

Comparison of Rhythmic Processing in Language and Music: An Interdisciplinary Approach

Cyrille Magne^{1*}, Mitsuko Aramaki¹, Corine Astesano¹, Reyna Leigh Gordon¹, Sølvi Ystad²,
Snorre Farner², Richard Kronland-Martinet² & Mireille Besson¹

¹ Institut de Neurosciences Cognitives de la Méditerranée (INCM) - CNRS, Marseille, France

² Laboratoire de Mécanique et d'Acoustique (LMA) - CNRS, Marseille, France

Acknowledgements:

This research was supported by a grant from the Human Frontier Science Program (HSFP #RGP0053) to Mireille Besson. Cyrille Magne benefited from a Cognitive Science fellowship from the French Ministry of Research and Industry, and Corine Astésano is a post-doctorate student supported by a grant from the Human Frontier Science Program.

Abstract

In this paper we describe an interdisciplinary collaboration between phoneticians, acousticians and neuroscientists that led to a study of rhythm in music and language. In the first part of the paper we discuss general aspects of rhythm, with a short overview of some earlier studies on the cultural influences of linguistic rhythm on musical rhythm. In the second part, we describe an experimental procedure aimed at comparing the perception of rhythmic and semantic violations in language with the perception of rhythmic and harmonic violations in music. Subjects listened to different sentences and melodies and were asked to focus on either rhythm or semantics/harmony to indicate whether or not the last word/arpeggio was acceptable or not in the context. The Event-Related Brain Potential method was used to study perceptual and cognitive processing related to the rhythmic and semantic/harmonic incongruities. The results indicated that the processing of rhythmic incongruities was associated with increased positive deflections in the Brain Potential in similar latency bands in both language and music. However, these positivities were present independently of the participants' focus in the music part while they were only present when the participants focused on semantics in the language part.

I RHYTHM IN LANGUAGE AND MUSIC

1. Introduction

Rhythm seems to be part of all human activities and may be considered as an organizing principle that structures events. In music, for instance, rhythm and harmony are the structural supports of a composition. Rhythm could be defined as the way in which one or more unaccented beats (i.e., the basic pulses of the music) are grouped together in relation to an accented beat (by means of loudness, increase of pitch, or lengthening). Rhythm is also a central feature of language and speech. We defined it as the “temporal organization of prominence” that partakes in the prosodic structuring of utterances (Astésano, 2001). In this sense, rhythm, together with intonation, constitute the framework of prosodic organization. Speech prosody fulfills both an emotional and a linguistic function. While the emotional function of prosody is clearly to convey emotions through the speech signal, the linguistic function is realized at different levels of language organization. Prosody not only functions so as to delimit the word at the lexical level, by means of accentuation and metrical markers (e.g., stress patterns), but also to structure the sentence at the phrasing level (e.g., grouping of words into relevant syntactic units and providing intonation contours, e.g. rising for questions and falling for declarations,...), and at the discourse level (e.g., prosodic focus, e.g. emphasizing new relevant information). Interestingly, prosody relies on several acoustic parameters: spectral characteristics, fundamental frequency (F0), intensity and duration that are similar in music. In this paper, we mainly focus on duration (rhythm) and we compare the perception of rhythm in linguistic and musical phrases. Before turning to the perceptual aspects, however, it seemed important to quickly mention some recent and very interesting work on rhythm production.

2. Cultural influences of linguistic rhythm on musical rhythm

In a seminal paper, Patel and Daniele (2003a) tested the very interesting hypothesis that the language of a culture influences the music of that culture. As a consequence, the prosody of a composer’s native language may influence the structure of his music. The authors chose to compare speech and music from England and France because English and French provide prototypical examples of stress-timed vs. syllable-timed languages (i.e., languages that rely more on the stress patterns vs. on syllabic duration). Though this rhythmic

distinction is widely debated in the linguistic literature, English seems to present more contrast in durational values between successive syllables than French.

