

The Music of Speech: Electrophysiological Approach

Schön D.^{1,2}, Magne C.², Schrooten M.³ & Besson M.²

¹ Dipartimento di Psicologia, Università di Trieste, Italie

² CRNC-CNRS, Marseille, France

³ Faculty of Psychology, Department of Neurocognition, Maastricht University, The Netherlands

danschon@lnf.cnrs-mrs.fr

Abstract

This experiment was aimed at studying the changes in the brain electrical activity associated with some aspects of prosodic processing and at comparing the results with those obtained for the melodic processing of short musical phrases, by manipulating the F0 of the final word/note. Both musicians and non-musicians were presented with French sentences, ending with prosodically congruous, weakly or strongly incongruous words, and with short musical phrases ending with congruous, weakly or strongly incongruous notes. Both Reaction Times (RTs) and Evoked Related Potentials (ERPs) were recorded. In both the prosodic and melodic conditions, RTs were shorter for strong incongruities than for weak incongruities and congruous completions. ERP recorded in both conditions showed that large positive components (P600s) were associated with incongruous stimuli. Moreover, in both cases, the amplitude of the P600 component was larger, and its latency was shorter, for the strong than for the weak incongruities. Interestingly, an early negative component (N150) develop in response to strong incongruities. For the musicians, this N150 component is clearly localized over right temporo-frontal electrodes, specifically in the melodic condition. For non-musicians, the N150 is more broadly distributed across scalp sites. These results point to strong similarities in the processing of prosodic and melodic incongruities, and a general cognitive process may be at play in both cases. However, they also point to some early differences which may reflect the effects of musical expertise.

1. Introduction

Some aspects of language perception and comprehension, such as semantic and syntactic processing, have been extensively studied using the Event-Related brain Potentials (ERP) method. This method allows to record the changes in the brain electrical activity that are time-locked to the presentation of an event of interest [1]. It has an excellent temporal resolution, so that it is possible to determine when the ERPs recorded in two experimental conditions, and by inference the processes involved in the two conditions, start to diverge. Previous ERP studies of semantic processing have, for example, shown that the ERPs associated with sentence final congruous and incongruous words start to diverge around 200 ms, in the visual modality, with a larger negative component, peaking around 400 ms after final word onset, to incongruous than congruous words: the N400 component (e.g., [6]). In contrast, research on syntactic processing has shown that violations of word order within a sentence or agreement violations, are associated with a positive component peaking around 600 ms post-final word onset ([7] ; [2]). These results, that seem to indicate that semantic

processing occurs earlier than syntactic processing, have led to hot, and yet unsolved, debates in the literature.

Mainly because of methodological problems, very few studies have examined prosodic processing using the ERP method. However, the recent development of powerful computers, linked with an increased quality of synthetic speech, now allows to apply the ERP method to the study of some aspects of prosodic processing. Prosody is defined as the melodic variations of the voice, the music of speech, the stress and intonation patterns, as well as the pauses that help to underline or clarify the meaning of sentences [3]. As such, prosodic processing certainly is an essential aspect of language perception and comprehension.

Few experiments have recently been aimed at studying the relationships between prosodic processing, on one side, and syntactic or semantic processing, on the other side. Steinhauer et al [8] have for instance shown that the position of pauses in syntactically ambiguous sentences determines the occurrence of a positive component that they called the "closure positive shift" or CPS. Their results therefore demonstrate that prosodic cues such as pauses are processed on-line to clarify ambiguous sentences. [9] ; [5]). In a series of experiments, Astesano, Alter et Besson (in preparation) have presented semantically and prosodically congruous or incongruous sentences to participants that were asked to focalize their attention on the semantic aspects of speech in order to detect the semantic violations, or to focalize their attention on the prosodic aspects of speech in order to detect the prosodic violations. Results showed that while an N400 component was associated with semantic violations, late positive components, peaking around 800 ms post-final word, were associated with prosodically incongruous words. Moreover, results also indicated that while participants were able to focus attention on the semantics of speech and ignore the prosody, the reverse was not possible. Thus, participants could not help but notice that the final word of the sentence was semantically incongruous even when they were asked to focus attention only on the prosody.

The aim of the present study was three-fold. First, in order to replicate and extend the results reported above we wanted to examine the changes in the brain electrical activity associated with other aspects of prosodic processing, that is, the intonation contour of the final word. Second, it was of interest to compare prosodic processing with melodic processing of short musical phrases. Thus, in both conditions, prosodic and melodic, we manipulated the F0 of the final word/note. For sentences, the intonation contour of the final word was congruous, weakly incongruous or strongly incongruous. Similarly, for musical phrases, the pitch of the final note was congruous, weakly or strongly incongruous. Finally, it was important to examine the influence of musical expertise on prosodic processing. Therefore, sentences and

musical phrases were presented to both musicians and non musicians who were asked to determine whether the final word or the final note was congruous or incongruous in the context.

