Neuromagnetic brain activities associated with perceptual categorization and sound-content incongruency: a comparison between monosyllabic words and pitch names

Chen-Gia Tsai [1,2], Chien-Chung Chen [2,3], Ya-Chien Wen [1] and Tai-Li Chou [2,3,4,5]*

[1] Graduate Institute of Musicology, National Taiwan University, Taipei, Taiwan, [2] Neurobiology and Cognitive Science Center, National Taiwan University, Taipei, Taiwan, [3] Department of Psychology, National Taiwan University, Taipei, Taiwan, [4] Graduate Institute of Brain and Mind Sciences, National Taiwan University, Taipei, Taiwan, [5] Graduate Institute of Linguistics, National Taiwan University, Taipei, Taiwan

In human cultures, the perceptual categorization of musical pitches relies on pitch-naming systems. A sung pitch name concurrently holds the information of fundamental frequency and pitch name. These two aspects may be either congruent or incongruent with regard to pitch categorization. The present study aimed to compare the neuromagnetic responses to musical and verbal stimuli for congruency judgments, for example a congruent pair for the pitch C4 sung with the pitch name *do* in a C-major context (the pitch-semantic task) or for the meaning of a word to match the speaker's identity (the voice-semantic task). Both the behavioral data and neuromagnetic data showed that congruency detection of the speaker's identity and word meaning was slower than that of the pitch and pitch name. Congruency effects of musical stimuli revealed that pitch categorization and semantic processing of pitch information were associated with P2m and N400m, respectively. For verbal stimuli, P2m and N400m did not show any congruency effect. In both the pitch-semantic task and the voice-semantic task, we found that incongruent stimuli evoked stronger slow waves with the latency of 500–600 ms than congruent stimuli. These findings shed new light on the neural mechanisms underlying pitch-naming processes.

Keywords: pitch name, speech, categorization, semantic, MEG

Introduction

The relationships between language and music, as well as their origins, have been the subject of intensive multidisciplinary research. There are studies comparing the neural substrates underlying the processing of music's and language's acoustic/structural features (Zatorre et al., 2002; Wong et al., 2008; Rogalsky et al., 2011), meanings (Koelsch et al., 2004; Steinbeis and Koelsch, 2008, 2011; Daltrozzo and Schön, 2009), combination rules (Patel, 2003; Koelsch et al., 2005; Maidhof and Koelsch, 2011), and motor expressions (Ozdemir et al., 2006; Hickok et al., 2009; Wan et al., 2011; Tsai et al., 2012). Among these aforementioned processes, the perceptual organizations of music and speech are little known in terms of their neural substrates for music and language. Like language, perceptual categorization of musical pitches also relies on symbols in numerous human cultures.

Unlike natural sounds or spoken sounds in which fundamental frequencies are continuously distributed, musical pitches are often categorized into discrete entities and are given labels. For example, the pitch names of the major mode scale in Western music are: *do, re, mi, fa, sol, la,* and *ti*. The current study aimed to examine the perceptual organizations of music (i.e., pitch) and speech (i.e., speaker identity).

Perceptual Organization of Music: Pitch

According to the mapping rules for associative transformation from a perceived frequency to a pitch name, systems of pitch naming (solmization) can be divided into two types, as represented by the fixed-*do* solmization and the moving-*do* solmization in Western music. The pitch names in the fixed-*do* solmization are determined by the fundamental frequency of auditory stimuli. On the other hand, the moving-*do* solmization relies on pitch relationships and is associated with the use of musical scales. The notes of a scale tend to be arranged unevenly within the octave, with some pitch steps bigger than others (Ball, 2008; Honingh and Bod, 2011). Different musical scales are characterized by different arrangements of pitch steps, namely, by different divisions of an octave. Each note in a scale has unique pitch relationships to other notes and these relationships allow a listener with good relative pitch to label perceived pitches using the moving-*do* solmization. Relative pitch relies on pitch relationships to label pitches. An interesting finding was that the solmization strategy in possessors of absolute pitch differs from that in possessors of relative pitch. Absolute pitch is a rare ability to identify a musical pitch without the use of an external reference pitch. On the other hand, relative pitch possessors identify a musical pitch with the use of an external reference pitch and/or a tonal context. Relative pitch possessors use the moving-*do* solmization, while absolute pitch possessors tend to use the fixed-*do* solmization (Miyazaki, 2000).

A sung pitch name informs us the fundamental frequency and pitch name that may be either congruent or incongruent with regard to pitch categorization, and a few experiments have used Stroop-like paradigms to study the congruency effect of pitch and pitch name (Itoh et al., 2005; Akiva-Kabiri and Henik, 2012; Schulze et al., 2013). With an auditory Stroop task, Miyazaki (2000) found that relative pitch possessors and absolute pitch possessors tended to use the moving-*do* solmization and the fixed-do solmization, respectively. Schulze et al. (2013) recruited musicians with absolute pitch and musicians with relative pitch (and without absolute pitch) to examine the neural substrates underlying solmization using functional magnetic resonance imaging (fMRI). They used tonal sequences as stimuli with half of these sequences being congruent (e.g., the pitch of C sung as *do*) and half incongruent (e.g., the pitch of C sung as *fa*). Their results showed that detecting verbal-tonal incongruencies activated the left superior temporal gyrus/sulcus (STG/STS) in absolute pitch possessors but not in relative pitch possessors, suggesting the involvement of semantic processing in conflict-monitoring for pitch-naming. Using the event-related potentials (ERP) technique and a Stroop-like paradigm, Itoh et al. (2005) showed that sung pitch names at incongruent

pitches evoked stronger positive slow waves 450–550 ms after the stimulus onset compared to congruent stimuli. They also found that absolute pitch possessors elicited a component 150 ms after the stimulus onset in both passive listening and pitch-naming conditions. This suggests the involvement of automatic semantic processing in absolute pitch possessors during music listening. This result is in line with a behavioral experiment of the Stroop effect by Akiva-Kabiri and Henik (2012), which reported that only absolute pitch possessors were unable to ignore the auditory tone when asked to read the note. These studies support the long-held belief that pitch identification in absolute pitch possessors is automatic and impossible to suppress. The present study focused on relative pitch, which is a common ability and its neural correlates are still not fully understood.

Perceptual Organization of Speech: Speaker Identity

In the language domain, the Stroop color-word test has been widely used to study the conflict processing in the visual modality, whereas the experimental data for spoken words is relatively scant. Haupt et al. (2009) explored the neural underpinning of an auditory Stroop task using fMRI. Their participants were presented with the words "high" or "low" in either high- or low-pitched voice, focusing either on tone pitch (relatively high or low) or on the meaning of a spoken word (high/low/good). The results showed greater activation in the anterior cingulate cortex and the pre-supplementary motor area due to task-related and sensory-level interference. Henkin et al. (2010) investigated the auditory conflict processing using ERP and behavioral measures during Stroop tasks. Their participants were asked to classify word meaning or speaker's gender while ignoring the irrelevant speaker's gender or word meaning, respectively. The results showed significantly reduced N1 amplitude and prolonged N4 in the speaker's gender task compared to the word meaning task. Using a similar auditory Stroop test, Christensen et al. (2011) demonstrated a significant interference effect with gender-typical nouns spoken by gender-mismatched voices, and the fMRI data showed that interference-related activation was localized ventrally in medial frontal areas. The methodology of these studies provides evidence of exploring the neural correlates of the perceptual organizations of speech.

