From Psychoacoustics to Brain Waves: A Longitudinal Approach to Novel Word Learning

Mylène Barbaroux, Arnaud J Noreña, Maud Rasamimanana, Eric Castet, Mireille R Besson

From Psychoacoustics to Brain Waves: A Longitudinal Approach to Novel Word Learning

Mylène Barbaroux, Arnaud Norena, Maud Rasamimanana, Eric Castet, and Mireille Besson

Abstract

■ Musical expertise has been shown to positively influence high-level speech abilities such as novel word learning. This study addresses the question whether low-level enhanced perceptual skills causally drives successful novel word learning. We used a longitudinal approach with psychoacoustic procedures to train 2 groups of nonmusicians either on pitch discrimination or on intensity discrimination, using harmonic complex sounds. After short (approximately 3 hr) psychoacoustic training, discrimination thresholds were lower on the specific feature (pitch or intensity) that was trained. Moreover, compared to the intensity group, participants trained on pitch were faster to categorize words varying in pitch. Finally, although the N400 components in both the word learning phase and in the semantic task were larger in the pitch group than in the intensity group, no between-group differences were found at the behavioral level in the semantic task. Thus, these results provide mixed evidence that enhanced perception of relevant features through a few hours of acoustic training with harmonic sounds causally impacts the categorization of speech sounds as well as novel word learning. These results are discussed within the framework of near and far transfer effects from music training to speech processing. ■

INTRODUCTION

Positive transfer effects from music practice to various levels of language processing have been demonstrated in many experiments (e.g., see Besson, Barbaroux, & Dittinger, 2017, for a review). Of most interest here, some studies focused on the benefits of musical practice on high-level speech ability such as learning a new language. Results typically showed that musicianship leads to enhanced proficiency in processing and learning nonnative words, at least when these words are presented orally. For instance, Alexander, Wang, and Bradlow (2005) showed higher Mandarin tone identification and discrimination abilities in native English speakers with musical experience than in nonmusicians. Wong and Perrachione (2007) also reported that amateur musicians with increased pitch patterns discrimination abilities (Mandarin tones superimposed on English pseudowords) learned new pseudowords better than nonmusicians. Cooper and Wang (2012) showed that English adult musicians outperformed nonmusicians when learning novel Cantonese words via picture–word associations. Word learning success was positively correlated with tone identification scores and, to a lesser extent, to musical aptitude scores. Moreover, directly comparing the effects of musicianship and of tone language experience (Thai), they found that both musical experience and tone language background positively influenced Cantonese word learning. However, combined musical and tone language knowledge was not associated with better word learning than each ability in isolation. Taken together, these different results provided evidence that auditory expertise improved word learning abilities in a tone language.

Recently, we examined the effects of musical expertise on Thai word learning across the life span, using both behavioral and electrophysiological measures (Dittinger, Scherer, Jäncke, Besson, & Elmer, 2019; Dittinger, Valizadeh, Jäncke, Besson, & Elmer, 2018; Dittinger, Chobert, Ziegler, & Besson, 2017; Dittinger et al., 2016). After a learning phase in which picture–word associations were presented several times, participants were tested on the efficiency of learning in a matching task and in a semantic task. In the matching task, they had to decide whether a picture–word association matched or mismatched the previously learned pairs and, in the semantic task, they had to decide whether the newly learned word was semantically related or not to a new picture (not seen before in the experiment). Results showed that musicians, children and younger adults, made overall fewer errors than nonmusicians in both the matching and semantic tasks. This was taken to suggest that musicians learned the picture–word association and generalized the meaning of the words to new, semantically related pictures more easily than nonmusicians. At the electrophysiological level, the N400, taken as a marker of semantic processing (Kutas & Hillyard, 1980, 1984), developed faster in musicians, children, younger, and older adults, than in nonmusicians in the word learning phase and musicians showed larger N400s over parietal sites to semantically unrelated compared to related words in the semantic task.

Université Publique de France

Although interesting, these results did not test for causal links between auditory training, pitch perception, tone identification, and word learning. The only way to determine causality in humans is to conduct longitudinal studies while training participants on specific dimensions. To our knowledge, only a few longitudinal studies have directly tested for causality in adults (for a pitch-based music program in children; see Patscheke, Degé, & Schwarzer, 2018). For instance, Wang, Spence, Jongman, and Sereno (1999) trained participants to identify Mandarin tones during the course of 2 weeks. They showed a robust improvement of tone perception in trained compared to untrained participants that generalized across tones and across talkers and that remained stable 6 months after training. Similarly, in the study by Cooper and Wang (2013), which was conducted as a follow-up of the study reported above (Cooper & Wang, 2012), nonmusician native English speakers were trained to identify monosyllabic Cantonese tones (with feedback and training of 3 × 30 min over 1 week). Trained nonmusicians outperformed untrained nonmusicians, and their level of performance was not significantly different from untrained musicians, showing that a specific short and intensive training provided, at least in the short-term, advantages similar to 10 years of musical practice. Moreover, tone identification ability was a significant predictor of word learning success. Based on these results, the authors proposed the "phonetic–phonological–lexical continuity" hypothesis (Cooper & Wang, 2013; Cooper & Wang, 2012; Wong & Perrachione, 2007), also called the "cascading hypothesis" (e.g., Besson et al., 2017; Dittinger et al., 2016) after which low-level training (e.g., tone identification training) improved high-level processes (e.g., associating a meaning to tone words). Importantly, in the Cooper and Wang (2013) study, nonmusician participants were directly trained on the discrimination of Cantonese tones. To our knowledge, the specific role of acoustic training with nonspeech stimuli on lexical tone perception has never been investigated. Because pitch is a relevant acoustic parameter for lexical tone discrimination, the question we asked here is whether acoustic training on pitch with complex harmonic tones positively impacts novel word learning of Thai words.

Previous results using psychoacoustics training methods demonstrated that perceptual abilities of nonexperts can be enhanced in a relatively short time. For instance, Micheyl et al. (Micheyl, Delhommeau, Perrot, & Oxenham, 2006) showed that, although pitch discrimination thresholds for nonmusicians were more than 6 times higher than for musicians before training, nonmusicians' thresholds decreased across training sessions to become as low as for musicians after four to eight training hours. However, whether such improvements reflected improved auditory perception or whether they were mediated by attention remains an open question. In an interesting study, Amitay et al. (Amitay, Irwin, & Moore, 2006) showed that participants trained in a pitch discrimination task in which they were asked to compare identical stimuli (0-Hz difference) were able to discriminate smaller pitch differences after than before training. Moreover, participants passively exposed to the stimuli used in the pitch discrimination task also improved their discrimination abilities, as well as participants not exposed to the auditory stimuli but playing visuospatial games. These results suggest that stimulus exposure, as well as focusing attention on irrelevant stimuli, positively influenced discrimination abilities, simply because participants are trained to perform a task. This aspect was considered in the auditory training protocol described below.

Based on previous literature (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016; Cooper & Wang, 2012, 2013; Wong & Perrachione, 2007; Wang et al., 1999), we aimed at testing the phonetic–phonological–lexical continuity or cascading hypothesis after which fine-grained perception of acoustic parameters drives more efficient novel word learning, via better phonological categorization of the novel words and strengthened associations to word meaning. To this end, two groups of participants were trained, using psychoacoustics methods, on auditory discrimination of complex harmonic sounds varying on pitch (experimental group) or on intensity (control group), to rule out the effects of exposure, attention, and arousal (Amitay et al., 2006). In a two-alternative forced-choice procedure, participants decided which one of two successive sounds was the highest (pitch task) or the loudest (intensity task). The difference in pitch frequency or intensity level ($\Delta\Delta f$ or Δi) was varied according to the participants' responses in a two-down one-up rule. Pitch and intensity discrimination thresholds were computed before and after training. After three training sessions (50 min each, at least separated by 2 days) were completed, the same protocol was used as in Dittinger et al. (2016). As described above, participants performed a pitch categorization task on monosyllabic Thai words (both before and after training), they learned the meaning of these words based on picture–word associations, and finally, they performed a matching and a semantic task, so that we could compare novel word learning efficiency between the two groups. Both behavioral data and event-related brain potentials were analyzed in these different tasks.

Being interested in transfer effects from auditory perception to word learning, we tested the general hypothesis that training pitch perception of nonlinguistic, complex harmonic sounds improves the ability to learn novel words in a tone language in which pitch is linguistically relevant. In detail, we expected decreased pitch discrimination thresholds in the group trained on pitch (pitch group [PG]) and decreased intensity discrimination thresholds in the group trained on intensity (intensity group [IG]) between pre and posttraining sessions. Moreover, if better auditory perception drives higher level cognition and contributes to explain the different patterns of results found for adult professional musicians and nonmusicians in the Dittinger et al. (2016) study, we predicted that the PG would outperform (lower error rates and/or shorter RTs) the IG in tasks requiring to categorize and to learn the meaning of Thai words.

Turning to the electrophysiological aspects, previous experiments in novel word learning experiments revealed several findings of interest. First, independently of whether novel words are presented in prime-target lexical decision tasks (McLaughlin, Osterhout, & Kim, 2004), along with their definitions (Balass, Nelson, & Perfetti, 2010; Perfetti, Wlotko, & Hart, 2005), in sentence contexts (Borovsky, Elman, & Kutas, 2012; Borovsky, Kutas, & Elman, 2010; Mestres-Missé, Rodriguez-Fornells, & Münte, 2007), in short story contexts (Batterink & Neville, 2011), or paired with pictures of known (Dittinger et al., 2016; Dobel, Lagemann, & Zwitserlood, 2009) or novel objects (Angwin, Phua, & Copland, 2014), an N400 component develops very fast (e.g., within a few minutes [Dittinger et al., 2016], with a single word exposure [Borovsky et al., 2010]) when novel words acquire meaning ("fast mapping"; Carey, 1978). Second, in the early phase of novel word acquisition, the N400 shows a frontal distribution (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016; Borovsky et al., 2010; Mestres-Missé et al., 2007) that is taken to reflect the formation of new associations in working or short-term memory and the initial building up of word representations in episodic memory (Rodríguez-Fornells, Cunillera, Mestres-Missé, & de Diego-Balaguer, 2009). Finally, when the meaning of the novel word has been consolidated through repeated exposures and is integrated within preexisting semantic networks, the N400 shows the typical centroparietal distribution (Kutas, Van Petten, & Besson, 1988). Note that in the Dittinger et al. (2016) experiment, results also showed that an N200 component, taken to reflect early contextual influences (van den Brink, Brown, & Hagoort, 2001) and phonological processing (Connolly & Phillips, 1994), always preceded the N400 component in the learning phase and in the matching and semantic tasks.

Based on these results and because we used the same experimental design as in Dittinger et al. (2016), we expected similar electrophysiological patterns to develop during word learning and testing. Specifically, we predicted that both N400 and N200 components would rapidly develop in the learning phase over frontocentral sites, with larger amplitudes in the second than in the first learning block and with larger amplitude in the PG than in the IG. Moreover, in the matching and semantic tasks, we expected larger N200 and N400 components to mismatch than to match words and to unrelated than to related words over centroparietal sites together with larger amplitude in the PG than in the IG.

METHODS

Participants

Twenty-eight nonmusicians, all right-handed, participated in the experiment. None of the participants had prior experience with psychoacoustic tasks or musical practice, except as part of the cursus in primary school. All participants had normal hearing, as defined by audiometric pure-tone absolute thresholds below 20 dB HL at octave frequencies between 500 and 8000 Hz. They were pseudorandomly assigned to one of two auditory training groups: the PG or the IG. Assignment was pseudorandom rather than random to control for age (PG = 23.4 years old; IG = 23.8 years old, $p = .84$), sex (eight women and five men in each group), and practice of foreign languages (not bilingual, and two foreign languages learned at school). None of the participants was familiar with tone languages, and none was dyslexic (based on their own knowledge). Two of them were excluded from the analyses, because they were outliers in several tasks and/or had noisy EEG traces. The study was conducted in accordance to the Declaration of Helsinki and with the ethical guidelines of Aix-Marseille University. Participants signed an informed consent document, and they were told that they could stop the experiment at any moment. All participants received a remuneration for their participation.

