Hindawi Publishing Corporation
EURASIP Journal on Audio, Speech, and Music Processing
Volume 2007, Article ID 30194, 13 pages
doi:10.1155/2007/30194

Research Article

Electrophysiological Study of Algorithmically Processed Metric/Rhythmic Variations in Language and Music

Sølvi Ystad,[1] Cyrille Magne,[2,3] Snorre Farner,[1,4] Gregory Pallone,[1,5] Mitsuko Aramaki,[2] Mireille Besson,[2] and Richard Kronland-Martinet[1]

[1] *Laboratoire de Mécanique et d'Acoustique, CNRS, Marseille, France*
[2] *Institut de Neurosciences Cognitives de la Méditerranée, CNRS, 13402 Marseille Cadex, France*
[3] *Psychology Department, Middle Tennessee State University, Murfreesboro, TN 37127, USA*
[4] *IRCAM, 1 Place Igor Stravinsky, 75004 Paris, France*
[5] *France Télécom, 22307 Lannion Cedex, France*

Received 1 October 2006; Accepted 28 June 2007

Recommended by Jont B. Allen

This work is the result of an interdisciplinary collaboration between scientists from the fields of audio signal processing, phonetics and cognitive neuroscience aiming at studying the perception of modifications in meter, rhythm, semantics and harmony in language and music. A special time-stretching algorithm was developed to work with natural speech. In the language part, French sentences ending with tri-syllabic congruous or incongruous words, metrically modified or not, were made. In the music part, short melodies made of triplets, rhythmically and/or harmonically modified, were built. These stimuli were presented to a group of listeners that were asked to focus their attention either on meter/rhythm or semantics/harmony and to judge whether or not the sentences/melodies were acceptable. Language ERP analyses indicate that semantically incongruous words are processed independently of the subject's attention thus arguing for automatic semantic processing. In addition, metric incongruities seem to influence semantic processing. Music ERP analyses show that rhythmic incongruities are processed independently of attention, revealing automatic processing of rhythm in music.

1. INTRODUCTION

The aim of this project associating audio signal processing, phonetics and cognitive neuroscience is twofold. From an audio point of view, the purpose is to better understand the relation between signal dilation and perception in order to develop perceptually ecological algorithms for signal modifications. From a cognitive neuroscience point of view, the aim is to observe the brain's reactions to modifications in duration of small segments in music and language in order to determine whether the perceptual and cognitive computations involved are specific to one domain or rely on general cognitive processes. The association of different expertise made it possible to construct precisely controlled stimuli and to record objective measures of the stimuli's impact on the auditor, using the event-related potential (ERP) method.

An important issue in audio signal processing is to understand how signal modification affects our perception when striving for naturalness and expressiveness in synthesized music and language. This is important in various applications such as designing new techniques to transcode audio tracks from cinema to video format and vice-versa. Specifically, the cinema format comprises a succession of 24 images per second, while the video format comprises 25 images per second. Transcoding between the two formats is realized by projecting the images at the same rate, inducing changes in the duration of the film. Consequently, the soundtrack duration needs to be modified to guarantee synchronization between sounds and images, thus requiring the application of time-stretching algorithms preserving the timbre content of the original soundtrack. A good understanding of how time-stretching can be used without altering perception, and how the quality of various algorithms can be evaluated, are thus of great importance.

A better understanding of how signal duration modifications influence our perception is also important for musical interpretation, since local rhythmic variations represent a key aspect of musical interpretation. A large number of authors (e.g., Friberg et al. [1]; Drake et al. [2]; Hirsh et al. [3]; Hoopen et al. [4]) have studied timing in

acoustic communication and the just noticeable difference for small perturbations of isochronous sequences. Algorithms that act on the duration of a signal without modifying its properties are important tools for such studies. Such algorithms have been used in recent studies to show how a mixture between rhythm, intensity and timbre changes influence the interpretation (Barthet et al. [5]).

From a neuro cognitive point of view, recording the brain's reactions to modifications in duration within music and language is interesting for several reasons. First, to determine whether metric cues such as final syllabic lengthening in language[1] are perceived by the listeners, and how these modifications alter the perception (and/or comprehension) of linguistic phrases. This was the specific aim of the language experiment that we conducted. Second, to better understand how musical rhythm is processed by the brain in relation with other musical aspects such as harmony. This was the aim of the music experiment.

Since the early 1980's, the ERP method has been used to examine and compare different aspects of language and music processing. This method has the advantage of allowing to record changes in the brain electrical activity that are time-locked to the presentation of an event of interest. These changes are, however, small in amplitude (of the order of 10 μV) compared to the background EEG activity (of the order of 100 μV). It is therefore necessary to synchronize EEG recordings to the onset of the stimulation (i.e., event of interest) and to average a large number of trials (20 to 50) in which similar stimulations are presented. The variations of potential evoked by the event of interest (therefore called event-related potentials, ERPs) then emerge from the background noise (i.e., the EEG activity). The ERPs comprise a series of positive and negative deflections, called components, relative to the baseline, that is, the averaged level of brain electrical activity within 100 or 200 ms before 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 maximum amplitude on the scalp) and their sensitivity to experimental factors.

So far, these studies seem to indicate that general cognitive principles are involved in language processing when aspects such as syntactic or prosodic processing are compared with harmonic or melodic processing in music (Besson et al. [6], Patel et al. [7]; Magne et al. [8]; Schön et al. [9]). By contrast, a language specificity seems to emerge when semantic processing in language is compared to melodic and harmonic processing in music (Besson and Macar [10], but see Koelsch et al. [11] for counter evidence). Until now, few electrophysiological studies have considered fine metric/rhythmic changes in language and music. One of these studies was related to the analysis of an unexpected pause before the last word of a spoken sentence, or before the last note of a musical phrase (Besson et al. [6]). Results revealed similar reactions to the pauses in music and language, suggesting similarities in rhythmic/metric processing across domain. However, since these pauses had a rather long duration (600 ms), such a manipulation was not ecological and results might reflect a general surprise effect. Consequently, more subtle manipulations are needed to consider rhythmic/metric processing in both music and language. This was the motivation behind the present study. In the language experiment, French sentences were presented, and the duration of the penultimate syllable of trisyllabic final words was increased to simulate a stress displacement from the last to the penultimate syllable. In the music experiment, the duration of the penultimate note of the final triplet of a melody was increased to simulate a rhythmic displacement.

Finally, it was of interest to examine the relationship between violations in duration and harmony. While several authors have used the ERPs to study either harmonic (Patel et al. [12]; Koelsch et al. [13]; Regnault et al. [14]) or rhythmic processing (Besson et al. [6]), to our knowledge, harmonic and rhythmic processing have not yet been combined within the same musical material to determine whether the effects of these fundamental aspects of music are processed in interaction or independently from one another. For this purpose, we built musical phrases composed of triplets, which were presented within a factorial design, so that the final triplet either was both rhythmically and harmonically congruous, rhythmically incongruous, harmonically incongruous, or both rhythmically and harmonically incongruous. Such a factorial design was also used in our language experiment and was useful to demonstrate that metric incongruities in language seems to hinder comprehension. Most importantly, we have developed an algorithm that can stretch the speech signal without altering its other fundamental characteristics (fundamental frequency/pitch, intensity and timbre) in order to use natural speech stimuli. The present paper is mainly devoted to the comparison of reactions to metric/rhythmic and semantic/harmonic changes in language and music, and to the description of the time-stretching algorithm applied to the language stimuli. A more detailed description of the behavioral and ERP data results of the language part is given in (Magne et al. [15]).

2. CONSTRUCTION OF STIMULI

2.1. Language experiment

Rhythm is part of all human activities and can be considered as the framework of prosodic organization in language (Astésano [16]). In French, rhythm (or meter, which is the term used for rhythm in language), is characterized by a final lengthening. Recent studies have shown that French words are marked by an initial stress (melodic stress) and a final stress or final lengthening (Di Cristo [17]; Astésano [16]). The initial stress is however secondary, and words or groups of words are most commonly marked by final lengthening. Similarly, final lengthening is a widespread musical phenomenon leading to deviations from the steady beat that is present in the underlying presentation. These analogies

[1] Final syllable lengthening is a widespread phenomenon across different languages by which the duration of the final syllable of the last word of sentences, or groups of words, is lengthened, supposedly to facilitate parsing/segmentation of groups of words within semantically relevant units.

between language and music led us to investigate rhythm perception in both domains.