Comparison of Music and Speech: Sound-Content Incongruency

While auditory Stroop-like tasks have been studied in the domains of music and speech, no experiment has directly compared the congruency effect that these two domains have on brain activities. This may be due to the distinct acoustical features of music and speech. All pitch names are monosyllabic and the processing units of Chinese words are also monosyllabic (i.e., characters; Tsang and Chen, 2009). The seven notes in Chinese traditional music are represented by monosyllabic Chinese words (*shang, che, gong, fan, liu, wu,* and *yi*). In the present study, we benefited from this characteristic of Chinese language to directly compare the congruency effect

of monosyllabic spoken words with that of monosyllabic pitch names. The specific aim of this study was two-fold: to compare the neuromagnetic activities associated with stimulus categorization of sung pitch names and spoken words, and to compare the neuromagnetic activities associated with the detection of sound-content incongruency of these stimuli. To achieve these goals, we presented participants with stimuli conveying two kinds of information for the same concept. A sung pitch name conveys: (1) the acoustic information of pitch in terms of fundamental frequency; and (2) the semantic information of pitch in terms of the pitch name. A spoken word conveys: (1) the phonetic cues of speaker's gender and age; and (2) the semantic information of gender and age. These auditory stimuli provide a novel opportunity to identify the common and distinct properties of the neural mechanisms underlying music and speech processing. We used magnetoencephalography (MEG) with a high temporal resolution to disentangle information processing stages of musical stimuli and verbal stimuli.

Regarding the hypotheses of the present study, previous electroencephalography (EEG) studies suggest that early processing of perceptual categorization enhances the amplitude of the P2 response around 200 ms after stimulus presentation (Cranford et al., 2004; Tong and Melara, 2007; Tong et al., 2009; Ross et al., 2013). We thus predicted that its neuromagnetic counterpart, P2m, is enhanced in response to musical stimuli of pitch congruent with pitch name compared to incongruent musical stimuli, because the congruent stimuli are associated with a rapid process of pitch categorization. The sources of P2m have been found to be localized in the auditory association cortices (Shahin et al., 2003; Kuriki et al., 2006; Thaerig et al., 2008; Tong et al., 2009; Liebenthal et al., 2010; Ross et al., 2013), which were defined as the regions of interests (ROIs) in this study.

Moreover, we expected the amplitude of late neuromagnetic components to be modulated by semantic conflict during both the music task and the speech task (Christensen et al., 2011). Congruency manipulation of our stimuli allows us to test the semantic N400 effect, which manifests in a larger negative potential with a latency of approximately 400 ms for words that are semantically incongruent to a given context compared to words that are congruent (for a recent review, see Kutas and Federmeier, 2011). We predicted that both musical and verbal stimuli show this N400 effect. Moreover, the conflict processing for incongruent stimuli may elicit late slow waves of magnetic fields (SWm) 500 ms after stimulus onset, as suggested by previous studies on musical pitch (Itoh et al., 2005; Elmer et al., 2013), sentence perception (van Herten et al., 2006; Frenzel et al., 2011), and color-naming Stroop tasks (Larson et al., 2009; Coderre et al., 2011).

Materials and Methods

Participants
Nineteen volunteers (20–28 years old, 11 females) were recruited by means of a public announcement, which stated the requirement of relative pitch capacity for participation in this study. All of the participants had taken musical lessons for more than 5 years, but none were professional musicians. All of them were right-handed and had normal hearing. The informed consent procedures were approved by the Institutional Review Board of Academia Sinica. Participants gave informed written consent and received monetary compensation for participating in the study. Three participants (two females and one male) were excluded from the data collection because of severe artifactual activity that could be noted in the MEG sensor array. One female participant was excluded on the basis of her low accuracy (<95%) of the pitch-semantic task in the pre-scan session (see Procedure). The data collected from the remaining fifteen participants with good relative pitch were used for the final analysis.

Stimuli
There were two tasks in this study, the pitch-semantic task and the voice-semantic task. In the pitch-semantic task, the stimuli were pitch names of *do*, *re*, *mi*, and *sol* sung by a semi-professional soprano. The sung pitches of these pitch names were restricted in four notes: C4, D4, E4, and G4. When a pitch matched the pitch name, it was a congruent stimulus in a C-major context, otherwise it was an incongruent stimulus. For example, the pitch C4 sung with the pitch name *do* was a congruent stimulus in a C-major context, whereas this pitch C4 sung with the pitch name *re* was an incongruent stimulus in a C-major context. We did not use the pitch F4 as stimulus because the E4-F4 interval is the minor second (semitone) which is not as large as the intervals between other pitches, and to discern between E4 and F4 might be particularly difficult. **Table 1** displays all the combinations of pitch and pitch names used as stimuli in this study. The fundamental frequency of these sung pitch names ranges from 261 to 392 Hz.

In the voice-semantic task, we used monosyllabic Chinese words, including: /Nan2/ (male), /Nü3/ (female), /Xiao3/ (child), and /Lau3/ (elder) spoken by a man, a woman, an 8-year-old girl, and a hoarse voice that exaggerated the voice characteristic of elderly men. If the meaning of a word matched the speaker's identity revealed by the acoustic characteristics of the voice, it was a congruent stimulus, otherwise it was an incongruent stimulus. For example, a word /Nü3/ (female) spoken by a woman was a congruent stimulus, whereas the words "male" or "elder" spoken by a woman were incongruent stimuli. The voices of these four speakers differed in acoustical features.

TABLE 1 | The sixteen stimuli of the pitch-semantic task.

Pitch \ Syllables	Pitch names			
	Congruent	Incongruent		
C4	Do	Re	Mi	Sol
D4	Re	Do	Mi	Sol
E4	Mi	Do	Re	Sol
G4	Sol	Do	Re	Mi

In congruent stimuli, the pitch and pitch name were matched in a C-major context.

The fundamental frequency of man's voice was within the range of 120–200 Hz, whereas the fundamental frequency of woman's and girl's voices was in the range of 250–400 Hz, with different formant frequencies to differentiate between a woman and a girl (Peterson and Barney, 1952; Fitch, 1997). Finally, an amateur actor mimicked the voice of an elderly man by highlighting the hoarse quality characteristic of vocal fold bowing, which is commonly found in the larynx of elderly men (Pontes et al., 2005). To facilitate the detection of sound-content incongruency, we excluded the male-elder and female-child combinations, because their voices share some acoustic characteristics, and to discern between them might be particularly difficult. **Table 2** displays all the combinations of speakers and words used as stimuli in this study.

