

Musicians Detect Pitch Violation in a Foreign Language Better Than Nonmusicians: Behavioral and Electrophysiological Evidence

Carlos Marques¹, Sylvain Moreno², São Luís Castro¹,
and Mireille Besson²

Abstract

■ The aim of this study was to determine whether musical expertise influences the detection of pitch variations in a foreign language that participants did not understand. To this end, French adults, musicians and nonmusicians, were presented with sentences spoken in Portuguese. The final words of the sentences were prosodically congruous (spoken at normal pitch height) or incongruous (pitch was increased by 35% or 120%). Results showed that when the pitch deviations were small and difficult to detect (35%: weak prosodic incongruities), the level of performance was higher for musicians than for nonmusicians. Moreover, analysis of the time course of pitch processing, as revealed by the event-related brain potentials to the prosodically congruous and incongruous

sentence-final words, showed that musicians were, on average, 300 msec faster than nonmusicians to categorize prosodically congruous and incongruous endings. These results are in line with previous ones showing that musical expertise, by increasing discrimination of pitch—a basic acoustic parameter equally important for music and speech prosody—does facilitate the processing of pitch variations not only in music but also in language. Finally, comparison with previous results [Schön, D., Magne, C., & Besson, M. The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, 41, 341–349, 2004] points to the influence of semantics on the perception of acoustic prosodic cues. ■

INTRODUCTION

An intuitive and widespread idea is that musicians are better than nonmusicians at learning foreign languages. Recent advances in the field of the neuroscience of music provide some indirect evidences related to this claim. First, musical expertise seems to influence the anatomo-functional organization of the brain. Research using magnetic resonance imaging has revealed morphological differences between musicians and nonmusicians in the Heschl's gyrus and secondary auditory cortex, motor and visuospatial brain areas (Gaser & Schlaug, 2003; Schneider et al., 2002), as well as in the size of the corpus callosum and planum temporale (Schlaug, Jancke, Huang, Staiger, & Steinmetz, 1995; Schlaug, Jancke, Huang, & Steinmetz, 1995). Moreover, at the functional level, research using functional magnetic resonance imaging and magnetoencephalography reveals that professional and/or amateur musicians show larger activation than nonmusicians in several brain regions, including the Heschl's gyrus (Schneider et al.,

2002), the planum temporale bilaterally (Ohnishi et al., 2001), the frontal operculum, and the anterior part of the superior temporal gyrus (STG; Koelsch et al., 2005), and increased somatosensory and motor representations with musical practice (Pantev et al., 1998; Elbert, Pantev, Wienbruch, Rockstroh, & Taub, 1995).

Second, and perhaps most importantly, these regions are not necessarily specifically devoted to music processing but are also involved in different types of perceptual and cognitive processing. For instance, the Heschl's gyrus, the secondary auditory cortex, the planum temporale, and the anterior part of the STG are known to be involved in different aspects of language processing (Koelsch, 2005; Koelsch & Siebel, 2005; Levitin & Menon, 2003; Griffiths & Warren, 2002; Meyer, Alter, Friederici, Lohmann, & von Cramon, 2002; Tzourio et al., 1997; and see Vigneau et al., 2006 for a meta-analysis of left hemisphere languages areas). Moreover, a key region for language processing, Broca's area, has been shown to be activated not only by syntactic processing of linguistic phrases but also by syntactic processing of musical phrases (Knoesche et al., 2005; Koelsch et al., 2002, 2005; Tillmann, Janata, & Bharucha, 2003; Maess, Koelsch, Gunter, & Friederici, 2001). Consequently, musical expertise, by increasing activation in brain regions

¹Universidade do Porto (FPCE-UP), Portugal, ²CNRS—Marseille Université, France

that may also be involved in language processing, may facilitate some aspects of language processing.

Because speech processing is extremely fast, high temporal resolution methods are needed to study on-line speech comprehension. Thus, the event-related brain potentials (ERPs) method, known for its excellent temporal resolution, is particularly well suited to this aim. Here, we used the ERPs to determine whether musicians are more efficient than nonmusicians to perceive prosody, a central aspect of spoken language comprehension (Frazier, Carlson, & Clifton, 2006). At the acoustic level, prosody is defined by four acoustic parameters: fundamental frequency (F0), duration, intensity, and spectral characteristics. At the phonological level, prosody can be defined by stress and intonation patterns. For instance, in English, as in Portuguese, most words carry a stress accent on a particular syllable, whose position within the word varies. In French, by contrast, word-final syllables are most likely to carry the stress accents (Astésano, 2001). In these languages, and many others, words are grouped into prosodic units or intonational phrases. Finally, at the functional level, prosody has both emotional and linguistic functions that have recently been studied using ERPs (for studies of the emotional function of prosody, see Kotz et al., 2003; Besson, Magne, & Schön, 2002; Schirmer, Kotz, & Friederici, 2002; and for the linguistic function, see Eckstein & Friederici, 2005; Isel, Alter, & Friederici, 2005; Magne et al., 2005; Pannekamp, Toepel, Alter, Hahne, & Friederici, 2005; Astésano, Besson, & Alter, 2004; Friedrich, Kotz, Friederici, & Alter, 2004; Friedrich, Alter, & Kotz, 2001; Böcker, Bastiaansen, Vroomen, Brunia, & De Gelder, 1999; Steinhauer, Alter, & Friederici, 1999).

Interestingly, prosody relies on the same acoustic parameters (i.e., pitch or F0, duration, intensity, and spectral characteristics) as the melody (e.g., the succession of ups and downs in pitch) of a musical phrase. Based upon these similarities, a series of experiments was recently conducted with French adults (Schön, Magne, & Besson, 2004) and children (Magne, Schön, & Besson, 2006) who were either musicians or nonmusicians. The aim was to test the following hypothesis: If musical practice increases pitch discrimination, musicians should be better than nonmusicians at detecting pitch violations not only in music but also in language (F0 variations). Musicians and nonmusicians listened to sentences with natural global pitch contour but with local pitch manipulations on the final words. The pitch was increased either by 35% (weak incongruity) or by 120% (strong incongruity) without modifying the F0 contour of the final words.