2. Material

2.1. Participants

Eighteen participants, nine musicians and 9 non-musicians (mean age: 31 years old), were paid to participate in the experiment that lasted for about 2 hours. All were right-handed, had normal audition and were native speakers of French.

2.2. Linguistic and musical material

A total of 120 sentences, taken from children's book, and ended by bisyllabic words (e.g., « *Dans la barque se tient l'ennemi de Peter Pan, le terrible pirate* » ; « *In the boat is the enemy of Peter Pan, the pirate* ») were selected for this experiment. To introduce some variability, the sentence final word was preceded by an adjective in half of the sentences (e.g., « *terrible pirate* »). Sentences were spoken at a normal speech rate by a native French female speaker, recorded and synthesized using the computer program *Winpitch*.

A total of 120 musical phrases were also used in the experiment. Half were selected from the children's repertoire (e.g., « Happy birthday to you » ...) and half were composed for the experiment by a professional musician following the same rules of composition as for familiar melodies. Musical phrases were recorded using the synthetic sound of a piano (KORG XDR5).

For both the prosodic and melodic materials, the experimental design included 3 conditions in which the final word or note was prosodically or melodically: 1) congruous, 2) weakly incongruous (35 % increase of the F0 of the final word and 1/5 of a tone increase of the pitch of the final note, 3) strongly incongruous (125% increase of the F0 of the final word and 1/2 tone increase of the pitch of the final note).

2.3. Procedure

Within an experimental session, participants were presented with two blocks of sentences, each comprising 120 sentences, and two block of musical phrases, each comprising 120 melodies. Block order was counterbalanced across participants so that some participants listened to the sentences first, while others listened to the musical phrases first. Participants were asked to pay attention to the linguistics or musical materials in order to decide whether the final word or note was congruous or incongruous. When the final stimulus was presented, they were required to press a response button as quickly and as accurately as possible. To familiarize the participants with the task, the materials and the response buttons, the experiment started with a practice session comprising 6 sentences and 6 musical phrases, with two trials in each of the 3 experimental conditions.

2.4. ERP recording

EEG was recorded for 2200 ms starting 200 ms before the onset of the stimulus, from 28 scalp electrodes, mounted on an elastic cap, and located according to the International 10/20 system [4]. These recording sites, plus an electrode placed on

the right mastoid, were referenced to the left mastoid electrode; the data were then re-referenced offline to the algebraic average of the left and right mastoids. Impedances of these electrodes never exceeded 3 kW. In order to detect blinks and vertical eye movements, the horizontal electrooculogram (EOG) was recorded from electrodes placed 1 cm to the left and right of the external canthi, and the vertical EOG was recorded from an electrode beneath the right eye, referenced to the left mastoid. Trials containing ocular artefacts, movement artefacts, or amplifier saturation were excluded from the averaged ERP waveforms. The EEG and EOG were amplified by a SA Instrumentation amplifier with a bandpass of 0.01-30 Hz, and were digitised at 250 Hz by a PC-compatible microcomputer (Compaq Prosignia 486).

ERP data were analysed by computing the mean amplitude in selected latency windows relative to a 200 ms pre-stimulus baseline. Analysis of variance (ANOVA) was 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 of freedom and the probability level after correction.

3. Results

3.1. Behavioral data

Overall RTs were shorter for prosodic than for melodic materials. Moreover, in both cases, musicians as well as non musicians were faster to detect weak and strong incongruities than to decide that the final word / note was congruous. They were also faster for the strong than for the weak incongruities (see Table 1).

Table 1: Reaction times in milliseconds (msec) for both musicians and non musicians listening to the prosody or melody of sentences/musical phrases in each condition.

Conditions	Musicians		Non-musicians	
	Prosody	Melody	Prosody	Melody
Congruous	1006	1232	926	1141
Weak Incongruity	935	853	904	1113
Strong Incongruity	784	864	798	869

3.2. Electrophysiological data

See Figure 1 for a summary of results.

3.2.1. Prosodic processing

Final words are associated with positive components in all conditions (P600s). The amplitude of the P600 component is larger for strongly incongruous words than for weakly incongruous words or congruous words ($p < .05$). Moreover, the latency of the P600 greatly differs across conditions: it is shorter for the strong than for the weak incongruities. Finally, it should be noted that these differences are larger for musicians than for non-musicians.