All stimuli were edited and digitized (44,100 Hz sampling rate, 16 bit mono) for presentation using Gold Wave Digital Audio Editor (Gold Wave Inc.). The sound level of the presentation was approximately 60 dB. The duration of these stimuli ranged from 450 to 500 ms. All the musical stimuli had a stable pitch, whereas all verbal stimuli had a gliding pitch.

Procedure

There were three sessions in this experiment. In the first session, a singing test used a newly-composed melody (B♭, 4 bars) to ensure that participants had a good capacity for relative pitch. They were asked to sing this melody in movable-*do* solmization after hearing it. The participants were admitted to the next stage only if they could sing fluently. All participants passed this singing test.

In the second session, participants listened to the stimuli and practiced using the device for responding. They were instructed to make a button-press response indicating whether the stimulus was sound-content congruent or not (a right button for "congruent" and a left button for "incongruent"). There were two runs (pitch-semantic task and voice-semantic task) in this session, each run comprising 20 trials. Participants would enter the third session (MEG scan) only if their mean response accuracy was higher than 95%. One participant was excluded from the third session as a result of this criterion.

Prior to the MEG data acquisition, each participant's head shape was digitized, and head position indicator coils were used to localize the position of the participant's head inside the MEG helmet. The MEG scan consisted of four runs, each lasting approximately 4.5 min. There were 1-min breaks interspersed between these runs, and the total duration of the MEG experiment was approximately 21 min. The tasks in the four runs were: voice-semantic,

pitch-semantic, voice-semantic, and pitch-semantic. Because our participants relied on relative pitch, we presented an upward scale and a tonic chord of C major immediately before the two runs for the pitch-semantic task to help participants establish a tonal schema (tonality), which determines the rules of the pitch-to-pitch-name associative transformations.

Each run consisted of 180 trials, including 80% sound-content congruent stimuli and 20% sound-content incongruent stimuli (see **Tables 1**, **2**). This proportion of incongruent stimuli was chosen for maintaining a relatively stable tonal schema. These trials were presented in a pseudorandom order within each run. Each trial lasted 1.5 s, consisting of a sound (a sung pitch name or a spoken word) with the duration of 500 ms, followed by an interstimulus silence. Sounds were delivered binaurally through silicon tubes. Participants were instructed to indicate the sound-content congruency of the stimulus with a button press (right for "congruent" and left for "incongruent") as quickly as possible.

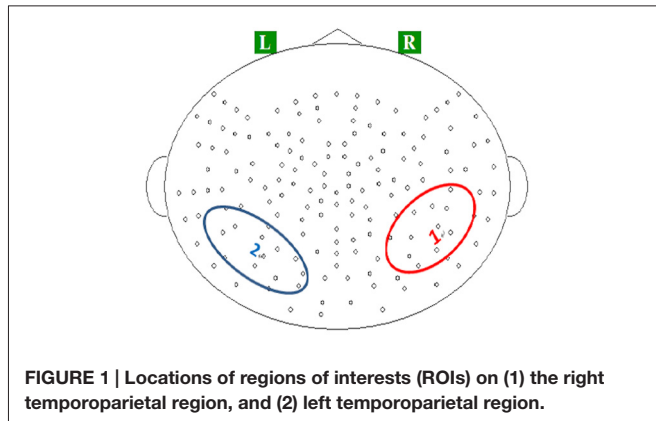## Data Acquisition and Analysis

E-Prime was used to present all stimuli and to collect the behavioral data of button-press responses. A two task by two congruency ANOVA with repeated measures was conducted to analyze the reaction time data for correct responses. Based on previous studies (Haupt et al., 2009; Christensen et al., 2011), reaction times exceeding 1000 ms were classified as outliers and did not enter the statistical analysis.

Neuromagnetic brain activities evoked by auditory stimuli were acquired using a 156-channel axial gradiometer whole-head MEG system (Kanazawa Institute of Technology, Kanazawa, Japan) at a sampling frequency of 1 kHz. A band-pass filter (DC to 100 Hz) was applied during the recording. MEG data were processed by a MEG laboratory 2.004A (Yokogawa Electric Corporation). The trials in which participants failed to respond or made incorrect responses were rejected from further analysis. MEG data were first noise reduced, and then epoched with 100 ms prestimulus intervals as well as 800 ms post-stimulus intervals. Trials with amplitude variations larger than 1.5 pT were excluded from further processing. Both the congruent stimuli and incongruent stimuli contained at least 72 artifact-free trials for each participant and each task. These trials were baseline-corrected using the prestimulus data. Each participant's MEG data within the epoch was averaged across the trials of the same condition, and was low-pass filtered at 30 Hz.

On the basis of previous studies (Itoh et al., 2005; Steinbeis and Koelsch, 2008), we focused on three components of the evoked magnetic fields: P2m, N400m, and SWm. Neuromagnetic responses recorded in temporoparietal channels were evaluated using an ROI analysis. These channels were selected due to their role in auditory categorization based on previous studies (Shahin et al., 2003; Kuriki et al., 2006; Thaerig et al., 2008; Tong et al., 2009; Liebenthal et al., 2010; Rueschemeyer et al., 2014). The ROIs were defined as the pronounced evoked fields in the average topographies for P2m (**Figure 1**). For each participant, we selected the three channels from the ROIs that recorded

TABLE 2 | The twelve stimuli of the voice-semantic task.

Syllables	Words		
Speaker	Congruent	Incongruent	
Adult man	男 /Nan2/ (man)	女 /Nü3/ (woman)	小 /Xiao3/ (young)
Adult woman	女 /Nü3/ (woman)	男 /Nan2/ (man)	老 /Lau3/ (elder)
Male elder	老 /Lau3/ (elder)	女 /Nü3/ (woman)	小 /Xiao3/ (young)
Girl	小 /Xiao3/ (young)	男 /Nan2/ (man)	老 /Lau3/ (elder)

FIGURE 1 | Locations of regions of interests (ROIs) on (1) the right temporoparietal region, and (2) left temporoparietal region.

FIGURE 2 | Reaction times for the pitch-semantic task and the voice-semantic task. Error bars indicate standard errors of the mean.

