

Interactive Processing of Timbre Dimensions: An Exploration with Event-related Potentials

Anne Caclin^{1,2,3,4}, Stephen McAdams^{4,5}, Bennett K. Smith^{4,5},
and Marie-Hélène Giard^{1,2,3}

Abstract

■ Timbre characterizes the identity of a sound source. On psychoacoustic grounds, it has been described as a multidimensional perceptual attribute of complex sounds. Using Garner's interference paradigm, we found in a previous behavioral study that three timbral dimensions exhibited interactive processing. These timbral dimensions acoustically corresponded to attack time, spectral centroid, and spectrum fine structure. Here, using event-related potentials (ERPs), we sought neurophysiological correlates of the interactive processing of these dimensions of timbre. ERPs allowed us to dissociate several levels of interaction, at both

early perceptual and late stimulus identification stages of processing. The cost of filtering out an irrelevant timbral dimension was accompanied by a late negative-going activity, whereas congruency effects between timbre dimensions were associated with interactions in both early sensory and late processing stages. ERPs also helped to determine the similarities and differences in the interactions displayed by the different pairs of timbre dimensions, revealing in particular variations in the latencies at which temporal and spectral timbre dimensions can interfere with the processing of another spectral timbre dimension. ■

INTRODUCTION

Sound sources are characterized by a number of perceptual attributes: pitch, loudness, duration, position, and timbre. Timbre perception is intrinsically linked to our ability to recognize the identity of sound sources (Handel, 1995; McAdams, 1993). Psychoacoustic studies have revealed that timbre is actually a multidimensional attribute of complex tones, which means that it is better described as a collection of dimensions constituting a timbre space than as a unitary perceptual attribute (see Hajda, Kendall, Carterette, & Harschberger, 1997, for a review). Each timbre dimension corresponds to a particular acoustic parameter. The number of timbre dimensions and their acoustic underpinnings continue to be debated, but most studies converge on the idea that timbre dimensions are of temporal (e.g., attack time), spectral (e.g., spectral centroid), and spectro-temporal natures (e.g., spectral flux) (see, for example, Caclin, McAdams, Smith, & Winsberg, 2005; McAdams, Winsberg, Donnadieu, De Soete, & Krimphoff, 1995; Grey, 1977). This multidimensional account of timbre raises the issue of the processing relationships among its various dimensions. In the following, we review some of the studies that

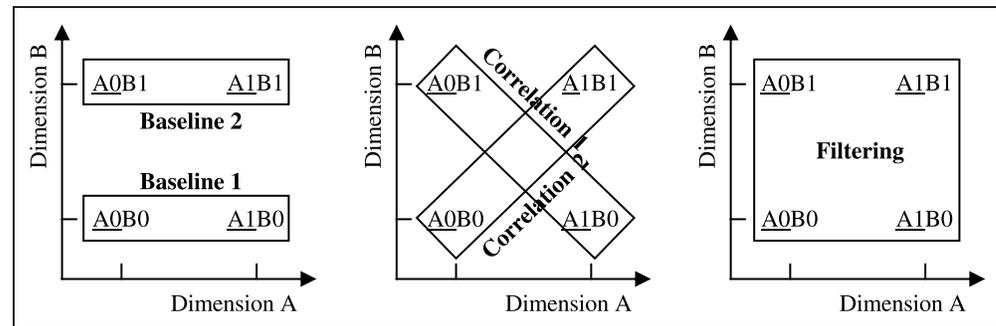
have analyzed the processing interactions among auditory attributes, in order to isolate some unresolved issues concerning the interactions involved in the processing of timbre dimensions.

Interactions among Auditory Attributes

Interactions in the processing of auditory attributes (dimensions) have been investigated in a number of studies, using both behavioral and electrophysiological measures. A first set of behavioral studies assessed Garner interference (Garner, 1974) between auditory attributes and found evidence for an interactive processing of auditory dimensions. The core idea in Garner's approach is that if the processing of two dimensions is separate, then performance in a classification task based on a particular dimension of the stimulation should be unaffected by variations along the other dimension (*separability* property). Should the reverse be observed, then there exist interactions during the processing of these two dimensions (referred to as *integrality* of the dimensions). Garner's classical paradigm involves creating a square of four stimuli by combining two values on each of the two dimensions of interest (Figure 1). Participants are asked to categorize the stimuli according to one dimension in five different types of sequences: the other (irrelevant) dimension is either kept at a constant value (two *baseline* conditions), is varied systematically with the dimension of interest (two *correlation* conditions), or is varied independently of the dimension of

¹INSERM U821 Brain Dynamics and Cognition, Lyon, France, ²Institut Fédératif des Neurosciences, Lyon, France, ³University of Lyon 1, Lyon, France, ⁴Institut de Recherche et Coordination Acoustique/Musique (STMS-IRCAM-CNRS), Paris, France, ⁵McGill University, Montréal, Québec, Canada

Figure 1. Principle of Garner's interference paradigm. Four stimuli are created by combining two values along the two dimensions of interest (A0 and A1 for Dimension A; B0 and B1 for Dimension B). In this example, the participant has to classify the stimuli according to Dimension A (as underlined), in five types of stimulus sequences represented by the rectangles.



interest (*filtering* condition). *Separability* is revealed by equal performance in baseline, correlation, and filtering conditions, whereas a decrease in performance in the filtering condition and/or a gain in correlation conditions (both relative to baseline) characterizes *integral* dimensions. Thus, when categorizing one of the dimensions of an integral pair, there is a cost of filtering out irrelevant random variations along the other dimension, and/or a benefit of redundant yet irrelevant variations along the other dimension. Studies of Garner interference between auditory attributes have always revealed an interactive processing (*integrality* property) of these attributes, in particular for pitch, loudness, and timbre (Krumhansl & Iverson, 1992; Melara & Marks, 1990b, 1990c; Grau & Kemler-Nelson, 1988), and for three timbre dimensions (Caclin, Giard, Smith, & McAdams, 2007).

In Garner interference studies, it is also possible to assess *congruency* effects (i.e., whether there is a difference in performance for the two diagonals of the square formed by the four stimuli). With the notations of Figure 1, this means that a congruency analysis explores whether performance is systematically better (or worse) for A0B1 and A1B0 than for A0B0 and A1B1, irrespective of the experimental condition (baseline, correlation, or filtering). Congruency effects thus share some similarities with Stroop interference (Stroop, 1935), in the sense that both assess whether the amount of interaction between two dimensions depends on the particular values they take (in the case of Stroop interference, one can generally establish a semantic correspondence between the values along the two dimensions). However, Stroop interference studies typically involve one linguistic dimension (color names) and one perceptual dimension (color). Further, in Stroop interference studies, contrary to Garner interference studies, baseline discriminabilities of the two dimensions are not equated, and this inequality could be necessary for Stroop interference to emerge (Melara & Mounts, 1993). In Garner interference studies, congruency effects have been observed for auditory attributes, in particular for pitch and loudness, with high-pitched/loud and low-pitched/soft tones classified faster than low-pitched/loud and high-pitched/soft tones (e.g., Melara & Marks, 1990c). Congruency effects

involving timbre (taken as a whole) are more complex and probably depend on the exact acoustic attributes that have been varied to create timbre differences (see Krumhansl & Iverson, 1992; Melara & Marks, 1990b). Finally, when studying Garner interference between three timbre dimensions corresponding acoustically to attack time, spectral center of gravity (SCG), and spectrum fine structure, we have observed congruency effects for the three possible pairs of dimensions (Caclin et al., 2007), which allowed us to distinguish two sets of congruent attributes: sharp attack, brilliant (high SCG), and neutral (regular spectrum), on the one hand, and slow attack, dull (low SCG), and hollow (jaggy spectrum), on the other hand.

The relationships between auditory attributes have also been investigated at the level of sensory memory, both in behavioral experiments and using the mismatch negativity (MMN) of the auditory event-related potential (ERP), which indexes the representation in auditory sensory memory (see Näätänen & Winkler, 1999, for a review). Separate sensory memory representations have been found for pitch, loudness, and duration (Clément, Demany, & Semal, 1999; Giard et al., 1995); for pitch and timbre (Krumhansl & Iverson, 1992; Semal & Demany, 1991); or for pitch and frequency pattern (Alain, Achim, & Woods, 1999). We have also recently reported evidence for separate representations of timbre dimensions in auditory sensory memory using the MMN (Caclin et al., 2006), which argues for a parallel processing of different timbre dimensions, at least during some processing stages. A possible interpretation that reconciles Garner interference studies (suggesting interactive processing of auditory attributes) and sensory memory studies (suggesting separate processing of auditory attributes) is to postulate that auditory attributes are processed within distinct channels. However, these channels can interact at various processing stages, which would account for the observed Garner interference (see also Caclin et al., 2007, for a discussion of this point).

An unresolved issue thus concerns the level of processing at which interactions between auditory attributes (and in particular between timbre dimensions) can occur. Although it has been argued that interactions at

early (perceptual) processing stages are likely to explain Garner interference between auditory attributes (e.g., Melara & Marks, 1990b), this has never been assessed directly. The present study aimed to clarify this point in the case of timbre dimensions using ERPs. More precisely, the study had three main objectives: (i) identify possible levels of processing at which timbre dimensions interact; (ii) assess whether filtering costs (the cost of ignoring an irrelevant dimension) and congruency effects observed in Garner interference studies reflect different interaction mechanisms, as has been argued on psychophysical grounds (e.g., Ben-Artzi & Marks, 1995; Melara & Mounds, 1993, 1994); and (iii) examine possible differences in processing among pairs of timbre dimensions as a function of their acoustic nature (spectral or temporal).