Stimuli

Monosyllabic Words

Eight monosyllabic consonant–vowel words derived from Thai were spoken by a Thai–French bilingual woman, who pronounced five versions of each word to reproduce the natural variations encountered within a language. To equate the different parameters, some words were resynthesized in duration or F0 using the Praat software (Boersma & Weenink, 2011). Words varied in pitch (fundamental frequency or F0), duration, and VOT. If all parameters were relevant to distinguish words, pitch was the more important regarding our hypothesis. Words could be either low-tone (/pa1/, /p^{h}a1/, /pa:1/, /p^{h}a:1/; mean F0 = 175 Hz) or mid-tone (/pa0/, /p^{h}a0/, /pa:0/, /p^{h}a:0/; mean F0 = 215 Hz), either short (/pa1/, /p^{h}a1/, /pa0/, /p^{h}a0/; mean vowel length = 195 msec) or long vowel duration (/pa:1/, /p^{h}a:1/, /pa:0/, /p^{h}a:0/; vowel = 458 msec), and either with (/p^{h}a1/, /p^{h}a:1/, /p^{h}a0/, /p^{h}a:0/; mean VOT = 58 msec) or without aspiration of the consonant (/pa1/, /pa:1/, /pa0/, /pa:0/; VOT = 18 msec; see Figure 1). Sound pressure level was normalized at 60 dB for all stimuli.

Complex Sounds

Stimuli were created as follows. First, a harmonic complex sound was built by adding 20 tones in phase at the harmonic frequencies. The harmonic complex was then multiplied by the envelope extracted from one of the representative monosyllabic Thaï word. The F0 of the standard stimulus was equal to 195 Hz (the F0 of the target or comparison tone was varied in the pitch discrimination task). The duration and level of the sounds were fixed and equal to 327 msec and 60 dB.

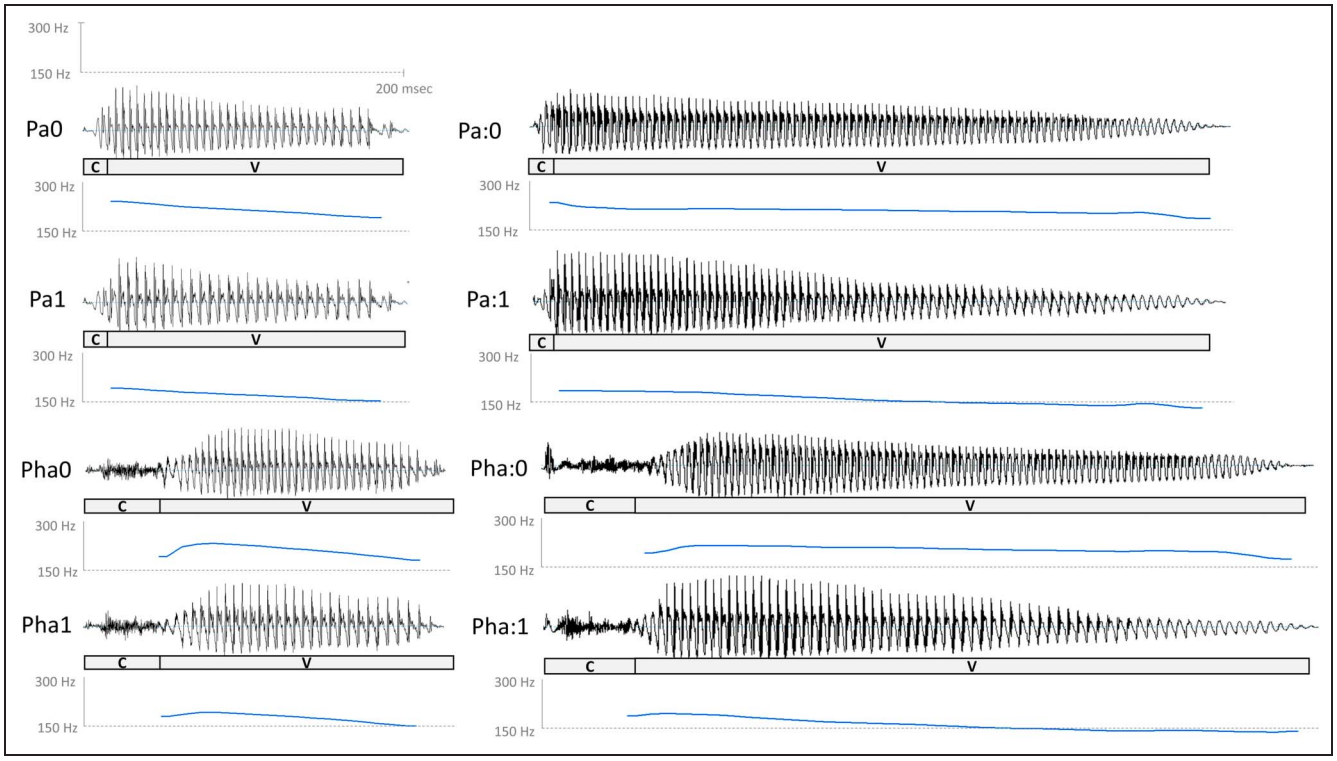

Figure 1. Illustration of the eight Thai stimuli. Waveforms are represented on the top (time in abscissa, amplitude in ordinate), and pitch contours are represented on the bottom (time in abscissa, F0 in ordinate). C = consonant; V = vowel.

Visual Stimuli

In the learning phase, eight pictures of familiar objects (i.e., *bear, flower, key, chair, bell, strawberry, train,* and *glass*) were chosen, based on the standardized set of pictures built by Snodgrass and Vanderwart (1980), that were controlled for familiarity and visual complexity. To avoid any interference effects, the items' names were always monosyllabic words in French (i.e., *ours, fleur, clef, chaise, cloche, fraise, train, verre*) and did not contain any of the Thaï syllable used in the experiment, e.g., /pa/, /p^{h}a/). The same pictures as in the learning phase were then presented in the matching task to test whether participants had learned the picture–word associations. By contrast, 48 novel pictures were presented in the semantic task that were chosen from the Internet (six novel pictures for each word), to be semantically related to each novel word. The semantic relatedness between new and old pictures was confirmed based on results of a pilot experiment with 60 university students (age range = 19–25 years) who were asked to rate the semantic relatedness between pairs of pictures, half a priori considered as semantically related and half a priori considered as semantically unrelated.

Procedure

Participants were involved in a longitudinal procedure based on a pretraining—training—posttraining design that comprised five experimental sessions separated by at least 1 day (see Figure 2A). The *pretraining session (on Day 1)* included the phonological categorization tasks and the psychoacoustic tests to measure acoustic thresholds using the Just Noticeable Difference (JND) method. Then, training comprised three *psychoacoustic training sessions on Day 2, Day 3, and Day 4.* Finally, the *posttraining session (Day 5)* included the same tests as in the pretraining plus a novel word learning phase as well as matching and semantic tests. These different tests are described in detail below. Two MMN experiments were also included, one during the pretraining session and the other during the posttraining session (i.e., on Day 1 and on Day 5) to study the preattentive processing of complex harmonic sounds and of syllables, and how preattentive processing is influenced by psychoacoustic training. Participants in both the pitch and intensity training groups were asked to watch a self-selected silent movie displayed on a DVD player screen, whereas complex harmonic sounds or syllables were presented through headphones. These experiments, that lasted for 12.5 min each, are not described further because their aims were different from the main topic of this article, and results will be reported elsewhere.

Auditory stimuli were played binaurally through headphones (Sennheiser, HD600) at 60 dB. Visual stimuli were presented on a computer screen with the Presentation software (NeuroBehavioral Systems, Version 11.0). MATLAB (The MathWorks, Inc.) was used for stimulus presentation in the psychoacoustic tasks.

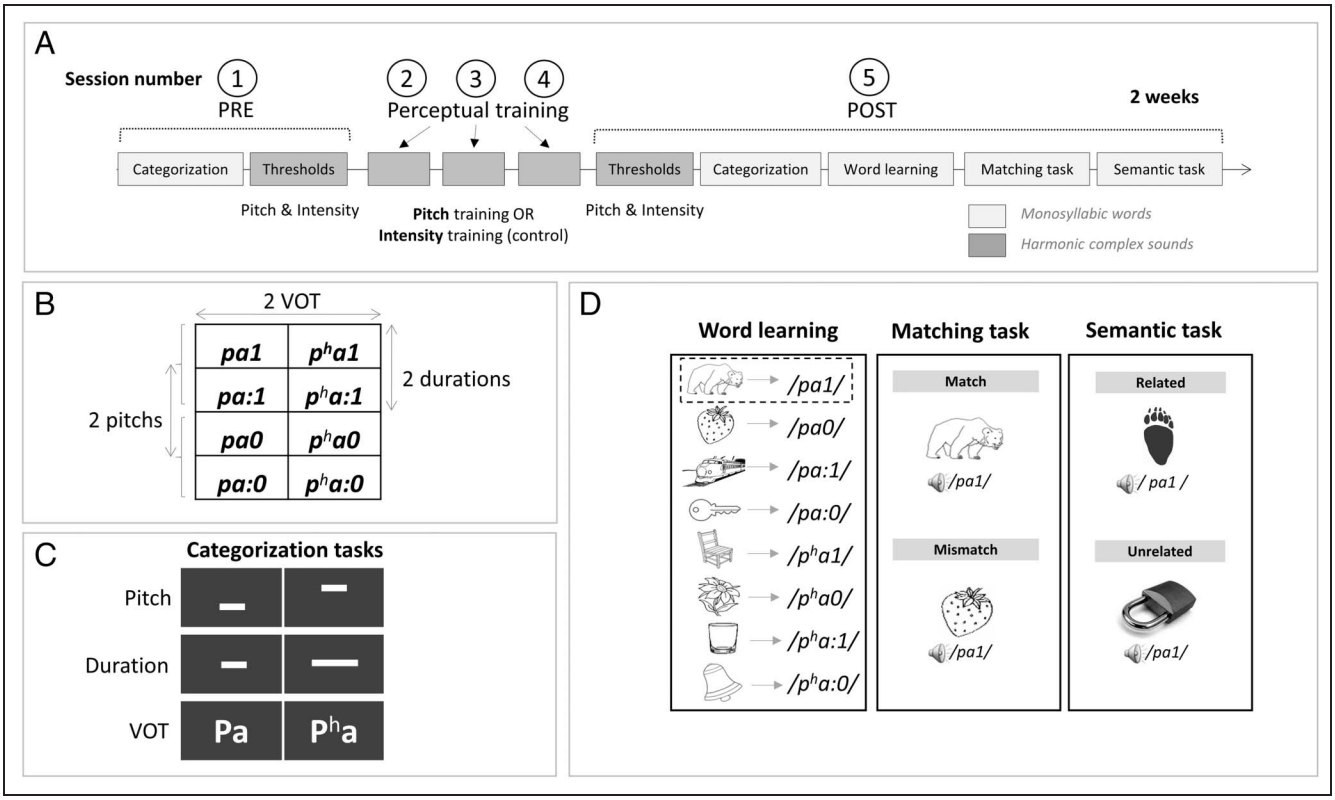

Figure 2. Experimental protocol. (A) Overview of the general experimental procedure that included five sessions taking place over 2 weeks. Pretraining (1) included phonological categorization tasks on speech stimuli, as well as pitch and intensity discrimination tasks on complex harmonic sounds. Perceptual training sessions (2, 3, and 4): Half of the participants were trained on pitch discrimination (PG), whereas the other half were trained on intensity discrimination (IG). Posttraining session (5) included pitch and intensity discrimination tasks, phonological categorization tasks, as well as three tasks evaluating novel word learning abilities (word learning phase, matching task, and semantic task). (B) Monosyllabic Thai words used in the different tasks that varied on pitch, duration, and VOT. (C) Illustration of the phonological categorization tasks as displayed on the screen for the participants. (D) In the word learning phase, participants were asked to learn the meaning of novel words presented auditorily via picture–word associations (e.g., /pa1/ means bear). In the matching task, participants had to decide whether the association matched (e.g., bear – /pa1/) or mismatched (e.g., strawberry – /pa1/) the one previously learned. In the semantic task, new pictures were presented and participants had to decide whether they were related (e.g., bear footprint – /pa1/) or unrelated (e.g., padlock – /pa1/) to the newly learned word.