strongest P2m. Time courses of the MEG signal in these spatial ROIs were obtained by averaging the waveforms of the event-related field within the epoch of 0–600 ms over these three channels.
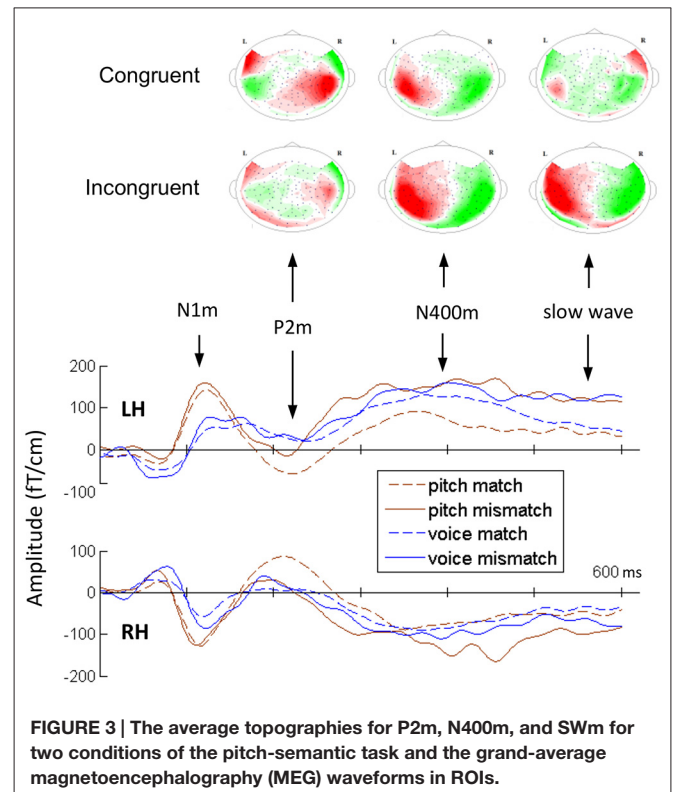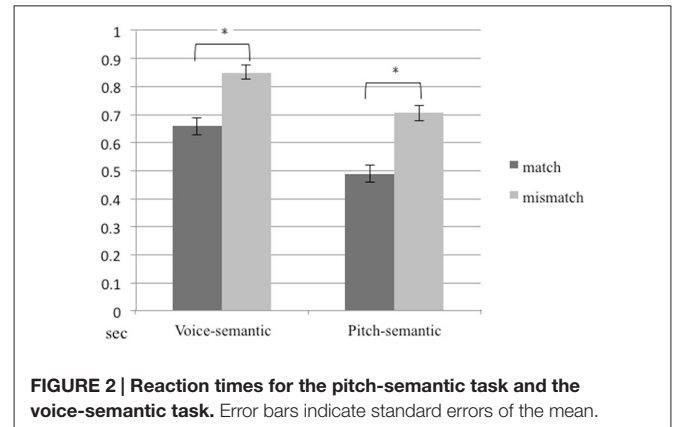
Appropriate time windows of measurement were determined by visual inspection of grand averages and individual participant data. A P2m within the 200–230 ms window, an N400m within the 330–430 ms window, and an SWm within the 500–600 ms window. We did not analyze N1m responses because it may be affected by acoustic features such as pitch glides (Mäkelä et al., 2004). The amplitudes of P2m, N400m, and SWm were estimated by averaging the field amplitude over these windows. The amplitude of each component was subjected to a repeated 2 (task) × 2 (stimulus congruency) × 2 (brain hemisphericity) ANOVA.

Results

The proportional correct for congruency/incongruency detection, averaged across participants, was 97.8%. A 2 (task) × 2 (congruency) repeated ANOVA on reaction time for trials with correct responses showed significant main effects for task ($F_{1,14} = 85.08$, $p < 0.01$) and congruency ($F_{1,14} = 622.27$, $p < 0.01$). The mean of reaction time in the pitch-semantic task was shorter than that in the voice-semantic task. The mean of reaction time for congruent stimuli was shorter than for incongruent stimuli (**Figure 2**).

Figure 3 shows the grand average MEG waveform components at 210, 380, and 550 ms for each condition and ROIs. Both the incongruent musical stimuli and incongruent verbal stimuli evoked significant N400m and SWm, whereas congruent stimuli of sung pitch name evoked significant P2m. The mean amplitudes of P2m, N400m, and SWm are presented in **Figure 4**. The results of ANOVA for peak amplitude are summarized as follows (only significant effects were reported).
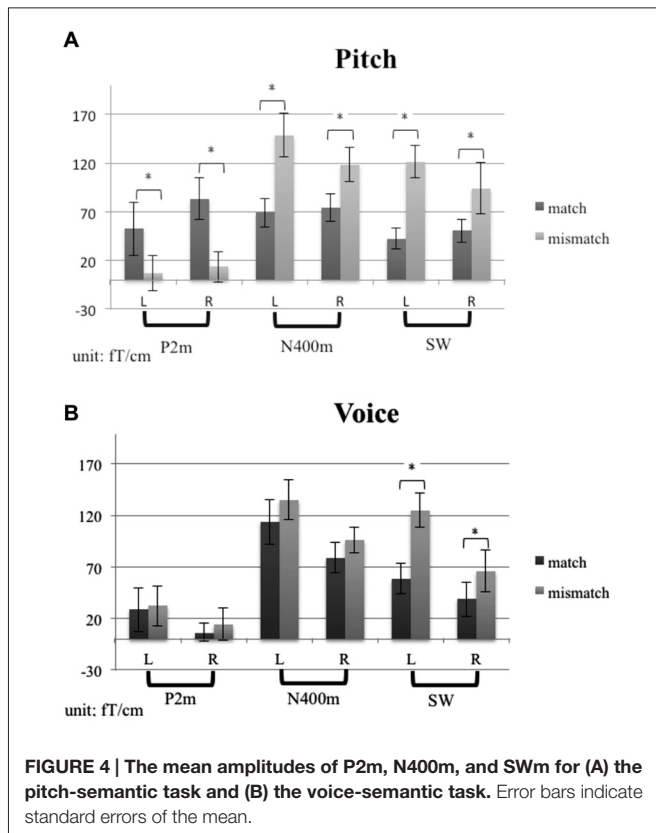
P2m. A 2 (task) × 2 (congruency) × 2 (hemisphere) repeated ANOVA revealed a main effect of congruency ($F_{1,14} = 8.33$, $p < 0.05$) and a significant interaction between congruency and task ($F_{1,14} = 4.51$, $p = 0.05$). A simple main effects analysis was further conducted, showing that congruent stimuli

FIGURE 3 | The average topographies for P2m, N400m, and SWm for two conditions of the pitch-semantic task and the grand-average magnetoencephalography (MEG) waveforms in ROIs.

evoked stronger P2m for the pitch-semantic task compared to incongruent stimuli ($F_{1,28} = 11.95$, $p < 0.01$).

N400m. A 2 (task) × 2 (congruency) × 2 (hemisphere) repeated ANOVA revealed a significant main effect of congruency ($F_{1,14} = 37.26$, $p < 0.01$) and an interaction between congruency and task ($F_{1,14} = 7.40$, $p < 0.05$). A simple main effects analysis was further conducted, showing that incongruent stimuli evoked stronger N400m for the pitch-semantic task compared to congruent stimuli ($F_{1,28} = 36.33$, $p < 0.01$).