Rationale of the Study

We considered three timbre dimensions (Figure 2A): a temporal dimension—attack time (ATT); and two spectral dimensions—spectral center of gravity (SCG, amplitude-weighted mean frequency of the spectral distribution) and spectrum fine structure (controlled by a selective attenuation of even harmonics, EHA for even-harmonic attenuation). All three dimensions have been confirmed as major acoustic determinants of timbre (Caclin et al., 2005), and we have already studied both their sensory memory representations, using the MMN, and their possible interactions using Garner's paradigm for the three possible pairs of dimensions. We found evidence for mostly separate representations of the dimensions in auditory sensory memory using the MMN paradigm (Caclin et al., 2006). Yet we obtained behavioral evidence of interactive processing of the very same dimensions using Garner's paradigm (Caclin et al., 2007): For all three pairs of dimensions, we observed large filtering costs, small redundancy gains, and congruency effects. These effects were observed when judging any of the dimensions of a pair, suggesting bidirectional crosstalk between the processing channels of these timbre dimensions (see Melara & Marks, 1990a). Because we found highly similar patterns of interactions (i.e., filtering costs, redundancy gains, and congruency effects of the same size, as measured with reaction times [RT]) for the two temporal/spectral pairs (ATT/SCG and ATT/EHA) in this last study, we only tested one of the temporal/spectral pairs (ATT/EHA) and the spectral/spectral pair (SCG/EHA) in the current ERP study.

In order to discern at which levels of processing interactions between these timbre dimensions occur, we contrasted the ERPs recorded for the very same sounds in the different conditions of the Garner interference paradigm (Figure 2B and C). Our interest was thus to find modulations in the auditory ERPs depending on the type of sequence in which the sound was presented

(baseline, correlation, or filtering), as well as modulations reflecting congruency effects, rather than modulations depending on the acoustic characteristics of the sounds per se. Special care was taken when constructing the stimuli to ensure that baseline discriminabilities for the three tested timbre dimensions were equal (i.e., the RTs in the baseline conditions were equal for the three dimensions; see Caclin et al., 2007). This was done for two reasons: firstly, because it is a necessary prerequisite to interpret Garner interference studies (Melara & Mounds, 1993), and secondly, because it guarantees that any difference observed between pairs of timbre dimensions is genuine and not the consequence of the timbre difference used along one dimension (e.g., $t0$ vs. $t1$ with the notations of Figure 2) being more pronounced than along another (e.g., $c0$ vs. $c1$).

METHODS

Participants

Eighteen listeners participated in this electroencephalogram (EEG) experiment (11 women, 1 left-handed, aged 20–30 years). Eight of them reported regular musical practice. All were naive as to the purpose of the experiment and were paid for their participation. They were randomly assigned to one of two groups of nine participants (with the constraints that each group include four musically trained participants and that sex was balanced between groups). Each group was tested with only one pair of dimensions (ATT/EHA or SCG/EHA). The protocol was approved by the regional Ethical Committee (RBM number: 02-08) and all participants gave written informed consent prior to the experiment.

Stimuli

Stimuli were identical to those of our behavioral study of Garner interference between timbre dimensions (Caclin et al., 2007) for the ATT/EHA and SCG/EHA pairs. Stimuli were created by additive (Fourier) synthesis of 20 harmonics with a 311-Hz fundamental frequency (Eb_4). The amplitude envelope of the sounds was composed of a linear attack (15 or 100 msec, labeled $t0$ and $t1$, respectively), followed by a plateau and an exponential decay (see Figure 2A). The duration of the plateau was adjusted to keep perceived sound duration constant. The global shape of the spectrum was manipulated through a power relationship between harmonic amplitude and harmonic rank, which determined the value of SCG. SCG was fixed at 3 ($c0$) or 4.5 ($c1$) in harmonic rank units. The local shape of the spectrum was controlled by selectively attenuating the even harmonics by 0 ($b0$) or 10 dB ($b1$) relative to the odd ones. A stimulus was thus characterized by one of two values for each of the three dimensions investigated. The two values taken by each

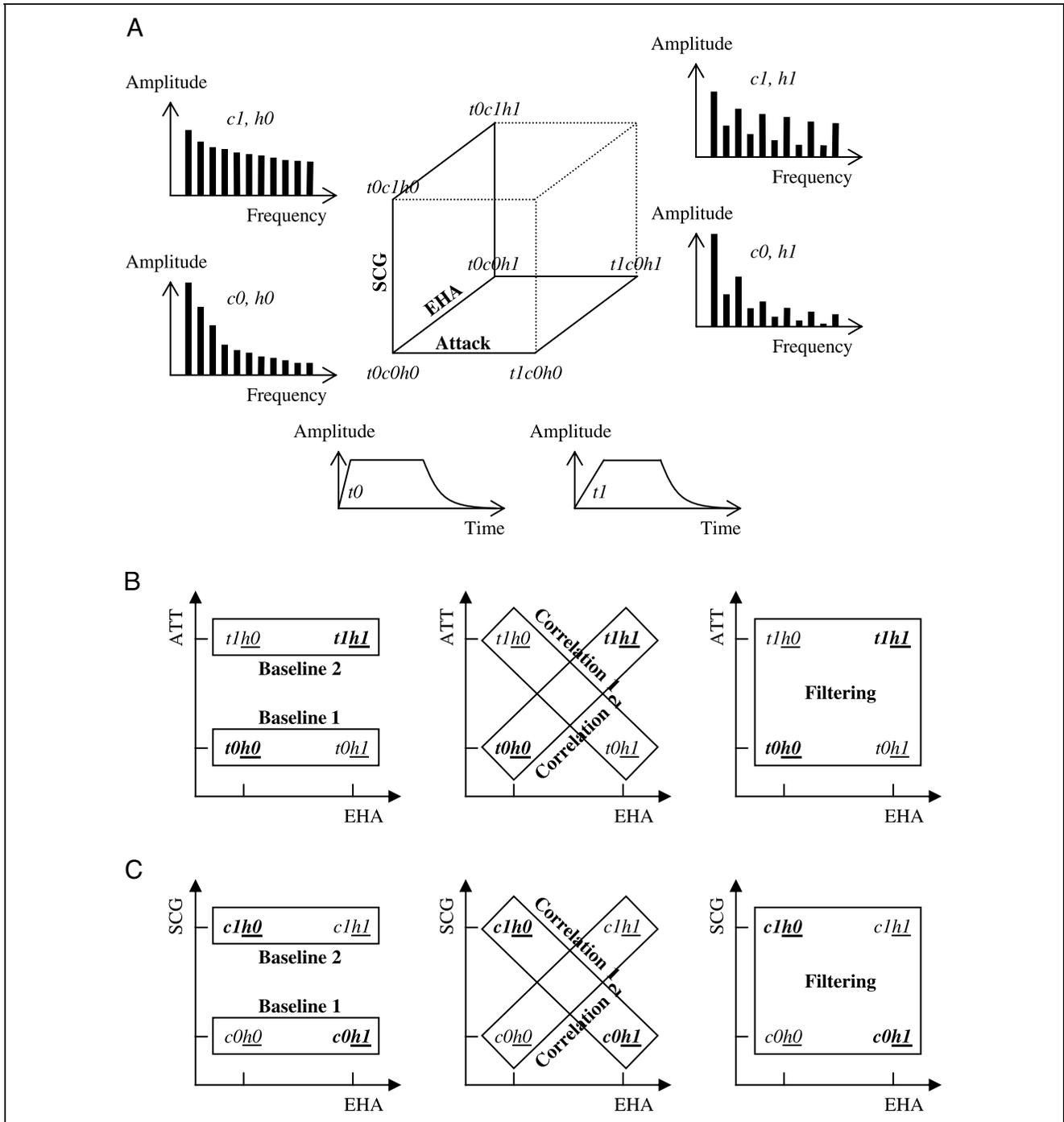


Figure 2. (A) Stimulus construction. All stimuli have the same pitch, loudness, and perceived duration but differ according to three timbre dimensions: ATT = attack ($t0 = 15$ msec, $t1 = 100$ msec); SCG = spectral center of gravity ($c0 = 3$, in harmonic rank units, $c1 = 4.5$); and EHA = even-harmonic attenuation (attenuation of even relative to odd harmonics: $b0 = 0$ dB, $b1 = -10$ dB). The two stimulus squares used in the experiment reported here are represented with solid lines. The two diagrams on the left represent the SCG variation (with $c0$ at the bottom and $c1$ at the top); the variation between the two leftmost diagrams and the two rightmost diagrams represents the spectrum fine structure variation (with $b0$ on the left and $b1$ on the right). (B,C) The five types of sequences of Garner's paradigm are represented as in Figure 1 for the ATT/EHA pair (B) and for the SCG/EHA pair (C). The participants were requested to judge the EHA dimension (as underlined). The congruent diagonals are highlighted in bold. For ATT/EHA (B), SCG was kept at a constant value ($c0$). For SCG/EHA (C), ATT was kept at a constant value ($t0$).

of the three dimensions were selected on the basis of preliminary testing in order to obtain equal performance (RT) when classifying any of these three dimensions in baseline conditions (i.e., with the other dimensions kept at constant values). Levels were adjusted to ensure constant loudness across stimuli. Details about stimulus synthesis can be found in Caclin et al. (2005).