Using a measure of durational variability recently developed for the study of speech rhythm (Grabe and Low, 2002), Ramus (2002) was able to demonstrate significant differences between spoken British English and French. Interestingly, Patel and Daniele (2003a) then demonstrated that, differences (albeit smaller) of the same sort were found between excerpts of English and French instrumental music. Thus, these results support the idea that the linguistic rhythm of a language leaves an imprint on its musical rhythm.

These results were then extended by Huron and Ollen (2003) to a larger sample of musical themes by Brown (2000), including not only a larger sample of English and French music than in the Patel and Daniele (2003a) study, but also music from many other countries (e.g., Norwegian, Polish, Russian and Italian). Overall, they replicated the differences between English and French music reported in the previous study. Moreover, they also found that the durational variability index value for Austrian and German music was closer to Spanish, Italian and French music than to American or English music. Insofar as Austrian and German are stress-timed languages, like English, and following the idea that linguistic rhythm influences musical rhythm, the reverse was expected (Austrian and German closer to English than to some syllable-timed languages, such as Italian). By examining these data from an historical perspective (as a function of the composer's birth year), Patel and Daniele (2003b) were able to show that the values of the durational variability index increase with time, possibly reflecting less influence from Italian musical style on Austrian and German music over the years. Note, however, that for most Austrian and German composers, values of the durational variability index are smaller than the averaged value for Italian composers. In any event, these results open the interesting possibility that historical influences of one musical culture on the other may sometimes outweigh linguistic influences on music within the same culture.

These results are interesting in that they demonstrate that the rhythmic structure of a language may exert an overall influence on musical composition, and therefore clearly emphasize the similarities between linguistic and musical rhythmic production. We now turn to the perception of rhythm in speech and music

II AN INTERDISCIPLINARY APPROACH TO THE PERCEPTION OF RHYTHM IN LANGUAGE AND MUSIC

1. Why is interdisciplinarity a necessity?

In order to design experiments aiming at a better understanding as to how the rhythmic structure of a language, or of a piece of music, influences or determines its perception, it is absolutely necessary to combine and use knowledge from different research domains. First, we need to find the most important characteristics of the rhythmic structure of our language or music stimuli in order to manipulate the relevant parameters. Consequently, both phoneticians and musicologists are strongly needed. Second, we need to be able to manipulate speech and musical signals to effectuate fine transformations and construct high-quality stimuli. Consequently, acousticians are strongly needed. Third, we need to determine how the manipulations that are introduced within the speech or musical signals are perceived by the volunteers who participate to the experiment. Thus, to be able to draw significant conclusions, it is necessary to build up clear experimental designs, based on the methods developed in experimental psychology. Consequently, both psycho-linguists and psycho-musicologists are strongly needed. Finally, if we are interested in the brain computations that support auditory perception and cognition, we strongly need neuroscientists. In the example reported below, we conducted a series of experiments that required the collaboration among researchers from the fields of phonetics, acoustics, psychology and neuroscience.

2. An example of interdisciplinary approach

2.1. Introduction

Recent results in the field of phonetics have led some authors to propose that, in French, words are marked by an initial stress (melodic stress) and a final stress or final lengthening (Di Cristo, 1999; Astésano, 2001). However, as the initial stress is secondary, the basic French metric pattern is iambic, meaning that words or groups of words are marked by a final lengthening. Therefore, final syllabic lengthening is very important for our perception and understanding of a spoken language, since it allows the marking of the end of groups of words and of sentences in an otherwise continuous flow of auditory information. Interestingly, based on research in the music domain, Repp (1992) also considered that final lengthening is a widespread musical phenomenon that manifests itself in musicians' performance in the form of deviations from the steady beat that is present in the underlying representation. As in

speech, final lengthening in music primarily allows a demarcation of structural boundaries between musical phrases. Based on these (phonetic and musical) considerations, we decided to manipulate final syllabic lengthening in language and final note lengthening in music.

