



Brain responses to spoken F_0 changes: Is H special?



Chun-Hsien Hsu, Jonathan P. Evans*, Chia-Ying Lee

Institute of Linguistics, Academia Sinica, No. 128, Section 2, Academia Road, 115 Taipei, Taiwan, ROC

ARTICLE INFO

Article history:

Received 28 April 2014

Received in revised form

31 December 2014

Accepted 9 February 2015

Available online 28 February 2015

Keywords:

F_0 elevation

Direction of deviants

Mismatch negativity

P300

Beta band power

ABSTRACT

Across languages and linguistic phenomena, rises in spoken pitch seem to occupy a privileged position compared to steady F_0 or pitch lowering. Speakers are more likely to use sudden rises in pitch to arouse listener attention, rather than using falls; e.g., contrastive stress, questions seeking a response, beginnings of units of discourse, expressions of intense emotion, etc. The study evaluated whether there are brain responses that are more sensitive to stepwise raised vs. lowered spoken F_0 , and whether any such responses were also caused by pure tone stimuli. Three types of brain response were evaluated. Mismatch negativity (MMN), which reflects pre-attentive mechanisms, was only sensitive to degrees of F_0 change, and not to direction. P300, which is an indicator of attention orientation, did not show sensitivity to F_0 direction to nonspeech stimuli, but showed greater sensitivity to raised F_0 than to lowered F_0 in speech. Time–frequency analysis of EEG data showed beta-band response in the right parietal area, when presented with spoken F_0 elevation. Taken together, the P300 and beta band results suggest that attention modulation mechanisms are triggered when listeners are exposed to sudden rises in spoken F_0 . However, there was a lack of attention orientation response when non-speech F_0 changes were presented, or when subjects were exposed to lowering of F_0 in speech.

© 2015 Published by Elsevier Ltd.

1. Introduction

Across the world's languages, sudden, or step-wise rises in spoken pitch are employed in linguistic situations in which a speaker might desire increased attention from a listener. Examples include contrastively focused syllables or words, beginnings of discourse segments, initiation of new topics, and lexical stress in the majority of languages that have it (Brown, Currie, & Kenworthy, 1980; Bolinger, 1986, 1989; Fernald & Mazzie, 1991; Swerts, 1997; Fujisaki, Ohno, & Wang, 1998; Gussenhoven, 2004 and many other studies). Even in languages with phonological tone, sudden rises in F_0 mark new discourse sections or sentences, as well as marking questions (Shih, 2000; Kochanski, Shih, & Jing, 2003; Yuan, 2011). Of relevance to the current study is the fact that when uttering a question, a speaker seeks a listener's attention, and desires a response.

Paralinguistically, expressions of passionate emotions such as happiness, anger, fear and surprise tend to be expressed by sustained elevated F_0 , across languages and cultures (Pell, Monetta, Paulmann, & Kotz, 2009; Schröder, 2001; Williams & Stevens, 1972). In infant-directed speech, speakers employ pitch peaks on stressed syllables, and exaggerated pitch peaks to indicate focus (Krahmer & Swerts, 2001).

A special status of phonologically high pitch is found in many tonal languages, particularly noted in Africa. In cases where the tone system can be reduced to an inventory of two tone levels, it is typically realized as an opposition between a neutral, unmarked tone on a lower default pitch, with less frequently occurring high-toned syllables. The opposite situation is much less frequent, although attested (Hyman, 2001; Sun, 1997). In addition, in these languages, H(igh) tones are often targets for phonological processes including insertion, removal and movement. One well-documented case is the deletion of the second H tone in a HH sequence (Meussen's rule; cf. Goldsmith, 1984, Yip, 2002:100–101). On the other hand, phonologically L(ow) tones typically emerge via default processes, and are not subject to insertion, removal or movement.

On the perception side, studies have used overt tasks, which require intentional motor output by the participants. For example, in pitch discrimination tasks, participants might press a button to indicate that one sound is higher or lower than another. It has been demonstrated that human pitch discrimination is often more sensitive to F_0 raising than to F_0 lowering. For example, Kishon-Rabin,

* Corresponding author. Tel.: +886 2 2652 5038; fax: +886 2 2785 6622.

E-mail address: jonathan@sinica.edu.tw (J.P. Evans).

Roth, Batya Van Dijk, Yinon, and Amir (2004) indicate that the threshold for detecting pitch rises is lower than that for detecting pitch lowering. This study also showed that sensitivity to rises in pitch was significantly affected by the height of the baseline pitch. That is, at a baseline pitch of 200 Hz, 15 out of 16 participants showed sensitivity to minor pitch rises. On the other hand, at a baseline pitch of 1000 Hz, only 11 out of 16 participants showed the same sensitivity. These two sets of results imply that there may be a heightened sensitivity to raising of pitch within the range of human speech *F0* (Honorof & Whalen, 2005).