Screening Measures

Several screening tests (musicality, psychometric, and audiometric) were presented to the participants.

Musicality tests. Participants had to decide whether 18 pairs of melodies from the Montreal Battery of Evaluation of Amusia (Peretz, Champod, & Hyde, 2003) were same or different, based either on rhythm or on melody.

Psychometric tests. Forward and backward digit span tasks from the WAIS-III battery (Wechsler, Coalson, & Raiford, 1997) were administered to the participants to measure auditory short-term and working memory. Matrix Reasoning (WAIS-III) was used to measure nonverbal intelligence using a 30-sec time limit for each matrix. The Verbal Fluency test (Cardebat, Doyon, Puel, Goulet, & Joanette, 1990) was also administered: Participants were asked to say aloud as many words as possible, that started with a specific letter (P, R), or that belonged to a specific category (animals, fruits) in 1 min. Visual attention was tested with the d2-R (Brickenkamp, Schmidt-Atzert, & Liepmann, 2015): Participants had to cross out as fast and as accurately as possible visual targets (d surrounded by 2 points) presented among distractors. Auditory attention was evaluated using the Response Set test, adapted from the developmental NEuroPSYchological (NEPSY-II) assessment for children by increasing time pressure (20% faster; Korkman, Kirk, & Kemp, 2007): Participants had to touch colored circles accordingly to a list of words presented orally that included targets and distractors (e.g., they had to touch the red circle when they heard the word "yellow," but touch nothing when they heard the word "black").

Finally, participants were given an *audiometric test* between 125 and 8000 Hz, in which they had to determine the minimal intensity in which they could hear the presented sound. All participants had normal hearing with thresholds below 20 dB HL.

Psychoacoustics

Psychoacoustical procedure. Discrimination thresholds (JNDs) were measured in separate blocks using complex sounds that varied either in pitch (F0) or in intensity. A

two-interval, two-alternative forced-choice procedure was used. On each trial, two tones, the standard tone and the comparison tone, were presented in random order. Depending upon the task, tones varied in frequency (standard tone F0 [195 Hz] + Δf Hz) or in intensity (standard tone intensity [60dB] +Δi dB). Participants were asked to press the left key if the first tone was higher—or louder—and the right key if the second tone was higher—or louder. The order of the pitch and intensity tasks was counterbalanced across participants.

For the JND measurements, a two-down, one-up procedure was used to estimate the frequency or intensity difference that corresponded to the 70% correct point on the psychometric function (Levitt, 1971). At the beginning of a run, the difference in frequency (Δf) between the standard and comparison tones was 20% of the standard tone. This difference decreased after two consecutive correct responses and increased after one incorrect response by a factor equal to 2 until 4 reversals occurred, and $\sqrt{2}$ after the fourth reversal (Micheyl et al., 2006). The adaptive procedure stopped after 12 reversals, and the threshold was taken as the geometric mean of the Δf values over the last eight reversal values. A trial-by-trial feedback for correct and incorrect responses was provided visually, and participants could see the time-course and evolution of the Δf of the adaptive procedure (see Figure 3). The procedure for the intensity discrimination training was identical to the frequency discrimination training except that the initial intensity difference was +10 dB and changed by 3 dB until the first 4th reversals and by 0.25 dB thereafter. The threshold was taken as the arithmetic mean of the Δi values over the last eight reversal values.

Twenty runs in each task (pitch and intensity) were performed in Session 1 to familiarize participants with the psychoacoustics procedure and to bypass procedural learning (typically associated with rapid improvement of the level of performance in the first trials). Then, participants were trained on pitch or on intensity performing 30 runs in each of the three training sessions. Finally, in the fifth and last session, participants performed 10 runs in both the pitch and intensity tasks. In total, participants were trained for approximately 3 hr on the dimension of interest.

In sum, each participant performed 120 runs in the trained task and 30 runs in the untrained task. A power function was adjusted on the resulting psychoacoustic thresholds, separately in the pitch and intensity tasks. Pre- and posttraining thresholds were computed as the intercept value of the curve, corresponding to Runs 20 and 120 in the trained task and to Runs 20 and 30 in the nontrained task (see Figure 3).

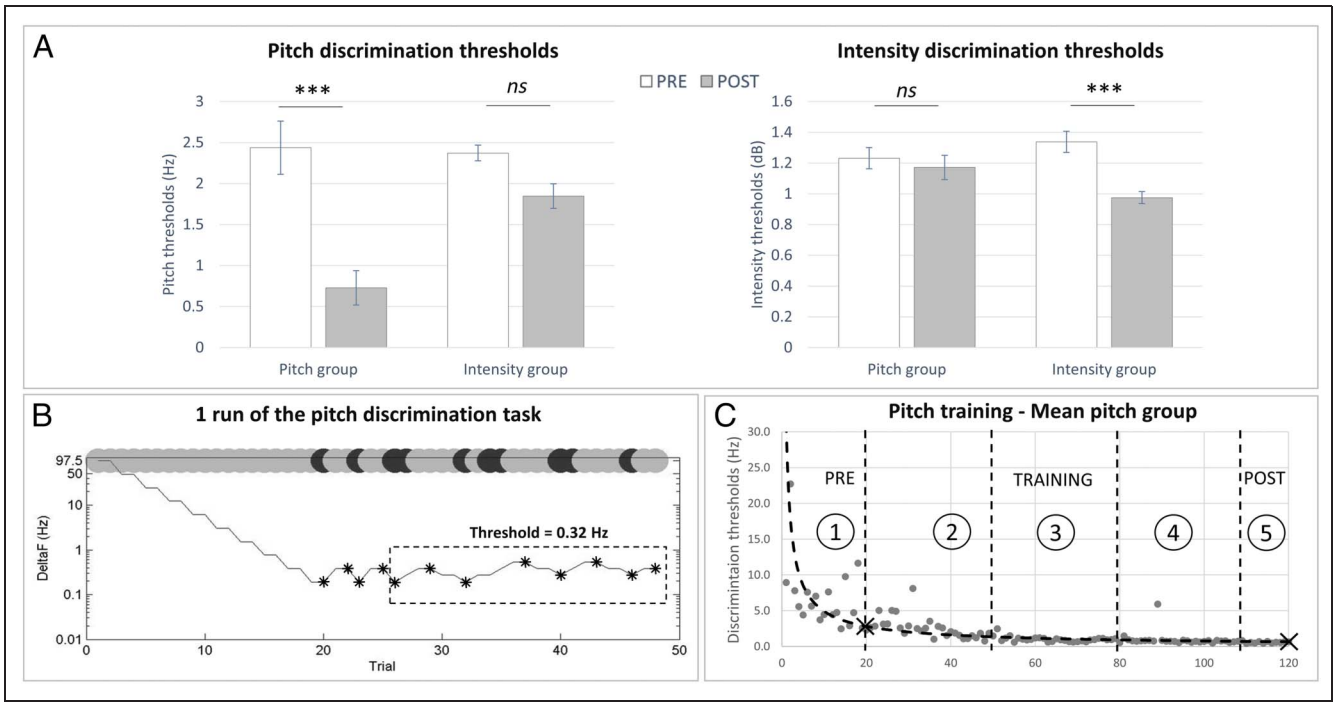

Figure 3. Psychoacoustics. (A) Pre- and posttraining thresholds in the pitch and intensity discrimination tasks. PG significantly improved in the pitch discrimination task, whereas the IG did not. IG significantly improved in the intensity discrimination task, whereas the PG did not. Error bars represent standard errors. (B) Example of one run in the pitch discrimination task for one participant. The main figure was displayed on the screen, as visual feedback for the participants. Gray dots represent correct responses; black dots represent incorrect responses. At the beginning of each run, the difference between sounds (DeltaF) is large; it decreases after two correct responses and increases after one incorrect response. The run stops after 12 curve reversals (black stars), and the threshold is computed across the eight last reversals (dotted rectangle). (C) Mean pitch threshold evolution in the PG (Sessions 1, 2, 3, 4, and 5). Each gray dot represents one threshold measurement (i.e., one run). Pre- and posttraining thresholds values were computed as the intercept of the adjusted power function (black dotted curve), corresponding to Runs 20 and 120 in the trained task (black X). ***$p < .001$. ns = not significant.

Psychoacoustical training sessions. Half of the participants were trained on pitch discrimination and the other half on intensity discrimination within three sessions of 50 min each (total: 2.5 hr) with at least 1 day between training sessions.

Pre- and Posttraining Sessions

Phonological categorization tasks. Participants performed three categorization tasks of the eight monosyllabic words: They were asked to categorize the words as 1) low-tone or mid-tone (pitch task), 2) short or long (duration task), or 3) with an aspiration (/p^ha/) or without (/pa/; VOT task), by clicking on one of the two response keys (see Figure 2B and C). The duration and VOT tasks were used to familiarize participants to the different parameter variations, but error rates and RTs were not analyzed. A visual representation of the auditory stimulus was displayed on a screen, to help participants remember the side of response. Examples were played before each categorization task to ensure that the participant understood the concept of high–low/short–long/with–without aspiration words. Response hand and task order were counterbalanced across participants. Each word was presented 10 times in each task, in a pseudorandom order (i.e., no immediate repetition of the same word, and no more than four successive same responses). Altogether, the duration of the three phonological categorization tasks was 6 min (2 min each).

Posttraining Session

Word learning phase. Participants were asked to learn the meaning of the eight words they previously categorized in the phonological categorization tasks through picture–word associations without overt response (the meaning of the words was pseudorandom; see Figure 2D). The picture was presented first, followed by the auditory word 750 msec later. Total trial duration was 2000 msec. Each of the eight picture–word association was presented 20 times, for 160 trials presented pseudorandomly (i.e., no immediate repetition) in two blocks of 3 min each. Two different association lists were constructed, so that each word was paired with a different picture across lists and participants. No behavioral response was required for this task, but participants were informed that they would then be tested on the associations and on word meaning.

Matching task. One of the eight pictures was presented first, followed 750 msec later by an auditory word matching or mismatching the previously learned association (see Figure 2D). Participants were asked to press one of the two response keys accordingly, as quickly and as accurately as possible. At the end of the trial, a row of Xs was presented during 1000 msec during which participants were asked to blink. Total trial duration was 3750 msec. The eight picture–word associations were presented pseudorandomly 20 times (i.e., no immediate repetition of the same association, and no more than four successive same responses), half in match and half in mismatch condition, for 160 trials presented in two blocks of 5 min each. Four familiarization trials were administered before starting the task.