SWm. A 2 (task) × 2 (congruency) × 2 (hemisphere) repeated ANOVA revealed a main effect of congruency ($F_{1,14} = 38.44$, $p < 0.01$). The hemisphere × task × congruency interaction ($F_{1,14} = 5.3$, $p < 0.05$), the hemisphere × task interaction ($F_{1,14} = 5.05$, $p < 0.05$), and task × congruency

FIGURE 4 | The mean amplitudes of P2m, N400m, and SWm for (A) the pitch-semantic task and (B) the voice-semantic task. Error bars indicate standard errors of the mean.

interaction ($F_{1,14} = 6.62$, $p < 0.05$) were also significant. To further understand the three-way interaction, a 2 (task) by 2 (congruency) repeated ANOVA was calculated for the left hemisphere and the right hemisphere separately. In the left hemisphere, the main effect of congruency was significant ($F_{1,14} = 21.92$, $p < 0.01$). In the right hemisphere, there was a significant interaction effect of task × congruency ($F_{1,14} = 15.08$, $p < 0.01$) and a main effect of congruency ($F_{1,14} = 34.63$, $p < 0.05$). A follow-up simple simple main effects analysis illustrated that incongruent stimuli evoked stronger SWm than congruent stimuli for both tasks (pitch-semantic task, $F_{1,28} = 49.40$, $p < 0.01$; voice-semantic task, $F_{1,28} = 5.04$, $p < 0.05$).

Discussion

The present MEG study aimed at comparing the neuromagnetic activities associated with perceptual categorization and congruency effects within the music and speech domains. We observed that the detection of semantic congruency/incongruency for musical stimuli occurred earlier than that for verbal stimuli. This pattern emerged in both behavioral data (reaction time) and event-related field components. The reaction time in the pitch-semantic task was significantly shorter than the voice-semantic task, and the reaction time of congruent stimuli was significantly shorter than for incongruent stimuli. Neuromagnetic data suggests that P2m evoked by congruent stimuli of sung pitches was

stronger than that evoked by incongruent pitches. In addition, we replicated the previously documented N400 effect of musical stimuli (Steinbeis and Koelsch, 2008, 2011; Daltrozzo and Schön, 2009) but we did not find the N400 effect of verbal stimuli. Furthermore, for both musical and verbal stimuli, SWm evoked by incongruent stimuli was stronger than that evoked by congruent stimuli, as predicted by previous studies on conflict processing (West, 2003; van Herten et al., 2006; Larson et al., 2009; Coderre et al., 2011; Frenzel et al., 2011) and a Stroop-like effect of pitch naming (Itoh et al., 2005).

The earliest neuromagnetic component showing the congruency effect of pitch naming was P2m, which was pronounced 200–230 ms after stimulus onset and enhanced for congruent stimuli. To the best of our knowledge, the present study is the first one reporting that musical pitches elicited a congruency-sensitive component with latency shorter than 230 ms. In the ERP experiment by Itoh et al. (2005), the P2 responses to the musical stimuli with congruent pitch and pitch name appeared stronger than incongruent stimuli when selective attention was focused on pitch (see Figure 5A of their paper). However, they did not perform statistical analysis on the P2 amplitude.

The enhanced P2m response to pitch congruent with pitch name may reflect the activation of short-term memory during a rapid perceptual categorization of auditory stimuli. On the other hand, the incongruent musical stimuli may fail to be classified due to conflicting information, thereby inducing a weaker P2m response. In EEG studies, P2 has been related to the processes of object identification and stimulus classification (Cranford et al., 2004; Tong and Melara, 2007; Tong et al., 2009; Ross et al., 2013). In a combined EEG-fMRI experiment, Liebenthal et al. (2010) compared the pattern of activation in the left STS between familiar phonemic patterns and unfamiliar nonphonemic patterns during categorization. They found stronger P2 responses to phonemic patterns than nonphonemic patterns before training and increased P2 response to nonphonemic patterns with training. The authors argued that P2 may reflect the activation of neural representations of the relevant (trained) sound features providing the basis for perceptual categorization, and the P2 training effect may be related to the activation of new short-term neural representations of novel auditory categories in the left STS (also see Hickok et al., 2003). In this study, participants may have kept short-term memories of sung pitch names in the C-major scale, which served as the tonal schema for pitch categorization. The musical stimuli of pitch congruent with pitch name might activate the short-term neural representations of musical pitch and elicit a prominent P2 response. This view is in line with Marie et al. (2011), who found that metrically incongruous words elicited larger P2 components in musicians compared to metrically congruous words. Their result shows P2 enhancement by a match between the auditory input and the metrical template. Our result also shows the similar effect by a match between the auditory input and the template of pitch names.

Previous findings of music-listening also support the idea that P2 may reflect the activation of neural representations of

the trained sound features providing the basis for perceptual categorization. Shahin et al. (2003, 2005) examined whether the auditory-evoked responses were modulated by the spectral complexity of musical sounds, finding larger P2 responses to instrumental sounds in musicians relative to non-musicians. Moreover, the musician's P2 response to instrumental sounds was enhanced relative to pure tones. Kuriki et al. (2006) also found P2 enhancement for harmonic progressions in musicians. Seppänen et al. (2012) used the oddball paradigm to examine training-induced neural plasticity, finding that P2 amplitude was enhanced after 15 min of passive exposure in both musicians and non-musicians. The P2 training effect was also found in discrimination experiments of amplitude modulated pure tones (Bosnyak et al., 2004) and speech sounds (Tremblay et al., 2009). Our hypothesis of P2 playing a role in categorization is also supported by its topography. Although we did not reconstruct the electromagnetic sources for MEG components, the grand-average topography of P2 shows direction inversion in the magnetic fields around the bilateral superior temporal regions (**Figure 3**). This provides strong evidence for the existence of current dipoles around these regions. Previous studies have converged to indicate that the sources of the auditory-evoked P2 were located in the mid-posterior regions of STG/STS (Verkindt et al., 1994; Godey et al., 2001; Shahin et al., 2003; Bosnyak et al., 2004; Kuriki et al., 2006; Thaerig et al., 2008; Liebenthal et al., 2010).

Regarding N400m, our finding of an enhanced N400m response to incongruent stimuli for the pitch-semantic task agrees with previous research on the semantic N400 effect, which is reflected in a larger N400 amplitude for words that are semantically incongruent to a given context than words that are congruent. ERP research over recent decades has indicated that N400 can be elicited by a wide range of stimulus types, and its amplitude is sensitive to semantic manipulations (for a recent review, see Kutas and Federmeier, 2011). In the music domain, N400 was mostly examined using an affective priming paradigm. Several studies have revealed that a target musical sound elicits stronger N400 when its prime word (Steinbeis and Koelsch, 2008, 2011; Daltrozzo and Schön, 2009) or facial expression (Kamiyama et al., 2013) has an incongruent emotional meaning. Recently, an N400 semantic priming effect for the congruency of pitch and pitch name was reported to be related to absolute pitch. Elmer et al. (2013) presented musical tones and visual labels of pitches to musicians with/without absolute pitch, finding an increased N400 effect in possessors of absolute pitch, in comparison with non-possessors of absolute pitch. The present study extends previous findings by demonstrating an N400 semantic effect for the concurrent information of acoustic pitch and pitch name in possessors of relative pitch.