The aim of the experiments that we conducted was three-fold. First, we wanted to determine whether such rhythmic cues as final syllabic and note lengthening are perceived in real time by the listeners. It was therefore of uttermost importance to build up linguistic and musical materials that matched this goal. This required a strong collaboration between phoneticians and acousticians. Second, it was of interest to determine whether the violation of final syllabic or note lengthening would hinder perception and/or comprehension of the linguistic and musical phrases. Moreover, we wanted to examine the relationship between rhythmic and semantic processing in language, on the one hand, and between rhythmic and harmonic processing in music, on the other hand. Therefore, we built an experimental design that comprised four experimental conditions (see below), and we recorded both the time it took for the participants to give their answers (i.e., Reaction Times, RTs) and the error percentage. Finally, we wanted to compare the effects obtained with speech and music in order to determine whether the perceptual and cognitive computations involved are specific to one domain or, by contrast, rely on general cognitive processes. Interestingly, Brown (2001) and others have also suggested that music and speech may have a common origin in a pre-semantic musilanguage. In this case, we could find similarities between speech and music that would be specific to both, but not necessarily dependent upon general perceptive or cognitive processes.

Note that while behavioral data such as RTs and error rates provide an excellent measure of the level of performance for a given task, they do not allow the study of perceptual and cognitive processing of final syllabic and note lengthening as they unfold in real time. Thus, we needed a method with an excellent time resolution. The Event-Related Brain Potentials (ERPs) method is perfectly suited for this purpose, because it allows for the time-tracking of information processing. Before summarizing the different steps of this interdisciplinary venture - building up the materials and the experimental design, analyzing the results and discussing their implications - we first quickly describe the main reasons for using the ERPs method.

2.2. The ERPs method

By placing electrodes on the scalp of human volunteers, it is possible to record the

variations of the brain electrical activity, known as the Electro-Encephalogram (EEG, Berger, 1929). The EEG being of small amplitude (of the order of 100 microvolts, μV), such analog signals are amplified and then converted to digital signals (see Figure 1). Interestingly, any stimulus, be it a sound, a light, a touch, a smell, etc... elicits changes, or variations of potentials, in the background EEG. However, such variations are very small (of the order of 10 μV) compared to the background EEG activity. It is therefore necessary to synchronize the EEG recording to the onset of the stimulation, and to sum and average a large number of trials in which similar stimulations were presented. As a consequence, the signal - that is, the variations of potential evoked by the stimulation (or Evoked Potentials or Event-Related Potentials, ERPs) - will emerge from the background noise linked to the EEG activity (see Figure 2). Typically, in perceptual and cognitive experiments, 20 to 30 stimulations belonging to the same experimental condition need to be averaged to obtain a good signal-to-noise ratio. The ERPs comprises a series of positive and negative deflections, called components, relative to the baseline, that is, the averaged level of brain electrical activity during 100 or 200 milliseconds (ms) before the stimulation. Components are defined by their polarity (negative, N, or positive, P), their latency from stimulus onset (100, 200, 300, 400 ms, etc...), their scalp distribution (location of the maximum amplitude on the scalp) and their sensitivity to experimental factors. Typically, the ERPs recorded in perceptual and cognitive experiments, comprise early components that reflect the sensory and perceptual stages of information processing (e.g., P100, N100 and P200) and later components that reflect cognitive stages of processing (e.g., P300, N400, etc...). ERPs are analyzed using different latency windows, centred around the maximum amplitude, or the peak amplitude, of the component of interest (e.g., 300-600 ms latency window to analyze the N400 component) and the mean amplitude within this latency band is computed (see Rugg & Coles, 1995 for an overview of the ERP method).

