phrasing, a corresponding hypothesis might be that the relevant acoustic properties of phrases in music differ from those of phrases in speech prosody (different representations), while neural resources associated with cognitive processes like memory and attention are involved in phrasing in both domains (same functions). Let us consider this as our working hypothesis.

First, it can be claimed that representations of musical phrases and of prosodic phrases do not completely differ; the fact that these two types of phrases share acoustic cues has indeed been the starting point for event-related potentials, ERP, studies of phrasing in music. For example, Knoesche et al. (2005) stated that the markers of prosodic phrase boundaries “mainly consist in changes in the temporal structure, especially in the insertion of pauses, but also in pitch and volume changes; therefore, they are very similar to the ones found in music.” (p. 260). Findings from this study led the same researchers to state that “the CPS seems to reflect closure processes that can be the consequence of the perception of phrase boundaries, rather than the perception of these boundaries itself” (ibd.).

In fact, a CPS for music has been observed in response to markers or cues that do not completely overlap with the acoustic markers of prosody that evoke a CPS for language. Studies on the music CPS have shown that this pattern of brain activity occurs in response to a pause between two four-bar phrases (phrased condition), and that it does not occur when the same two four-bar phrases are connected by musically-adequate tones (unphrased condition). Pauses can be reduced in length and still allow the CPS to be evoked (Neuhaus et al., 2006), but in this case there has to be a trade-off between pause length and another marker of phrase boundary, the lengthening the final tone (with the note-to-note interval at the boundary being maintained). In other words, the CPS for music has been observed in response to a slowing down at the end of a four-bar phrase, but not in response to the end of the same four-bar segment when there is no such slowing down. On the other hand, studies on the CPS for language have shown that this brain response is evoked by naturally produced speech when the phrase boundary is marked by a pause, lengthening of the final syllable, or a salient pitch movement corresponding to accent. The CPS has been observed when the pause was removed (Steinhauer, 1999), as well as after an accentuated word within the phrase (Hruska & Alter, 2004; Toepel & Alter, 2003). So, regarding *local cues* for phrasing, i.e., critical events arising within a time window shorter than the whole phrase, it seems that these local cues are quite close, though not strictly the same, in music and in language for a CPS brain response to occur.

Regarding *relational cues*, i.e., cues related to the processing of the whole phrase as a coherent configuration, it appears that they do not have to be strictly matched in music and language stimuli in order to evoke the CPS response: while musical phrases in CPS studies last on average 5 seconds, prosodic groups last 1 second; musical phrases contain motivic repetitions within and between phrases, something that does not exist in speech prosody; similarly, while musical phrases were designed to be four-bar segments, no temporal template can be defined for prosodic groups.

If there is no compelling evidence that the brain response to phrase boundaries in music and in speech is similar because they are elicited by similar acoustic events, what about common processes or functions? Can it be argued that these are shared? Based upon a technique to detect the brain regions originating the music CPS, Knoesche et al. (2005) have suggested that this brain response might arise from regions involved in storing the phrase as a “unified entity” (a coherent configuration), and in redirecting the focus of attention to the next incoming event, or phrase.

In order to contribute to this discussion, we present the results of two comparative studies on phrasing in music and in speech prosody.

Methodology

In Study 1, our aim was to test whether available cues guided boundary setting in the same manner for music and for speech prosody. Two groups of participants (between-subjects design) heard an ambiguous auditory stream, which was either labelled by the experimenter as music (N = 20) or as 'modified speech' without discernable words (N = 20). Half of the participants in each group were professional musicians. Participants were asked to set boundary marks for phrases while hearing the streams. In comparison with participants in the modified speech group, those hearing the stream labelled as music defined less and longer segments, which were more variable in length; they also attended less to pauses. Furthermore, no differences were observed between professional musicians and the other participants.

In Study 2, our aim was to gain better understanding on the cognitive processes that occur at the boundaries of musical phrases used as stimuli in CPS studies. We examined ERPs and event-related power changes in the electroencephalogram of musicians who listened to melodies composed by two musical phrases. The phrases were either separated by a pause or by filling tones. After the pause, there was a positive ERP peak ranging from 400 to 700 milliseconds post pause onset, and an alpha band (8-12 Hz) power increase. The filling tones evoked a theta band (4-7 Hz) power increase at their onset and an alpha band decrease at their offset. The ERPs after the pause are evidence of a CPS component.

Conclusions

The findings from Study 1 suggest that acoustic events alone do not determine the perception of boundaries: in face of exactly the same acoustic materials, listeners picked up cues to place boundaries in music which were different from the cues used to place boundaries in speech prosody. Expectations played a role, indicating that the mental representation of phrases in both domains is different, at least concerning some dimensions.

The findings from Study 2 support the hypothesis that the music CPS reflects a redirection of the attention focus from one phrase to the next. A possible interpretation of the results is that listeners anticipate the end of the phrase in both conditions, based perhaps upon the typical four-bar length. But whereas the pause confirms the listener's expectation of closure, the lack of pause precludes such a confirmation and inhibits the perception of closure. So, while a phrase followed by a pause would lead to redirect the focus of attention to the next time window, a phrase continued by short, filling tones, would not elicit such an attention shift.

Taken together, the results of both studies support the claim that phrases in music and in speech do not share representations (study 1), even though these can be arrived at through similar, possible the same, cognitive resources dealing with events that unfold in time (study 2). In this sense, Patel's proposal is likely to apply to the comparative study of phrasing in music and in speech.

Finally, the fact that musical expertise is necessary to evoke a music CPS (this brain response is not evoked when musically naive adults listen to phrased melodies) leaves room for several interpretations. It is possible that musicians become equipped with better resources for directing the attention focus over unfolding events than non-musicians. However, it is also possible that the online management of attention over musical materials requires explicit knowledge of musical structures that is acquired through formal musical training and/or continued practice.

Research on phrasing has started recently, and the question for the cognitive sciences remains open: are phrases in music *just like* phrases in speech, or is it that our mind/brain deals with both as time-dependent objects that have to be parsed in order to be perceived and, each in its own way, understood?

Bibliography

Hruska, C., & Alter, K. (2004). How prosody can influence sentence perception. In A. Steube, (ed). *Information structure: theoretical and empirical evidence*. Berlin: Walter de Gruyter.

Knösche, T.R., Neuhaus, C., Haueisen, J., Alter, K., Maess, B.,Witte, O.W., & Friederici, A.D. (2005). The perception of phrase structure in music. *Human Brain Mapping*, 24, 259–273.

Nan, Y., Knösche, T. R., & Friederici, A.D. (2006). The perception of musical phrase structure: a cross-cultural ERP study. *Brain Research*, 1094, 179-191.

Neuhaus, C., Knösche, T.R. & Friederici, A. D. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience*, 18, 1–22.

Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A.D. (2005). Prosody driven sentence processing: an event-related brain potential study. *Journal of Cognitive Neuroscience*, 17 (3), 407-421.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Steinhauer K., Alter K., & Friederici A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, 2, 191–196.

Toepel U., & Alter K. (2003). How mis-specified accents can distress our brain. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, Spain, August 2003, pp. 619–622.