It is not trivial to explain why humans have different sensitivities to pitch changes at high versus normal pitch ranges. One possible reason is the trade-off between pitch and loudness in pitch perception (Fletcher & Munson, 1933; Robinson & Dadson, 1956). Although the loudness level might play a role in pitch perception, interestingly, brain wave studies, summarized below, indicate that acoustic perception of pitch and loudness might be independently supported by different neural mechanisms (Giard et al., 1995; Schröger, 1996). Furthermore, it is possible that sensitivity to detection of pitch raises might be grounded in patterns present in the speech stream to which language learners are exposed. This study seeks to discover whether speaker/listeners are in fact more sensitive to spoken pitch rises than to lowering, and, if so, what aspects of the brain's processing of sound inputs might play a role.

The present study aimed to investigate the processing of raised versus lowered *F0* by measuring three brain responses, including mismatch negativity, P300, and time–frequency responses. The functional role of each index is described below. Mismatch negativity (MMN) is a component of event-related brain potentials (ERPs) that are used to index the detection of automatic and pre-attentive auditory changes (Näätänen, 1992; Näätänen, Astikainen, Ruusuvirta, & Huotilainen, 2010). In studies of attention processing, pre-attentive processing is a fast, parallel, and effortless processing that is not under the direct control of participants' intentions or motivations (Näätänen, 1990). In the experimental setting, MMN is elicited with an auditory oddball paradigm in which rare sounds (deviant stimuli) occasionally occur in a sequence of repetitive sounds (standard stimuli). Brain response is measured by subtracting the ERPs that are elicited in response to the standard stimulus from those elicited in response to the deviant stimulus. In adults, the MMN typically peaks between 100 and 200 ms after the onset of deviation in frontal–central electrodes. MMN can be elicited even when participants do not attend to the stimuli, such as while they are reading a book or watching a silent movie. The contrast between standard and deviant sounds is manipulated by modifying a variety of acoustic features, such as pitch (Moreau, Jolicoeur, & Peretz, 2013), duration (Chobert, Francois, Habib, & Besson, 2012), intensity (Schröger, 1996), etc. Across these studies, a robust effect of the size of deviance has been demonstrated in MMN. That is, as the discriminability between the standard and the deviant stimuli increases, MMN amplitudes increase, and the MMN response sometimes occurs earlier (decreased peak latency) (e.g., Hsu, Lin, Hsu, & Lee, 2014). Thus, it is to be expected that acoustic changes in a salient feature would generally yield a large MMN response.

While MMN reflects pre-attentive mechanisms, the amplitude of P300 response, a positive wave and peaks at 300 ms or later after onset of a stimulus, is proportional to the amount of attentional resources engaged in processing stimulus (Duncan et al., 2009). In general, P300 is traditionally assessed using an “oddball paradigm”, in which a rare stimulus elicits P300 responses (Donchin, 1981). Specifically, under exposure to novel stimuli, P300-like waveforms tend to have a frontal distribution. Studies have showed that the frontal P300, which is called the P300a component, reflects an orientation of attention to novel or salient stimuli. For example, in a dichotic listening study, self-relevant cues presented to the ignored ear also elicited P300 responses, while irrelevant stimulus presented to the ignored ear did not (Bargh, 1982).

In addition to measurements of ERP components, time–frequency responses of EEG data also indicate mechanisms in both sensory-based and endogenous processing of auditory stimuli. For example, studies have found that the pre-attentive mechanism supporting MMN responses is associated with theta band activity (5–7 Hz) (Bishop, Hardiman, & Barry, 2011; Liu, Shi, Zhao, & Yang, 2008). On the other hand, suppressed activity in the lower beta band (13–20 Hz) was related to increased attentional processes (Engel & Fries, 2010; Minami, Noritake, & Nakauchi, 2014; Shahin, Picton, & Miller, 2009).

Given that MMN shows an effect of deviance size, a separate question regards deviance direction; that is, whether subjects are more sensitive to raising or lowering of pitch. In fact, few studies investigate deviance direction. Peter, McArthur, and Thompson (2010) used a fixed range (100 Hz) to create a contrast of lowered pitch (standard sound: 350 Hz; deviant sound: 250 Hz) versus raised pitch (standard sound: 250 Hz; deviant sound: 350 Hz). The results showed that pitch rises elicited a larger MMN activity than did pitch lowering. Although the results seem to suggest that the mechanism of pre-attentive change detection might account for sensitivity to rises, their results were confounded with the proportional size of deviance. Pitch discrimination performance has been found to be scaled to the proportional difference between pairs of pure tones (De Looze & Hirst, 2010; Stevens, Volkman, & Newman, 1937). Similarly, in the findings of Peter et al., (2010), the percentage between the pitch difference and the pitch of the standard stimuli was 40% (100 Hz: 250 Hz) in the rising condition, but only 28% (100 Hz: 350 Hz) in the lowering condition. The larger percentage contrast elicited a larger MMN response, which was predicted by deviance size.