Recording brain waves

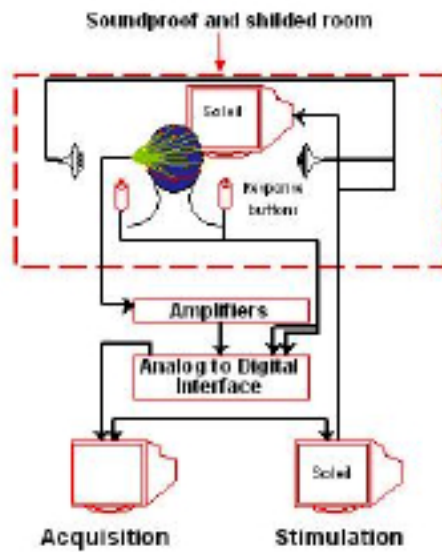


Figure 1: Illustration of the experimental room and materials used to record the brain electrical activity (see text for details)

Averaging method

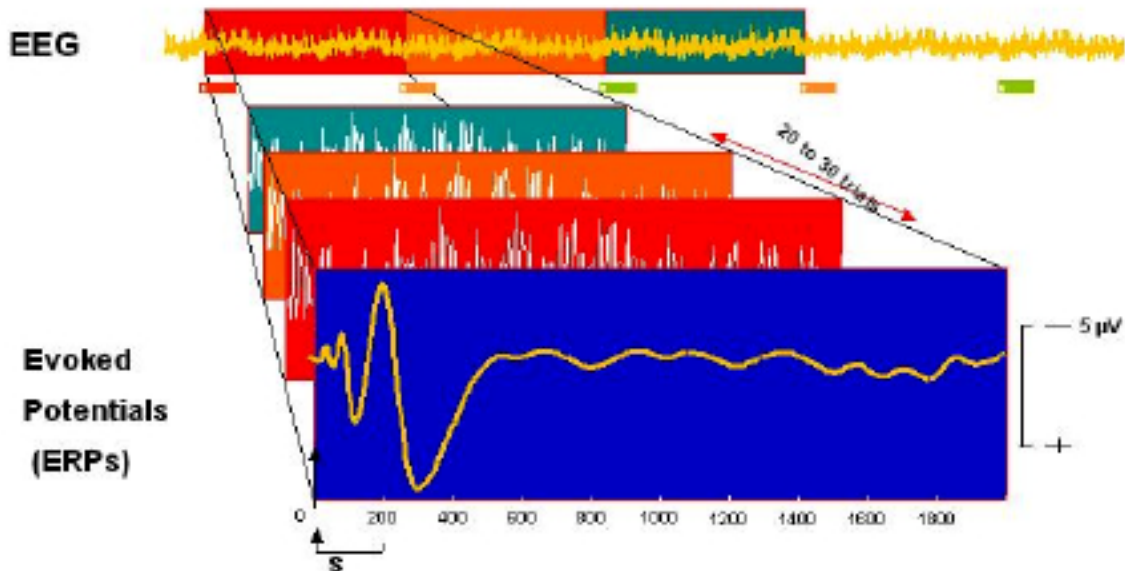


Figure 2: Averaging method used to extract the Event Related brain Potentials (ERPs) from the background Electroencephalogram (EEG; see text for details). On this and subsequent figures, the amplitude of the effects is represented on the ordinate (microvolts, μV ; negativity is up), time from stimulus onset on the abscissa (milliseconds, ms).

2.3. Construction of stimuli and design of the experiments

For the language part of the experiment, we created 128 sentences ending with trisyllabic words (i.e. “*Les musiciens ont composé une mélodie*” [“*The musicians composed a melody*”]). The sentences were read by a native French man and recorded in an anechoic room. The last word of each sentence was further segmented in three syllables, making it possible to measure the final syllabic lengthening for each final word, which in our case was found to be 1.7 times the average length of the non-stressed syllables. This ratio is within the range, 1.7 – 2.1, previously reported as average final syllable lengthening of words in prosodic groups in French (Astésano, 2001). In order to determine whether syllabic lengthening is processed in real time, that is, while participants listened to the sentences, we created rhythmic incongruities by lengthening the penultimate syllable, instead of the final