To our surprise, we did not find the semantic N400 effect for verbal stimuli. In reality, we seldom feel odd when hearing the word "male" spoken by a female, and *vice versa*. Whereas the detection of the incongruency of the speaker's identity and word meaning was instructed by the experimenter, detection of the incongruency of pitch and pitch name may be relatively automatic. In the past decade, accumulated evidence suggests

that the N400 effects seem to occur implicitly and may be associated with relatively automatic processes (Rolke et al., 2001; Deacon et al., 2004; Kiefer and Brendel, 2006; Kelly et al., 2010; Schendan and Ganis, 2012). In our view, the N400 effect observed in the pitch-semantic task might be attributed to the automaticity of incongruency detection of pitch and pitch name. Previous experiments suggest that N400 amplitude is likely to vary with many of the same factors that influence the reaction time (Gomes et al., 1997; Chwilla and Kolk, 2005; Kutas and Federmeier, 2011; Lehtonen et al., 2012). Moreover, the dipole source of N400 has been consistently suggested to be located in the temporal lobe (Dien et al., 2010; Dobel et al., 2010; Hirschfeld et al., 2011; Kutas and Federmeier, 2011) and is likely associated with sensory and automatic processes.

Given that the mean reaction time in the voice-semantic task was longer than pitch-semantic task for more than 100 ms, the incongruency detection in the voice-semantic task may not be automatic, and therefore this task did not show the N400 effect. It should be noted that the right anterior STG/STS is involved in integrative processing of several acoustical features necessary for speaker identification (Belin and Zatorre, 2003; von Kriegstein et al., 2003; Lattner et al., 2005; Bonte et al., 2014). This may partially explain that the mean reaction time in the pitch-semantic task was shorter than voice-semantic task.

We found that the incongruency detection of the speaker's identity and mismatched word meaning was significantly slower than the congruency detection of the speaker's identity and matched word meaning. This suggests the existence of an incongruity and conflict process in the verbal domain. Our view that the voice-semantic task mainly involves controlled processes is supported by a prior study of verbal processing. An interference effect with gender-typical nouns spoken by gender-mismatched voices (e.g., "father" spoken by a woman) was related to a controlled process (Christensen et al., 2011).

As to the final component of SWm (500–600 ms) in both the pitch-semantic task and the voice-semantic task, we found that incongruent stimuli evoked stronger SWm than congruent stimuli. The congruency effect of the voice-semantic task manifests in the left hemisphere, whereas the congruency effect of the pitch-semantic task manifests in both hemispheres. In EPR studies, the late positive slow wave after 500 ms is related to conflict detection and resolution (West, 2003; van Herten et al., 2006; Larson et al., 2009; Coderre et al., 2011; Frenzel et al., 2011). Our result of the congruency effect in the pitch-semantic task is consistent with Itoh et al. (2005), who found enhanced parietal late slow waves for auditory stimuli of pitch incongruent with pitch name in non-possessors of absolute pitch, in comparison to congruent stimuli. In ERP studies, the conflict slow potential reflects greater positivity for incongruent trials than for congruent trials over the parietal region 500 ms after stimulus onset (Liotti et al., 2000; West and Alain, 2000). We hypothesized that SWm may be the magnetic counterpart of the parietal conflict slow potential.

A limitation of this study was that the observed differences in the reaction time and the strength of event-related field

components between music and speech may relate to the acoustical properties of these two types of stimuli. While categorization of musical pitch relies on the fundamental frequency of voice, categorization of speaker identity relies on the fundamental frequency and voice quality. The voice quality of speech stimuli varied across speakers, whereas the voice quality of music stimuli did not change. The larger variation of voice quality of speech stimuli may increase the difficulty of the congruency-detection task compared to music stimuli and affects the amplitude of P2m, N400m, and SWm.

The present study benefited from the monosyllabic Chinese words in minimizing the acoustical differences between the speech and music stimuli. While all stimuli were monosyllabic, however, the spoken Chinese words differ from the sung pitch names in the gliding fundamental pitch. Future investigations should assess the effect of the gliding fundamental pitch of lexical tones on semantic processing.

It should be noted that our participants relied on relative pitch, and the enhanced P2m responses to pitch congruent with pitch name reflect the categorization of musical pitches in a given tonal context, which was established by an upward scale and a tonic chord immediately before the two runs for the pitch-semantic task. In contrast, absolute pitch possessors tend to categorize musical pitches with the use of the fixed-do solmization and without the use of an external reference pitch or a tonal context. A detailed comparison between relative pitch and absolute pitch awaits future research.

Conclusion

We compared the neuromagnetic responses to musical stimuli and verbal stimuli, with the sound-content congruency of these stimuli being manipulated. Detection of the incongruency of the speaker's identity and word meaning was slower than the detection of the incongruency of pitch and pitch name, as revealed by reaction time and event-related field components. We reported the novel finding of enhanced P2m elicited by pitch congruent with pitch name, which suggests that perceptual categorization of musical pitches occurs earlier than the detection of semantic incongruency reflected by N400m. For verbal stimuli, P2m and N400m did not show any congruency effect. Our results allow for the attribution of the nature and use of musical scales in numerous human cultures. Although the fundamental frequencies of sounds distribute continuously, our cognitive system tends to categorize musical pitches into discrete entities and to label each categorized pitch with a name. We suggest that pitch categorization with the use of the moving-do solmization occurs 200–230 ms after stimulus onset.

Acknowledgments

References

Akiva-Kabiri, L., and Henik, A. (2012). A unique asymmetrical stroop effect in absolute pitch possessors. *Exp. Psychol.* 59, 272–278. doi: 10.1027/1618-3169/a000153

Ball, P. (2008). Science and music: facing the music. *Nature* 453, 160–162. doi: 10.1038/453160a

Belin, P., and Zatorre, R. J. (2003). Adaptation to speaker' voice in right anterior temporal lobe. *Neuroreport* 14, 2105–2109. doi: 10.1097/00001756-200311140-00019

Bonte, M., Hausfeld, L., Scharke, W., Valente, G., and Formisano, E. (2014). Task-dependent decoding of speaker and vowel identity from auditory cortical response patterns. *J. Neurosci.* 34, 4548–4557. doi: 10.1523/JNEUROSCI.4339-13.2014

Bosnyak, D. J., Eaton, R. A., and Roberts, L. E. (2004). Distributed auditory cortical representations are modified when non-musicians are trained at pitch discrimination with 40 Hz amplitude modulated tones. *Cereb. Cortex* 14, 1088–1099. doi: 10.1093/cercor/bhh068

Christensen, T. A., Lockwood, J. L., Almryde, K. R., and Plante, E. (2011). Neural substrates of attentive listening assessed with a novel auditory Stroop task. *Front. Hum. Neurosci.* 4:236. doi: 10.3389/fnhum.2010.00236

Chwilla, D. J., and Kolk, H. H. (2005). Accessing world knowledge: evidence from N400 and reaction time priming. *Brain Res. Cogn. Brain Res.* 25, 589–606. doi: 10.1016/j.cogbrainres.2005.08.011

Coderre, E., Conklin, K., and van Heuven, W. J. (2011). Electrophysiological measures of conflict detection and resolution in the Stroop task. *Brain Res.* 1413, 51–59. doi: 10.1016/j.brainres.2011.07.017