Results were in line with the hypothesis: When the incongruity was most difficult to detect (weak incongruity), musicians, whether adults or children, performed better than nonmusicians. Moreover, at the ERP level, two results are of main interest for our present aims. First, for both musician and nonmusician adults, early negative components developed around 150 msec, which were

larger to strong than to weak incongruities and congruous endings. The scalp distribution of these negative components differed as a function of musical expertise, being larger over temporal sites for musicians, and more widely distributed for nonmusicians. These qualitative differences may reflect more specialized brain regions for pitch processing for musicians than for nonmusicians. Second, both weak and strong incongruities elicited larger late parietally distributed positive components than congruous endings in both groups, although with a longer onset latency for nonmusicians than for musicians. Taken together, these findings suggest that musical training facilitates the discrimination of pitch changes in one's native language and influences the spatio-temporal dynamics of language processing, as reflected by the differences in the patterns of brain waves.

Based upon these results, we hypothesized that French musicians should detect local pitch (F0) variations in a foreign language, Portuguese, that they do not understand, better than French nonmusicians. Portuguese (European Portuguese) was chosen because, as French, it is a Romance language. Thus, structural (syntactic and morphosyntactic) properties of both languages are related, yet they sound prosodically very dissimilar from one another. One reason is that there is extensive vowel reduction of unstressed syllables in spoken European Portuguese, which typically does not occur in French (e.g., Frota & Vigário, 2001; Ramus, Nespore, & Mehler, 1999). Another reason is that stress accents are distributed across different syllabic positions in European Portuguese but are most often located on word-final syllables in French (Astésano, 2001). By using a language not understood by the participants, we also hoped to tease apart the respective contribution of semantic and prosodic processing.

We used a design similar to the one in Schön et al. (2004), with a local parametric manipulation of the F0 on the sentence-final words. Sentences were spoken in Portuguese and the F0 of the final word was increased by 35% to create weak prosodic incongruities or by 120% to create strong prosodic incongruities (see Methods). We hypothesized that musicians should detect weak incongruities, which are most difficult to discriminate, better than nonmusicians but that no differences between groups should be found for both prosodically congruous and strongly incongruous endings. At the ERP level, we predicted that in line with previous results (Magne et al., 2006; Schön et al., 2004), strong incongruities should elicit larger positive components than congruous endings in both musicians and nonmusicians, although with a longer latency in nonmusicians. Moreover, weak incongruities should elicit larger positive components than congruous endings only in musicians. Finally, we asked whether early negativities would be associated to strong incongruities when participants listen to a language that they do not understand and whether their scalp distribution will match the scalp distribution

reported by Schön et al. when French participants listened to French sentences.

METHODS

Subjects

Twenty-eight native French speakers participated in this experiment, but six subjects were excluded from final data analysis because too many trials (more than 20%) were rejected due to artifact problems. Thus, the final group included 11 musicians and 11 nonmusicians. The mean age of musicians (2 women) was 38.33 years ($SD = 14$), whereas the mean age of nonmusicians (5 women) was 26.61 years ($SD = 2.08$). Musicians had at least 14 years of musical training and were regularly practicing their instruments. Nonmusicians never received formal musical training and were not music-lovers. All musician and nonmusician participants were right-handed and had normal hearing according to self-report. None of them spoke or understood Portuguese or Spanish.¹ All were paid to participate in the experiment that lasted for about 1 hour.

Materials

Stimuli comprised 120 Portuguese declarative sentences taken from children's books and ending with disyllabic words. Examples are: "A Ana tinha junto da janela a flor azul colhida no pinhal," *Ana had by the window the blue flower [that had been] picked up in the pine forest*; "Ninguém lhe tinha dado sequer um tostão," *Nobody had given him/her [not] even a cent*. In French, these sentences would be, respectively: *Ana avait près de la fenêtre la fleur bleue cueillie dans la pinède*, and *Personne ne lui avait donné ni même un centime*. Sentences were spoken by a native Portuguese

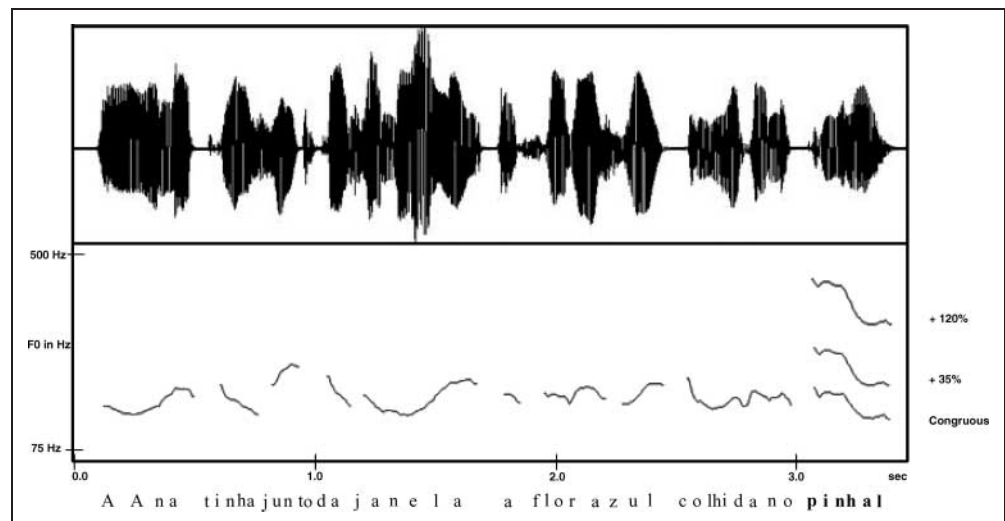
female speaker in a soundproof chamber and digitized (48 kHz/16-bit sampling rate) using ProTools LE Version 5.1.1. The mean duration of the sentences was 3.39 sec ($SD = 0.8$) and the mean duration of the final word was 531 msec ($SD = 81$).

A total of 120 sentences were auditorily presented with an equal number of sentences (40) being presented in three experimental conditions: The final word was prosodically congruous, weakly incongruous (35% increase in F_0), or strongly incongruous (120% increase in F_0). Local changes in the pitch (F_0) of final words were done using Praat (Boersma & Weenink, 2004) without changing the original pitch contour of the words (see Figure 1). We used the same F_0 manipulations as in the experiment of Schön et al. (2004) for two reasons: First, because results of a pilot study with five Portuguese participants had indicated that the error rate in the three experimental conditions was similar to what was found with the French materials, and, second, to be able to compare results between experiments.