syllable, of the last word of the sentence. For this purpose, we developed a time-stretching algorithm that allowed us to manipulate the duration of the acoustic signal, without altering timbre or frequency (Pallone, 1999, Ystad et al., 2005). For the incongruity to remain natural, the lengthening was based on the ratio 1.7. Moreover, to examine the relationship between rhythmic and semantic processing, we also created semantic incongruities by replacing the last word of the congruous sentences by a word that was unexpected (i.e., it did not make sense) in the sentence context (i.e. “*Les musiciens ont composé une bigamie.*” [“*The musicians composed a bigamy*”]).

By independently manipulating rhythmic and semantic congruency of the trisyllabic final words, we created four experimental conditions in which the last word of the sentence was rhythmically and semantically congruous (R+S+); rhythmically congruous and semantically incongruous (R+S-); rhythmically incongruous et semantically congruous (R-S+); and, rhythmically and semantically incongruous (R-S-).

In the music experiment, we used sequences of five to nine major and minor chords that were diatonic in the given key and followed the rules of Western harmony. Simple melodies were composed (by one French researcher, two Norwegian researchers and one American researcher with different musical backgrounds) out of these chord progressions by playing each chord as a 3-note arpeggio. This choice was related to the language experiment in which the last words in the sentences always were three-syllabic words. Indeed, while we certainly do not regard a syllable in language as being equal to a note in music, it was useful, for the purposes of the experiment, to try to isolate aspects of language that it would make sense to postulate as corresponding to aspects of musical. Therefore, the rhythmic and harmonic incongruities were obtained from modifications of the penultimate note of the last triad of the melody. The rhythmic incongruity was created by lengthening the penultimate note by a ratio of 1.7 as in the language experiment. Finally, to be consistent with the language experiment in which the first syllable of the last word did not necessarily indicate whether or not the last word was incongruous, the first note of the last triad was harmonically congruent with both the beginning and the end of the melody. Thus, disharmony could only be perceived from the second note of the last triad (sound examples illustrating these stimuli are available at <http://www.lma.cnrs-mrs.fr/~kronland/JMM/sounds.html>).

To summarize, as in the language experiment, we independently manipulated rhythmic and harmonic congruencies to create four experimental conditions, in which the last three notes of the sentence were 1) both rhythmically and harmonically congruous (R+H+), 2)

rhythmically incongruous and harmonically congruous (i.e., the penultimate note was lengthened, R-H+), 3) harmonically incongruous and rhythmically congruous (i.e., the penultimate note was transposed, R+H-) and 4) both rhythmically and harmonically incongruous (i.e., both lengthened and transposed, R-H-).

Within both the language and music experiments, the participants were asked to focus their attention on the rhythm of the last word/arpeggio in half of the trials, and to decide whether its rhythm was acceptable or not (Rhythmic Task). In the other half of the trials, the participants were asked to focus their attention on the meaning (Harmonic structure) of the sentence (melody), and to decide whether the last word/arpeggio was acceptable in the context or not (Semantic/Harmonic Task).

2.4. Results and discussion

A total of 14 participants were tested in the language experiment and a total of 8 participants were tested in the music experiment. Volunteers were students from the Aix-Marseille Universities and were paid to participate in the experiments that lasted for about 2 hours (including the installation of the Electro-Cap). They were 23-years-old on the average, and right-handed with no known auditory deficit.

Results of the language experiment showed that rhythmic processing seems to be task-dependent (see Figure 3). Indeed, when participants focused their attention on the rhythmic aspects (left column, Meter), words that were semantically congruous and rhythmically incongruous (S+R-) elicited a larger positive component, peaking around 700 ms post-stimulus onset (P700), than semantically and rhythmically congruous words (S+R+). In contrast, when participants listened to the very same sentences, but focused their attention on the semantic aspects (right column, Semantic), words that were semantically congruous and rhythmically incongruous (S+R-) elicited a larger negative component, peaking around 400 ms post-stimulus onset (N400) than semantically and rhythmically congruous words (R+S+).