Semantic task. A new picture that participants had not seen before in the experiment was presented, followed after 1500 msec by one of the eight auditory newly learned words (see Figure 2D). The picture could be either semantically related or unrelated to the word. At the end of the trial, a row of Xs was presented for 1000 msec during which participants were asked to blink. Total trial duration was 4500 msec. Each of the 48 pictures was presented twice, once in the related and once in the unrelated condition. Each word was presented 12 times for 96 picture–word trials, presented pseudorandomly (i.e., no immediate repetition of the same association, and no more than four successive same responses), in two blocks of approximately 4 min each. Four familiarization trials were administered before starting the task.

EEG Data Acquisition

The EEG was recorded during MMN and phonological categorization (results not reported here), as well as during the learning, matching, and semantic tasks, but not during the psychoacoustic and psychometric tests.

The EEG was continuously recorded at a sampling rate of 512 Hz using a Biosemi amplifier system (Biosemi Active 2) from 32 active AgCl electrodes (Biosemi Pintype) at standard position of the International 10/20 System (Jasper, 1958). Flat-type active electrodes were placed on the left and right mastoids as well as on the nose (reference electrodes). The EOG was recorded from electrodes placed 1 cm to the left and right of the external canthi (to record horizontal ocular movements) and from an electrode beneath the left eye (for blinks). Electrode impedance was kept below 5 kΩ.

EEG data were analyzed using the Brain Vision Analyzer software (Version 1.05.0005; Brain Products). All data were rereferenced off-line to the average of the left and right mastoids, filtered with a 0.1- to 40-Hz bandpass filter (12 dB/oct). Independent component analysis and inverse independent analysis were computed to remove components associated to horizontal and vertical eye movements. Recordings were segmented epochs (1200 msec for the learning phase and 1700 for the matching and semantic tasks, including a 200 msec baseline), time-locked to stimulus onset. DC-detrend was automatically applied using a 100-msec duration window for both the start and the end intervals. DC-detrend and baseline corrections were applied to the segmented EEG signal, in addition to automatic removal of epochs containing artifacts individually for each channel (electrical activity exceeding +/− 75 μV around the baseline). Epochs were finally averaged within each condition to obtain individual averages, and then

Table 1. Results of the ANOVAs, Corrected for Multiple Comparisons with the Benjamini–Hochberg Method

Task	*ANOVA*	*Factors*	*p Value*	*Rank i*	*(i/I)*FDR*	*Significant*
Pychoacoustics Pitch thresholds	Group (G) × Session (S)	G	0.33	3	0.10	no
		S	**0.001**	1	0.03	yes
		SxG	**0.04**	2	0.07	yes
Pychoacoustics Intensity thresholds	Group × Session	G	0.80	3	0.10	no
		S	**0.001**	1	0.03	yes
		SxG	**0.02**	2	0.07	yes
Pitch categorization task (%Err)	Group × Session	G	0.28	2	0.07	no
		S	0.08	1	0.03	no
		SxG	0.44	3	0.10	no
Pitch categorization task (RT)	Group × Session	G	0.39	2	0.07	no
		S	0.85	3	0.10	no
		SxG	**0.03**	1	0.03	yes
Matching task (%Err)	Group (G) × Condition (C)	G	**0.008**	2	0.07	yes
		C	**0.001**	1	0.03	yes
		GxC	0.25	3	0.10	no
Matching task (RT)	Group × Condition	G	0.36	2	0.07	no
		C	0.04	1	0.03	no
		GxC	0.83	3	0.10	no
Semantic task (%Err)	Group × Condition	G	0.16	2	0.07	no
		C	0.08	1	0.03	no
		GxC	0.56	3	0.10	no
Semantic task (RT)	Group × Condition	G	0.38	3	0.10	no
		C	0.19	1	0.03	no
		GxC	0.28	2	0.07	no
Learning task (N200)	Group (G) × Block (B) × Laterality (L) × ROI (R)	G	0.72	14	0.09	no
		B	**0.003**	3	0.02	yes
		B*G	0.37	10	0.07	no
		L	**0.001**	2	0.01	yes
		L*G	0.32	8	0.05	no
		R	**0.001**	1	0.01	yes
		R*G	0.87	15	0.10	no

Table 1. (*continued*)

Task	*ANOVA*	*Factors*	*p Value*	*Rank i*	*(i/I)*FDR*	*Significant*
		B*L	0.29	7	0.05	no
		B*L*G	0.35	9	0.06	no
		B*R	0.09	4	0.03	no
		B*R*G	0.56	13	0.09	no
		L*R	0.44	11	0.07	no
		L*R*G	0.23	5	0.03	no
		B*L*R	0.44	12	0.08	no
		B*L*R*G	0.25	6	0.04	no
Leaning task (N400)	Group × Block × Laterality × ROI	G	0.10	6	0.04	no
		B	**0.007**	3	0.02	yes
		B*G	0.91	15	0.10	no
		L	**0.001**	2	0.01	yes
		L*G	0.35	8	0.05	no
		R	**0.001**	1	0.01	yes
		R*G	**0.02**	4	0.03	yes
		B*L	0.58	12	0.08	no
		B*L*G	0.43	10	0.07	no
		B*R	0.53	11	0.07	no
		B*R*G	0.82	14	0.09	no
		L*R	0.07	5	0.03	no
		L*R*G	0.22	7	0.05	no
		B*L*R	0.70	13	0.09	no
		B*L*R*G	0.38	9	0.06	no
Matching task (N200)	Group (G) × Condition (C) × Laterality (L) × ROI (R)	G	0.83	13	0.09	no
		C	**0.004**	1	0.01	yes
		C*G	0.04	3	0.02	no
		L	**0.01**	2	0.01	yes
		L*G	0.55	10	0.07	no
		R	0.98	14	0.09	no
		R*G	0.69	11	0.07	no
		C*L	0.99	15	0.10	no
		C*L*G	0.54	9	0.06	no
		C*R	0.45	7	0.05	no
		C*R*G	0.07	4	0.03	no
		L*R	0.33	6	0.04	no
		L*R*G	0.26	5	0.03	no
		C*L*R	0.49	8	0.05	no

Table 1. (*continued*)

Task	*ANOVA*	*Factors*	*p Value*	*Rank i*	*(i/I)*FDR*	*Significant*
		C*L*R*G	0.80	12	0.08	no
Matching task (N400)	Group × Condition × Laterality × ROI	G	0.32	7	0.05	no
		C	0.10	5	0.03	no
		C*G	0.20	6	0.04	no
		L	**0.005**	2	0.01	yes
		L*G	0.62	12	0.08	no
		R	0.39	9	0.06	no
		R*G	**0.02**	4	0.03	yes
		C*L	0.50	11	0.07	no
		C*L*G	0.91	15	0.10	no
		C*R	**0.001**	1	0.01	yes
		C*R*G	0.47	10	0.07	no
		L*R	**0.007**	3	0.02	yes
		L*R*G	0.33	8	0.05	no
		C*L*R	0.64	13	0.09	no
		C*L*R*G	0.71	14	0.09	no
Semantic task (N200)	Group × Condition × Laterality × ROI	G	0.61	9	0.06	no
		C	0.12	5	0.03	no
		C*G	0.64	10	0.07	no
		L	0.06	3	0.02	no
		L*G	0.95	15	0.10	no
		R	**0.001**	1	0.01	yes
		R*G	0.74	13	0.09	no
		C*L	0.05	2	0.01	no
		C*L*G	0.73	12	0.08	no
		C*R	0.31	8	0.05	no
		C*R*G	0.82	14	0.09	no
		L*R	0.07	4	0.03	no
		L*R*G	0.18	6	0.04	no
		C*L*R	0.19	7	0.05	no
		C*L*R*G	0.70	11	0.07	no
Semantic task (N400)	Group × Condition × Laterality × ROI	G	**0.01**	3	0.02	yes
		C	0.67	11	0.07	no
		C*G	0.80	14	0.09	no
		L	0.03	4	0.03	no
		L*G	0.20	6	0.04	no

Table 1. (*continued*)

Task	*ANOVA*	*Factors*	*p Value*	*Rank i*	*(i/I)*FDR*	*Significant*
		R	**0.001**	1	0.01	yes
		R*G	0.68	12	0.08	no
		C*L	0.28	7	0.05	no
		C*L*G	0.42	9	0.06	no
		C*R	**0.005**	2	0.01	yes
		C*R*G	0.62	10	0.07	no
		L*R	0.73	13	0.09	no
		L*R*G	0.32	8	0.05	no
		C*L*R	0.10	5	0.03	no
		C*L*R*G	0.91	15	0.10	no

Benjamini–Hochberg critical value is computed with the following formula: (i/I)*FDR, where i is the rank of the p value in the ANOVA in ascending order, I is the total number of tests, and FDR is the false discovery rate, set as 0.1. If the p value is inferior to the critical value, it is considered as significant. Bold = statistically significant.

averaged together across participants to obtain the grand average.

Statistical Analyses

Repeated-measures ANOVAs were computed using Statistica software (Version 12.0, StatSoft, Inc). For all tasks, percentage of errors (%Err, that included both misses and false alarms) and RTs were analyzed using Group (Pitch vs. Intensity) as between-subjects factor. For the psychoacoustic and categorization tasks, Session (PRE vs. POST) was included as within-subject factor. The Matching and Semantic tasks were analyzed including Condition (Match vs. Mismatch or Related vs. Unrelated) as within-subject factor.

Regarding electrophysiological data, for each component of interest and for each participant, the maximum of amplitude (peak amplitude) was semi-automatically detected in latency ranges chosen based on averaged traces (N200: 250–350 msec and N400: 350–500 msec). For each participant, the peak value corresponded to the mean amplitude within 10 msec surrounding the peak. Peak amplitude was analyzed for each component of interest with ANOVAs that included Group (pitch vs. intensity) as a between-subject factor and Block (Block 1 vs. Block 2) or Condition (match vs. mismatch or related vs. unrelated), Laterality (left hemisphere vs. midline vs. right hemisphere) and anterior/posterior positions (ROIs: frontal vs. central vs. parietal) as within-subject factors.

Multiple comparisons corrections were computed on ANOVAs, by using the Benjamini–Hochberg test.

RESULTS

As expected, the two groups showed no differences in any of the standard psychometric tests presented before training to assess nonverbal intelligence (Matrix Reasoning from WAIS-III, $t(24) = 0.66, p = .51$), concentration capacity (d2-R, $t(24) = 0.09, p = .93$), auditory attention (NEPSY-II, $t(24) = -1.11, p = .28$), verbal fluency (Verbal Fluency Test, $t(24) = 1.6, p = .13$), musicality (Montreal Battery of Evaluation of Amusia, $t(24) = 1.18, p = .25$), short-term (Forward Digit span, $t(24) = 0.27, p = .79$), and working memory (Backward Digit span, $t(24) = 0.97, p = .34$).

Behavioral Data

Psychoacoustics

Pitch discrimination task. The main effect of Group was not significant ($F < 1$) but the Group × Session interaction was significant, $F(1, 24) = 4.57, p = .04$; see Table 1. In the PG, pitch thresholds decreased from pre- (2.44 Hz) to post-training (0.73 Hz; Tukey Honestly Significant Difference (HSD): $p = .001$, Cohen's $d = 0.99$), with no significant effect in the IG (pre: 2.37 Hz; post: 1.84 Hz; Tukey HSD: $p = .55$; see Figure 3).

Intensity discrimination task. The main effect of Group was not significant ($F < 1$) but the Group × Session interaction was significant, $F(1, 24) = 6.52, p = .02$, Cohen's $d = 0.40$; see Table 1). In the IG, intensity thresholds decreased from pre- (1.34 db) to postsession (0.98 dB; $p = .001$), with no significant effect in the PG (pre: 1.23 dB; post: 1.17 dB; $p = .55$; see Figure 3).