Procedure

Participants were seated in a small, dark, sound-attenuated room, with the sounds presented over headphones using Presentation software (Neurobehavioral Systems, Albany, NY, USA). In each group of participants, there were five different types of sound sequences as described in the Introduction (two baselines, two correlations, and filtering; see Figure 2B and C). Whatever the sequence, the task of the participant was always the same: Classify the stimuli according to the EHA dimension. Verbal labels were used to describe the sounds and explain the task: Their approximate English translations would be “flat/neutral” (*plat* in French, corresponding to *b0*: no attenuation of the even harmonics) and “hollow” (*creux* in French, corresponding to *b1*: even harmonics attenuated relative to odd ones). A response was to be made after each sound, with the right hand, using one of the two buttons of a computer mouse (which was usual for the left-handed participant), and feedback was given after each response. Key assignment was balanced across participants. Participants were encouraged to respond as quickly and accurately as possible. The task was thus the same in both groups: a speeded two-alternative forced-choice decision regarding the EHA dimension.

Each sequence included 80 sound stimuli, with each stimulus occurring 800 to 1200 msec after the response to the previous one. The different sounds composing a particular sequence were equiprobable and appeared in random order (thus, in each baseline and correlation sequence, there were two different sounds presented 40 times each, and in filtering sequences, there were four sounds presented 20 times each). Each type of sequence was presented four times, except for the filtering condition, which was presented eight times in order to have the same number (160) of presentations of each sound in each experimental condition (baseline, correlation, and filtering). The order of the conditions was balanced across participants using a Latin-square design.

The experiment started with a familiarization with the sounds and a short training phase in a baseline condition. The procedure was thus exactly the same as that of behavioral experiments using Garner’s interference, except that it was a much longer experiment than usual (in such behavioral experiments, each type of sequence is typically presented only once, with sequences of about 50 sounds). Increasing the length of the experiment was necessary to allow reliable averaging of EEG data.

Statistical Analyses of the Behavioral Data

For each pair of dimensions, RTs were averaged separately for each diagonal of the stimulus square, in each experimental condition (baseline, correlation, and filtering), and were analyzed using a two-way analysis of variance (ANOVA) with condition (3 levels) and congruency (2 levels corresponding to the diagonals of the stimulus square) as within-subject factors. We also compared the RTs for the two pairs of timbral dimensions using a three-way ANOVA with condition and congruency as within-subject factors and pair-of-dimensions as a between-subjects factor. ANOVA results were corrected with the Greenhouse–Geisser procedure where applicable (epsilon and corrected *p* are reported). Significant effects (criterion alpha was set at .05) were explored using Tukey–Kramer HSD post hoc tests. Note that we term the diagonal exhibiting the shortest mean RT the “Congruent” diagonal.

EEG Recording

EEG was recorded continuously during the experiment using 34 Ag/AgCl electrodes referenced to the nose. Horizontal electrooculogram (EOG) was acquired through an additional electrode placed at the outer canthus of the right eye. Electrode impedances were kept below 5 k Ω . EEG was amplified (Neuroscan, El Paso, USA), filtered online (0.1–200 Hz), and digitized with a sampling frequency of 1000 Hz. Trials with values exceeding ± 70 μ V at any electrode site (including the EOG channel) were excluded. Averaging was done separately for each of the four sounds in each of the three experimental conditions in a time window from –200 to 800 msec around stimulus onset, with a –100 to 0 msec baseline. At least 126 trials were averaged for each participant, for each sound and condition. Evoked potentials thus obtained were digitally filtered with a low-pass filter (cutoff frequency 25 Hz, slope 24 dB/octave) and only the –100 to 700 msec time window was retained for further analysis.

Statistical and Topographical Analyses of the ERPs

ERPs for sounds belonging to the same diagonal of the stimulus square (see Figure 2B and C) were averaged in each condition, leading to a total of six different ERPs per participant (2 diagonals \times 3 conditions). With this procedure, the average acoustic content of the sounds was the same for all ERPs in each group of participants. We used an exploratory approach for the analysis of the ERP data that did not make any a priori assumption regarding the possible ERP correlates of the behavioral effects: We conducted an ANOVA on ERP amplitude at each time point and each electrode with condition (3 levels) and congruency (2 levels) as within-subject factors. We then analyzed the spatio-temporal patterns

of the main effects of the factors as well as that of their interaction. In order to correct for multiple comparisons, only effects lasting for more than 25 msec (i.e., 25 consecutive samples) with a stable topography (at least 6 electrodes) were considered significant (see Guthrie & Buchwald, 1991, for a method to correct for multiple comparison when using *t* tests at every time sample of ERPs).

ERP scalp topographies were computed using spherical spline interpolation (Perrin, Pernier, Bertrand, & Echallier, 1989; Perrin, Pernier, Bertrand, Giard, & Echallier, 1987). In order to interpret the results of the ANOVAs described above, we computed the topographies of the following ERPs in each of the two groups of participants: (i) the average of the ERPs for the four stimuli in each of the three conditions (baseline, correlation, and filtering), as well as of the following difference waves: “Filtering – Baseline” and “Correlation – Baseline,” in order to evaluate the effects of the condition factor; and (ii) the ERPs for each diagonal of the stimulus square in each of the three conditions, as well as of the “Congruent diagonal – Incongruent diagonal” difference wave in each condition, in order to evaluate the effects of the congruency factor and the Condition × Congruency interaction.

RESULTS

Behavioral Results

Behavioral results (RTs) for the two pairs of timbre dimensions are depicted in Figure 3. The results of the Condition (3 levels) × Congruency (2 levels) repeated-measures ANOVA on RTs in each group of participants are reported in Table 1. A significant effect of the condition factor was found for both pairs of dimensions. A significant effect of the congruency factor, as well as an interaction between condition and congruency, were found for the SCG/EHA pair.

Table 1. Behavioral Results: ANOVAs Exploring the Effect of Condition and Congruency on RTs for Categorizing Even-Harmonic Attenuation for Both Pairs of Dimensions

Factor (<i>df</i>)	ATT/EHA			SCG/EHA		
	<i>F</i>	ϵ (<i>GG</i>)	<i>p</i>	<i>F</i>	ϵ (<i>GG</i>)	<i>p</i>
Condition (2, 16)	9.8	0.713	.005	32.3	0.572	.0002
Congruency (1, 8)	1.7	1	.22	21.3	1	.002
Condition × Congruency (2, 16)	1.8	0.586	.22	10.2	0.723	.005

ATT = attack; SCG = spectral center of gravity; GG = Greenhouse-Geisser; EHA = even-harmonic attenuation.

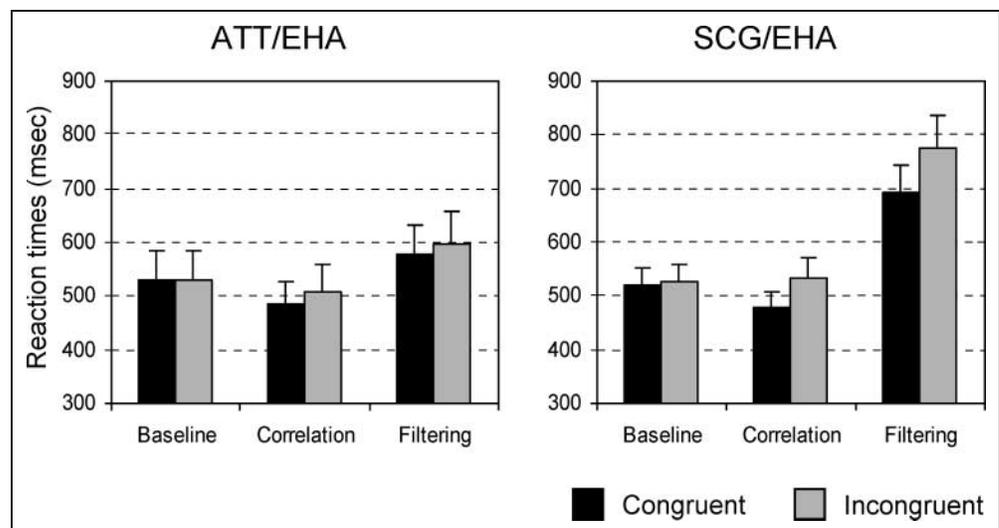
ATT/EHA Group

The effect of the condition factor was due to slower RTs in the filtering condition relative to both baseline ($p = .04$) and correlation ($p = .001$). There was no difference between baseline and correlation ($p = .24$). In our previous study with a larger sample of participants, we found a significant effect of the congruency factor for this pair of dimensions, which was mostly evident in the correlation and filtering conditions (Caclin et al., 2007). In the present study, we therefore sought such congruency effects using a nonparametric paired-value Wilcoxon test comparing RTs for the two diagonals of the stimulus square in each condition. There was no difference between the two diagonals for baseline ($p = .77$), but there was a trend for such a difference both in correlation ($p = .09$) and filtering ($p = .09$).