Procedure

Participants were seated in a comfortable chair in an electrically shielded room. They were required to listen attentively to the sentences presented through headphones and to decide, as quickly and accurately as possible, whether the last word of the sentence seemed normal or strange (i.e., something was wrong). They used a two-button box and pressed one button for "normal" (congruous condition) and another button for "strange" (both weak and strong incongruity conditions). The hand of response was counterbalanced across participants. Moreover, participants were forewarned that the sentences were spoken in Portuguese and that, even if they were not expected to understand their meaning, they should try to focus their attention on how the final word was spoken.

Figure 1. An example of the sentences used in the experiment: *A Ana tinha junto da janela a flor azul colhida no pinhal* (Literal translation: Ana had by the window the blue flower [that had been] picked up in the pine forest). The top panel shows the time waveform of the utterance. The fundamental frequency (F_0 in Hz) is shown below for the natural utterance (Congruous) and for the changes in pitch of the final word (+35% and 120%).



Before the experimental session, participants were given a practice block of six trials to familiarize them with the task, and to train them to blink during the intertrial interval. They were given feedback on the correctness of their responses in the practice session but not in the experimental session. After the training session, the experimental trials started; they comprised four blocks of 30 sentences. Sentences from each experimental condition occurred equally often within a block and were presented in a pseudorandom order (with the constraint that no more than three successive sentences required a “Yes” or “No” response).

ERP Recordings

Electroencephalogram (EEG) was continuously recorded from 32 BIOSEMI pin-type active electrodes (Amsterdam University) mounted on an elastic headcap and located at standard left and right hemisphere positions over frontal, central, temporal, parietal, and occipital areas (Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F7, F8, F3, F4, FC1, FC2, FC5, FC6, C3, C4, T7, T8, CP5, CP6, CP1, CP2, P7, P8, P3, P4, PO3, PO4, O1, O2). These recording sites were referenced off-line to the algebraic average of the activity over the left and right mastoids. In order to detect horizontal eye movements and blinks, the electrooculogram (EOG) was recorded from flat-type active electrodes placed 1 cm to the left and right of the

external canthi, and from an electrode beneath the right eye. EEG and EOG were recorded at a sampling rate of 512 Hz using BIOSEMI amplifiers and were filtered with a bandpass of 0–40 Hz. Impedances of the electrodes never exceeded 5 k Ω . Data were segmented in single trials of 2200 msec starting 150 msec before the onset of the sentence-final word and were analyzed using the Brain Vision Analyzer software (Brain Products, Munich). Trials containing ocular movement artifacts, or amplifier saturation, were excluded from the averaged ERP waveforms.

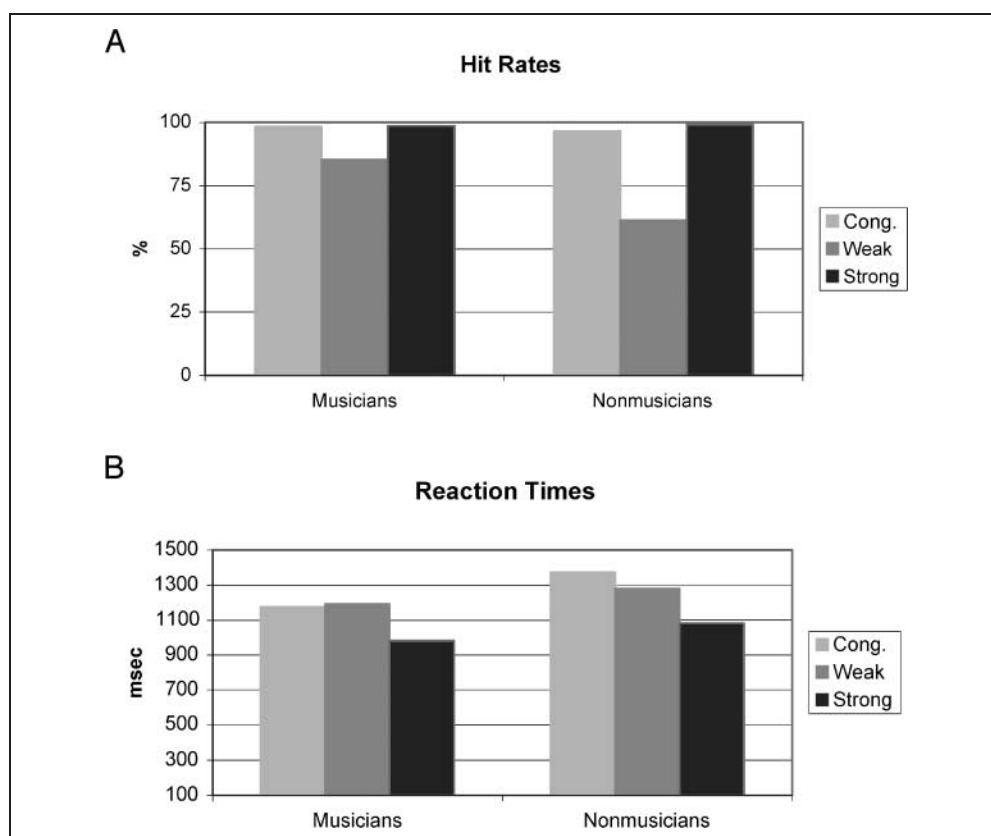
RESULTS

Behavioral Data

Hit rates (%) and reaction times (RTs) were averaged across participants and repeated-measures analyses of variance (ANOVAs) were computed that included expertise (musicians vs. nonmusicians) as a between-subject factor, and congruity (congruous, weak and strong incongruities) as a within-subject factor. Tukey tests were used for all post hoc comparisons.