Pitch Categorization Task

Neither the main effects of Group and Session, $F(1, 24) = 1.24, p = .28$, and $F(1, 24) = 3.33, p = .08$, respectively, nor the Group × Session interaction ($F < 1$) were significant on error rates (see Figure 4 and Table 1). Regarding RTs, neither the main effect of Group nor the main effect of

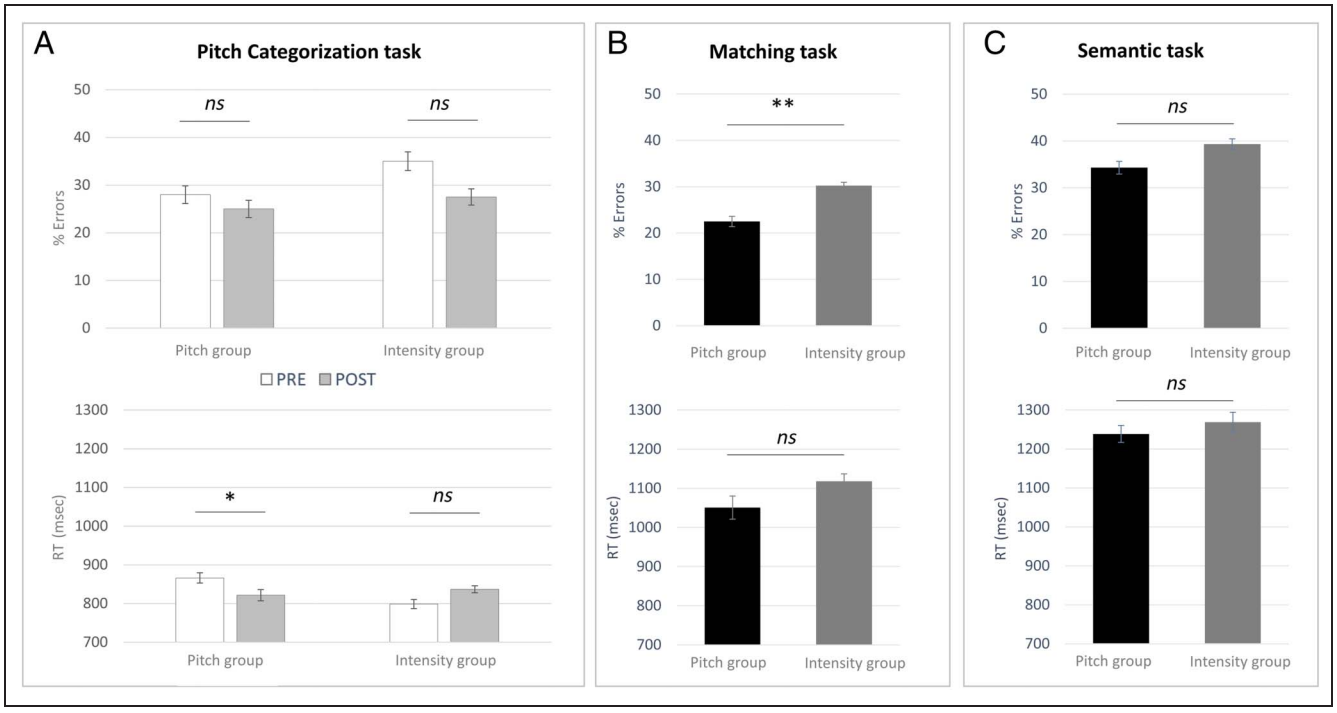

Figure 4. Between-group differences in the categorization, matching, and semantic tasks (error rates, %Err and RTs). Error bars represent standard errors. (A) Pitch categorization task: no pre- to postimprovement on %Err in either group but faster RTs after training in the PG. (B) In the matching task (posttraining session), participants in the PG (black) made fewer errors than in the IG (dark gray) with no difference on RTs. (C) In the semantic task (posttraining session), no significant between-group differences were found either on %Err or on RTs. *$p < .05$. **$p < .01$. ns = not significant.

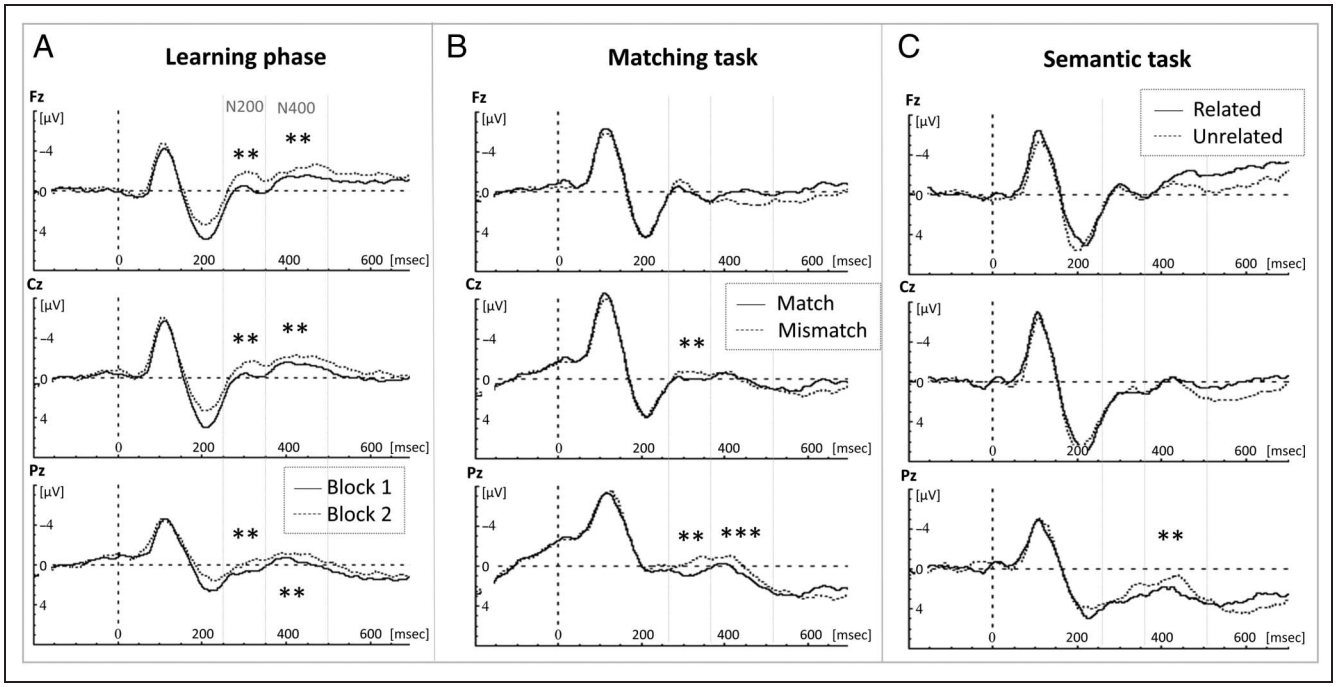

Figure 5. Main effects of blocks and conditions in the word learning phase, matching task, and semantic task (N200 and N400 components). Grand-average ERPs to the novel words are illustrated for midline electrodes (Fz, Cz and Pz). (A) Word learning phase: larger N400 and N200 components in Block 2 (dotted line) than in Block 1 (solid line) across scalp sites. (B) Matching task: larger N400 and N200 components to mismatching words (dotted line) than to matching words (solid line) over centroparietal sites. (C) Semantic task: larger N400 component to unrelated words (dotted line) than to related words (solid line) at parietal sites. **$p < .01$. ***$p < .001$.

Session were significant (both $F < 1$), but the Group × Session interaction effect was significant, $F(1, 24) = 5.67$, $p = .03$. Separate analyses for each group revealed that RTs were shorter after pitch training (PG: pre: 866 msec and post: 822 msec; $t(12) = 2.32$, $p = .04$, Cohen's $d = 0.44$) but not after intensity training (IG: pre: 799 msec and post: 837 msec; $t(12) = -1.32$, $p = .21$).

Matching Task

Participants trained in pitch made significantly fewer errors (22.47%) than participants trained in intensity (30.13%; main effect of Group: $F(1, 24) = 8.34$, $p = .008$, Cohen's $d = 1.13$; see Figure 4 and Table 1). Participants also made fewer errors to match (18.61%) than to mismatch words (33.99%; main effect of Condition: $F(1, 24) = 26.02$, $p < .001$), with no significant Group × Condition interaction, $F(1, 24) = 1.36$, $p = .25$. Regarding RTs, neither the main effect of Group nor any effects involving the Group factor were significant (all $F < 1$). The main effect of Condition was not significant after multiple comparison correction (see Table 1; Match = 1068 msec; mismatch = 1109 msec; Condition: $F(1, 24) = 4.59$, $p = .04$).

Semantic Task

The main effect of Group and the Group × Condition interaction were not significant on error rates, $F(1, 24) = 2.10$, $p = .16$ and $F < 1$, respectively, but participants tended to make fewer errors to related (32.99%) than to unrelated words (40.63%; Condition: $F(1, 24) = 3.34$, $p = .08$; see Figure 4 and Table 1). Regarding RTs, neither the main effect of Group ($F < 1$), nor Condition, $F(1, 24) = 1.86$, $p = .19$, nor the Group × Condition interaction were significant, $F(1, 24) = 1.23$, $p = .28$.

Electrophysiological Data

Learning Phase

See Figures 5A and B, Figure 6A, and Table 1.

N200 component. The N200 was significantly larger in Block 2 (−1.58 μV) than in Block 1 (−0.60 μV; $F(1, 24) = 10.83$, $p < .01$). Neither the main effect of Group nor interactions involving the Group factor were significant (all $p > .23$).