SCG/EHA Group

The effect of the condition factor was due to slower RTs in the filtering condition relative to both baseline

Figure 3. Behavioral data in the two groups of participants (left panel: ATT/EHA, right panel: SCG/EHA). RTs for categorizing the EHA dimension are plotted separately for each condition (baseline, correlation, and filtering) and each diagonal (congruent and incongruent), with error bars representing the standard error of the mean.



($p = .0002$) and correlation ($p = .0002$). There was no difference between baseline and correlation ($p = .85$). The Congruency \times Condition interaction was explored using a one-way ANOVA with condition as a within-subject factor on congruency scores (RT differences between the incongruent and congruent diagonals in each of the three conditions). A significant Congruency \times Condition interaction on RT data is equivalent to a significant effect of the condition factor on congruency scores. The effect of the condition factor on congruency scores arose from smaller congruency scores in baseline relative to correlation ($p = .03$) and filtering ($p = .001$), with no reliable difference between correlation and filtering ($p = .24$).

Comparison between Pairs of Dimensions

RTs were compared between the two groups using a three-way ANOVA with condition (3 levels) and congruency (2 levels) as within-subject factors and pair-of-dimensions (2 levels) as a between-subjects factor. The main effects of the condition and congruency factors and all the interactions were significant [$F(2, 32) = 41.1$, $\epsilon = 0.616$, $p < .0001$, for condition; $F(1, 16) = 18.1$, $\epsilon = 1$, $p = .0006$, for congruency; $F(1, 16) = 0.6$, $p = .45$, for pair-of-dimensions; $F(2, 32) = 10.5$, $\epsilon = 0.961$, $p = .0004$, for Condition \times Congruency; $F(2, 32) = 10.0$, $\epsilon = 0.616$, $p = .003$, for Condition \times Pair-of-dimensions; $F(1, 16) = 6.0$, $\epsilon = 1$, $p = .03$, for Congruency \times Pair-of-dimensions; and $F(2, 32) = 4.0$, $\epsilon = 0.961$, $p = .03$, for the three-way interaction]. The effect of the condition factor was expected from the analysis performed separately for each pair of dimensions. The Congruency and Condition \times Congruency effects mirror those observed for the SCG/EHA pair. More interesting here are the interactions involving the pair-of-dimensions factor. The Congruency \times Pair-of-dimensions interaction was expected because a significant effect of the congruency factor was found only for the SCG/EHA pair. To explore the Condition \times Pair-of-dimensions interaction and because baseline performance was the same in both groups by virtue of the stimulus construction (see also Caclin et al., 2007), we computed the RT difference between correlation and baseline, and between filtering and baseline for the two pairs of dimensions and compared them with t tests. This approach directly assesses redundancy gains and filtering costs, respectively. For Correlation minus Baseline, there was no difference between the two pairs of dimensions [$t(16) = 1.0$, $p = .34$], whereas a difference was observed for Filtering minus Baseline [$t(16) = 3.7$, $p = .002$]. Thus, the Condition \times Pair-of-dimensions interaction was due to a larger filtering cost for the SCG/EHA pair.

Summary

Behavioral results were similar to those of our previous behavioral experiment (Caclin et al., 2007). The fact that

the present experiment was much longer than the previous one emphasizes the robustness of the effects. As expected, both pairs of dimensions exhibited interactive behavior, in particular, with an increase of RT in the filtering condition (filtering cost) and congruency effects. Congruency effects were less pronounced for the ATT/EHA pair in the present experiment but were otherwise similar to those of the previous experiment: The fastest diagonal of the stimulus square was the same, and the difference in RTs between diagonals tended to increase from baseline to correlation and filtering (see Figure 3). As in our previous experiment, redundancy gains were quite small (and present only as numerical trends) and were asymmetrical (i.e., concerned only the congruent correlation). As expected, the SCG/EHA pair exhibited the largest cost of filtering and the strongest congruency effects due to the interaction of two spectral dimensions.

ERP Results

In both groups of participants and in all conditions, the succession of ERP components is rather similar (see Figures 4 and 5): Among the first distinguishable responses is a small positivity around 55 msec (P50) followed at 100 msec by a fronto-central negativity with a polarity reversal over the mastoids (N1). After the N1, a parieto-central positivity emerges, peaking around 170 msec (P2), followed by a frontal negativity peaking at about 220 msec (N2). A large parieto-central positivity is then observed, with a maximum around 350 msec (P3). Finally, the topography remains rather stable after 500 msec with a parietal positivity and a frontal negativity (late slow waves). This sequence of auditory ERP waves, which comprises “obligatory” auditory responses such as P50 and N1 waves (occurring even when the participant is not responding to the sound) and “non-obligatory” responses including a P3 wave, was expected as the participants were requested to make a response to each of the tones. As can be seen in Figures 4 and 5, the differences between the experimental conditions seem to consist mostly of modulations of the amplitudes of those ERPs rather than in differences in their latencies. These variations of ERP amplitudes were analyzed separately in each group of participants using two-way ANOVAs with condition (3 levels) and congruency (2 levels) as within-subject factors at each time sample and each electrode. The statistical significance of the main effects of these two factors, as well as their interaction, is presented in Figures 4 and 5.

Effect of the Condition Factor on the ERPs

Significant effects of the condition factor on the ERPs are observed in both groups of participants, in several latency ranges (Figure 4). Early ERP effects (before 200 msec) of this factor are observed for the SCG/EHA group only.

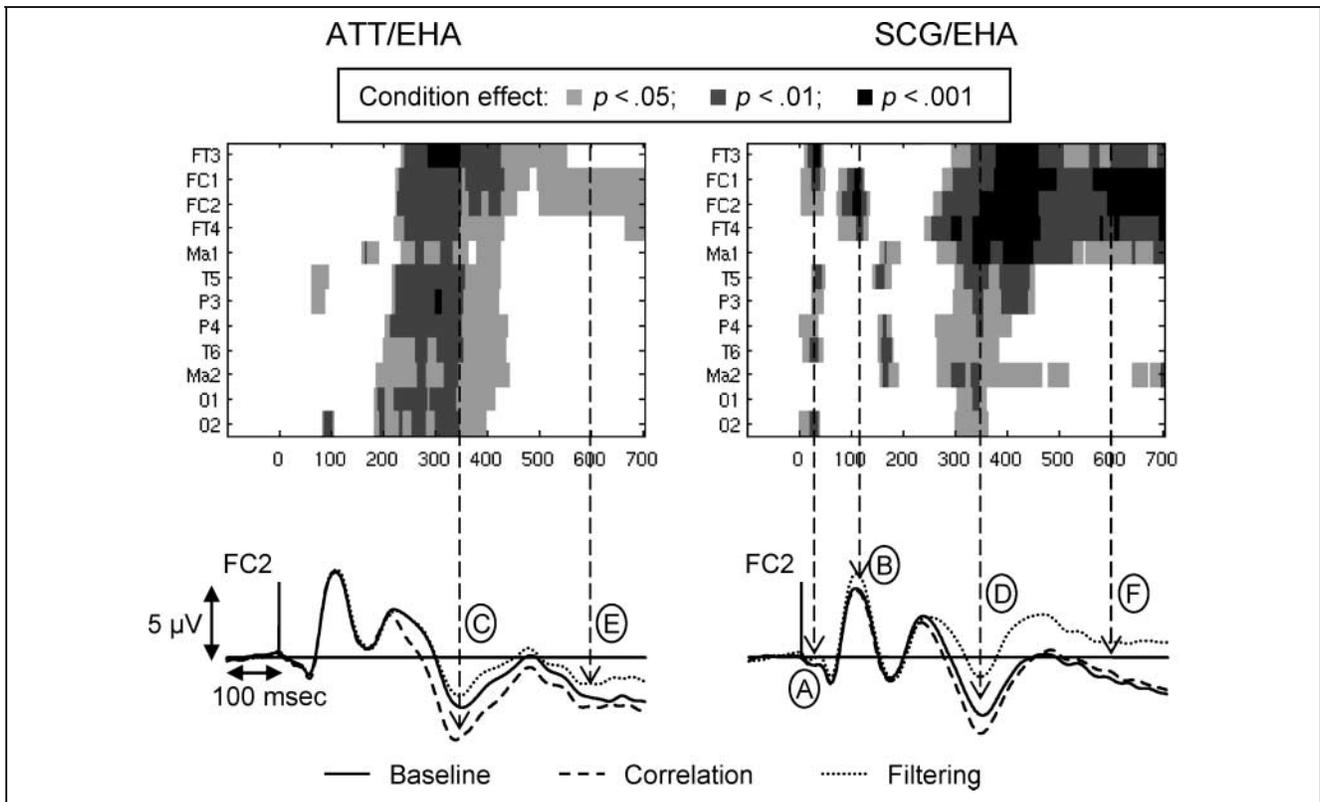


Figure 4. Statistics for the Condition \times Congruency ANOVAs performed at each time sample and each electrode on ERP amplitude, in the two groups of participants (left: ATT/EHA, right: SCG/EHA). p values for the main effect of the condition factor are reported across time at representative electrodes (left frontal sites: FT3, FC1; left temporo-parietal sites: Ma1, T5, P3; left occipital site: O1; and their right hemisphere counterparts) in the upper panels. Note that only effects lasting more than 25 msec are reported. The ERPs in the three different conditions (baseline, correlation, and filtering) are represented in the lower panels at an illustrative electrode (FC2), with negativity upward.