Analysis of hit rates showed main effects of expertise [$F(1, 20) = 12.06, p < .002$] and congruity [$F(2, 40) = 38.90, p < .001$; see Figure 2A]. Musicians (94%) showed higher hit rates than nonmusicians (86%), and weak incongruities (73%) were the most difficult to detect

Figure 2. (A) Hit rates for congruous (Cong.) final words, weak and strong prosodic incongruities, for musicians and nonmusicians. Clearly, the hit rate was lowest for nonmusicians in response to weak incongruities. (B) Reaction times (msec) for congruous (Cong.) final words, weak and strong prosodic incongruities are presented for musicians and nonmusicians. Overall, musicians were faster than nonmusicians. Moreover, for nonmusicians the stronger the incongruity, the shorter the RTs, whereas for musicians no differences between congruous items and weak incongruities were found; only RTs to strong incongruities were faster than in the other two conditions.



(congruous: 97%, and strong incongruities: 99%). Most importantly, the Expertise \times Congruity interaction was significant [$F(2, 40) = 8.84, p = .001$]: In line with our hypothesis, musicians (85%) detected the weak incongruities better than nonmusicians (61%), with no between-group differences for strong incongruities and congruous endings.

Analysis of RTs to correct responses showed main effects of expertise [$F(1, 20) = 4.83, p < .04$] and congruity [$F(2, 40) = 92.23, p < .001$]. Musicians (1116 msec) were overall faster than nonmusicians (1245 msec), and strong incongruities (1031 msec) were faster to detect than both weak incongruities (1237 msec) and congruous endings (1275 msec; see Figure 2B). Moreover, the Expertise \times Congruity interaction was significant [$F(2, 40) = 4.92, p = .01$]: For nonmusicians, all conditions differed significantly from one another (Congruous = 1375 msec, Weak = 1280 msec and Strong = 1081 msec), whereas for musicians RTs were shorter for the strong incongruities (980 msec) than for both congruous endings (1176 msec) and weak incongruities (1192 msec), which did not differ from each other.

Electrophysiological Data

ERP data were analyzed by computing mean amplitudes, relative to a 150-msec baseline, in successive 50-msec latency windows from the onset of the sentence-final word until 1200 msec afterward. When the effects were similar in successive 50-msec windows, latency ranges were pooled together. Thus, six successive 200-msec

duration latency bands were determined. ANOVAs were computed for midline and lateral electrodes separately and included expertise (musicians vs. nonmusicians) as a between-subjects factor; congruity (congruous, weakly incongruous, and strongly incongruous) and electrodes (Fz, Cz, and Pz) as within-subjects factors for midline analysis; and congruity (3), hemisphere (2), region (3 regions of interest [ROIs]: fronto-central, temporal, and parietal), and electrodes (3 for each ROI: F7, F3, FC5 and F8, F4, FC6; C3, T7, CP5 and C4, T8, CP6; CP1, P3, P7 and CP2, P4, P8) as within-subjects factors for lateral analysis. All p values were adjusted with the Greenhouse–Geisser epsilon correction for nonsphericity. The uncorrected degrees of freedom and the probability level after correction are reported in Tables 1 and 2. When the effects of expertise or congruity were significant or interacted with the effects of other factors, ANOVAs were computed for musicians and nonmusicians separately. Tukey tests were used for all post hoc comparisons. To simplify the presentation of the results, significant effects in the different latency ranges are reported in Table 1 for midline electrodes and in Table 2 for lateral electrodes. Mean amplitude values are reported in Table 3. When the main effects or interactions are significant, results of post hoc comparisons are presented in text.

No main effects or interactions were found in the 0–200 msec latency band. In both the 200–400 and 400–600 msec latency bands, the main effect of congruity was significant at midline and at lateral electrodes. Most importantly, the Expertise \times Congruity \times Electrodes interaction was also significant. Although for musicians strong

Table 1. Results of Main ANOVAs at Midline Electrodes

Latency Bands	Factors	Between Groups	Musicians	Nonmusicians
200–400	C (2, 40)/(2, 20)	6.20**	7.06**	<i>ns</i>
	E \times C \times Elec. (4, 80)	3.12**	(C \times Elec.) <i>ns</i>	(C \times Elec.) <i>ns</i>
400–600	C (2, 40)/(2, 20)	21.15***	30.42***	<i>ns</i>
	E \times C (2, 40)	3.54*	–	–
	E \times C \times Elec. (4, 80)/(4, 40)	3.49**	(C \times Elec.) 8.80***	(C \times Elec.) <i>ns</i>
600–800	C (2, 40)/(2, 20)	30.50***	30.30***	7.35**
	E \times C \times Elec./(4, 40)	<i>ns</i>	(C \times Elec.) 4.71**	(C \times Elec.) <i>ns</i>
800–1000	C (2, 40)/(2, 20)	17.32***	9.29***	8.41**
	E \times C \times Elec./(4, 40)	<i>ns</i>	(C \times Elec.) 3.10*	(C \times Elec.) <i>ns</i>
1000–1200	C–	<i>ns</i>	<i>ns</i>	<i>ns</i>
	E \times C \times Elec./(4, 40)	<i>ns</i>	(C \times Elec.) 4.28***	(C \times Elec.) <i>ns</i>

E = expertise; C = congruity; and Elec. = electrode.

Degrees of freedom are given in parentheses (between groups/musicians or nonmusicians).

*Differences were statistically significant at $<.05$.

**Differences were statistically significant at $<.01$.

***Differences were statistically significant at $<.001$.

Table 2. Results of Main ANOVAs at Lateral Electrodes

Latency Bands	Factors	Between Groups	Musicians	Nonmusicians
200–400	C (2, 40)/(2, 20)	2.43, $p = .10$	4.86**	<i>ns</i>
400–600	C (2, 40)/(2, 20)	12.64***	28.10***	<i>ns</i>
	E × C (2, 40)	5.42**	–	–
	E × C × H × ROI (4, 80)/(4, 40)	2.59*	(C × H × ROI) 3.92**	(C × H × ROI) <i>ns</i>
600–800	C (2, 40)/(2, 20)	28.73***	25.59***	10.96***
	E × C × H × ROI/(4, 40)	<i>ns</i>	(C × H × ROI) 3.21*	(C × H × ROI) <i>ns</i>
800–1000	C (2, 40)/(2, 20)	19.08***	8.27**	10.82***
	E × C × H × ROI/(4, 40)	<i>ns</i>	(C × H × ROI) 3.24*	(C × H × ROI) <i>ns</i>
1000–1200	C (2, 40)/(2, 20)	5.51**	7.96**	<i>ns</i>
	E × C × H × ROI (4, 80)	2.80*	(C × ROI) 4.74**	(C × H × ROI) <i>ns</i>

E = expertise; C = congruity; H = hemisphere; ROI = region of interest.