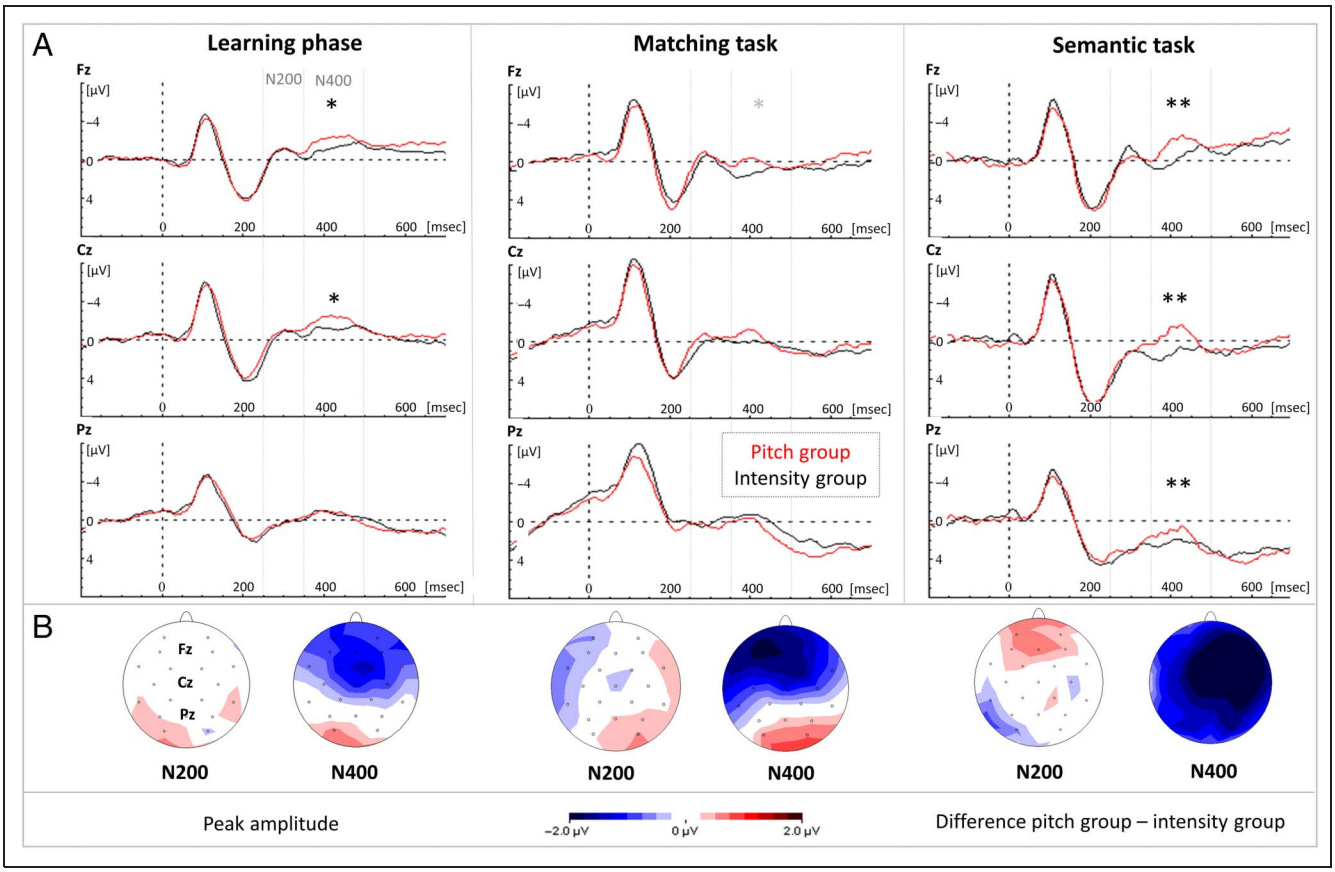

Figure 6. Between-group differences in the word learning phase, matching task, and semantic task (N200 and N400 components). (A) Grand-average ERPs to the novel words are illustrated for midline electrodes (Fz, Cz, and Pz). In the word learning phase, the N400 component was larger in the PG (red line) than in the IG (black line) over frontocentral sites. In the matching task, the between-group differences were not significant. In the semantic task, between-group differences were significant across all scalp sites, again with larger N400 component in the PG (red line) than in the IG (black line). No between-group differences were observed on the N200 component. (B) Topographic maps illustrating the between-group differences (PG − IG) for the N200 and N400 components. Gray *$p = .07$. *$p < .05$ and **$p < .01$.

Table 2. Results of the Mediation Analyses

Independent Variable: Pitch or IG Mediator: Pre/Post pitch threshold					
Dependent Variable		*ACME*	*ADE*	*Total Effect*	*Proportion Mediated*
Behavior					
Mean RT Pitch Categorization task	Estimate 95%	0.0	0.1	0.1	−0.3
	p value	.06	**.004** **	**.01** *	.07
Mean %Err Matching task	Estimate 95%	−2.2	−5.3	−7.5	0.3
	p value	.52	.07	**.006** **	.52
Mean %Err Semantic task	Estimate 95%	−1.6	−3.4	−5.0	0.3
	p value	.69	.34	.13	.67
N400 peak amplitude					
Learning task: frontal	Estimate 95%	0.1	−1.4	−1.3	−0.1
	p-value	.68	.06	**.02** *	.68
Learning task: central	Estimate 95%	0.4	−0.6	−0.2	−0.4
	p value	.28	**.04** *	**.04** *	.31
Matching task: frontal	Estimate 95%	−0.4	−1.4	−1.8	0.2
	p value	.26	.20	**.04** *	.29
Semantic task: frontal	Estimate 95%	0.3	−2.0	−1.7	−0.2
	p value	.65	.06	**.04** *	.65
Semantic task: central	Estimate 95%	0.7	−2.6	−1.9	0.3
	p value	.61	**< .001** *	**.02** *	.63
Semantic task: parietal	Estimate 95%	1.0	−2.5	−1.5	−0.7
	p value	.27	**.006** **	.07	.34

Group (Pitch vs. Intensity) was taken as the independent variable, pre/post pitch threshold ratio was taken as mediator, and several behavioral and electrophysiological measures (that showed significant between-group differences) were taken as dependent variables. ACME = average causal mediation effect; ADE = average direct effect. Bold = statistically significant.

N400 component. The N400 was significantly larger in Block 2 (−2.93 μV) than in Block 1 (−1.98 μV; $F(1, 24) = 8.77, p < .01$). The main effect of Group was not significant, $F(1, 24) = 2.99, p = .10$), but the Group × ROI interaction was significant, $F(2, 48) = 4.44, p = .02$, with a larger N400 in the PG than in the IG over frontocentral regions (*t* tests: frontal: PG = −3.64 μV and IG = −2.27 μV, $t(24) = -2.23$, $p < .05$; central: PG = −3.25 μV, IG = −2.19 μV, $t(24) = -2.03, p < .05$: PG = −1.75 μV and IG = −1.60 μV, $t(24) = -0.28, p = .78$).

Matching Task

See Figures 5A and B, Figure 6B, and Table 1.

N200 component. The N200 was significantly larger to mismatch (−2.09 μV) than to match words (−1.20 μV; $F(1, 24) = 9.87$, $p < .01$). Neither the main effect of Group ($F < 1$) nor any interaction involving the Group factor were significant after multiple comparisons correction.

N400 component. The N400 was significantly larger to mismatch (−2.32 μV) than to match words (−1.76 μV) over parietal regions (Condition × ROI: $F(2, 48) = 20.41$, $p < .001$; t test: frontal: $t(25) = -0.86$, $p = .40$; central: $t(25) = 1.70$, $p = .10$; parietal: $t(25) = 3.79$, $p < .001$). The main effect of Group was not significant, $F(1, 24) = 1.05$, $p = .32$, but the Group × ROI interaction was significant, $F(2, 48) = 4.24$, $p = .02$, with relatively larger N400 in the PG than in the IG over frontal regions (t tests: frontal: PG = −2.69 μV, IG = −0.88 μV, $t(24) = -1.90$, $p = .07$; central: PG = −2.75 μV, IG = −1.86 μV, $t(24) = -1.03$, $p = .31$; parietal: PG = −1.83 μV, IG = −2.23 μV, $t(24) = 0.49$, $p = .62$).

Semantic Task

See Figures 5A and B, Figure 6C, and Table 1.

N200 component. After multiple comparison correction, neither the main effect of Condition, nor any interaction involving the Condition factor were significant (see Table 1). Neither the main effect of Group nor any interaction involving the Group factor were significant (all $p > .18$).

N400 component. The N400 was significantly larger to unrelated (−2.84 μV) than to related words (−2.60 μV) over parietal regions (Condition × ROI: $F(2, 48) = 5.98$, $p < .01$); t test: frontal: $p = .45$; central: $p = .88$; parietal: $p = .04$). The N400 component was also significantly larger in the pitch (−3.60 μV) than in the intensity (−1.84 μV) group (Group: $F(1, 24) = 6.58$, $p < .01$).

Mediation Analyses

To further examine whether between-group differences on behavior and/or on the N400 component were directly linked to improved pitch perception abilities (i.e., decrease in pitch threshold after pitch training but not after intensity training; see Figure 3A), we conducted mediation analyses that included group (pitch vs. intensity) as an independent variable, change in pitch discrimination from pre- to posttraining as a mediator variable, and behavioral and N400 data as dependent variables. Results were in line with results of the ANOVAs in showing that the average direct effect and/or the total effect were significant in the categorization task (mean RT, $p < .004$ and $p < .01$, respectively) and in the matching task (%Err $p < .006$) as well as on the N400 component in the word learning task (frontal, $p < .02$ and central sites, $p < .04$), in the matching task (frontal, $p < .04$) and in the semantic task (frontal, $p < .04$; central, $p < .02$, and parietal, $p < .006$; see Table 2). However, the improvement in pitch threshold from pre- to posttraining did not significantly mediate the significant between-group differences (i.e., the average causal mediation effect was not significant).

DISCUSSION

Improved Pitch or Intensity Discrimination Abilities after Relevant Training

After approximately 3 hr of pitch discrimination training, distributed across three training sessions (see Figure 2A), participants discriminated pitch differences that were about 3.5 times smaller than before pitch training, with no significant improvement in intensity discrimination thresholds. After the same amount of intensity discrimination training, participants discriminated intensity differences that were about 1.5 times smaller than before training with no significant improvement in pitch discrimination thresholds. Thus, after relatively short psychoacoustic training, discrimination thresholds were specifically lower on the feature (pitch or intensity) that was trained. These results are in line with those of Micheyl et al. (2006), showing that 4–8 hr of psychoacoustics training on pitch was sufficient to decrease pitch discrimination thresholds by a factor of 6. The improvement in our experiment was smaller probably because training was shorter. However, our results stand in contrast to those of Amitay et al. (2006) showing that participants discriminated smaller pitch differences even after passive exposure to sounds. This is possibly because, in our training sessions, but not in some conditions of Amitay et al. (2006), participants were explicitly asked to pay attention to one specific dimension.

Faster Pitch Categorization in Word Context after Nonspeech Pitch Training

Participants were faster to categorize Thai monosyllabic words based on pitch after than before pitch training, with no significant differences after intensity training. These results are in line with the transfer hypothesis, after which improved perception of pitch differences in complex harmonic sounds transfers to speech sounds, thereby allowing to build clear phonological representations of nonfamiliar words that vary on pitch. Categorization of words based on pitch was faster post- than pretraining in the PG, possibly because the auditory system was better trained to judge pitch as a feature in this group.

Previous results have shown transfer effects from music expertise to segmental and phonological processing (Bidelman, Gandour, & Krishnan, 2011; Marie, Delogu, Lampis, Belardinelli, & Besson, 2011; Delogu, Lampis, & Belardinelli, 2010; Slevc & Miyake, 2006; Anvari, Trainor, Woodside, & Levy, 2002; see Gordon et al., 2015, for a meta-analysis). For instance, Bidelman et al. (Bidelman, Weiss, Moreno, & Alain, 2014) reported that musicians were faster to categorize speech sounds than nonmusicians and showed more robust sound encoding at the brainstem and cortical levels. These results were taken to show that increased auditory sensitivity in musicians drive enhanced categorical perception. However, the causal link between music training and categorical perception was not directly tested because musicians and nonmusicians were

compared in their study. Here, we show that music training may not be a necessary condition. Short duration low-level acoustic training with complex harmonic tones using a longitudinal procedure seems to be sufficient to causally influence the speed of speech sounds categorization.

Better Learning of Novel Words Meaning with Nonspeech Pitch Training

Results above are taken as evidence for near transfer effects from low-level acoustic perception to phonological categorization of speech sounds. The next question is to determine whether far transfer effects from low-level acoustic perception to higher cognitive processes can also be demonstrated when learning the meaning of novel words.

In the word learning phase, no overt response was required, but participants were asked to learn associations between pictures and words varying in pitch. As found in our previous experiments with adults and children (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016), results showed differences in brain electrical activity after only 3 min of word learning: The amplitude of the N200 and N400 components to novel words was significantly larger in the second than in the first block of trials (see Figure 5A). This is taken as evidence that participants were actively involved in learning the meaning of novel words (Bakker, Takashima, van Hell, Janzen, & McQueen, 2015; Batterink & Neville, 2011; Borovsky et al., 2010; Mestres-Missé et al., 2007). However, in contrast to previous results showing frontocentrally distributed N400 components (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016), the differences between the two blocks were widely distributed across the antero-posterior dimension. This may reflect the mobilization of more distributed neural resources in this experimental design (that involved several sessions before the word learning phase) than in previous experiments using a simpler procedure.