As early as 30 msec poststimulus onset, a significant condition effect is observed at 20 electrodes (Figure 4, marked A). The topographies of the responses around this latency in each condition (Figure 6, marked A) clearly show that the effect is due to more negative (or less positive) potentials in the filtering condition than in the baseline or correlation conditions. This first effect is followed by a second one at the latency of the N1, consisting of a more pronounced fronto-central negativity, also in the filtering condition (Figures 4 and 6, marked B).

After 200 msec, the effects of the condition factor on the ERPs are similar in *both* groups. A major effect of the condition factor on ERP amplitude starts around 250 msec after stimulus onset (Figure 4, marked C and D), and lasts until 450 to 500 msec. Compared to the responses in the baseline conditions, both groups exhibited (Figure 6, marked C and D): (i) a reduced frontal negativity and an enhanced posterior positivity in the correlation conditions, which was manifested as a widely distributed positive field in the Correlation minus Baseline maps; and (ii) an enhanced frontal negativity and a reduced posterior positivity in the filtering condition, which appeared as a wide negative potential field in the Filtering minus Baseline maps.

Finally, after 500 msec, a significant effect of the condition factor is observed for both groups over fronto-central sites, due to a prolonged frontal negativity in the filtering condition (see Figures 4 and 6, marked E and F). This effect is more pronounced in the SCG/EHA group.

Effect of the Congruency Factor on the ERPs

The effects on the ERPs of the congruency factor, as well as the Congruency \times Condition interactions, are depicted in Figure 5. For the *ATT/EHA* group, we observed a significant effect of the congruency factor at temporal and posterior sites around 70 msec (55–90 msec, Figure 5, marked G). Topographically (Figure 7, marked G), this effect corresponded to a more pronounced posterior positivity for the congruent stimuli, mostly evident in the correlation and filtering conditions. No other significant effect of the congruency factor, nor Condition \times Congruency interaction, was observed for this pair of timbre dimensions.

For the *SCG/EHA* group, there was virtually no effect of the congruency factor before 200 msec, but a significant Condition \times Congruency interaction, accompanied by a significant effect of the condition factor, was observed around 170 msec (150–190 msec, Figure 5, marked H) at

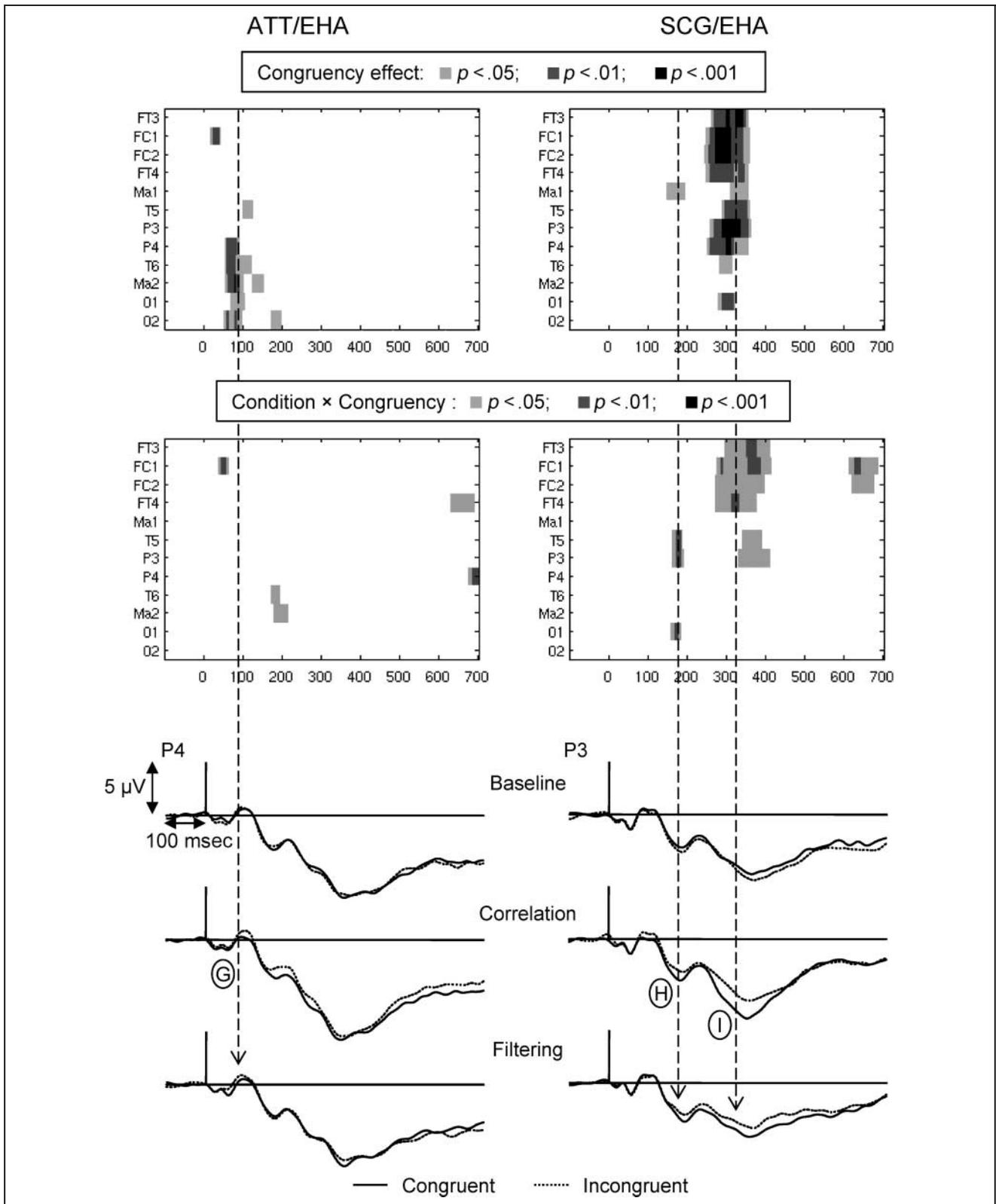


Figure 5. Statistics for the Condition \times Congruency ANOVAs performed at each time sample and each electrode on ERP amplitude, in the two groups of participants (left: ATT/EHA, right: SCG/EHA). p values for the main effect of the congruency factor (top) and the Condition \times Congruency interaction (middle) are reported across time at representative electrodes (as in Figure 4). The ERPs for congruent (plain lines) and incongruent (dotted lines) stimuli in the three different conditions (baseline, correlation, and filtering) are represented in the lower panels, at illustrative electrodes: P4 (ATT/EHA) and P3 (SCG/EHA), with negativity upward.