Degrees of freedom are given in parentheses (between groups/musicians or nonmusicians).

*Differences were statistically significant at $<.05$.

**Differences were statistically significant at $<.01$.

***Differences were statistically significant at $<.001$.

incongruities were associated with increased positivities compared to either congruous words or weak incongruities ($p < .001$ in both cases; see Figure 3), no differences were found for nonmusicians (see Figure 4). Moreover, for musicians, results of Tukey tests in the 400–600 msec range revealed that strong incongruities were significantly different from congruous endings at all scalp sites, and that weak incongruities were significantly different from congruous endings at parietal sites (Congruity × Electrodes interaction) and over the centroparietal regions of the left hemisphere (Congruity × Hemisphere × ROI interaction; see Figure 3). No such differences were found for nonmusicians (see Figure 4).

In both the 600–800 msec and 800–100 msec latency bands, the main effect of congruity was significant for

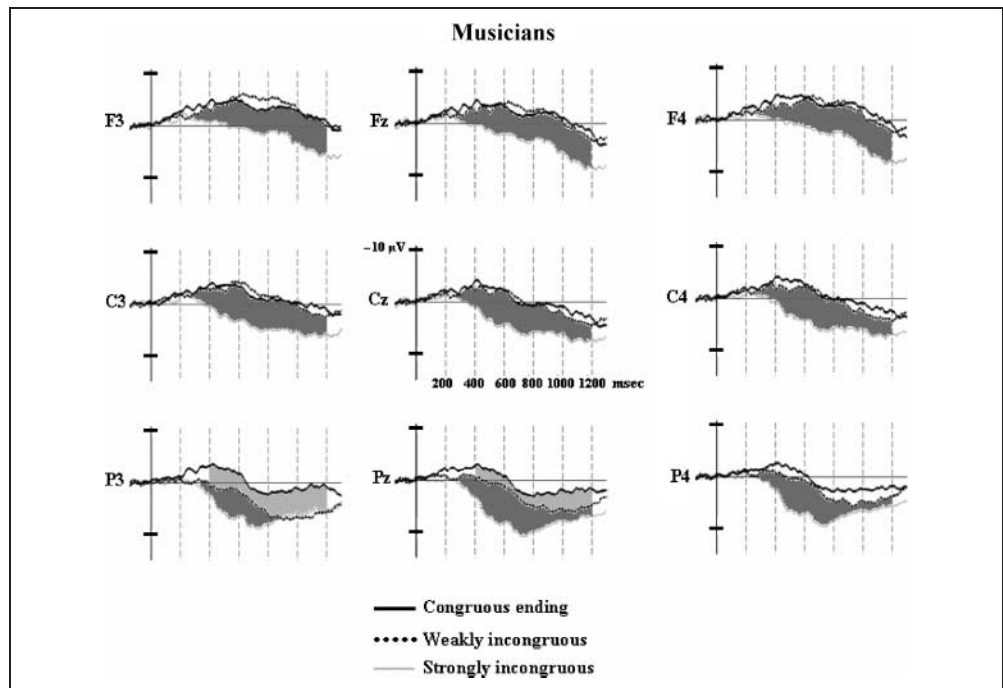
both musicians and nonmusicians at midline and at lateral electrodes. For musicians, strong incongruities differed from congruous endings at all scalp sites, and weak incongruities differed from congruous endings at parietal sites and over parietal regions of the left hemisphere (Congruity × Hemisphere × ROI interaction). For nonmusicians, both strong and weak incongruities differed from congruous endings at midline electrodes, but only strong incongruities differed from weak incongruities and congruous endings at lateral electrodes (Figure 4).

Finally, in the 1000–1200 msec latency band, the main effect of congruity was only significant for musicians and at lateral electrodes. Although strong incongruities differed from congruous endings over all regions, weak incon-

Table 3. Mean Amplitude Values (μV) in the Different Latency Bands for Musicians and Nonmusicians in the Three Conditions

Latency Bands	Electrodes	Musicians			Nonmusicians		
		Congruous	Weak	Strong	Congruous	Weak	Strong
200–400	Midline	–1.78	–1.51	–0.29	–2.08	–1.98	–1.51
	Lateral	–2.40	–1.50	–0.19	–0.77	–0.75	–1.05
400–600	Midline	–1.58	–0.14	2.38	–2.66	–1.94	–0.98
	Lateral	–2.68	–1.86	2.58	–1.36	–1.67	–0.43
600–800	Midline	–0.46	2.14	5.55	–0.94	0.95	2.86
	Lateral	–0.48	–0.83	4.25	–1.20	–1.05	4.59
800–1000	Midline	0.60	2.96	5.52	0.75	4.29	5.78
	Lateral	0.08	0.71	4.28	0.64	1.37	6.26

Figure 3. Averaged brain electrical activity in musicians time-locked to the onset of prosodically congruous endings (black solid line), weak incongruities (black dotted line), and strong incongruities (gray solid line). Selected traces from nine electrodes are presented. The latency ranges within which statistical analyses revealed significant differences between strong incongruities and congruous endings are shaded in dark gray and between weak incongruities and congruous endings in light gray. In this figure, as in the following ones, the amplitude (μV) is plotted on the ordinate (negative up) and the time (msec) on the abscissa.