A total of 128 sentences with similar number of words and durations, and ending with tri-syllabic words were spoken by a native male French speaker and recorded in an anechoic room. The last word of each sentence was segmented into syllables and the duration of the penultimate syllable was increased. As the lengthening of a word or a syllable in natural speech mainly is realized on the vowels, the artificial lengthening was also done on the vowel (which corresponds to the stable part of the syllable). Words with nasal vowels were avoided, since the segmentation of such syllables into consonants and vowels generally is ambiguous. The lengthening factor (dilation factor) was applied to the whole syllable length (consonant + vowel) for the following reasons:

(1) the syllable is commonly considered as the perceptual unit
(2) an objective was to apply a similar manipulation in both language and music, and the syllabic unit seems closer to a musical tone than the vowel itself. Indeed, musical tones consist of an attack and a sustained part, which may respectively be compared to the syllable's consonant and vowel.

The duration of the penultimate syllable of the last word was modified by a time-stretching algorithm (described in Section 2.1.2). Most importantly, this algorithm made it possible to preserve both the pitch and the timbre of the syllable without introducing audible artifacts. Note that the time-stretching procedure did not alter the F0 and amplitude contours of the stretched syllable, and simply caused these contours to unfold more slowly over time (i.e., the rate of F0 and amplitude variations differ between the metrically congruous and incongruous conditions). This is important to be aware of when interpreting the ERP effect, since it means that the syllable lengthening can be perceived soon after the onset of the stretched second syllable. Values of the mean duration of syllables and vowels in the tri-syllabic words are given in Table 1. The mean duration of the tri-syllabic words was 496 ms and the standard deviation was 52 ms.

Since we wanted to check possible cross-effects between metric and semantic violations, the tri-syllabic word was either semantically congruent or incongruous. The semantic incongruity was obtained by replacing the last word by an unexpected tri-syllabic word, (e.g., "Mon vin préféré est le karaté"—my favorite wine is the karate). The metric incongruity was obtained by lengthening the penultimate syllable of the last word of the sentence ("ra" in "karaté") by a dilation factor of 1.7. The choice of this factor was based on the work of Astésano (Astésano [16]), revealing that the mean ratio between stressed and unstressed syllables is approximately 1.7 (when sentences are spoken using a journalistic style).

2.1.1. Time-stretching algorithm

In this section, we describe a general time-stretching algorithm that can be applied to both speech and musical signals. This algorithm has been successfully used for cinema to video transcoding (Pallone [18]) for which a maximum of 20% time dilation is needed. We describe how this general algorithm has been adapted to allow up to 400% time dilation on the vowel part of speech signals.

Changing the duration of a signal without modifying its frequency is an intricate problem. Actually, if $s(\omega)$ represents the Fourier transform of a signal $s(t)$, then $(1/\alpha)s(\omega/\alpha)$ is the Fourier transform of $s(\alpha t)$. This obviously shows that compression (resp., lengthening) of a signal induces transposition to higher (resp., lower) pitches. Moreover, the formant structure of the speech signal—due to the resonances of the vocal tract—is modified, leading to an altered voice (the so-called "Donald Duck effect"). To overcome this problem, it is necessary to take into account the specificities of our hearing system.

Time-stretching methods can be divided into two main classes: frequency-domain and time-domain methods. Both methods present advantages and drawbacks, and the choice depends on both the signal to be modified and the specificities of the application.

2.1.2. Frequency domain methods

In the frequency domain approach, temporal "grains" of sound are constructed by multiplying the signal by a smooth and compact function (known as a window). These grains are then represented in the frequency domain and are further processed before being transformed back to the time domain. A well-known example of such an approach is the phase vocoder (Dolson [19]), which has been intensively used for musical purposes. The frequency-domain methods have the advantage of giving good results for high stretching ratios. In addition, they do not cause any anisochrony problems, since the stretching is equally spread over the whole signal. Moreover, these techniques are compatible with an inharmonic structure of the signal. They can however cause transient smearing since transformation in the frequency domain tends to smooth the transients (Pallone et al. [20]), and the timbre of a sound can be altered due to phase unlocking (Puckette [21]), although this has been improved later (Laroche and Dolson [22]). Such an approach is consequently not optimal for our purpose, where ecological transformations of sounds (i.e., that could have been made by human beings) are necessary. Nevertheless, they represent valuable tools for musical purpose, when the aim is to produce sound effects, rather than perfect perceptual reconstructions.

2.1.3. Time-domain methods

In the time-domain approach, the signal is time-stretched by inserting or removing short, non-modified segments of the original time signal. This approach can be considered as a temporal reorganization of non-modified temporal grains. The most obvious time-stretching method is the so-called "blind" method, which consists in regularly duplicating and inserting segments of constant duration (French and Zinn [23]). Such a method has the advantage of being very simple. However, even by using crossfades, synchronization discontinuities often occur, leading to a periodic alteration of the sound.

Table 1: Mean values (ms), and standard deviation (Sd) in brackets, of vowel(V) and syllable(S) lengths of the tri-syllabic words.

Segments	V1	V2	V3	V3/V2	S1	S2	S3	S3/S2
Meanva and Std	65 (24)	69 (17)	123 (36)	1.79(0.69)	150 (28)	145 (28)	202 (42)	1.39(0.45)

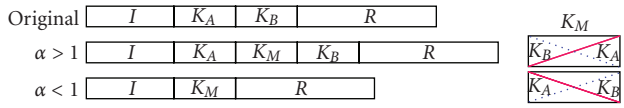

Figure 1: Insertion of a segment K_M to time-stretch a signal frame. The upper stripe represents the original signal. The second one illustrates how the signal is lengthened by adding an element K_M, and the third one illustrates how the signal can be shortened by replacing elements K_A and K_B by the element K_M. I is the initial delay, while R is the residual segment allowing to assure the correct dilation ratio before the next frame is processed.

Other time-domain approaches are based on adaptive methods aiming at matching the length of the inserted segments to the fundamental period (Roucos and Wilgus [24]). These methods give high quality sounds for dilation factors less than 20%. However, a doubling of transients might occur in this case as well as synchronization discontinuities on inharmonic and polyphonic sounds.

Finally, the problem of transient doubling has been addressed by Pallone [18]), whose work has been applied in a commercial product for real-time stretching of movie sound tracks between different playing speeds for instance between video (25 pictures/sec) and cinema (24 pictures/sec) format. The algorithm selects the best segment to insert, optimizes its duration and selects the best location for insertion. It was derived from so-called SOLA (WSOLA and SOLAFS) methods (Verhelst and Roelands [25], Hejna et al. [26]).

In our specific situation it was extremely important that the chosen signal processing method did not cause any audible sound quality modification. The algorithm used by Pallone [18] was found to be extensible to very strong dilation ratios, so we decided to adopt and optimize it for our purpose. We also foresee its usage on stretching of musical signals although we have settled on using MIDI in the music part of this study. In the following section, we briefly describe the algorithm in its completeness before presenting the optimizations that made us able to stretch vowels more than four times without audible defects.

2.1.4. A specific time-based algorithm

The principle of the time-domain algorithm is illustrated in Figure 1. The original signal is sequentially decomposed into a series of consecutive frames. Each frame is cut into 4 segments defined by 2 main parameters:

(1) the segment I, whose length I represents an initial delay, which can be adjusted in order to choose the best area of the frame for manipulation, and
(2) the segment K_M, whose length K is also the length of both K_A and K_B.

Letting α be the stretching factor, a lengthening of the signal ($\alpha > 1$) can be obtained by crossfading elements K_B and K_A, and inserting the resulting segment K_M between K_A and K_B. A similar procedure can be used to shorten the signal ($\alpha < 1$): by replacing K_A and K_B by a crossfaded segment K_M obtained from K_B and K_A. The crossfading prevents discontinuities because the transitions at the beginning and the end of K_M correspond to the initial transitions.