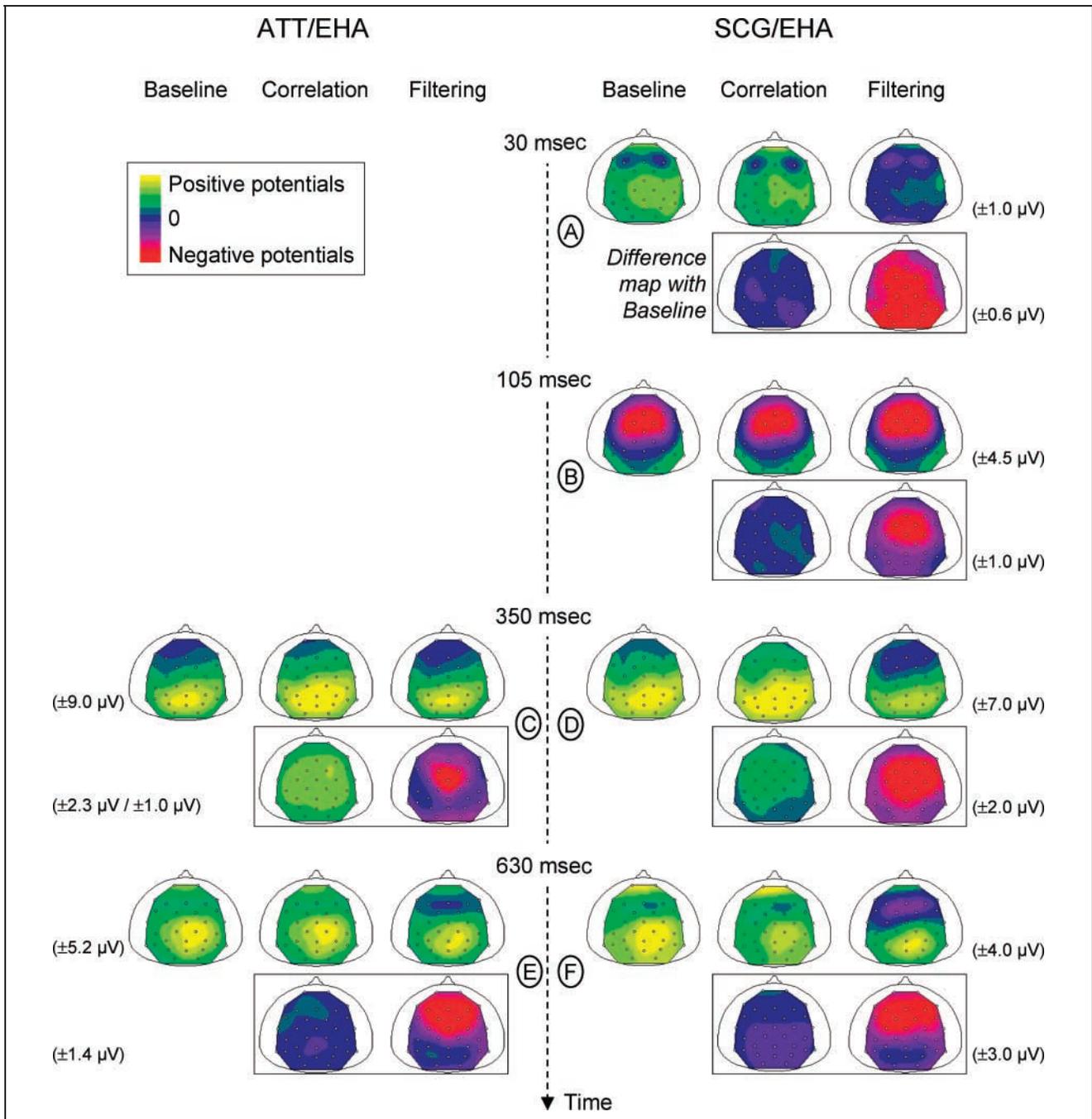


Figure 6. Topographies of the effects of the condition factor, for both pairs of timbre dimensions (left: ATT/EHA; right: SCG/EHA). Each of the effects reported in Figure 4 (marked A to F) is illustrated at a representative latency, with top views of the topographies in the three conditions (baseline, correlation, and filtering) and difference maps (Correlation minus Baseline and Filtering minus Baseline). The extreme values of the color scale are indicated in parentheses next to the maps: for each pair of dimensions, at a given latency, the scale is the same for all three conditions and the same for the two difference maps except for ATT/EHA at 350 msec.

posterior sites. This was due to more pronounced posterior positivities for congruent than for incongruent stimuli in correlation and filtering conditions (Figure 7, marked H). In the 250 to 350 msec latency range, a significant congruency effect is observed (see Figure 5, marked I), and slightly later (280–400 msec), a significant Condition \times Congruency interaction arises for this pair of dimensions. The topographies are depicted in Figure 7

(marked I), and reveal more positive potential fields for congruent stimuli in correlation and filtering conditions, with a rather broad central distribution.

DISCUSSION

This study replicates previous behavioral findings highlighting an interactive processing of timbre dimensions

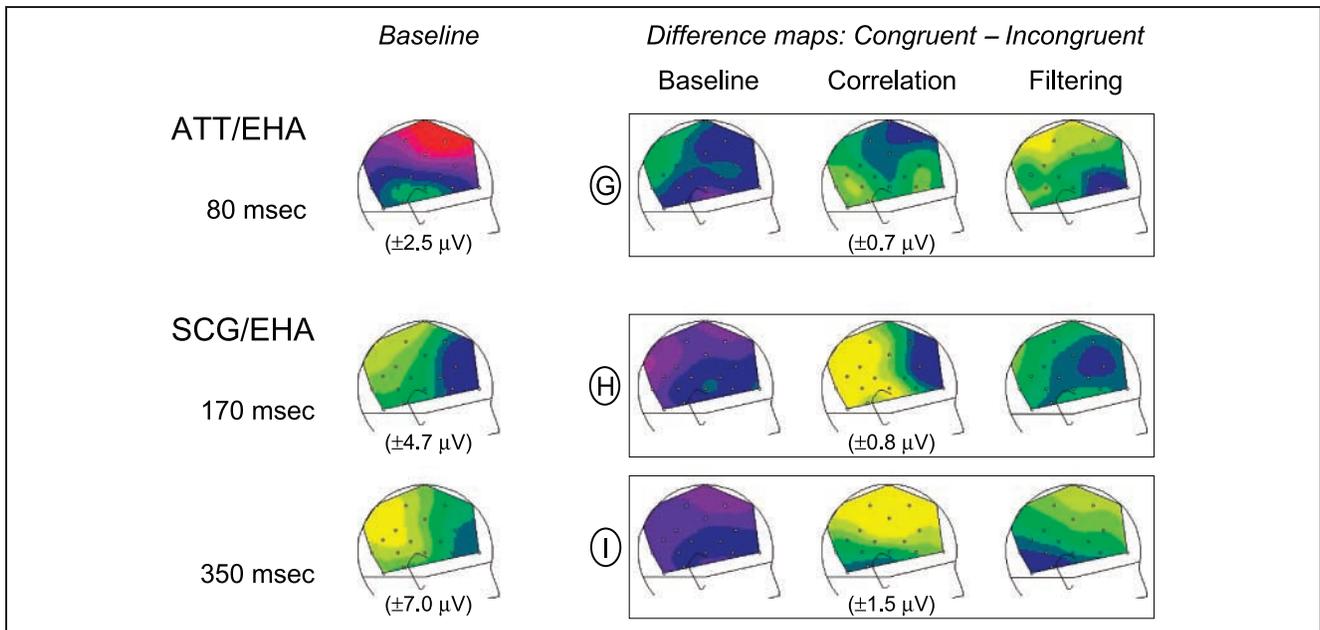


Figure 7. Topographies of the effects of the congruency factor, for both pairs of timbre dimensions (top: ATT/EHA; middle and bottom: SCG/EHA). Each of the effects reported in Figure 5 (marked G to I) is illustrated at a representative latency, with right views of the topographies in the baseline condition for reference and difference maps (Congruent diagonal minus Incongruent diagonal in the three conditions: baseline, correlation, and filtering). The extreme values of the color scale are indicated in parentheses below the maps: For each pair of dimensions, at a given latency, the scale is the same for the three difference maps.

using the Garner interference paradigm (Caclin et al., 2007), and extends them by providing insights into the neurophysiological correlates of these interactions. Behaviorally, interactions between timbre dimensions were manifested by filtering costs (i.e., the presence of irrelevant variations of a timbral dimension was accompanied by an increase in RTs for categorizing another timbre dimension), small redundancy gains (i.e., irrelevant variations of one timbral dimension correlated with the variations of the timbral dimension to be judged were accompanied by decreased RTs), and congruency effects (i.e., the pairings of the values along two timbre dimensions influenced RTs). These behavioral effects had different ERP expressions that were partly dependent on the pair of timbre dimensions being presented. The implications of these results for our understanding of the processing of timbre dimensions, and more generally for models of the interactive processing of perceptual attributes within the auditory system, are discussed below.

Processing of Timbre Dimensions

Timbre Classification in Different Timbre Contexts (Effects of the Condition Factor)

Comparing ERPs in the baseline, correlation, and filtering conditions (Figures 4 and 6) allows us to uncover neurophysiological correlates of Garner interference between timbre dimensions. For the *two pairs of timbre dimensions* investigated in this study, we observed two

types of effects of the condition factor after 200 msec, distinguishable by their scalp topographies: a broadly distributed central one in the 250–450 msec latency range and a frontal one after 500 msec.

The first effect consists of modulations of the ERP amplitude in the P3 latency range (see Figures 4 and 6, marked C and D), with the largest (most positive) amplitudes in the easiest conditions and the smallest amplitude in the most difficult conditions (as revealed with RTs). The exact functional significance of the P3 component is still being debated (see Kok, 2001; Verleger, 1997; Polich & Kok, 1995, for a review), but it has generally been associated with attention, response selection, working memory, and long-term memory operations. The topographies of the differences between the ERPs in the filtering (or correlation) and baseline conditions were, however, not that of a P3, but rather had a more central distribution, which suggests that only part of the components active in the P3 latency range were modulated. The P3 has, indeed, been described as resulting from multiple overlapping generators (e.g., Polich & Criado, 2006; Johnson, 1993).

Several factors might account for this modulation of the P3 amplitude at central sites (see Kok, 2001, for a review). One would involve response-selection processes. The stimulus–response mapping is indeed more complex in the filtering condition where each response button corresponds to two stimuli (as opposed to one in the other cases). Such an interpretation can be discarded on several grounds: (i) because this increase of stimulus–response

mapping complexity occurs similarly for both pairs of dimensions, ERP effects of comparable amplitudes would be expected in both cases, whereas the difference between filtering and baseline is larger for the SCG/EHA group than for the ATT/EHA group;¹ (ii) it would not account for the larger ERP amplitude observed in correlation relative to baseline; and (iii) it has been argued that in the case of auditory choice reaction tasks (i.e., speeded categorization such as used here), response-selection processes are reflected in a parietal subcomponent of P3 (Falkenstein, Hohnsbein, & Hoormann, 1993, 1994; Hohnsbein, Falkenstein, Hoormann, & Blanke, 1991, this component was termed the P-CR). The involvement of attentional and/or memory processes is therefore more likely here.