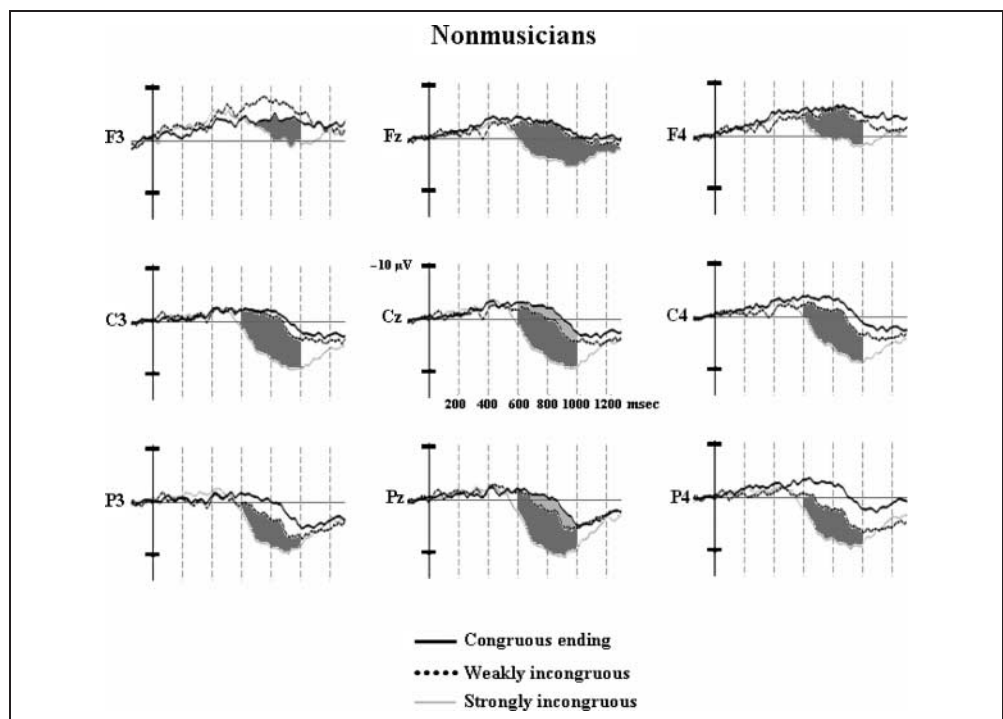
gruities differed from congruous endings only over parietal regions (Congruity \times ROI interaction; see Figure 3).

DISCUSSION

Results showed that French musicians and nonmusicians were able to accurately decide that the final words of sentences spoken in Portuguese were prosodically congruous

or strongly incongruous. Thus, participants were sensitive to some prosodic features of words (local pitch increase) and of sentence contour (reduction of the declination-lowering in pitch-typical of declarative sentences) without understanding their meaning. However, when the local pitch changes were small and difficult to detect (weak prosodic incongruities), musicians performed better than nonmusicians. These results extend those previously found

Figure 4. Averaged brain electrical activity in nonmusicians time-locked to the onset of prosodically congruous endings (black solid line), weak incongruities (black dotted line), and strong incongruities (gray solid line). Selected traces from nine electrodes are presented. Significant differences between strong incongruities and congruous endings are shaded in dark gray and between weak incongruities and congruous endings in light gray.



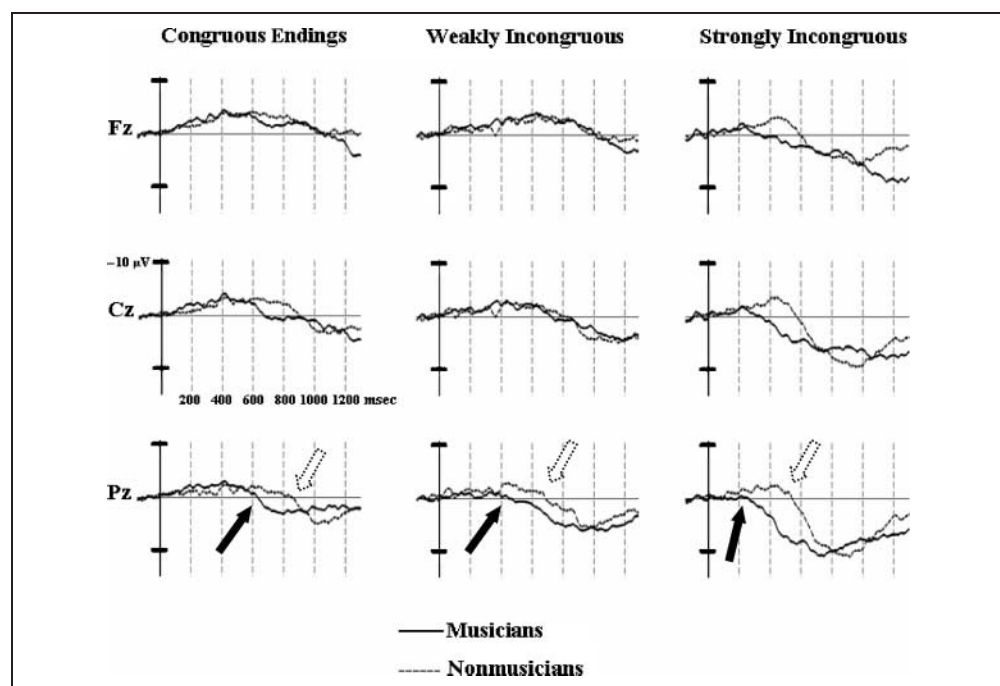
with both adults and children listening to prosodic manipulations in their own language (Magne et al., 2006; Schön et al., 2004) and are in line with the hypothesis that musical expertise, by increasing discrimination of pitch—a basic acoustic parameter which is equally important for music and speech prosody—facilitates the processing of pitch variations not only in music (Thompson, Schellenberg, & Husain, 2003, 2004; Besson & Faïta, 1995) but also in language. This conclusion is also supported by the RT data showing that musicians were faster overall than nonmusicians. Before considering the implications of this finding for foreign language learning, we first discuss the electrophysiological data.

ERP latency analyses revealed that musical expertise had a large influence on the time course of pitch processing. For musicians, the ERPs to the strong incongruity differed from the other two conditions as early as 200–400 msec and the difference between weak incongruities and congruous endings was significant within 400 and 600 msec at some scalp sites. By contrast, for nonmusicians, the ERPs to strong incongruities did not differ from congruous endings before 600–800 msec, and the difference between weak incongruities and congruous endings was only significant between 800 and 1000 msec. As found in previous studies using different types of prosodic manipulations (Magne et al., 2006; Eckstein & Friederici, 2005; Pannekamp et al., 2005; Astésano et al., 2004; Schön et al., 2004; Steinhauer et al., 1999), these differences were reflected by increased positivities to the deviant stimuli. Typically, the latency of late positivities is known to vary as a function of the difficulty to process task-relevant events (Duncan-Johnson & Donchin, 1977; Kutas, McCarthy, & Donchin, 1977) and is taken to reflect categorization and

decision processes. Based on this assumption, the present results show that musicians are, on average, 300 msec faster than nonmusicians to categorize prosodically congruous and incongruous endings (see Figure 5). Despite such large latency differences, the ERPs to the different types of endings do not seem qualitatively different for musicians and nonmusicians. In both groups, the late positivities showed a parietal distribution, as previously reported in a number of prosody experiments (Magne et al., 2006; Eckstein & Friederici, 2005; Pannekamp et al., 2005; Astésano et al., 2004; Schön et al., 2004; Steinhauer & Friederici, 2001; Steinhauer et al., 1999). Even if further experiments are needed to determine whether the positive components reported in these different prosody experiments reflect similar or different processes (such as the specific processing of sentence prosodic boundaries, closure positive shift [CPS]; Steinhauer et al., 1999), the results reported here show that musical expertise facilitates the time course of pitch processing in language.