Each signal frame should be modified so that the dilation ratio is respected within the frame. The relation linking the length of R with the length of I, K_A, K_B, and K_M is thus given by the equation:

$$\alpha(I + K_A + K_B + R) = (I + K_A + K_M + K_B + R). \quad (1)$$

For $\alpha < 1$ (signal shortening), the segments K_A and K_B are set to zero at the right-hand side. Although this process seems simple and intuitive in the case of a periodic signal (as the length K should correspond to the fundamental period), the choice of the segments K_A and K_B is crucial and may be difficult if the signal is not periodic. The difficulty consists in adapting the duration of these segments-and consequently of K_M-to prevent the time-stretching process from creating any audible signal modifications other than the perceptual dilation itself. On one hand, a segment that is too long might, for instance, provoke the duplication of a localized energetic event (for instance a transient) or create a rhythmic distortion (anisochrony). Studies on anisochrony have shown that for any tempo, the insertion of a segment of less than 6 ms remains inaudible unless it contains an audible transient (Friberg and Sundberg [1]). On the other hand, a short segment might cause discontinuities in a low-frequency signal, because the inserted segment does not correspond to a complete period of the signal. This also holds for polyphonic and inharmonic signals in the case that a (long) common period may be found. Consequently, the length of the inserted segment must be adapted to the nature of the signal so that a long segment can be inserted when stretching a low-frequency signal and a short segment can be inserted when the signal is non-stationary.

To calculate the location and length of the inserted element K_M, different criteria were proposed for determining the local periodicity of the signal and the possible presence of transients. These criteria are based on the behavior of the autocorrelation function and of the time-varying energy of the signal, leading to an improvement of the sound quality obtained using WSOLA.

Choice of the length K of the inserted segment

The main issue here consists in determining the length K that gives the strongest similarity between two successive segments. This condition assures an optimal construction of the segment K_M and continuity between the inserted segment

and its neighborhood. We have compared three different approaches for the measurement of signal similarities, namely the average magnitude difference function, the autocorrelation function, and the normalized autocorrelation function. Due to the noise sensitivity of the average magnitude function (Verhelst and Roelands [25] and Laroche [27]) and to the autocorrelation function's sensibility to the signal's energy level, the normalized autocorrelation function given by

$$CN(k) = \frac{\sum_{n=0}^{N_c-1} s(n)s(n+k)}{\sqrt{\sum_{n=0}^{N_c-1} s^2(n) \sum_{n=0}^{N_c-1} s^2(n+k)}} \quad (2)$$

was applied. This function takes into account the energy of the analyzed chunks of signal. Its maximum is given by $k = K$, as for the autocorrelation function $C(k)$, and indicates the optimal duration of the segment to be inserted. For instance, if we consider a periodic signal with a fundamental period T_0, two successive segments of duration T_0 have a normalized correlation maximum of 1. Note that this method requires the use of a "forehand criterion" in order to compare the energy of the two successive elements K_A and K_B, otherwise, the inserted segment K_M might create a doubling of the transition between a weak and a strong sound level. Using a classical energy estimator easily allows to deal with this potential problem.

2.1.5. Modifications for high dilation factors

As mentioned in Section 2.1.1, our aim was to work with natural speech and to modify the syllable length of the second-last syllable of the last word in a sentence by a factor 1.7.

The described algorithm works very well for dilation factors up to about 20% ($\alpha = 1.2$) for any kind of audio signal, but for the current study higher dilation factors were needed. Furthermore, since vowels rather than consonants are stretched when a speaker slows down the speed in natural speech, only the vowel part of the syllable was stretched by the algorithm. Consequently, the local dilation factor applied on the vowel was necessarily greater than 1.7, and varied from 2 to 5 depending on the vowel to consonant ratio of the syllable. To achieve such stretching ratios, the above algorithm had to be optimized for vowels. Since the algorithm was not designed for dilation ratios above $\alpha = 1.2$, it could be applied iteratively until the desired stretching ratio was reached. Hence, applying the algorithm six times would give a stretching ratio of $\alpha = 1.2^6 \approx 3$. Unfortunately, we found that after only a few repetitions, the vowel was perceived as "metallic," probably because the presence of the initial segment I (see Figure 1) caused several consecutive modifications of some areas while leaving other ones unmodified.

Within a vowel, the correlation between two adjacent periods is high, so the initial segment I does not have to be estimated. By setting its length I to zero and allowing the next frame to start immediately after the modified element K_M, the dilation factor can be increased to a factor 2. The algorithm inserts one modified element K_M of length K between the two elements K_A and K_B, each of the same length K, and then lets K_B be the next frame's K_A. In the above described algorithm, this corresponds to a rest segment R of length-K for $\alpha = 2$.

The last step needed to allow infinite dilation factors, consists in letting the next segment start inside the modified element K_M (i.e., allowing for $-2K < R < -K$). This implies re-modifying the already modified element and this is a source for adding a metallic character to the stretched sound. However, with our stretching ratios, this was not a problem. In fact, as will be evident later, no specific perceptual reaction to the sound quality of the time-stretched signal were elicited, as evidenced by the typical structure of the ERP components.

Sound examples of speech signal stretched by means of such a technique can be found at http://www.lma.cnrs-mrs.fr/~ystad/Prosem.html, together with a small computer program to do the manipulations.

2.2. Music experiment

Rhythmic patterns like long-short alternations or final lengthening can be observed in both language and music (Repp [28]). In this experiment, we constructed a set of melodies comprising 5–9 triplets issued from minor or major chords. The triplets were chosen to roughly imitate the language experiment, since the last word in each sentence always was tri-syllabic. As mentioned above, the last triplet of the melody was manipulated either rhythmically or harmonically, or both, leading to four experimental conditions. The rhythmic incongruity was obtained by dilating the second-last note of the last triplet by a factor 1.7, like in the language experiment. The first note of the last triplet was always harmonically congruous with the beginning of the melody, since in the language part the first syllable of the last word in the sentences did not indicate whether or not the last word was congruous or incongruous. Hence, this note was "harmonically neutral," so that the inharmonicity could not be perceived before the second note of the last triplet was presented. In other words, the first note of an inharmonic triplet was chosen to be harmonically coherent with both the beginning (harmonic part) and the end (inharmonic part) of the melody.

A total of 128 melodies were built for this purpose. Further, the last triplet in each melody was modified to be harmonically incongruous (R+H−), rhythmically incongruous (R−H+), or both (R−H−). Figure 2 shows a harmonically congruous (upper part) and harmonically incongruous (lower part) melody. Each of these 4 experimental conditions comprised 32 melodies that were presented in pseudo-random order (no more than 4 successive melodies for the same condition) in 4 blocks of 32 trials. Thus, each participant listened to 128 different melodies. To ensure that each melody was presented in each of the four experimental conditions across participants, 4 lists were built and a total of 512 stimuli were created.

Piano tones from a sampler (i.e., prerecorded sounds) were used to generate the melodies. Frequencies and durations of the notes in the musical sequences were modified by altering the MIDI codes (Moog [29]). The time-stretching algorithm used in the language experiment could also have

been used here. However, the use of MIDI codes considerably simplified the procedure and the resulting sounds were of very good quality (http://www.lma.cnrs-mrs.fr/~ystad/Prosem.html, for sound examples). To facilitate the creation of the melodies, a MAX/MSP patch (Puckette et al. [30]) has been developed so that each triplet was defined by a chord (see Figure 3). Hereby, the name of the chord (e.g., C3, G4...), the type (minor or major), the first and following notes (inversions) can easily be chosen. For instance, to construct the first triplet of the melody in Figure 3 (notes G1, E1 and C2), the chord to be chosen is C2 with inversions -1 (giving G1 which is the closest chord note below the tonic), -2 (giving E1 which is the second closest note below the tonic) and 1 (giving C2 which is the tonic). A rhythmic incongruity can be added to any triplet. In our case, this incongruity was only applied to the second note of the last triplet, and the dilation factor was the same for all melodies ($\alpha = 1.7$). The beat of the melody can also be chosen. In this study, we used four different beats: 70, 80, 90, and 100 triplets/minute, so that the inter-onset-interval (IOI) between successive notes varied from 200 ms to 285 ms, with an increase of IOI, due to the rhythmic modifications, that varied from 140 ms to 200 ms.[2] Finally, when all the parameters of the melodies were chosen, the sound sequences were recorded as wave files.