In order to pose a more precise hypothesis regarding the processes involved, it would help to be able to identify the particular ERP subcomponent that is modulated here. A first possibility comes from the studies cited above using choice reaction tasks (Falkenstein et al., 1993, 1994; Hohnsbein et al., 1991), in which a central subcomponent of the P3 (termed the P-SR) was found to be related to stimulus evaluation (identification), and exhibited the smallest amplitudes over central areas in the most difficult conditions (but see Dien, Spencer, & Donchin, 2004). Alternatively, the effect observed here could result from an overlap of the P3 with an activity of negative polarity (see Kaganovich, Francis, & Melara, 2006; Kok, 2001; Falkenstein et al., 1994), which is compatible with the topography of the ERPs in the latency range of the P3: A negative fronto-central component starting in the N2 latency range could persist more strongly in the filtering condition relative to baseline, and in the baseline condition relative to correlation. A family of negativities related to attention or working memory processes can indeed overlap the P3 (see Kok, 2001). For example, in the visual modality, the late part of the “NA” wave described by Ritter, Simson, and Vaughan (1983, 1988) and Ritter, Simson, Vaughan, and Macht (1982) has been associated with stimulus identification. In the auditory modality, a long-lasting processing negativity (or “PN”; see Näätänen, 1990 for a review) would reflect a matching process between the incoming stimulus and an “attentional trace,” which is defined as a voluntarily maintained representation of the target tone (see Ritter et al., 1988, for a discussion of a possible similarity between NA and PN). Altogether, this modulation of ERP amplitudes in the P3 latency range is likely to reflect a continuum in the engagement of attentional resources and working memory operations, ranging from correlation to filtering (through baseline), in order to achieve accurate stimulus recognition whatever the task difficulty. A related finding supporting this hypothesis is that varying the difficulty of a timbre deviant detection task modulates the amplitude of the ERPs in the P3 latency range (Crummer, Walton, Wayman, Hantz, & Frisina, 1994; see also Goydke, Altenmüller, Möller, & Münte, 2004).

The latest effect, a prolonged frontal negativity in the filtering condition relative to baseline and correlation after 500 msec (Figures 4 and 6, marked E and F), could be related to additional processing in the condition with the longest RTs (see Figure 3), because in conditions other than filtering, the response has already been produced before these latencies. This effect exhibits a topography and latency close to those of late frontal negative waves (Spencer, Dien, & Donchin, 2001; Loveless, Simpson, & Näätänen, 1987). However, given the late latency of this effect, a possible contamination by motor artifacts cannot be excluded.

Besides these effects observed for the two pairs of dimensions, for the *SCG/EHA pair* only, we observed a widespread negativity around 30 msec (Figure 6, marked A) and an enhanced frontal negativity in the latency range of the N1 component (Figure 6, marked B) in the filtering condition relative to baselines (and correlations). Such effects could be attributed to increased attention to the acoustic input in the most difficult experimental condition (filtering costs were significantly higher for the SCG/EHA pair than for ATT/EHA, see Figure 3). An effect of attention on the auditory N1 amplitude has indeed been reported repeatedly (see Giard, Fort, Mouchetant-Rostaing, & Pernier, 2000, for a review). Because the topography of the effect in the N1 latency range is similar to that of the N1 itself, it can be viewed as an attentional amplification of evoked responses to increase the signal-to-noise ratio (Alcaini, Giard, Echallier, & Pernier, 1995). Regarding the earliest effect (around 30 msec), it is known that auditory attention can modify neuronal responses as early as 20 msec poststimulus onset (Woldorff et al., 1993; Woldorff & Hillyard, 1991) and up to the cochlear level (Giard, Collet, Bouchet, & Pernier, 1994).

To summarize, these ERP results are compatible with Garner interference between timbre dimensions arising as a consequence of various difficulties in identifying the particular stimulus presented in the different types of sound sequences. ERP modulations as a function of the experimental condition are observed mostly in fairly late latency ranges (after 250 msec), suggesting that early sensory analysis is not much affected. Nevertheless, when it becomes particularly difficult to dissociate the relevant from the irrelevant dimension, which is the case in the filtering condition for the SCG/EHA pair, participants seem to engage additional attentional resources that affect the early stages of sensory processing, putatively to improve the quality of the information extracted from the acoustic input.

Processing Different Pairings of Values along Timbre Dimensions (Effects of the Congruency Factor)

A second type of effect observed behaviorally in Garner interference studies is that of congruency, which appears as an RT difference between the two diagonals of the stimulus square (corresponding to different pairings

of the values taken by the two dimensions). These effects are typically observed in correlation and filtering conditions only (see Caclin et al., 2007; Ben-Artzi & Marks, 1995; Melara & Mounts, 1994), that is, when the irrelevant dimension varies in the sound sequence, irrespective of whether these variations are redundant (correlation) or not (filtering) with the variations of the dimension to be judged. Such behavioral effects of congruency were observed in the present study (although only as a trend for the ATT/EHA group), with ERP manifestations accordingly observed mostly in the correlation and filtering conditions (Figure 7).

An effect of the congruency factor on the ERPs was first observed in the *ATT/EHA group*, starting approximately at 55 msec after stimulus onset (Figures 5 and 7, marked G). Interestingly, the effect exhibited a rather stable topography in the 55 to 100 msec latency range, whereas the general topography of the ERPs evolved from that of a P50 to that of an N1. This result suggests that congruency does not modulate the main generators of the P50 and N1, but rather affects some other early auditory activity. In the *SCG/EHA group*, a significant Condition \times Congruency interaction was observed from about 150 msec poststimulus onset (Figures 5 and 7, marked H). It is interesting to note that despite the fact that these early congruency effects arise at different latencies for the two pairs of timbre dimensions tested, they, nevertheless, have similar (albeit not strictly identical) topographies. This suggests that similar mechanisms with strongly overlapping neural substrates operate in the two cases. The latency differences of the congruency effects are likely to reflect genuine differences between the two pairs of timbre dimensions, rather than spurious differences due to stimulus construction. Indeed, the stimuli were constructed so that the baseline discriminabilities of all three dimensions were identical: RTs for discriminating the two attack times, the two SCGs, or the two EHA values used in this study are virtually equal (see Caclin et al., 2007). Therefore, our results suggest that attack time can interfere with the processing of EHA (a measure of spectrum fine structure) earlier than can SCG (controlling the global shape of the spectrum). To further interpret this result, it is important to remember that congruency effects in Garner interference studies are contextual effects arising only when the irrelevant dimension varies in the stimulus set (i.e., in correlation and filtering but not in baseline conditions). With the hypothesis that such optional influences of one dimension over the processing of another are likely to require some prior degree of separation between their respective processing channels (Melara, Marks, & Potts, 1993), our results may indicate a segregation of temporal and spectral processing channels that occurs earlier than does the segregation of two different spectral processing channels.

Finally, for the *SCG/EHA pair* only, there was a significant Condition \times Congruency interaction around

350 msec (Figures 5 and 7, marked I). The topography of this effect, as well as its relationship with RTs, mirrors the effect of the condition factor observed in the same latency range, that is, more negative (positive) potential fields are observed fronto-centrally when the longest (shortest) RTs are obtained. It is therefore likely that the effects of condition (discussed above) and congruency factors on the ERP in the P3 latency range reflect similar processes.

Implications for a Model of Interactive Processing of Auditory Attributes

As discussed in the Introduction, one way to conciliate the results of Garner interference studies, which have revealed interactive processing of auditory dimensions, with the results of sensory memory studies, which argue for separate processing of the very same dimensions, is to postulate a model where several auditory attributes are processed within separate channels, yet with a possibility of crosstalk between those channels (see also Hansen & Hillyard, 1983, for a model postulating parallel, yet not independent, attentional processing of different auditory attributes). Our findings allow us to specify a number of properties that such a model would have to incorporate. Firstly, we have observed that different behavioral effects (filtering costs, redundancy gains, and congruency effects) have distinct ERP signatures, with each effect having a particular spatio-temporal expression over the scalp. This observation supports the view that several mechanisms of interaction are operating, as already suggested on behavioral grounds (Ben-Artzi & Marks, 1995; Melara & Mounts, 1993, 1994). The latencies of the different ERP effects mark an upper bound for the timing of these interactions.

We know from MMN studies that around 100 to 250 msec poststimulus onset, there already exist in the auditory cortex individualized representations pertaining to a variety of elementary perceptual attributes: pitch, loudness, duration, timbre dimensions, and so forth (see Introduction). Yet, in the present study, we found that filtering costs were accompanied by a reduction of the P3 amplitude over central sites, starting at about 250 msec, possibly reflecting the overlap with an activity of negative polarity. It therefore seems that additional processing of the initially available dimensional information is necessary to categorize the stimuli adequately. The few other studies exploring the ERP correlates of Garner interference report similar findings, although none of these previous studies used basic auditory dimensions as was done here (Kaganovich et al., 2006, used the dimensions of vowel and talker; Lew, Chmiel, Jerger, Pomerantz, & Jerger, 1997, used Stroop-like pairs of dimensions, with one sensory and one linguistic dimension). It is quite unlikely, however, that such a negative component would be a specific correlate of Garner interference; it would rather reflect fairly

general cognitive operations, implicating attentional or working memory processes (see also Kok, 2001).