Most importantly, and in line with our prediction, the N400 over frontocentral regions was larger in the PG than in the IG (see Figure 6A and 6B), thereby suggesting that novel word learning was more efficient in the pitch trained group. Interestingly, ERPs traces in the two groups overlapped very well for the early components (N100, P200, and N200) and the between-group difference, albeit small, was centered on the N400 component that appeared most sensitive to training differences. Taken together, these results are in line with the phonetic–phonological–lexical continuity and cascading hypothesis (Besson et al., 2017; Dittinger et al., 2016; Cooper & Wang, 2012, 2013; Wong & Perrachione, 2007), after which participants trained on pitch with complex harmonic tones, categorized words varying on pitch faster (see above), and were able to build more efficient picture–word associations (as reflected by larger N400 amplitude in the learning phase) than participants trained on intensity. We take these results as evidence for far transfer effects from low-level acoustic training to high-level novel word learning. Importantly, and although we did not directly test for verbal learning abilities, it is unlikely that between-group differences in verbal learning or verbal IQ account for faster word pitch categorization and more efficient word learning in the pitch than in the intensity trained group because participants to this longitudinal study were randomly assigned to one of the two groups. Finally, results also showed that neural plasticity related to learning the meaning of novel words developed very fast (in 20 repetitions of picture–word associations, that is, in between 3 min [Block 1] and 6 min [Block 2]).

Mixed Evidence for Far Transfer Effects in the Matching Task

In the matching task, which tested for the efficacy of learning picture–word associations, the level of performance was clearly higher than chance level in the two groups of participants (78% and 70% correct responses in the PG and IG, respectively). As in previous studies, participants made fewer errors to match than to mismatch words (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016). These results at the behavioral level are clearly in line with those found at the electrophysiological level, showing, as predicted, that the N200 and N400 components were larger to mismatch than to match words over centroparietal sites (see Figure 5B). We take these findings as clear evidence that 1) all participants had successfully learned picture–words associations, 2) that they processed match and mismatching words differently, and 3) that the representations of the newly learned words was being integrated into semantic networks as reflected by the centroparietal distribution of the N400 that was similar to the typical distribution found for known words (Kutas et al., 1988).

Most importantly, the PG made significantly fewer errors than the IG, with no speed accuracy tradeoff (no significant effect on RTs; see Figure 4B). These results extend those of Cooper and Wang (2013), showing that a group trained on tone identification made fewer errors than their untrained counterparts in a tone word learning phase. Interestingly, participants in this study were not trained with lexical tones but with complex harmonic tones, thereby providing evidence at the behavioral level for far transfer effects from low acoustic pitch training to higher level matching task. However, results were not as clear-cut at the electrophysiological level because they only showed a tendency ($p <$ .07) for the frontal N400 to be larger in the PG compared to the IG (see Figure 6A and B). Thus, the evidence for far transfer effects from acoustic training with harmonic complex tones to the processes involved in the matching task, as reflected by changes in the amplitude of the N400 component, was not as strong as in the word learning phase.

Mixed Evidence for the Phonetic–Phonological–Lexical Continuity and Cascading Hypothesis in the Semantic Task

In the semantic task, which tested for the generalization of word learning, all participants performed higher than

chance level (66% and 61% correct responses in the PG and IG, respectively), thereby suggesting that they were able to generalize the word meaning they had learned before to new pictures. However, behavioral data provided no clear indications that participants processed semantically related and unrelated words differently. By contrast, and in line with the well-known N400 semantic priming effect (Dittinger et al., 2016; Angwin et al., 2014; Holcomb & Neville, 1990; Bentin, McCarthy, & Wood, 1985), the N400 component was larger to semantically unrelated than to semantically related words over parietal sites (see Figure 5C). Thus, in contrast to behavior, ERPs revealed evidence that participants processed unrelated and related words differently and that novel word meaning was integrated into semantic networks, as shown by the parietal scalp distribution of the N400 component, typical of known words (Kutas et al., 1988).

Turning to the effects of training, no between-group differences were found in the semantic task at the behavioral level, neither on error rates nor on RTs (see Figure 4C), thereby suggesting a limit to transfer effects. By contrast, the N400 component in the semantic task was clearly larger in the PG than in the IG across all scalp sites (see Figure 6A and B). Based on the interpretation of the frontocentral N400 components proposed above, this suggests that participants trained on pitch were possibly still forming new picture–word associations, thereby mobilizing more frontal resources than participants in the IG. In parallel and based on the interpretation of the parietal N400 components, participants in the PG had possibly already integrated the meaning of some novel words into semantic networks while this effect was less pronounced for participants in the IG.

Taken together, results in the semantic task again provided mixed evidence for far transfer effects, from low level acoustic training to higher level semantic processing, that were significant on N400 amplitude but not significant at the behavioral level. These findings are discussed below in light of results in the different tasks of the present experiment.

Near and Far Transfer Effects

In this experiment, we used a longitudinal procedure with nonmusician participants involved in short psychoacoustic training with nonspeech sounds to test for near and far transfer effects. Results were clear-cut in showing that discrimination thresholds were lower on the specific feature (pitch or intensity) that was trained, thereby showing that training was efficient. Moreover, between-group differences were reflected in behavior with enhanced level of performance in the PG than in the IG in the word pitch categorization task (faster RTs, near transfer) and in the matching task (lower error rate, far transfer). Between-group differences were also found in the ERPs, with larger N400 in the PG than in the IG in the word learning phase (larger frontocentral N400s, far transfer) and in the semantic task (larger N400s across scalp sites, far transfer) but not in the matching task (no significant difference on the N400, far transfer). In sum, the results reported here with two groups of nonmusicians show evidence for both near and far transfer effects (from training with harmonic sounds to processing speech stimuli) in each one of the various tasks used in this experiment. It is also important to note that, when they emerged, between-group differences were always found on the N400 (and not on other ERP components), as predicted based on previous results from our group comparing musicians and nonmusicians (Dittinger et al., 2016) and on other reports in the literature (Bakker et al., 2015; Batterink & Neville, 2011; Borovsky et al., 2010; Mestres-Missé et al., 2007).

An important issue, however, is why between-groups differences were sometimes found in behavior and sometimes in the ERPs. We propose some interpretations that will need to be tested in further experiments. First, it is likely that training duration was too short (approximately 3 hr distributed across 3 training sessions) in this study to lead to consistent between group differences. Although short-term psychoacoustic pitch training with nonspeech sounds can offset 10 years of music training in a pitch discrimination task (Micheyl et al., 2006) and can influence RTs in the pitch phonological task with monosyllabic Thai words that was used here (near transfer), and response correctness in the matching task (far transfer), it does not seem sufficient to facilitate generalization of priming effects to new pictures in the semantic task at the behavioral level. By contrast, significant semantic priming effects, together with significant learning effects, were found on the N400 component. Importantly, for each one of the several tasks used in this experiment, results of the mediation analyses considering RTs, error rates or N400 amplitude as dependent variables paralleled the main results of the ANOVAs (significant direct or total effects on behavior in the categorization and matching tasks and on the N400 component in the word learning and in the semantic tasks). However, these effects were not mediated by the improvement in pitch threshold from pre- to posttraining. It may be that the auditory system was nevertheless trained to process pitch information deeper and/or more efficiently than before training. An interesting question for further research will be to determine if acoustic training with longer duration can elicit stronger and more consistent far transfer effects and clarify the role of pitch training compared, for instance, to more general auditory attention training.

Second, it may be that low level acoustic training improved auditory pitch perception, which, in turn, helped performing tasks in which pitch is an important feature, as when categorizing words based on pitch or when building up associations between pictures and words varying in pitch (learning and matching task), but had lesser impact when other functions than pitch perception (e.g., working memory, attention, motivation) may be more important for the task at hand, as in the semantic task. This interpretation is in line with the bottom-up, "phonetic–phonological–lexical

continuity" hypothesis (Cooper & Wang, 2012, 2013; Wong & Perrachione, 2007), also called the "cascading hypothesis" (e.g., Besson et al., 2017; Dittinger et al., 2016), after which improved auditory perception in musicians (or with acoustic training) drives some aspects of word learning. Moreover, it is also in line with the top–down hypothesis (or multidimensional hypothesis; Dittinger et al., 2016), after which other functions such as short-term memory (George & Coch, 2011), verbal working memory (Chan, Ho & Cheung, 1998; Franklin et al., 2008), attention (Strait, Slater, O'Connell & Kraus, 2015), and executive functions (Diamond, 2013; Zuk, Benjamin, Kenyon & Gaab, 2014) are likely to play a role in the semantic task (Bakker et al., 2015; see Robinson, 2005, for a review). These factors were not necessarily enhanced by the psychoacoustic training procedure used here, thereby possibly explaining why no effect was found on behavior in the semantic task. In line with this explanation, Wong and Perrachione (2007) showed that whereas half of the variance in the tone word learning phase was explained by pitch identification, thereby showing the importance of this factor, the other half of the variance was explained by other factors that contributed to word learning success (e.g., verbal working memory, motivation, length, intensity and quality of training).

Third, and more generally, behavioral and neural responses not always go hand in hand. Such findings are difficult to interpret based on current state of knowledge because several, not exclusive explanations, can be proposed. For instance, it may be that learning effects first need to stabilize at the neural level before being observed at the behavioral level, as may be the case in the semantic task (evidence for semantic priming in N400 amplitude but not in behavior). However, the reverse was found in the matching task, with significant effects at the behavioral level and not in the ERPs. In this case, it is possible that the between-group differences at the neural level were not strong enough to be significant. In sum, different, not exclusive accounts of the lack of brain–behavior correlations can be proposed. To better understand these results, more experiments are needed to specify the levels (behavior, ERPs, and so forth) and the conditions under which far transfer effects are present.

Conclusion

By using a longitudinal approach with nonmusician participants and by contrasting two types of training distributed across 2 weeks, we aimed at testing the causal role of improved processing of low acoustic features on several aspects of novel word learning. We expected the effects to be small, because of the short duration of acoustic training. In the ERPs, the N400 effects were localized in a small latency band within the entire time window of interest, thereby being susceptible to noise in the EEG waveform. However, because these effects showed the expected latency, polarity, and parietal distribution and because they replicated and extended previous results in the literature, we are confident that these are N400 effects. Moreover, we were able to make specific predictions based on previous results (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016: Balass et al., 2010; Borovsky et al., 2010, 2012; Batterink & Neville, 2011; Dobel et al., 2009, 2010; Mestres-Missé et al., 2007; Perfetti et al., 2005; McLaughlin et al., 2004). Some predictions were verified, demonstrating the causal role of improved pitch training in the pitch categorization task and in the word learning phase. However, causality was only partially demonstrated (either on behavior or on the N400 component) in the matching and semantic tasks, and it was not demonstrated using mediation analyses. Thus, the present results provide mixed evidence that enhanced perception of relevant features through a few hours of acoustic training with harmonic sounds causally impacts the different processing stages involved in word learning.

One final aspect deserves comments. The novel words learned in most experiments (Dittinger, Chobert, et al., 2017; Dittinger et al., 2016; Cooper & Wang, 2012; Wong & Perrachione, 2007) were words from tone languages (Cantonese, Thai and Mandarin Chinese), in which pitch variations are linguistically contrastive. It would be of interest in further experiments to determine whether results similar to those reported here after pitch training are also found after training other acoustic features. For instance, duration training may be most important when learning the meaning of novel words in quantitative languages, such as Finnish, in which the duration of segmental features is linguistically contrastive.

Acknowledgments

This work was funded by the CNRS and supported by Aix-Marseille University, by the Labex BLRI (ANR-11-LABX-0036) and by the ILCB, managed by the French National Agency for Research, under the program "Investissements d'Avenir" (ANR-11-IDEX-0001-02). M.Ba. was supported by a doctoral fellowship from the French Ministry of Research and Education. We would like to thank all the students who participated in this experiment.