LANGUAGE

Rhythmic incongruity effect

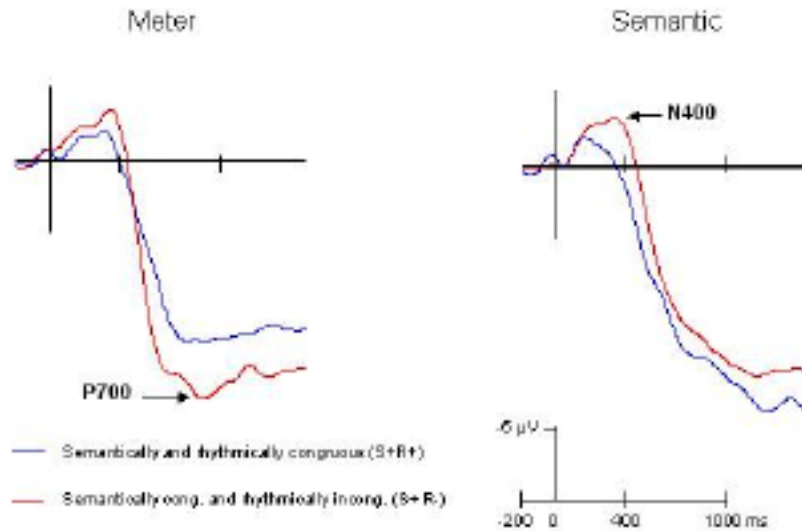


Figure 3. Event-Related Potentials (ERPs) evoked by the presentation of the semantically congruous words when rhythmically congruous (S+R+, blue trace) or rhythmically incongruous (S+R-, red trace). Results when participant focused their attention on the rhythmic aspects are illustrated in the left column (Meter) and when they focused their attention on the semantic aspects in the right column (Semantic). The electrophysiological data are presented for one representative central electrode (Cz).

In contrast, preliminary results of the music experiment, suggest that the effects elicited by the rhythmic incongruity are task-independent (see Figure 4). Indeed, regardless of the specific aspect, rhythmic or harmonic, on which participants focused their attention, harmonically congruous and rhythmically incongruous musical sequences (R-H+) elicited a larger biphasic negative-positive complex between 200 and 900 ms (N300-P600) than harmonically and rhythmically congruous sequences (R+H+).

MUSIC

Rhythmic incongruity effect

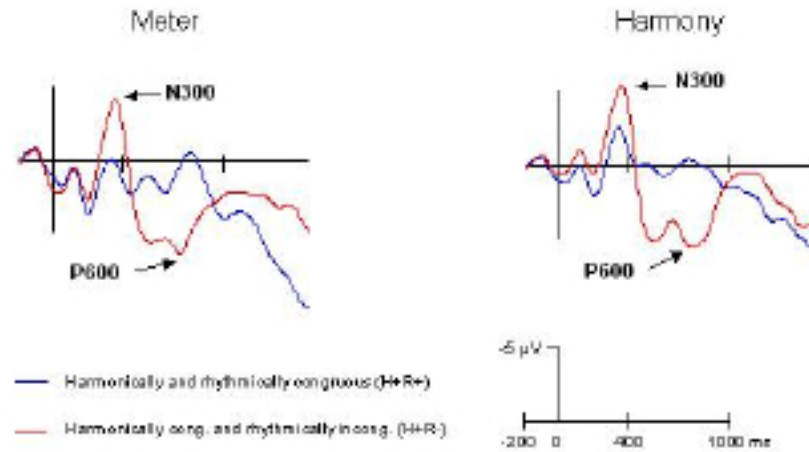


Figure 4. Event-Related Potentials (ERPs) evoked by the presentation of the harmonically congruous musical sequence when rhythmically congruous ($H+R+$, blue trace) or rhythmically incongruous ($H+R-$, red trace). Results when participant focused their attention on the rhythmic aspects are illustrated in the left column (Meter) and when they focused their attention on the harmonic aspects in the right column (Harmony). The electrophysiological data are presented for one representative central electrode (Cz).