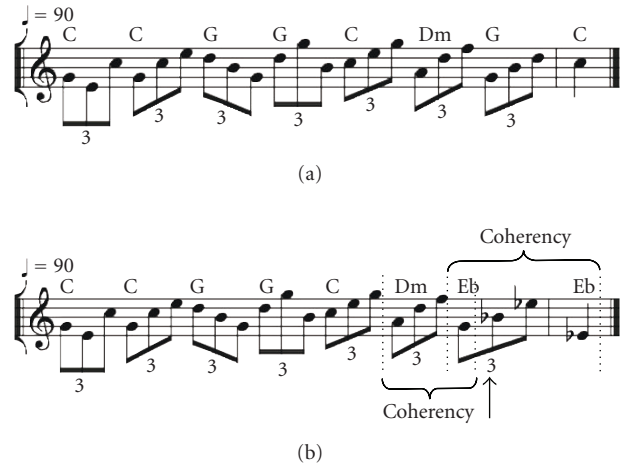

Figure 2: Upper part of the figure corresponds to a harmonically congruous melody, while the lower part corresponds to a harmonically incongruous melody. In the rhythmically incongruous conditions, the duration of the second last notes of the last triplet (indicated by an arrow in the lower part) was increased by a factor 1.7.

2.3. Methods

Subjects

A total of 14 participants (non-musicians, 23-years-old on the average) participated in the language part, of which 8 participated in the music part of the experiment. Volunteers were students from the Aix-Marseille Universities and were paid to participate in the experiments that lasted for about 2 hours. All were right-handed native French speakers, without hearing or neurological disorders. Each experiment began with a practice session to familiarize participants with the task and to train them to blink during the interstimulus interval.

Procedure

In the present experiment, 32 sound examples (sentences or melodies) were presented in each experimental condition, so that each participant listened to 128 different stimuli. To make sure a stimulus was presented only once in the four experimental conditions, 512 stimuli were created to be used either in the language or in the music experiment. Stimuli were presented in 4 blocks of 32 trials.

The experiment took place in a Faradized room, where the participants, wearing an Electro Cap (28 electrodes), listened to the stimuli through headphones. Within two blocks of trials, participants were asked to focus their attention on the metric/rhythmic aspects of the sentences/melodies to decide whether the last syllable/note was metrically/rhythmically acceptable or not. In the other two blocks, participants were asked to focus their attention on the semantic/harmony in order to decide whether the last syllable/note was semantically/harmonically acceptable or not. The responses are given by pressing one of two response buttons as quickly as possible. The side (left or right hand) of the response was balanced across participants.

In addition to the measurements of the electric activity (EEG), the percentage of errors, as well as the reaction times (RTs), were measured. The EEG was recorded from 28 active electrodes mounted on an elastic head cap and located at standard left and right hemisphere positions over frontal, central, parietal, occipital and temporal areas (International 10/20 system sites; Jasper [31]). EEG was digitized at a 250 Hz sampling rate using a 0.01 to 30 Hz band pass. Data were re-referenced off-line to the algebraic average over the left and right mastoids. EEG trials contaminated by eye-, jaw- or head movements, or by a bad contact between the electrode and the skull, were eliminated (approximately 10%). The remaining trials were averaged for each participant within each of the 4 experimental conditions. Finally, a grand average was obtained by averaging the results across all participants.

Error rates and reaction times were analyzed using Analysis of Variance (ANOVAs) that included Attention (Rhythmic versus Harmonic), Harmonics (2 levels) and Rhythmic (2 levels) within-subject factors.

ERP data were analyzed by computing the mean amplitude in selected latency windows, relative to a baseline, and determined both from visual inspection and on the basis of previous results. Analysis of variance (ANOVAs) were used for all statistical tests, and all P-values reported below were adjusted with the Greenhouse-Geisser epsilon correction for non-sphericity. Reported are the uncorrected degrees

[2] A simple statistical study of syllable lengths in the language experiment showed that an average number of around 120 tri-syllabic words per minute were pronounced. Such a tempo was however too fast for the music part.

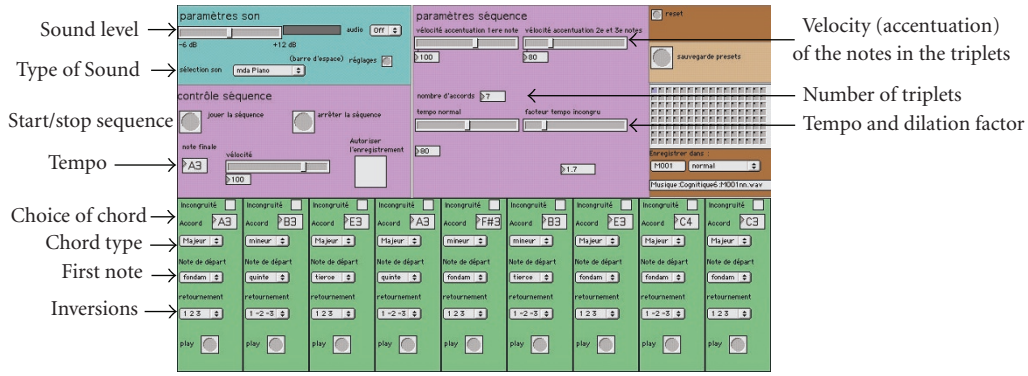

FIGURE 3: Real-time interface (Max/MSP) allowing for the construction of the melodies. In the upper left corner, the sound level is chosen (here constant for all the melodies) and underneath a sequence control allowing to record the melodies suitable for the experiment. In the upper right part, the tempo, number of triplets and the incongruity factor are chosen. Finally, the chords defining each triplet are chosen in the lowest part of the figure.

of freedom and the probability level after correction. Separate ANOVAs were computed for midline and lateral sites separately.

Separate ANOVAs were conducted for the Metric/Rhythmic and Semantic/Harmonic task. Harmony (2 levels), Rhythmic (2 levels) and Electrodes (4 levels) were used as within-subject factors for midline analysis. The factors Harmony (2 levels) and Rhythm (2 levels) were also used for the lateral analyses, together with the factors Hemisphere (2 levels), Anterior-Posterior dimension (3 regions of interest-ROIs): fronto-central (F3, Fc5, Fc1; F4, Fc6, Fc2), temporal (C3, T3, Cp5; C4, T4, Cp6) and temporo-parietal (Cp1, T5, P3; Cp2, T6, P4) and Electrodes (3 for each ROI), as within-subject factors, to examine the scalp distribution of the effects. Tukey tests were used for all post-hoc comparisons. Data processing was conducted with the Brain Vision Anayser software (Version 01/04/2002; Brain Products, Gmbh).

3. RESULTS

3.1. Language experiment

We here summarize the main results of the experiment conducted with the linguistic stimuli, mainly focusing on the acoustic aspects. A more detailed description of these results can be found in (Magne et al. [15]).

3.1.1. Behavioral data

Results of a three-way ANOVA on a transformed percentage of errors showed two significant effects. The meter by semantics interaction was significant ($F(1,12) = 16.37$, $P < .001$): the participants made more errors when one dimension, Meter (19.5%) or Semantics (20%) was incongruous than when both dimensions were congruous (12%) or incongruous (16.5%). The task by meter by semantics interaction was also significant ($F(1,12) = 4.74$, $P < .05$): the participants made more errors in the semantic task when semantics was congruous, but meter was incongruous (S+M−), (24%), than in the other three conditions.

The results of the three-way ANOVA on the RTs showed a main effect of semantics ($F(1,12) = 53.70$, $P < .001$): they always were significantly shorter for semantically congruous (971 ms) than for incongruous words (1079 ms).

3.1.2. Electrophysiological data

Results revealed two interesting points. First, independently of the direction of attention toward semantics or meter, semantically incongruous (but metrically congruous) final words (M+S−) elicited larger N400 components than semantically congruous words (M+S+). Thus, semantic processing of the final word seems task-independent and automatic. This effect was broadly distributed over the scalp.