Comparison of the present results with those of Schön et al. (2004) reveals that RTs were 280 msec shorter, on average, when French participants detected local strong F0 variations in French than in Portuguese. Moreover, the onset of the positivity to strong prosodic incongruities was around 100 msec earlier when French participants listened to French sentences as compared to when they listened to Portuguese ones. Although participants in the French and Portuguese experiments were different, it is unlikely that between-group differences are responsible for such strong effects. A more likely interpretation is that understanding the sentence context allows the French listeners to anticipate when the final word is going to be presented, thereby speed-

Figure 5. Averaged brain electrical activity recorded from musicians and nonmusicians at midline electrodes (Fz, Cz, and Pz) is overlapped for each experimental condition. The black filled arrow shows when effects start to be significant for musicians and the dashed empty arrow shows when effects start to be significant for nonmusicians (300 msec later on average).



ing up the detection of the strong incongruities when they are actually heard. Thus, understanding the semantics influences the perception of acoustic prosodic cues. Related to this issue, French participants made more errors to detect local strong F0 variations in French (10%) than in Portuguese (1%). Again, although we must be careful in making direct comparisons between experiments (because both participants and sentences materials were different), the higher error rate in French may reflect some interference from semantic onto prosodic processing, thereby pointing to the nonindependence of these two aspects of language processing. Although this issue is still open, Astésano et al. (2004) reported an interaction between semantic and prosodic processing in an experiment in which the semantics and the modality (declarative vs. interrogative) of French sentences were manipulated so as to obtain semantically and/or prosodically congruous and incongruous final words. Results showed that the prosodic manipulation was only significant for semantically incongruous sentences, thereby pointing to an interaction between semantic and prosodic processing.

Except for the strong incongruity, inspection of the ERPs reveals that, relative to baseline, negative components develop in the 300–600 msec latency band for musicians (see Figure 3) and in the 400–800 msec range for nonmusicians (see Figure 4). Therefore, the question arises whether these negative components are N400-like components (Kutas & Hillyard, 1980) or are rather directly linked to prosodic processing? Eckstein and Friederici (2005) recently showed that early negative components are elicited by German sentence-final words that carry the prosodic features of sentence penultimate words, and are, consequently, prosodically incongruous. However, aside from latency differences (but see below), the amplitude of this negative component is larger over right anterior regions of the scalp, a distribution not found here.

The N400 component is typically interpreted as reflecting access to word meaning and meaning expectancy derived from words or sentence contexts (DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984; see Kutas, Van Petten, & Kluender, in press, for a review). Because Portuguese is a language linguistically close to French (Azevedo, 2005; Mateus & Andrade, 2000), participants may notice some similarities between the two languages and engage in a search for meaning. Lexical decision task results have shown that the N400 is typically larger to pseudowords than to words, a result taken to reflect increased difficulties in accessing word meaning (Holcomb & Neville, 1990; Bentin, McCarthy, & Wood, 1985). Following this interpretation, the negative components found here might be N400-like components. However, their widespread scalp distribution differs from the typical centro-parietal distribution of the N400 component and they are also temporally more extended than when participants understand the lan-

guage (see, for instance, Figure 5 of Schön et al., 2004). Thus, the functional significance of these negative components remains open to further investigation.

Finally, it is also interesting to point out that, by contrast to previous results in auditory sentence experiments, no early N1–P2 component can be identified on the ERP traces that remained at baseline level until around 200 msec for musicians and 400 msec for nonmusicians (see Figures 3 and 4). Because participants did not understand the language, they were probably unable to decide that the final word of the sentence was being presented. Consequently, the sensory/perceptual processes were not time-locked to final word presentation and no N1–P2 components were generated. Finally, the finding that the negative components start to develop around 200 msec later for nonmusicians than for musicians raises the intriguing possibility that musicians were able to pick up sentence-final prosodic cues that were not perceived by nonmusicians. This possibility will be explored further in future experiments.

In conclusion, these results demonstrate that musical expertise influences pitch processing in a foreign language and the time course of semantic and prosodic processing. In further experiments, it will be interesting to track the effects of musical expertise on the processing of other acoustic parameters of prosody such as duration or spectral characteristics. Moreover, it will also be important to determine whether the influence of musical expertise extends to the phonological aspects of prosody such as the detection of word stress (Böcker et al., 1999), intonation (Astésano et al., 2004), or prosodic phrase boundaries (Steinhauer & Friederici, 2001; Steinhauer et al., 1999). Such results, together with the ones presented here, should have important consequences for the pedagogical methods used for foreign language learning. Based upon the growing evidence pointing to the importance of prosody for language comprehension (Frazier et al., 2006), it may prove useful to develop learning programs relying on the spoken form of languages. Musical training may be particularly useful in this context.

Acknowledgments

This work was supported by a grant from the Human Frontiers Science Program (HFSP, RG P0053/2002) to Mireille Besson (PI), and by a grant from the Fundação para a Ciência e Tecnologia (FCT/POCTI/34806/PSI/2000) to São Luís Castro (PI). Sylvain Moreno is a PhD student supported by a fellowship from the French Ministry of Research, Action “Ecole et Sciences Cognitives.” We thank Monique Chiambretto, Mitsuko Aramaki, and Marjorie Geva for methodological assistance, and Daniele Schön and two reviewers for insightful comments on a previous version of the manuscript.