In addition, our data shed light on the neuronal expression of congruency effects, which may also occur in fairly late processing stages (after 200 msec), as observed here in the case of the SCG/EHA pair. However, importantly, congruency effects appear to be primarily associated with modulations of the early auditory responses (before 200 msec). As mentioned in the Introduction, congruency effects (as measured in Garner interference studies) share some similarities with Stroop interference. ERP studies of the Stroop effect have generally reported differences between the ERPs for congruent and incongruent trials in the 300 to 500 msec latency range (e.g., West, 2003; Liotti, Woldorff, Perez, & Mayberg, 2000; Ilan & Polich, 1999; Lew et al., 1997), as observed here for the SCG/EHA pair. The results suggest that there may be similarities in the neural mechanisms for processing congruency in Garner and Stroop interference tasks. Nevertheless, these Stroop studies typically do not report congruency effects as early as those described here (i.e., before 200 msec, although see Lew et al., 1997). These earlier congruency effects observed here using the Garner interference paradigm might actually be due to the nature of the dimensions manipulated (perceptual dimensions here vs. perceptual and semantic dimensions in Stroop interference experiments). These early interactions in auditory areas might thus be a fairly specific correlate of congruency effects resulting from interactions among basic auditory dimensions.

Finally, the strength and latencies of these different interactive processes seem to depend partly on the particular pair of dimensions considered. Here we observed more pronounced behavioral and ERP signatures of interactive processing for two spectral timbral dimensions (SCG/EHA) than for a pair with temporal and spectral timbral dimensions (ATT/EHA). The latencies of the congruency effects in the ERPs also depended on the pair of timbral dimensions, and we have proposed that this could be related to the stage of auditory analysis from which the processing channels for the two dimensions considered would actually be separate. The generality of these findings to other pairs of auditory (or visual) dimensions remains to be tested.

Conclusion

We have argued that the processing of timbre dimensions could be understood as occurring in mostly separate channels, with, in particular, separate representations of these dimensions in sensory memory, but which could interact at other processing stages than that of sensory memory. The current study shows that such interactions are reflected in the ERPs during at least two distinct processing stages: sensory analysis (as revealed by early congruency effects on the ERPs), and later stages of pro-

cessing probably involved in stimulus identification (as revealed by both congruency and condition effects on the ERPs). Separate or interactive processing between the various perceptual dimensions of an auditory object appears to depend on the processing levels (and partly on the particular dimensions) considered and might furthermore depend on the task at hand.

Acknowledgments

This work was supported by a CNRS (Centre National de la Recherche Scientifique) grant (CTI01-28). Author S. M. is supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

Reprint requests should be sent to Anne Caclin, INSERM U821, 69675 Bron Cedex, France, or via e-mail: caclin@lyon.inserm.fr.

Note

1. This hypothesis was verified in a two-way ANOVA on the mean amplitude of the ERP in the 300 to 400 msec range, with condition as a within-subject factor (2 levels: baseline and filtering) and pair-of-dimensions as a between-subject factor (2 levels). The interaction between the two factors was significant [$F(1, 16) = 18.7, p = .03$].

REFERENCES

- Alain, C., Achim, A., & Woods, D. L. (1999). Separate memory-related processing for auditory frequency and patterns. *Psychophysiology*, *36*, 737–744.
- Alcaini, M., Giard, M. H., Echallier, J. F., & Pernier, J. (1995). Selective auditory attention effects in tonotopically organized cortical areas: A topographic ERP study. *Human Brain Mapping*, *2*, 159–169.
- Ben-Artzi, E., & Marks, L. E. (1995). Congruence effects in classifying auditory stimuli: A review and a model. In C. A. Possamaï (Ed.), *Fechner Day 95, Cassis, France* (pp. 145–150). Cassis, France: International Society for Psychophysics.
- Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M. H., et al. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, *18*, 1959–1972.
- Caclin, A., Giard, M. H., Smith, B. K., & McAdams, S. (2007). Interactive processing of timbre dimensions: A Garner interference study. *Brain Research*, *1138*, 159–170.
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *Journal of the Acoustical Society of America*, *118*, 471–482.
- Clément, S., Demany, L., & Semal, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America*, *106*, 2805–2811.
- Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., & Frisina, R. D. (1994). Neural processing of musical timbre by musicians, nonmusicians and musicians possessing absolute pitch. *Journal of the Acoustical Society of America*, *95*, 2720–2727.
- Dien, J., Spencer, K. M., & Donchin, E. (2004). Parsing the late positive complex: Mental chronometry and the ERP

- components that inhabit the neighborhood of the P300. *Psychophysiology*, *41*, 665–678.
- Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1993). Late visual and auditory ERP components and choice reaction time. *Biological Psychology*, *35*, 201–224.
- Falkenstein, M., Hohnsbein, J., & Hoormann, J. (1994). Effects of choice complexity on different subcomponents of the late positive complex of the event-related potential. *Electroencephalography and Clinical Neurophysiology*, *92*, 148–160.
- Garner, W. R. (1974). *The processing of information and structure*. New York: Erlbaum.
- Giard, M. H., Collet, L., Bouchet, P., & Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain Research*, *633*, 353–356.
- Giard, M. H., Fort, A., Mouchetant-Rostaing, Y., & Pernier, J. (2000). Neurophysiological mechanisms of auditory selective attention in humans. *Frontiers in Bioscience*, *5*, D84–D94.
- Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., et al. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, *7*, 133–143.
- Goydke, K. N., Altenmüller, E., Möller, J., & Münte, T. F. (2004). Changes in emotional tones and instrumental timbre are reflected by the mismatch negativity. *Cognitive Brain Research*, *21*, 351–359.
- Grau, J. W., & Kemler-Nelson, D. G. (1988). The distinction between integral and separable dimensions: Evidence for integrality of pitch and loudness. *Journal of Experimental Psychology: General*, *117*, 347–370.
- Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, *61*, 1270–1277.
- Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology*, *28*, 240–244.
- Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harschberger, M. L. (1997). Methodological issues in timbre research. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 253–307). Hove, UK: Psychology Press.
- Handel, S. (1995). Timbre perception and auditory object identification. In B. C. J. Moore (Ed.), *Hearing* (pp. 425–461). San Diego: Academic Press.
- Hansen, J. C., & Hillyard, S. A. (1983). Selective attention to multidimensional auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *9*, 1–19.
- Hohnsbein, J., Falkenstein, M., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components: I. Simple and choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*, 438–446.
- Ilan, A. B., & Polich, J. (1999). P300 and response time from a manual Stroop task. *Clinical Neurophysiology*, *110*, 367–373.
- Johnson, R. J. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*, 90–97.
- Kaganovich, N., Francis, A. L., & Melara, R. D. (2006). Electrophysiological evidence for early interaction between talker and linguistic information during speech perception. *Brain Research*, *1114*, 161–172.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577.
- Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 739–751.
- Lew, H., Chmiel, R., Jerger, J., Pomerantz, J. R., & Jerger, S. (1997). Electrophysiologic indices of Stroop and Garner interference reveal linguistic influences on auditory and visual processing. *Journal of the American Academy of Audiology*, *8*, 104–118.
- Liotti, M., Woldorff, M. G., Perez, R., III, & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color–word interference effect. *Neuropsychologia*, *38*, 701–711.
- Loveless, N. E., Simpson, M., & Näätänen, R. (1987). Frontal negative and parietal positive components of the slow wave dissociated. *Psychophysiology*, *24*, 340–345.
- McAdams, S. (1993). Recognition of sound sources and events. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 146–198). Oxford: Oxford University Press.
- McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., & Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: Common dimensions, specificities and latent subject classes. *Psychological Research*, *58*, 177–192.
- Melara, R. D., & Marks, L. E. (1990a). Dimensional interactions in language processing: Investigating directions and levels of crosstalk. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 539–554.
- Melara, R. D., & Marks, L. E. (1990b). Interaction among auditory dimensions: Timbre, pitch and loudness. *Perception & Psychophysics*, *48*, 169–178.
- Melara, R. D., & Marks, L. E. (1990c). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 398–414.
- Melara, R. D., Marks, L. E., & Potts, B. C. (1993). Early-holistic processing or dimensional similarity? *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1114–1120.
- Melara, R. D., & Mounts, J. R. W. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition*, *21*, 627–645.
- Melara, R. D., & Mounts, J. R. W. (1994). Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty. *Perception & Psychophysics*, *56*, 73–90.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive functions. *Behavioral and Brain Sciences*, *13*, 201–288.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*, 826–859.
- Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*, 184–187.
- Perrin, F., Pernier, J., Bertrand, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface splines interpolation. *Electroencephalography and Clinical Neurophysiology*, *66*, 75–81.
- Polich, J., & Criado, J. R. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology*, *60*, 172–185.
- Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, *45*, 103–146.

- Ritter, W., Simson, R., & Vaughan, H. G. J. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. *Psychophysiology*, *20*, 168–179.
- Ritter, W., Simson, R., & Vaughan, H. G. J. (1988). Effects of the amount of stimulus information processed on negative event-related potentials. *Electroencephalography and Clinical Neurophysiology*, *69*, 244–258.
- Ritter, W., Simson, R., Vaughan, H. G. J., & Macht, M. (1982). Manipulation of event-related potential manifestations of information processing stages. *Science*, *218*, 909–911.
- Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, *89*, 2404–2410.
- Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, *38*, 343–358.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.
- Verleger, R. (1997). On the utility of P3 latency as an index of mental chronometry. *Psychophysiology*, *34*, 131–156.
- West, R. (2003). Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. *Neuropsychologia*, *41*, 1122–1135.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., et al. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *90*, 8722–8726.
- Woldorff, M. G., & Hillyard, S. A. (1991). Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalography and Clinical Neurophysiology*, *79*, 170–191.