These preliminary results on the processing of rhythm in both speech and music already point to interesting similarities and differences. Considering the similarities first, it is clear that the rhythmic structure is processed on-line in both language and music. Indeed, rhythmic incongruities elicited a different pattern of brain waves than rhythmically congruous events. Moreover, it is worth noting that the processing of these rhythmic incongruities was associated with increased positive deflections in the Brain Potential in similar latency bands in both language (P700) and music (P600). However, while these positive components were present in music under both attentional conditions (rhythmic and harmonic tasks), they were

only present in the rhythmic task in the language experiment. Thus, while the processing of rhythm may be obligatory when listening to the short melodic sequences presented here, it seems to be modulated by attention when listening to the linguistic sentences.

Turning to the differences, it is interesting to note that while the positive components were preceded by negative components (N300) in music under both attentional conditions, no such negativities were found in language in the rhythmic task. Thus, again, rhythmic violations seem to elicit earlier (N300) and larger (N300-P600) brain responses in music than in language (only P700). Interestingly, in the semantic task of the language experiment, negative components (N400) were elicited by semantically congruent but rhythmically incongruous words (S+R-). This finding may seem surprising in the context of a large body of literature showing that semantically congruous words (i.e., words that are expected within a sentence context such as “candidates” in *Le concours a regroupé mille candidats*) typically elicit much smaller N400 components than semantically incongruous words, a result generally interpreted as reflecting larger integration difficulties for incongruous words (e.g., “*Le concours a regroupé mille bigoudis*”). Therefore, our results seem to indicate that the lengthening of the penultimate syllable disturbs access to word meaning and consequently increases difficulties in the integration of the word meaning within the sentence context. These findings, by pointing to the influence of metrical structure, may have important consequences for our understanding of the factors that contribute to lexical access and word meaning. More generally, they demonstrate the important role of lexical prosody in speech processing.

In conclusion, by briefly presenting our research project on the perception of rhythm in both language and music, we tried to illustrate the strong need for interdisciplinarity in order to conduct such experiments. Moreover, our results open interesting perspectives: for instance, using the functional Magnetic Resonance Imaging (fMRI) method to determine whether similar or different networks of brain structures are involved in the processing of the rhythmic structures in language and music. More generally, following up our previous work (Schön et al., 2004; Magne et al., 2003) showing that musical training facilitates the perception of changes in fundamental frequency, not only in music, but also in language, we will test transfer of learning from musical to linguistic rhythm. Recent experiments with dyslexic children tend, for instance, to demonstrate that classroom music lessons had beneficial effects on both rhythm copying and phonological and spelling abilities (Overy, 2003).