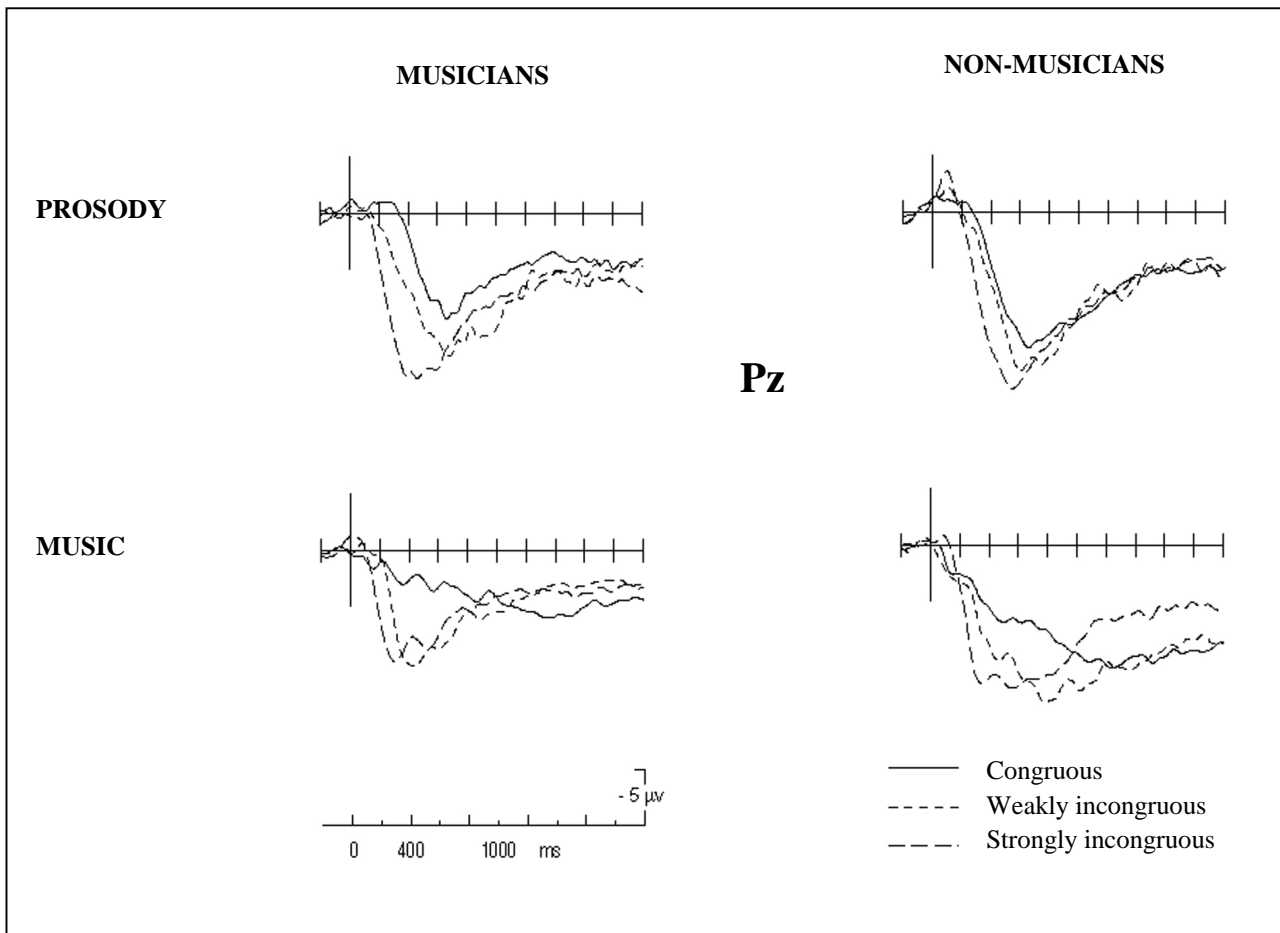


Figure 1 : ERPs evoked by the presentation of the last word/note in each condition. Data are presented for the centro-parietal electrode. The amplitude is on the ordinate, time on the abscissa (msec). Negativity is up.

Interestingly, an early negative component, peaking around 130 ms post-final word onset (N130), develop over temporo-anterior sites in response to the strong incongruity. Its scalp distribution is more localized for musicians than for non-musicians.