Second, some aspects of metric processing also seemed task independent because metrically incongruous words also elicited an N400-like component in both tasks (see Figure 4). As opposed to the semantically incongruous case, the meter by hemisphere interaction was almost significant ($P < .06$): the amplitude of the negative component was somewhat larger over the right hemisphere (metrically congruous versus incongruous: $F(1,13) = 15.95$, $P = .001$; $d = -1.69\,\mu$V) than over the left hemisphere (metrically congruous versus incongruous: $F(1,13) = 6.04$, $P = .03$; $d = -1.11\,\mu$V). Finally, a late positivity (P700 component) was only found for metrically incongruous words when participants focused their attention on the metric aspects, which may reflect the explicit processing of the metric structure of words.

No differences in low-level acoustic factors between the metrically congruous and incongruous stimuli were observed. This result is important from an acoustical point of view, since it confirms that no spurious effect due to a non-ecological manipulation of the speech signal has been created by the time-stretching algorithm described in Section 2.1.2.

3.2. Music experiment

3.2.1. Behavioral data

The percentages of errors and the RTs in the four experimental conditions (R+H+, R+H−, R−H+, and R−H−) in

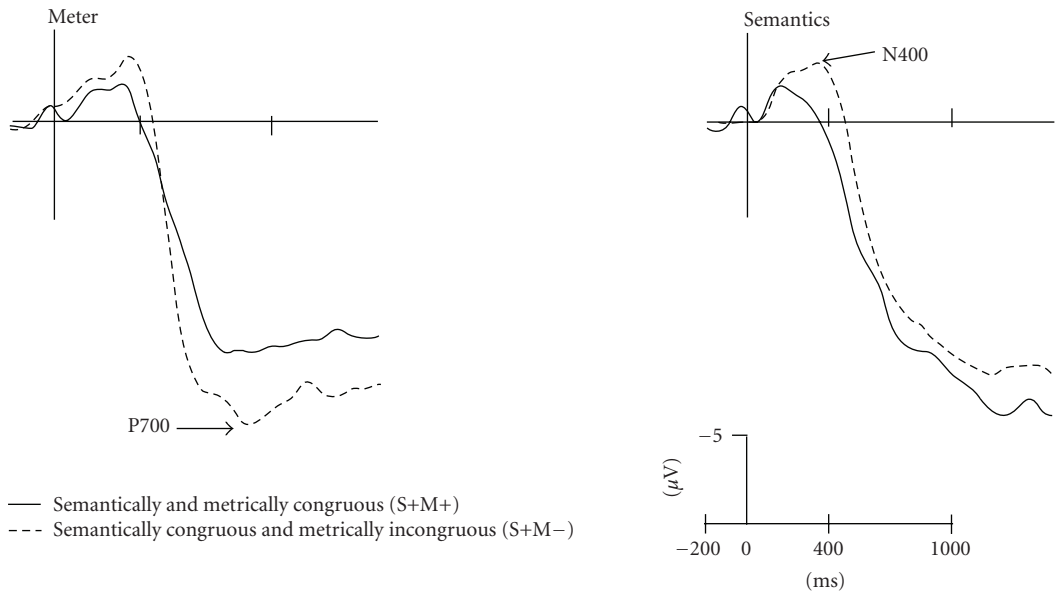

Figure 4: Event-related potentials (ERP) evoked by the presentation of the semantically congruous words when metrically congruous (S+M+) or metrically incongruous (S+M−). Results when participant focused their attention on the metric aspects are illustrated in the left column (Meter) and when they focused their attention on the semantic aspects in the right column (Semantic). The averaged electrophysiological data are presented for one representative central electrode (C_z).

the two attentional tasks (Rhythmic and Harmonic) are presented in Figures 5 and 6.

Results of a three-way ANOVA on the transformed percentages of errors showed a marginally significant main effect of Attention [$F(1,7) = 4.14$, $P < .08$]: participants made somewhat more errors in the harmonic task (36%) than in the rhythmic task (19%). There was no main effect of Rhythmic or Harmonic congruity, but the Rhythmic by Harmonic congruity interaction was significant [$F(1,7) = 6.32$, $P < .04$]: overall, and independent of the direction of attention, participants made more errors when Rhythm was congruous, but Harmony was incongruous (i.e., condition R+H−) than in the other three conditions.

Results of a three-way ANOVA on RTs showed no main effect of Attention. The main effect of Rhythmic congruity was significant [$F(1,7) = 7.69$, $P < .02$]: RTs were shorter for rhythmically incongruous (1213 ms) than for rhythmically congruous melodies (1307 ms). Although a similar trend was observed in relation to Harmony, the main effect of Harmonic congruity was not significant.

3.2.2. Electrophysiological data

The electrophysiological data recorded in the four experimental conditions (R+H+, R+H−, R−H+, and R−H−) in the two tasks (Rhythmic and Harmonic) are presented in Figures 7 and 8. Only ERPs to correct response were analyzed.

Attention to rhythm

In the 200–500 ms latency band, the main effect of Rhythmic congruity was significant at midline and lateral electrodes [Midlines: $F(1,7) = 11.01$, $P = .012$; Laterals: $F(1,7) = 21.36$, $P = .002$]: Rhythmically incongruous notes (conditions R−H+ and R−H−) elicited more negative ERPs than rhythmically congruous notes (conditions R+H+ and R+H−−). Moreover, the main effect of Harmonic congruity was not significant, but the Harmonic congruity by Hemisphere interaction was significant [$F(1,7) = 8.47$, $P = .022$]: Harmonically incongruous notes (conditions R+H− and R−H−) elicited more positive ERPs than harmonically congruous notes (conditions R+H+ and R−H+) over the right than the left hemisphere.

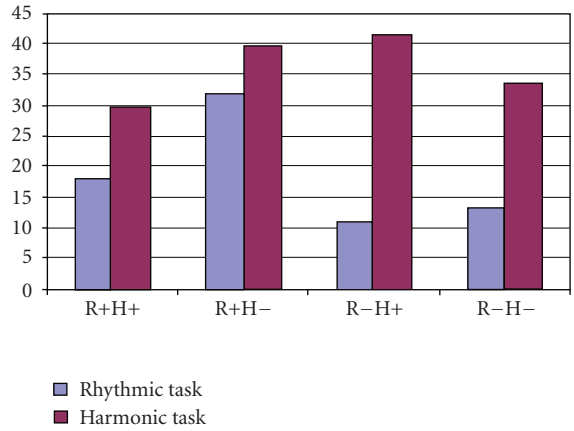

Figure 5: Percentages of error.

In the 500–900 ms latency band, results revealed a main effect of Rhythmic congruity at midline and lateral electrodes [midlines: $F(1,7) = 78.16$, $P < .001$; laterals: $F(1,7) = 27.72$, $P = .001$]: Rhythmically incongruous notes (conditions R−H+ and R−H−) elicited more positive ERPs

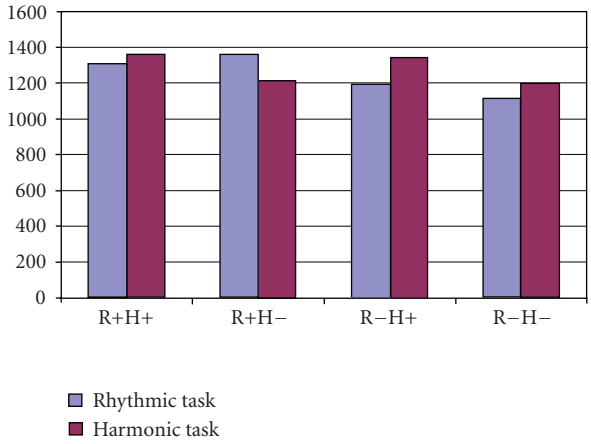

Figure 6: Reaction times (RTs).

than rhythmically congruous notes (conditions R+H+ and R+H−). This effect was broadly distributed over the scalp (no significant rhythmic congruity by Localization interaction). Finally, results revealed no significant main effect of Harmonic congruity, but a significant Harmonic congruity by Localization interaction at lateral electrodes [$F(2,14) = 10.85$, $P = .001$]: Harmonically incongruous notes (conditions R+H− and R−H−) elicited more positive ERPs than harmonically congruous notes (conditions R+H+ and R−H+) at frontal electrodes. Moreover, the Harmonic congruity by Hemisphere interaction was significant [$F(1,7) = 8.65$, $P = .02$], reflecting the fact that this positive effect was larger over the right than the left hemisphere.