References

- Astésano C. (2001) *Rythme et accentuation en français. Invariance et variabilité stylistique*. Editions L'Harmattan, Collection Langue & Parole.
- Berger, H. (1929). Uber das Electrenkephalogramm das menchen. *Archiv für Psychiatrie*, 87:527-570.
- Besson, M., F. Faïta, C. Czernasty & M. Kutas. (1997). What's in a pause: Event-Related Potential Analysis of Temporal Disruptions in Written and Spoken Sentences. *Biological Psychology* 46:3-23.
- Besson, M., & Macar, F. (1987). An Event-Related Potentials Analysis of Incongruity in Music and Other Non-linguistic Contexts. *Psychophysiology*, 24:14-25.
- Besson, M. & D. Schön. (2002). Comparison Between Language and Music. In *The Biological Foundations of Music*, R. Zatorre & I. Peretz, Eds., Science and Medical Publications, Oxford University Press, Oxford.
- Brown (2000). The "Musilanguage" Model of Music Evolution. In *The Origins of Music*. Eds N.L. Wallin, B. Merker and S. Brown, MIT Press, Cambridge, Massachusetts.
- Di Cristo, A. (1999). Le cadre accentuel du français contemporain: essai de modélisation (1^{ère} partie). *Langues*, 2 (3), 184-205.
- Duez, D. & Nishinuma, Y. (1985) "Le rythme en français: alternance des durées syllabiques", *Travaux de l'Institut de Phonétique d'Aix*, Vol. 10, 151-169.
- Gabrielsson, A. (1982) "Perception and Performance of Musical Rhythm", in *Music, Mind and Brain. The neuropsychology of Music*. 'An expanded version of proceedings from the Third Workshop on the Physical and Neuropsychological Foundations of Music, held in August 8-12, 1980, in Ossiach, Austria', Manfred Clynes (ed.), Plenum Press. New York., 159-169.
- Grabe, E., & Low, E. L. (2002). Durational Variability in Speech and the Rhythm Class Hypothesis. In C. Gussenhoven & N. Warner, *Laboratory phonology* (pp. 515–546). 7. Berlin: Mouton de Gruyter.
- Huron, D. & Ollen, J. (2003). Agogic Contrast in French and English Themes: Further Support for Patel and Daniele (2003). *Music Perception*, 21: 267-272.
- Magne, C., Schön, D., & Besson, M. (2003). Prosodic and Melodic Processing in Adults and Children: Behavioural and Electrophysiological Approaches. *Annals of the New York Academy of Sciences*, 999:461-476.

- Nishinuma, Y. & Duez, D. (1987) "Étude perceptive de l'organisation temporelle de l'énoncé en français", *Travaux de l'Institut de Phonétique d'Aix*, 11, 184-201.
- Overy, K. (2003) *Dyslexia and Music: From Timing Deficits to Musical Intervention*. Ann. N.Y. Acad.Sci. 999:497-505.
- Pallone, G., Boussard, P., Daudet, L., Guillemain, P., Kronland-Martinet, R. (1999) A Wavelet Based Method for Audio Video Synchronization in Broadcasting Applications. In proceedings of the DAFX'99, Trondheim, Norway, Dec 9-11.
- Patel, A.D. & Daniele, J.R. (2003a). An Empirical Comparison of Rhythm in Language and Music. *Cognition*, 87: B35-B45.
- Patel, A.D. & Daniele, J.R. (2003b). Stressed-timed vs. Syllable-timed Music? A Comment on Huron and Ollen. *Music Perception*, 21: 273-276.
- Patel, A., Gibson, E., Ratner, J., Besson, M. & Holcomb, P. (1998). Processing Syntactic Relations in Language and Music: An Event-Related Potential study. *Journal of Cognitive Neuroscience*, 10:717-733.
- Ramus, F. (2002). Acoustic Correlates of Linguistic Rhythm: Perspectives. Proceedings of Speech Prosody 2002, Aix-en-Provence (pp. 115–120). Aix-en-Provence: Laboratoire Parole et Langage.
- Repp, B. (1992) "Probing the Cognitive Representation of Musical Time: Structural Constraints on the Perception of Timing Perturbations.", *Haskins Laboratories Status Report on Speech Research*, 111/112, New Haven, Connecticut, 293-320.
- Rugg, M & Coles, M.G.H. (1995). *Electrophysiology of Mind: Event-Related Brain Potentials and Cognition*. Oxford Psychology Series, 25. Oxford.
- Schön, D., Magne, C., & Besson, M. (2004). The Music of Speech: Electrophysiological Study of Pitch Perception in Language and Music. *Psychophysiology*, 41, 341-349.
- Touati, P. (1987) "Structures prosodiques du suédois et du français : profils temporels et configurations tonales.", *Travaux de l'Institut de Linguistique de Lund*, Vol. 21, Lund University Press.