3.2.2. Melodic processing

Final notes are also associated with the occurrence of positive components (P600s). Again the amplitude of the P600 component, and its latency, is larger / shorter for the strongly incongruous note than for the weakly incongruous note and for the congruous note ($p < .05$). Finally, as found with the prosodic materials, these differences are larger for musicians than for non-musicians. Again, an early negativity (N150) is associated with the presentation of the strong incongruities over right temporo-anterior sites, and its scalp distribution is clearly more localized for musicians than for non-musicians.

3.2.3. Comparison between prosody and melody

Both prosodic and melodic processing elicit positive components in the 180 – 800 ms range post-final stimulus onset ($p < .05$). The amplitude and latency of these positive components are very sensitive to the type of incongruities. Moreover, early negative components also develop in both cases in response to strong incongruities. However, the

amplitude of the N130 seems larger to the music materials than to the linguistic materials. The scalp distribution of this negative component is more localized for the musical phrases than for the linguistic sentences.

4. Discussion

These results, that are still preliminary, tend to underline the similarities between prosodic and melodic processing. Indeed, in both cases, weak and strong violations are associated with the occurrence of positive components. The finding that similar ERP effects are associated with prosodic and melodic processing would indicate that similar processes are called into play in both cases. Thus, F0 manipulations seem to be processed similarly when they are included in speech or in music.

It is difficult to determine whether these positive components should be called P300, P600 or P800 because their latency is quite different across experimental conditions. Thus, for musicians, the positivity peaks at 275 ms after strong musical violations and at 350 ms after strong prosodic violations. For non-musicians, the positivity peaks at 320 ms after strong musical violations and at 380 ms after strong prosodic violations. Therefore, it seems that melodic processing was faster than prosodic processing. It may be that prosodic processing takes intrinsically longer than melodic processing or that the prosodic incongruities were more

difficult to detect. However, in contrast with the latency analyses based on the ERP data, behavioural analyses showed that RTs were faster for the prosodic than melodic materials. It may be that the sensory/perceptual processing of the melodic materials is faster but that the discrimination between the congruous and weak incongruity is more difficult for the melodic than the prosodic materials. Detailed analyses are still in progress to further test this interpretation.

It is important to note that musicians were overall faster than non musicians. While this last result was expected, since musical expertise should facilitate the detection of incongruous notes, it is interesting that musicians were also faster in detecting the prosodic incongruities. It is as if musicians have developed a “musical ear” that allow them to detect prosodic violations faster than non musicians.

Again underlying the similarities between the processing of F0 violations in the language and in the music, a negative component develop in response to strong incongruities in both cases. However, while this N130 is present for both musicians and non musicians, its scalp distribution is very different in both groups. For musicians it is clearly localized over right temporo-anterior sites, while it is spread across scalp sites for non-musicians. It is as if musical expertise leads to the specialization of some brain areas for some aspects of music processing.

Taken together these results clearly underline the similarities, more than the differences between prosodic and melodic processing. It should be kept in mind, however, that one specific parameter of the prosodic and melodic contour of sentences and musical phrase was manipulated in the experiment, namely the F0. Therefore, it will be of interest in future experiments to determine whether similar results are found when other parameters such as intensity or duration are manipulated. Finally, results also point to the influence of musical training, both in the efficiency and specialization for musical processing, but also, and this may open new lines of research for some aspects of language processing as prosody.

5. References

- [1] Berger, H., 1929. Uber das Electrenkephalogramm das menchen. *Archiv für Psychiatrie*, 87, 527-570.
- [2] Friederici, A.D., 1998. The neurobiology of language comprehension. In A.D. Friederici (Ed.) *Language comprehension: A biological approach*, 263-301. Springer, New York.
- [3] Hirst, D.J.; Di Cristo, A., 1998. In D.J. Hirst, & A. Di Cristo (Eds), *Intonation systems: a survey of twenty languages*, 1-44. Cambridge University Press.
- [4] Jasper, H., 1958. The 10-20 electrode system of the International Federation. *Electroencephalographie and Clinical Neurophysiology*, 10, 370-375.
- [5] Jescheniak, J.D.; Hahne, A.; Friederici, A.D., 1998. Brain activity patterns suggest prosodic influences on syntactic parsing in the comprehension of spoken sentences. *Music Perception*, 16 (1), 55-62.
- [6] Kutas, M.; Hillyard, S. A., 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- [7] Osterhout, L.; Holcomb, P.J., 1992. Event-Related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language* 31, 785-804.
- [8] Steinhauer, K.; Alter, K.; Friederici, A.D., 1999. Brain potentials indicate immediate use of prosodic cues in

natural speech processing. *Nature Neuroscience*, 2 (2), 191-196.