Reprint requests should be sent to Carlos Marques, Laboratório de Fala, Faculdade de Psicologia e Ciências da Educação, Universidade do Porto, rua Doutor Manuel Pereira Silva, 4200 — 392 Porto Portugal, or via e-mail: labfala@fpce.up.pt.

Note

1. We controlled that participants neither spoke nor understood Portuguese and Spanish because these two languages are linguistically very close, and knowing one of them might allow to guess the meaning of words in the other. This, however, is not the case of Portuguese and French due to differences in the lexica, in the phonetic properties of the phonemes, and in the position of word stress.

REFERENCES

- Astésano, C. (2001). *Rythme et accentuation en français. Invariance et variabilité stylistique*. Paris: L'Harmattan.
- Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research*, *18*, 172–184.
- Azevedo, M. M. (2005). *Portuguese. A linguistic introduction*. Cambridge: Cambridge University Press.
- Bentin, S., McCarthy, G., & Wood, C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 353–355.
- Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with non-musicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1278–1296.
- Besson, M., Magne, C., & Schön, D. (2002). Emotional prosody: Sex differences in sensitivity to speech melody. *Trends in Cognitive Sciences*, *6*, 405–407.
- Böcker, K. B. E., Bastiaansen, M. C. M., Vroomen, J., Brunia, C. H. M., & De Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology*, *36*, 706–720.
- Boersma, P., & Weenink, D. (2004). Praat: Doing phonetics by computer (Version. 4.1.15) [Computer software]. Retrieved from www.praat.org/ on 12 July 2004.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.
- Duncan-Johnson, C., & Donchin, E. (1977). On quantifying surprise, the variation of event-related potentials with subjective probability. *Psychophysiology*, *14*, 456–467.
- Eckstein, K., & Friederici, A. D. (2005). Late interaction of syntactic and prosodic processes in sentence comprehension as revealed by ERPs. *Cognitive Brain Research*, *25*, 130–143.
- Elbert, T., Pantev, C., Wienbruch, C., Rockstroh, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science*, *270*, 305–307.
- Frazier, L., Carlson, K., & Clifton, C., Jr. (2006). Prosodic phrasing is central to language comprehension. *Trends in Cognitive Sciences*, *6*, 244–249.
- Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *NeuroReport*, *12*, 3189–3191.
- Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Journal of Cognitive Brain Research*, *20*, 300–308.
- Frota, S., & Vigário, M. (2001). On the correlates of rhythmic distinctions: The European/Brazilian Portuguese case. *Probus*, *13*, 247–275.
- Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience*, *23*, 9240–9245.
- Griffiths, T. G., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*, 348–353.
- Holcomb, P. J., & Neville, H. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*, 281–312.
- Isel, F., Alter, K., & Friederici, A. D. (2005). Processing of syntactic and prosodic information in German particle verbs: Evidence from ERPs. *Journal of Cognitive Neuroscience*, *17*, 154–164.
- Knoesche, T., Neuhaus, C., Hauelsen, J., Alter, K., Maess, B., Friederici, A., et al. (2005). The perception of phrase structure in music. *Human Brain Mapping*, *24*, 259–273.
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, *15*, 1–6.
- Koelsch, S., Fritz, T., Schulze, K., Alsup, D., Schröger, E., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *Neuroimage*, *25*, 1068–1076.
- Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical “language-network” serves the processing of music. *Neuroimage*, *17*, 956–966.
- Koelsch, S., & Siebel, W. A. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences*, *6*, 578–584.
- Kotz, S., Meyer, M., Alter, K., Besson, M., von Cramon, Y., & Friederici, A. (2003). On the lateralization of affective prosody: An event-related functional MR investigation. *Brain and Language*, *86*, 366–376.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*, 792–795.
- Kutas, M., Van Petten, C., & Kluender, R. (2006). Psycholinguistics electrified II: 1994–2005. In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 659–724). New York: Elsevier.
- Levitin, D. J., & Menon, V. (2003). Musical structure is processed in “language” areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, *20*, 2142–2152.
- Maess, B., Koelsch, S., Gunter, T., & Friederici, A. D. (2001). Musical syntax is processed in Broca’s area: An MEG study. *Nature Neuroscience*, *4*, 540–545.
- Magne, C., Astésano, C., Lacheret-Dujour, A., Morel, M., Alter, K., & Besson, M. (2005). On-line processing of “pop-out” words in spoken French dialogues. *Journal of Cognitive Neuroscience*, *17*, 740–756.
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than non-musician children: Behavioural and electrophysiological approaches. *Journal of Cognitive Neuroscience*, *18*, 199–211.
- Mateus, M. H., & Andrade, E. (2000). *The phonology of Portuguese*. Oxford: Oxford University Press.
- Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human Brain Mapping*, *17*, 73–88.
- Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., et al. (2001). Functional anatomy of musical perception in musicians. *Cerebral Cortex*, *11*, 754–760.

- Pannekamp, A., Toepel, U., Alter, K., Hahne, A., & Friederici, A. D. (2005). Prosody-driven sentence processing: An event-related brain potential study. *Journal of Cognitive Neuroscience*, *17*, 407–421.
- Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature*, *392*, 811–814.
- Ramus, F., Nespors, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 265–292.
- Schirmer, A., Kotz, S. A., & Friederici, A. D. (2002). Sex differentiates the role of emotional prosody during word processing. *Cognitive Brain Research*, *14*, 228–233.
- Schlaug, G., Jancke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia*, *33*, 1047–1055.
- Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*, 699–701.
- Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience*, *5*, 688–694.
- Schön, D., Magne, C., & Besson, M. (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology*, *41*, 341–349.
- Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature Neuroscience*, *2*, 191–196.
- Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The Closure Positive Shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research*, *30*, 267–295.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech: Effects of music lessons. *Annals of the New York Academy of Sciences*, *999*, 530–532.
- Thompson, W. F., Schellenberg, E. G., & Husain, G. (2004). Decoding speech prosody: Do music lessons help? *Emotion*, *4*, 46–64.
- Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research*, *16*, 145–161.
- Tzourio, N., El Massioui, F., Crivello, F., Joliot, M., Renault, B., & Mazoyer, B. (1997). Functional anatomy of human auditory attention studied with PET. *Neuroimage*, *5*, 63–77.
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage*, *30*, 1414–1432.