Cranford, J. L., Rothermel, A. K., Walker, L., Stuart, A., and Elangovan, S. (2004). Effects of discrimination task difficulty on N1 and P2 components of late

auditory evoked potential. *J. Am. Acad. Audiol.* 15, 456–461. doi: 10.3766/jaaa.15.6.6

Daltrozzo, J., and Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *J. Cogn. Neurosci.* 21, 1882–1892. doi: 10.1162/jocn.2009.21113

Deacon, D., Grose-Fifer, J., Yang, C. M., Stanick, V., Hewitt, S., and Dynowska, A. (2004). Evidence for a new conceptualization of semantic representation in the left and right cerebral hemispheres. *Cortex* 40, 467–478. doi: 10.1016/s0010-9452(08)70140-0

Dien, J., Michelson, C. A., and Franklin, M. S. (2010). Separating the visual sentence N400 effect from the P400 sequential expectancy effect: cognitive and neuroanatomical implications. *Brain Res.* 1355, 126–140. doi: 10.1016/j.brainres.2010.07.099

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. *J. Cogn. Neurosci.* 22, 1251–1261. doi: 10.1162/jocn.2009.21297

Elmer, S., Sollberger, S., Meyer, M., and Jäncke, L. (2013). An empirical reevaluation of absolute pitch: behavioral and electrophysiological measurements. *J. Cogn. Neurosci.* 25, 1736–1753. doi: 10.1162/jocn_a_00410

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *J. Acoust. Soc. Am.* 102, 1213–1222. doi: 10.1121/1.421048

Frenzel, S., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2011). Conflicts in language processing: a new perspective on the N400–P600 distinction. *Neuropsychologia* 49, 574–579. doi: 10.1016/j.neuropsychologia.2010.12.003

Godey, B., Schwartz, D., de Graaf, J. B., Chauvel, P., and Liégeois-Chauvel, C. (2001). Neuromagnetic source localization of auditory evoked fields and intracerebral evoked potentials: a comparison of data in the same patients. *Clin. Neurophysiol.* 112, 1850–1859. doi: 10.1016/s1388-2457(01)00636-8

Gomes, H., Ritter, W., Tartter, V. C., Vaughan, H. G., Jr., and Rosen, J. J. (1997). Lexical processing of visually and auditorily presented nouns and verbs: evidence from reaction time and N400 priming data. *Brain Res. Cogn. Brain Res.* 6, 121–134. doi: 10.1016/s0926-6410(97)00023-2

Haupt, S., Axmacher, N., Cohen, M. X., Elger, C. E., and Fell, J. (2009). Activation of the caudal anterior cingulate cortex due to task-related interference in an auditory Stroop paradigm. *Hum. Brain Mapp.* 30, 3043–3056. doi: 10.1002/hbm.20731

Henkin, Y., Yaar-Soffer, Y., Gilat, S., and Muchnik, C. (2010). Auditory conflict processing: behavioral and electrophysiologic manifestations of the stroop effect. *J. Am. Acad. Audiol.* 21, 474–486. doi: 10.3766/jaaa.21.7.6

Hickok, G., Buchsbaum, B., Humphries, C., and Muftuler, T. (2003). Auditorymotor interaction revealed by fMRI: speech, music and working memory in area Spt. *J. Cogn. Neurosci.* 15, 673–682. doi: 10.1162/jocn.2003.15.5.673

Hickok, G., Okada, K., and Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *J. Neurophysiol.* 101, 2725–2732. doi: 10.1152/jn.91099.2008

Hirschfeld, G., Zwitserlood, P., and Dobel, C. (2011). Effects of language comprehension on visual processing – MEG dissociates early perceptual and late N400 effects. *Brain Lan.* 116, 91–96. doi: 10.1016/j.bandl.2010.07.002

Honingh, A., and Bod, R. (2011). In search of universal properties of musical scales. *J. New Music Res.* 40, 81–89. doi: 10.1080/09298215.2010.543281

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K., and Nakada, T. (2005). Electrophysiological correlates of absolute pitch and relative pitch. *Cereb. Cortex* 15, 760–769. doi: 10.1093/cercor/bhh177

Kamiyama, K. S., Abla, D., Iwanaga, K., and Okanoya, K. (2013). Interaction between musical emotion and facial expression as measured by event-related potentials. *Neuropsychologia* 51, 500–505. doi: 10.1016/j.neuropsychologia.2012.11.031

Kelly, S. D., Creigh, P., and Bartolotti, J. (2010). Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J. Cogn. Neurosci.* 22, 683–694. doi: 10.1162/jocn.2009.21254

Kiefer, M., and Brendel, D. (2006). Attentional modulation of unconscious "automatic" processes: evidence from event-related potentials in a masked priming paradigm. *J. Cogn. Neurosci.* 18, 184–198. doi: 10.1162/089892906775783688

Koelsch, S., Gunter, T. C., Wittfoth, M., and Sammler, D. (2005). Interaction between syntax processing in language and in music: an ERP Study. *J. Cogn. Neurosci.* 17, 1565–1577. doi: 10.1162/089892905774597290

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., and Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nat. Neurosci.* 7, 302–307. doi: 10.1038/nn1197

Kuriki, S., Kanda, S., and Hirata, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *J. Neurosci.* 26, 4046–4053. doi: 10.1523/jneurosci.3907-05.2006

Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu. Rev. Psychol.* 62, 621–647. doi: 10.1146/annurev.psych.093008.131123

Larson, M. J., Kaufman, D. A., and Perlstein, W. M. (2009). Neural time course of conflict adaptation effects on the Stroop task. *Neuropsychologia* 47, 663–670. doi: 10.1016/j.neuropsychologia.2008.11.013

Lattner, S., Meyer, M. E., and Friederici, A. D. (2005). Voice perception: Sex, pitch and the right hemisphere. *Hum. Brain Mapp.* 24, 11–20. doi: 10.1002/hbm.20065

Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuominen, J., and Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: behavioral and ERP evidence. *Neuropsychologia* 50, 1362–1371. doi: 10.1016/j.neuropsychologia.2012.02.021

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., and Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cereb. Cortex* 20, 2958–2970. doi: 10.1093/cercor/bhq045

Liotti, M., Woldorff, M. G., Perez, R., and Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia* 38, 701–711. doi: 10.1016/S0028-3932(99)00106-2

Maidhof, C., and Koelsch, S. (2011). Effects of selective attention on syntax processing in music and language. *J. Cogn. Neurosci.* 23, 2252–2267. doi: 10.1162/jocn.2010.21542

Mäkelä, A. M., Alku, P., Mäkinen, V., and Tiitinen, H. (2004). Glides in speech fundamental frequency are reflected in the auditory N1m response. *NeuroReport* 15, 1205–1208. doi: 10.1097/00001756-200405190-00025

Marie, C., Magne, C., and Besson, M. (2011). Musicians and the metric structure of words. *J. Cogn. Neurosci.* 23, 294–305. doi: 10.3410/f.1453958.932056

Miyazaki, K. (2000). "Interaction in musical-pitch naming and syllable naming: an experiment on a Stroop-like effect in hearing," in *Integrated Human Brain Science: Theory, Method, Application (Music)*, ed. T. Nakada, (Amsterdam: Elsevier), 415–423.