Reprint requests should be sent to Mireille Besson, Université Publique de France, Laboratoire de Neuroscience Cognitives, CNRS and Université Aix-Marseille, Centre Saint Charles, 3 Place Victor Hugo, 13331, Marseille, France, or via e-mail: mireille.besson@univ-amu.fr.

Author Contributions

Mylène Barbaroux: Conceptualization; Data curation; Formal analysis; Methodology; Writing-Original Draft. Arnaud Norena: Formal analysis; Methodology; Software; Visualization. Maud Rasamimanana: Data curation; Formal analysis. Eric Castet: Conceptualization; Formal analysis; Methodology; Software; Visualization. Mireille Besson: Conceptualization; Funding acquisition; Investigation; Methodology; Project administration; Resources; Supervision; Validation; Writing - Original Draft; Writing - Review & Editing.

REFERENCES

Alexander, J. A., Wang, P. C. M., & Bradlow, A. R. (2005). Lexical tone perception in musicians and non-musicians. 397–400. Paper presented at 9th European Conference on Speech Communication and Technology, Lisbon, Portugal.

Amitay, S., Irwin, A., & Moore, D. R. (2006). Discrimination learning induced by training with identical stimuli. *Nature Neuroscience*, *9*, 1446–1448. **DOI:** https://doi.org/10.1038/nn1787, **PMID:** 17028582

Angwin, A. J., Phua, B., & Copland, D. A. (2014). Using semantics to enhance new word learning: An ERP investigation. *Neuropsychologia*, *59*, 169–178. **DOI:** https://doi.org/10.1016/j.neuropsychologia.2014.05.002, **PMID:** 24846835

Anvari, S. H., Trainor, L. J., Woodside, J., & Levy, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, *83*, 111–130. **DOI:** https://doi.org/10.1016/S0022-0965(02)00124-8, **PMID:** 12408958

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41. **DOI:** https://doi.org/10.1016/j.neuropsychologia.2015.10.020, **PMID:** 26476370

Balass, M., Nelson, J. R., & Perfetti, C. A. (2010). Word learning: An ERP investigation of word experience effects on recognition and word processing. *Contemporary Educational Psychology*, *35*, 126–140. **DOI:** https://doi.org/10.1016/j.cedpsych.2010.04.001, **PMID:** 22399833, **PMCID:** PMC3293175

Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: An event-related potential study. *Journal of Cognitive Neuroscience*, *23*, 3181–3196. **DOI:** https://doi.org/10.1162/jocn_a_00013, **PMID:** 21452941, **PMCID:** PMC3129368

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, *60*, 343–355. **DOI:** https://doi.org/10.1016/0013-4694(85)90008-2, **PMID:** 2579801

Besson, M., Barbaroux, M., & Dittinger, E. (2017). Music in the brain: Music and language processing. In R. Ashley & R. Timmers (Eds.), *The Routledge companion to music cognition* (pp. 37–48). New York and London: Taylor & Francis Group.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*, 425–434. **DOI:** https://doi.org/10.1162/jocn.2009.21362, **PMID:** 19925180

Bidelman, G. M., Weiss, M. W., Moreno, S., & Alain, C. (2014). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *European Journal of Neuroscience*, *40*, 2662–2673. **DOI:** https://doi.org/10.1111/ejn.12627, **PMID:** 24890664

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.1. 2.9) [Computer software]. Retrieved from www.fon.hum.uva.nl/praat.

Borovsky, A., Elman, J. L., & Kutas, M. (2012). Once is enough: N400 indexes semantic integration of novel word meanings from a single exposure in context. *Language Learning and Development*, *8*, 278–302. **DOI:** https://doi.org/10.1080/15475441.2011.614893, **PMID:** 23125559, **PMCID:** PMC3484686

Borovsky, A., Kutas, M., & Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition*, *116*, 289–296. **DOI:** https://doi.org/10.1016/j.cognition.2010.05.004, **PMID:** 20621846, **PMCID:** PMC2904319

Brickenkamp, R., Schmidt-Atzert, L., & Liepmann, D. (2015). *d2-R: Test d'attention concentrée [d2-R: test of attention]*. Éditions Hogrefe France.

Cardebat, D., Doyon, B., Puel, M., Goulet, P., & Joanette, Y. (1990). Formal and semantic lexical evocation in normal subjects. Performance and dynamics of production as a function of sex, age and educational level. *Acta Neurologica Belgica*, *90*, 207–217. **PMID:** 2124031

Carey, S. (1978). The child as word learner. Linguistic theory and psychological reality. In M. Halle, J. Bresnan, & G. A. Miller (Eds.), *Linguistic theory and psychological reality* (pp. 264–293). Cambridge, MA: MIT Press.

Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature*, *396*, 128. **DOI:** https://doi.org/10.1038/24075, **PMID:** 9823892

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, *6*, 256–266. **DOI:** https://doi.org/10.1162/jocn.1994.6.3.256, **PMID:** 23964975

Cooper, A., & Wang, Y. (2012). The influence of linguistic and musical experience on Cantonese word learning. *Journal of the Acoustical Society of America*, *131*, 4756–4769. **DOI:** https://doi.org/10.1121/1.4714355, **PMID:** 22712948

Cooper, A., & Wang, Y. (2013). Effects of tone training on Cantonese tone-word learning. *Journal of the Acoustical Society of America*, *134*, EL133–EL139. **DOI:** https://doi.org/10.1121/1.4812435, **PMID:** 23927215

Delogu, F., Lampis, G., & Belardinelli, M. O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology*, *22*, 46–61. **DOI:** https://doi.org/10.1080/09541440802708136

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. **DOI:** https://doi.org/10.1146/annurev-psych-113011-143750, **PMID:** 23020641, **PMCID:** PMC4084861

Dittinger, E., Barbaroux, M., D'Imperio, M., Jäncke, L., Elmer, S., & Besson, M. (2016). Professional music training and novel word learning: From faster semantic encoding to longer-lasting word representations. *Journal of Cognitive Neuroscience*, *28*, 1584–1602. **DOI:** https://doi.org/10.1162/jocn_a_00997, **PMID:** 27315272

Dittinger, E., Chobert, J., Ziegler, J. C., & Besson, M. (2017). Fast brain plasticity during word learning in musically-trained children. *Frontiers in Human Neuroscience*, *11*, 233. **DOI:** https://doi.org/10.3389/fnhum.2017.00233, **PMID:** 28553213, **PMCID:** PMC5427084

Dittinger, E., Scherer, J., Jäncke, L., Besson, M., & Elmer, S. (2019). Testing the influence of music training on novel word learning across the lifespan using a cross-sectional approach in children, young adults and older adults. *Brain and Language*, *198*, 104678. **DOI:** https://doi.org/10.1016/j.bandl.2019.104678, **PMID:** 31450024

Dittinger, E., Valizadeh, S. A., Jäncke, L., Besson, M., & Elmer, S. (2018). Increased functional connectivity in the ventral and dorsal streams during retrieval of novel words in professional musicians. *Human Brain Mapping*, *39*, 722–734. **DOI:** https://doi.org/10.1002/hbm.23877, **PMID:** 29105247, **PMCID:** PMC6866404

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., et al. (2010). New Names for known things: On the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*, 1251–1261. **DOI:** https://doi.org/10.1162/jocn.2009.21297, **PMID:** 19583468

Dobel, C., Lagemann, L., & Zwitserlood, P. (2009). Non-native phonemes in adult word learning: Evidence from the N400m. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *364*, 3697–3709. **DOI:** https://doi.org/10.1098/rstb.2009.0158, **PMID:** 19933141, **PMCID:** PMC2846316

Franklin, M. S., Moore, K. S., Yip, C.-Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, *36*, 353–365. **DOI:** https://doi.org/10.1177/0305735607086044

George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, *49*, 1083–1094, **DOI:** https://doi.org/10.1016/j.neuropsychologia.2011.02.001, **PMID:** 21315092

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Science*, *18*, 635–644. **DOI:** https://doi.org/10.1111/desc.12230, **PMID:** 25195623

Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, *5*, 281–312. **DOI:** https://doi.org/10.1080/01690969008407065

Jasper, H. H. (1958). The ten–twenty system of the International Federation. *Clinical Neurophysiology*, *10*, 371–375.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-Second Edition (NEPSY-II)*. San Antonio, TX: The Psychological Corporation. **DOI:** https://doi.org/10.1037/t15125-000

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. **DOI:** https://doi.org/10.1126/science.7350657, **PMID:** 7350657

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161–163. **DOI:** https://doi.org/10.1038/307161a0, **PMID:** 6690995

Kutas, M., Van Petten, C., & Besson, M. (1988). Event-related potential asymmetries during the reading of sentences. *Electroencephalography and Clinical Neurophysiology*, *69*, 218–233. **DOI:** https://doi.org/10.1016/0013-4694(88)90131-9, **PMID:** 2450003

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477. **DOI:** https://doi.org/10.1121/1.1912375, **PMID:** 5541744

Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, *23*, 2701–2715. **DOI:** https://doi.org/10.1162/jocn.2010.21585, **PMID:** 20946053

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, *7*, 703–704. **DOI:** https://doi.org/10.1038/nn1264, **PMID:** 15195094

Mestres-Missé, A., Rodriguez-Fornells, A., & Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cerebral Cortex*, *17*, 1858–1866. **DOI:** https://doi.org/10.1093/cercor/bhl094, **PMID:** 17056648

Micheyl, C., Delhommeau, K., Perrot, X., & Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, *219*, 36–47. **DOI:** https://doi.org/10.1016/j.heares.2006.05.004, **PMID:** 16839723

Patscheke, H., Degé, F., & Schwarzer, G. (2018). The effects of training in rhythm and pitch on phonological awareness in four-to six-year-old children. *Psychology of Music*, *47*, 376–391. **DOI:** https://doi.org/10.1177/0305735618756763

Peretz, I., Champod, A. S., & Hyde, K. L. (2003). Varieties of musical disorders. *Annals of the New York Academy of Sciences*, *999*, 58–75. **DOI:** https://doi.org/10.1196/annals.1284.006, **PMID:** 14681118

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1281–1292. **DOI:** https://doi.org/10.1037/0278-7393.31.6.1281, **PMID:** 16393047

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. **DOI:** https://doi.org/10.1017/S0267190505000036

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & de Diego-Balaguer, R. (2009). Neurophysiological mechanisms involved in language. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *364*, 3711–3735. **DOI:** https://doi.org/10.1098/rstb.2009.0130, **PMID:** 19933142, **PMCID:** PMC2846313

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*, 675–681. **DOI:** https://doi.org/10.1111/j.1467-9280.2006.01765.x, **PMID:** 16913949

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215. **DOI:** https://doi.org/10.1037/0278-7393.6.2.174, **PMID:** 7373248

Strait, D. L., Slater, J., O'Connell, S., & Kraus, N. (2015). Music training relates to the development of neural mechanisms of selective auditory attention. *Developmental Cognitive Neuroscience*, *12*, 94–104. **DOI:** https://doi.org/10.1016/j.dcn.2015.01.001

van den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, *13*, 967–985. **DOI:** https://doi.org/10.1162/089892901753165872, **PMID:** 11595099

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649–3658. **DOI:** https://doi.org/10.1121/1.428217, **PMID:** 10615703

Wechsler, D., Coalson, D. L., & Raiford, S. E. (1997). *WAIS-III: Wechsler Adult Intelligence Scale*. San Antonio, TX: Psychological Corporation. **DOI:** https://doi.org/10.1037/t49755-000

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics*, *28*, 565–585. **DOI:** https://doi.org/10.1017/S0142716407070312

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PLoS One*, *9*, e99868. **DOI:** https://doi.org/10.1371/journal.pone.0099868, **PMID:** 24937544, **PMCID:** PMC4061064