Attention to Harmony

In the 200–500 ms latency band, both the main effects of Harmonic and Metric congruity were significant at midline electrodes [$F(1,7) = 5.16$, $P = .05$ and $F(1,7) = 14.88$, $P = .006$, resp.] and at lateral electrodes [$F(1,7) = 5.55$, $P = .05$ and $F(1,7) = 11.14$, $P = .01$, resp.]: Harmonically incongruous musical notes (conditions H−R+ and H−R−) elicited more positive ERPs than harmonically congruous notes (conditions H+R+ and H+R−). By contrast, rhythmically incongruous notes (conditions H+R− and H−R−) elicited more negative ERPs than Rhythmically congruous notes (conditions H+R+ and H−R+). These effects were broadly distributed over the scalp (no Harmonic congruity or Rhythmic congruity by Localization interactions).

In the 500–900 ms latency band, the main effect of Harmonic congruity was not significant, but the Harmonic congruity by Localization interaction was significant at lateral electrodes [$F(2,14) = 4.10$, $P = .04$]: Harmonically incongruous musical notes (conditions H−R+ and H−R−) still elicited larger positivities than harmonically congruous notes (conditions H+R+ and H+R−) over the parieto-temporal sites of the scalp. Finally, results revealed a main effect of Rhythmic congruity at lateral electrodes [$F(1,7) = 5.19$, $P = .056$]: Rhythmically incongruous notes (conditions H+R− and H−R−) elicited more positive ERPs than rhythmically congruous notes (conditions H+R+ and H−R+). This effect was broadly distributed over the scalp (no significant Rhythmic congruity by Localization interaction).

4. DISCUSSION

This section is organized around three main points. First, we discuss the result of the language and music experiments, second we compare the effects of metric/rhythmic and semantic/harmonic incongruities in both experiments, and finally, we consider the advantages and limits of the algorithm that was developed to create ecological, rhythmic incongruities in speech.

4.1. Language and music experiment

In the language part of the experiment, two important points were revealed. Independently of the task, semantically incongruous words elicited larger N400 components than congruous words. Longer RTs are also observed for semantically incongruous than congruous words. These results are in line with the literature and are usually interpreted as reflecting greater difficulties in integrating semantically incongruous compared to congruous words in ongoing sentence contexts (Kutas and Hillyard [32]; Besson et al. [33]). Thus participants seem to process the meaning of words even when instructed to focus attention on syllabic duration. The task independency results are in line with studies of Astésano (Astésano et al. [34]), showing the occurrence of N400 components independently of whether participants focused their attention on semantic or prosodic aspects of the sentences. The second important point of the language experiment is related to the metric incongruence. Independently of the direction of attention, metrically incongruous words elicit larger negative components than metrically congruous words in the 250–450 ms latency range. This might reflect the automatic nature of metric processing. Such early negative components have also been reported in the literature when controlling the influence of acoustical factors as prosody. In a study by Magne (Magne et al. [35]), a N400 component was observed when prosodically incongruous final sentence words were presented. This result might indicate that the violations of metric structure interfere with lexical access and thereby hinder access to word meaning. Metric incongruous words also elicited late positive components. This is in line with previous findings indicating that the manipulation of different acoustic parameters of the speech signal such as F0 and intensity, is associated with increased positivity (Astésano et al. [34], Magne et al. [35], Schön et al. [9]).

In the music part of the experience, analysis of the percentage of errors and RTs revealed that the harmonic task was somewhat more difficult than the rhythmic task. This may reflect the fact, pointed out by the participants at the end of the experiment, that the harmonic incongruities could be interpreted as a change in harmonic structure possibly continued by a different melodic line. This interpretation is coherent with the high error rate in the harmonically incongruous, but rhythmically congruous condition (R+H−) in both attention tasks. Clearly, harmonic incongruities seem

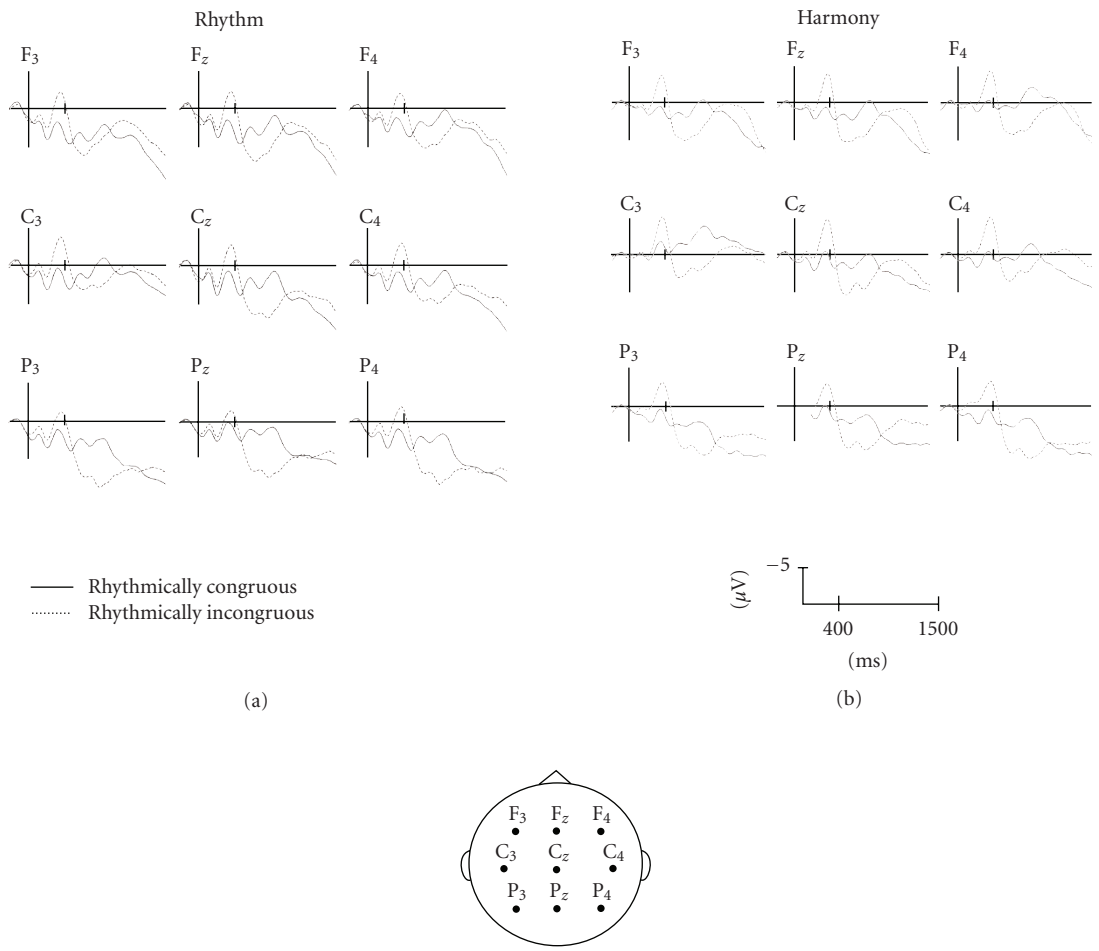

FIGURE 7: Event-related potentials (ERPs) evoked by the presentation of the second note of the last triplet when rhythmically congruous (solid trace; conditions H+R+ and H−R+) or rhythmically incongruous (dashed trace, conditions H+R− and H−R−). Results when participant focused their attention on the rhythmic aspects are illustrated in the left column (a) and when they focused their attention on the harmonic aspects in the right column (b). 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).

more difficult to detect than rhythmic incongruities. Finally, RTs were shorter for rhythmically incongruous than congruous notes, probably because participants in the last condition waited to make sure the length of the note was not going to be incongruous.