Ozdemir, E., Norton, A., and Schlaug, G. (2006). Shared and distinct neural correlates of singing and speaking. *Neuroimage* 33, 628–635. doi: 10.1016/j.neuroimage.2006.07.013

Patel, A. D. (2003). Language, music, syntax and the brain. *Nat. Neurosci.* 6, 674–681. doi: 10.1038/nn1082

Peterson, G. E., and Barney, H. L. (1952). Control methods used in a study of vowels. *J. Acoust. Soc. Am.* 24, 175–184. doi: 10.1121/1.1906875

Pontes, P., Brasolotto, A., and Behlau, M. (2005). Glottic characteristics and voice complaint in the elderly. *J. Voice* 19, 84–94. doi: 10.1016/j.jvoice.2004.09.002

Rogalsky, C., Rong, F., Saberi, K., and Hickok, G. (2011). Functional anatomy of language and music perception: temporal and structural factors investigated using functional magnetic resonance imaging. *J. Neurosci.* 31, 3843–3852. doi: 10.1523/JNEUROSCI.4515-10.2011

Rolke, B., Heil, M., Streb, J., and Hennighausen, E. (2001). Missed prime words within the attentional blink evoke an N400 semantic priming effect. *Psychophysiology* 38, 165–174. doi: 10.1111/1469-8986.3820165

Ross, B., Jamali, S., and Tremblay, K. L. (2013). Plasticity in neuromagnetic cortical responses suggests enhanced auditory object representation. *BMC Neurosci.* 14:151. doi: 10.1186/1471-2202-14-151

Rueschemeyer, S. A., Ekman, M., van Ackeren, M., and Kilner, J. (2014). Observing, performing and understanding actions: revisiting the role of cortical motor areas in processing of action words. *J. Cogn. Neurosci.* 26, 1644–1653. doi: 10.1162/jocn_a_00576

Schendan, H. E., and Ganis, G. (2012). Electrophysiological potentials reveal cortical mechanisms for mental imagery, mental simulation and grounded (embodied) cognition. *Front. Psychol.* 3:329. doi: 10.3389/fpsyg.2012.00329

Schulze, K., Mueller, K., and Koelsch, S. (2013). Auditory stroop and absolute pitch: An fMRI study. *Hum. Brain Mapp.* 34, 1579–1590. doi: 10.1002/hbm.22010

Seppänen, M., Hämäläinen, J., Pesonen, A. K., and Tervaniemi, M. (2012). Music training enhances rapid neural plasticity of n1 and p2 source activation for unattended sounds. *Front. Hum. Neurosci.* 6:43. doi: 10.3389/fnhum.2012.00043

Shahin, A., Bosnyak, D. J., Trainor, L. J., and Roberts, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *J. Neurosci.* 23, 5545–5552.

Shahin, A., Roberts, L. E., Pantev, C., Trainor, L. J., and Ross, B. (2005). Modulation of P2 auditory-evoked responses by the spectral complexity of musical sounds. *Neuroreport* 16, 1781–1785. doi: 10.1097/01.wnr.0000185017.29316.63

Steinbeis, N., and Koelsch, S. (2008). Comparing the processing of music and language meaning using EEG and FMRI provides evidence for similar and distinct neural representations. *PLoS One* 3:e2226. doi: 10.1371/journal.pone.0002226

Steinbeis, N., and Koelsch, S. (2011). Affective priming effects of musical sounds on the processing of word meaning. *J. Cogn. Neurosci.* 23, 604–621. doi: 10.1162/jocn.2009.21383

Thaerig, S., Behne, N., Schadow, J., Lenz, D., Scheich, H., Brechmann, A., et al. (2008). Sound level dependence of auditory evoked potentials: simultaneous EEG recording and low-noise fMRI. *Int. J. Psychophysiol.* 67, 235–241. doi: 10.1016/j.ijpsycho.2007.06.007

Tong, Y., and Melara, R. D. (2007). Behavioral and electrophysiological effects of distractor variation on auditory selective attention. *Brain Res.* 1166, 110–123. doi: 10.1016/j.brainres.2007.06.061

Tong, Y., Melara, R. D., and Rao, A. (2009). P2 enhancement from auditory discrimination training is associated with improved reaction times. *Brain Res.* 1297, 80–88. doi: 10.1016/j.brainres.2009.07.089

Tremblay, K. L., Shahin, A. J., Picton, T., and Ross, B. (2009). Auditory training alters the physiological detection of stimulus-specific cues in humans. *Clin. Neurophysiol.* 120, 128–135. doi: 10.1016/j.clinph.2008.10.005

Tsai, C. G., Fan, L. Y., Lee, S. H., Chen, J. H., and Chou, T. L. (2012). Specialization of the posterior temporal lobes for audio-motor processing – evidence from a functional magnetic resonance imaging study of skilled drummers. *Eur. J. Neurosci.* 35, 634–643. doi: 10.1111/j.1460-9568.2012.07996.x

Tsang, Y. K., and Chen, H. C. (2009). Do position-general radicals have a role to play in processing Chinese characters? *Lang. Cognitive Proc.* 24, 947–966. doi: 10.1080/01690960802154615

van Herten, M., Chwilla, D. J., and Kolk, H. H. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *J. Cogn. Neurosci.* 18, 1181–1197. doi: 10.1162/jocn.2006.18.7.1181

Verkindt, C., Bertrand, O., Thevenet, M., and Pernier, J. (1994). Two auditory components in the 130–230 ms range disclosed by their stimulus frequency dependence. *Neuroreport* 5, 1189–1192. doi: 10.1097/00001756-199406020-00007

von Kriegstein, K., Egerm, E., Kleinschmidt, A., and Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res. Cogn. Brain Res.* 17, 48–55. doi: 10.1016/s0926-6410(03)00079-x

Wan, C. Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., et al. (2011). Auditory-motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: a proof of concept study. *PLoS One* 6:e25505. doi: 10.1371/journal.pone.0025505

West, R. (2003). Neural correlates of cognitive control and conflict detection in the Stroop and digit-location tasks. *Neuropsychologia* 41, 1122–1135. doi: 10.1016/s0028-3932(02)00297-x

West, R., and Alain, C. (2000). Effects of task context and fluctuations of attention on neural activity supporting performance of the Stroop task. *Brain Res.* 873, 102–111. doi: 10.1016/S0006-8993(00)02530-0

Wong, P. C., Warrier, C. M., Penhune, V. B., Roy, A. K., Sadehh, A., Parrish, T. B., et al. (2008). Volume of left Heschl's Gyrus and linguistic pitch learning. *Cereb. Cortex* 18, 828–836. doi: 10.1001/archpsyc.60.8.766

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends. Cogn. Sci.* 6, 37–46. doi: 10.1016/s1364-6613(00)01816-7