Interestingly, while rhythmic incongruities elicited an increased negativity in the early latency band (200–500 ms), harmonic incongruities were associated with an increased positivity. Most importantly, these differences were found independently of whether participants paid attention to rhythm or to harmony. Thus, different processes seem to be involved by the rhythmic and harmonic incongruities and these processes seem to be independent of the task at hand. By contrast, in the later latency band (500–900 ms) both types of incongruities elicited increased positivities compared to congruous stimuli. Again, these results were found independently of the direction of attention. Note, however, that the scalp distribution of the early and late positivity to harmonic incongruities differs depending upon the task: while it was larger over right hemisphere in the rhythmic task, it was largely distributed over the scalp and somewhat larger over the parieto-temporal regions in the harmonic task. While this last finding is in line with many results in the literature (Besson and Faïta [36]; Koelsch et al. [13, 37]; Patel et al. [12]; Regnault et al. [14]), the right distribution is more surprising. It raises the interesting possibility that the underlying process varies as a function of the direction of attention, a hypothesis that already has been proposed in the literature (Luks et al. [38]). When harmony is processed implicitly, because irrelevant for the task at hand (Rhythmic task), the right hemisphere seems to be more involved, which is in line with brain imaging results showing that pitch processing seems to be lateralized in right frontal regions (e.g., Zatorre et al. [39]). By contrast, when harmony is processed explicitly (Harmonic task), the typical centro-parietal distribution is found which may reflect the influence of decision-related processes. Taken together, these results are important because they show that different processes are responsible for the processing of rhythm and harmony when listening to the short musical sequences used here. Moreover, they open the

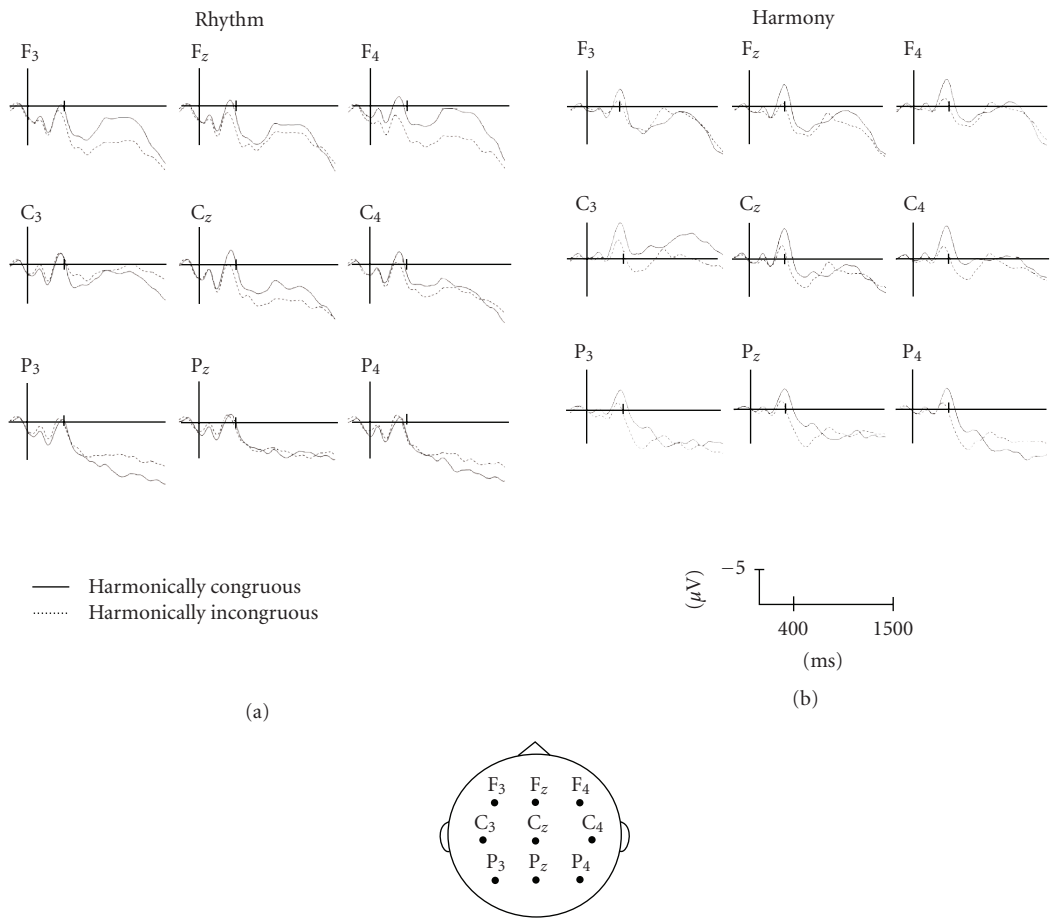

Figure 8: Event-related potentials (ERPs) evoked by the presentation of the second note of the last triplet when harmonically congruous (solid trace, conditions H+R+ and H−R+) or harmonically incongruous (dashed trace; conditions H+R− and H−R−). Results when participant focused their attention on the rhythmic aspects are illustrated in the left column (a) and when they focused their attention on the harmonic aspects in the right column (b).

intriguing possibility that the distribution of these processes vary as a function of attention. These issues clearly need to be pursued in future experiments.

4.1.1. Comparison between language and music

These results on the processing of rhythm in both speech and music 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 rhythmic incongruities elicited increased positive deflections 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 rhythmic incongruity elicited positive components that were preceded by negative components in music, under both attentional conditions, no such early negativities were found in language in the rhythmic task. Thus, rhythmic violations seem to elicit earlier and larger effects in music than in language, which points to a larger influence on rhythm in music than in speech. Finally, one important difference between speech and music is that while semantic incongruities elicited clear N400 components (Kutas and Hillyard [32]) in the language experiment (Magne et al. [15]), as expected from a large body of results in the ERPs and language litterature, no such N400s were associated to the presentation of harmonic incongruities in the music experiment reported here. While semantic processing may not be restricted to linguistic stimuli as demonstrated by (Koelsch et al. [11]) with the occurrence of an N400 component to words that did not match the musical content of the preceding musical phrase, it nevertheless "remains" that the type of harmonic violations used here did not seem to involve semantic processing. Therefore, the intriguing issue of musical semantic processing, remains open for future research.

4.1.2. Algorithm

The time-stretching algorithm that was adapted to dilation of syllables in the language experiment, allowed up to 400% time dilation on the vowel part of the speech signals. Most importantly for our purpose, and in spite of these high stretching ratios, the application of the algorithm did not induce any modifications in sound quality as evidenced by the typical structure of the electrophysiological data in this experiment. No spurious effect, due to a non-ecological manipulation of the speech signal (producing differences in low-level acoustic factors between the rhythmically congruous and incongruous stimuli) was observed in the ERPs. These results are important from an acoustic point of view, since they show the ecological validity of the time-stretching algorithm described in section 2.1.3.

Taken together, from an interdisciplinary perspective, our results demonstrate the influence of metrical structure in language and its important consequences for lexical access and word meaning. More generally, they highlight the role of lexical prosody in speech processing. These are important results from an audio signal processing point of view, since they imply that dilating signals can affect the comprehension of the utterances. Speech sequences in an audio signal must therefore be modified with care and the context should be taken into account, meaning that, for instance, sequences containing important information or sequences with a lot of expressiveness should be less modified than other sequences.

In the music part of the experiment, MIDI codes were used to modify the note duration. The algorithm could have been applied for this purpose, and results of this type are important for future applications of the algorithm on musical signals. Finally, note that our findings imply that the rhythmic modifications should be applied to music and language according to different rules. In the language part, semantics and context should be taken into account, as mentioned earlier, while in the music part interpretation rules according to instrument types and musical genre should determine eventual modifications of sound sequences.

ACKNOWLEDGMENTS

We would like to thank Thierry Voinier for developing the MAX/MSP patch used for constructing the melodies in the music part of the experiment. This research was partly supported by a grant from the Human Frontier Science Program (HSFP #RGP0053) to Mireille Besson and a grant from the French National Research Agency (ANR, JC05-41996, "sensons") to Sølvi Ystad. Mitsuko Aramaki was supported by a post-doctoral grant from the HFSP and the ANR.

REFERENCES

[1] A. Friberg and J. Sundberg, "Time discrimination in a monotonic, isochronous sequence," *Journal of the Acoustical Society of America*, vol. 98, no. 5, pp. 2524–2531, 1995.

[2] C. Drake and M. C. Botte, "Tempo sensitivity in auditory sequences: Evidence for a multiple-look model," *Perception and Psychophysics*, vol. 54, pp. 277–286, 1993.

[3] I. J. Hirsh, C. B. Monahan, K. W. Grant, and P. G. Singh, "Studies in auditory timing : I, simple patterns," *Perception and Psychophysics*, vol. 74, no. 3, pp. 215–226, 1990.

[4] G. ten Hoopen, L. Boelaarts, A. Gruisen, I. Apon, K. Donders, N. Mul, and S. Aker-boom, "The detection of anisochrony in monaural and interaural sound sequences," *Perception and Psychophysics*, vol. 56, no. 1, pp. 210–220, 1994.

[5] M. Barthet, R. Kronland-Martinet, S. Ystad, and Ph. Depalle, "The effect of timbre in clarinet interpretation," in *Proceedings of the International Computer Music Conference (ICMC '07)*, Copenhagen, Denmark, August 2007.

[6] M. Besson, F. Faïta, C. Czternasty, and M. Kutas, "What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences," *Biological Psychology*, vol. 46, no. 1, pp. 3–23, 1997.

[7] A. D. Patel and J. R. Daniele, "An empirical comparison of rhythm in language and music," *Cognition*, vol. 87, no. 1, pp. B35–B45, 2003.

[8] C. Magne, D. Schön, and M. Besson, "Prosodic and melodic processing in adults and children: behavioral and electrophysiologic approaches," *Annals of the New York Academy of Sciences*, vol. 999, pp. 461–476, 2003.

[9] D. Schön, C. Magne, and M. Besson, "The music of speech: music training facilitates pitch processing in both music and language," *Psychophysiology*, vol. 41, no. 3, pp. 341–349, 2004.

[10] M. Besson and F. Macar, "An event-related potential analysis of incongruity in music and other non-linguistic contexts," *Psychophysiology*, vol. 24, no. 1, pp. 14–25, 1987.

[11] S. Koelsch, E. Kasper, D. Sammler, K. Schulze, T. Gunter, and A. D. Friederici, "Music, language and meaning: brain signatures of semantic processing," *Nature Neuroscience*, vol. 7, no. 3, pp. 302–307, 2004.

[12] A. D. Patel, E. Gibson, J. Ratner, M. Besson, and P. J. Holcomb, "Processing syntactic relations in language and music: an event-related potential study," *Journal of Cognitive Neuroscience*, vol. 10, no. 6, pp. 717–733, 1998.

[13] S. Koelsch, T. Gunter, A. D. Friederici, and E. Schröger, "Brain indices of music processing: "nonmusicians" are musical," *Journal of Cognitive Neuroscience*, vol. 12, no. 3, pp. 520–541, 2000.

[14] P. Regnault, E. Bigand, and M. Besson, "Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: evidence from auditory event-related brain potentials," *Journal of Cognitive Neuroscience*, vol. 13, no. 2, pp. 241–255, 2001.

[15] C. Magne, C. Astésano, M. Aramaki, S. Ystad, R. Kronland-Martinet, and M. Besson, "Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence," *Cerebral Cortex*, 2007, Oxford University Press, January 2007.

[16] C. Astésano, *Rythme et accentuation en français: Invariance et variabilité stylistique*, Collection Langue & Parole, L'Harmattan, Paris, France, 2001.

[17] A. Di Cristo, "Le cadre accentuel du français contemporain: essai de modélisation: premiére partie," *Langues*, vol. 2, no. 3, pp. 184–205, 1999.

[18] G. Pallone, *Dilatation et transposition sous contraintes perceptives des signaux audio: application au transfert cinéma-vidéo*, Ph.D. thesis, University of Aix-Marseille II, Marseilles, France, 2003.

[19] M. Dolson, "The phase vocoder: a tutorial," *Computer Music Journal*, vol. 10, no. 4, pp. 14–27, 1986.

[20] G. Pallone, P. Boussard, L. Daudet, P. Guillemain, and R. Kronland-Martinet, "A wavelet based method for audio-video synchronization in broadcasting applications," in *Proceedings of the 2nd COST-G6 Workshop on Digital Audio Effects (DAFx '99)*, pp. 59–62, Trondheim, Norway, December 1999.

[21] M. Puckette, "Phase-locked vocoder," in *Proceedings of IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 222–225, New Paltz, NY, USA, October 1995.

[22] J. Laroche and M. Dolson, "Improved phase vocoder time-scale modification of audio," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 3, pp. 323–332, 1999.

[23] N. R. French and M. K. Zinn, "Method of an apparatus for reducing width of trans-mission bands," US patent no. 1,671,151, May 1928.

[24] S. Roucos and A. Wilgus, "High quality time-scale modification for speech," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '85)*, vol. 10, pp. 493–496, Tampa, Fla, USA, April 1985.

[25] W. Verhelst and M. Roelands, "An overlap-add technique based on waveform similarity (WSOLA) for high quality time-scale modification of speech," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '93)*, vol. 2, pp. 554–557, Minneapolis, Minn, USA, April 1993.

[26] D. J. Hejna, B. T. Musicus, and A. S. Crowe, "Method for time-scale modification of signals," US patent no. 5,175,769, December 1992.

[27] J. Laroche, "Time and pitch scale modification of audio signals," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., pp. 279–309, Kluwer Academic Publishers, Norwell, Mass, USA, 1998.

[28] B. Repp, "Probing the cognitive representation of musical time: structural constraints on the perception of timing perturbations," *Haskins Laboratories Status Report on Speech Research*, vol. 111-112, pp. 293–320, 1992.

[29] B. Moog, "MIDI: musical instrument digital interface," *Journal of Audio Engineering Society*, vol. 34, no. 5, pp. 394–404, 1986.

[30] M. S. Puckette, T. Appel, and D. Zicarelli, "Real-time audio analysis tools for Pd and MSP," in *Proceedings of the International Computer Music Conference*, pp. 109–112, International Computer Music Association, Ann Arbor, Mich, USA, October 1998.

[31] H. H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalography and Clinical Neurophysiology*, vol. 10, pp. 371–375, 1958.

[32] M. Kutas and S. A. Hillyard, "Reading senseless sentences: brain potentials reflect semantic incongruity," *Science*, vol. 207, no. 4427, pp. 203–205, 1980.

[33] M. Besson, C. Magne, and P. Regnault, "Le traitement du langage," in *L'imagerie fonctionnelle électrique (EEG) et magnétique (MEG): Ses applications en sciences cognitives*, B. Renault, Ed., pp. 185–216, Hermés, Paris, France, 2004.

[34] C. Astésano, M. Besson, and K. Alter, "Brain potentials during semantic and prosodic processing in French," *Cognitive Brain Research*, vol. 18, no. 2, pp. 172–184, 2004.

[35] C. Magne, C. Astésano, A. Lacheret-Dujour, M. Morel, K. Alter, and M. Besson, "On-line processing of "pop-out" words in spoken French dialogues," *Journal of Cognitive Neuroscience*, vol. 17, no. 5, pp. 740–756, 2005.

[36] M. Besson and F. Faïta, "An event-related potential (ERP) study of musical expectancy: comparison of musicians with non-musicians," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 21, no. 6, pp. 1278–1296, 1995.

[37] S. Koelsch, T. Gunter, E. Schröger, and A. D. Friederici, "Processing tonal modulations: an ERP study," *Journal of Cognitive Neuroscience*, vol. 15, no. 8, pp. 1149–1159, 2003.

[38] T. L. Luks, H. C. Nusbaum, and J. Levy, "Hemispheric involvement in the perception of syntactic prosody is dynamically dependent on task demands," *Brain and Language*, vol. 65, no. 2, pp. 313–332, 1998.

[39] R. J. Zatorre, "Neural specializations for tonal processing," in *The Biological Foundations of Music*, R. J. Zatorre and I. Peretz, Eds., vol. 930 of *Annals of the New York Academy of Sciences*, pp. 193–210, New York Academy of Sciences, New York, NY, USA, June 2001.