With the ratio of pitch changes controlled, three studies found no significant difference between MMN responses to pitch raising and pitch lowering (Jacobsen & Schröger, 2001; Maess, Jacobsen, Schröger, & Friedericia, 2007; Ruusuvirta & Astikainen, 2012). However, their results were not sufficient to evaluate the effect of the direction of deviance in speech for at least three reasons. First, the pitch of the standard stimuli were either at the top limit (400 Hz and 550 Hz), or above the top limit (1000 Hz) of the range of human *F0* (Honorof & Whalen, 2005). Thus, it is not possible to apply their findings to responses to spoken *F0*. Furthermore, lower frequencies have been found to elicit a direction-based response (Kishon-Rabin et al., 2004). A second limitation in applying the findings of the abovementioned ERP studies to speech perception is the harmonic structure of the stimuli. These studies used pure tones, which lack the spectral and experiential characteristics of the human voice. If sensitivity to detection of pitch change direction is related to the language experience, it is possible that an effect of direction of change could be found while hearing human voice stimuli, but not while hearing pure tones.

Finally, previous studies of deviance direction only reported MMN response results. Since MMN responses only reflect one aspect of auditory information processing, the nature of processing F_0 raising might be associated with other EEG measurements. If sensitivity to pitch rises is related to the pre-attentive mechanism of change detection, then rises should elicit a larger MMN response than lowering does. Another possible account for sensitivity to rises might be the mechanism of overt processing of the stimulus that follows the time window of MMN (Collard, Corley, MacGregor, & Donaldson, 2008; Macdonald & Campbell, 2011). That is, attention recourse might be captured after deviant stimuli are detected. Thus, in addition to measuring the MMN response, the present study also measured other brain responses, such as P300 and time–frequency responses. In our study, powers of theta and beta band were extracted by a trial-by-trial based wavelet analysis, and theta and beta band activities were used to evaluate effects of deviance direction.

The hypothesis of the present study is that sudden rises in the fundamental frequency of the human voice arouse listeners' attention to a greater extent than sudden lowering of human voice pitch does, and that the change in attention status is detectable in listeners' brain responses. Furthermore, it is also expected that responses to human F_0 rises will be different from responses to similar changes in pure tones. Linguistic justification for this hypothesis comes from the distribution and function of pitch elevations observed cross-linguistically.

2. The experiment

2.1. Participants

Twelve Mandarin-speaking adults (7 males) aged 18–30 years (mean: 22 years) participated in this study. All participants were right-handed with no history of neurological or psychological disorders. The current study was approved by the Human Subject Research Ethics Committee/Institutional Review Board of Academia Sinica, Taiwan. Written consent forms were obtained from all participants. All participants were involved in four MMN paradigm sessions (described below).

2.2. Stimuli

Four sets of auditory stimuli were designed and used separately in four experimental sessions. Speech stimuli were employed in two sessions, and pure tones of sinusoid sounds were employed in another two sessions. Speech stimuli were created by asking a female native Mandarin speaker to produce [a] in the level tone. The speaker was asked to begin at a comfortable habitual pitch and then keep increasing fundamental frequency until modal phonation could no longer be sustained. The speaker would begin again if she or the experimenter judged that the voice had taken on a sung quality. F_0 and formants were estimated with Praat (Boersma & Weenink, 2014). The speaker's production of [a] began with F_0 at 244 Hz (F_1 : 1187 Hz; F_2 : 1694 Hz; F_3 : 3362 Hz) and she reached her upper limit at 503 Hz (F_1 : 1235 Hz; F_2 : 1879 Hz; F_3 : 4109 Hz). The 503 Hz voice was further modified in Praat by changing F_0 to 600 Hz, using standard Praat synthesis settings. The purpose of raising the spoken pitch to 600 Hz was to preserve auditory characteristics of the human voice (keeping F_0 well below F_1), but to present a sound that is too high to be of true human origin. In spite of synthetically raising F_0 , the extra-high human voice still sounded like [a]. In each experiment session, four types of deviant stimuli were created by increasing or decreasing F_0 in ratios relative to the F_0 of the standard stimuli. The four types of deviant stimuli consisted of raising or lowering F_0 by 5% or 15%. In trials where the standard pitch was 244 Hz, changes of 5% and 15% correspond to 12.2 Hz and 36.6 Hz. At 600 Hz, the deviant sounds differed from the standard by 30 and 90 Hz. All stimuli were normalized to 70 dB SPL (A-weighting curve) and 180 ms in Praat. Although the results of Fletcher and Munson (1933) might suggest changing the intensity to preserve constant loudness at the two different pitches, human voice sensitivity may not follow the same pattern as pure tone sensitivity; thus the dB level was kept constant. At 244 Hz, shifts of 5% and 15% yield changes of 12.2 Hz and 36.6 Hz. These values were chosen because they correspond to shifts that are both easily produced and easily perceived, hence could be used to convey linguistic content.

2.3. Procedure

The stimuli were presented over two loudspeakers at 70 dB SPL. Four experimental sessions, including human voices with normal F_0 (244 Hz), human voices with extra-high F_0 (600 Hz), and pure tones at the same two frequencies, were presented in four separate experimental sessions for each participant. The orders of sessions were counterbalanced across participants. Each experimental session began with 20 standard trials, followed by a thousand trials comprised of 40% deviant (10% for each type of deviant) and 60% standard. The order of stimuli was pseudo-randomized with at least two successive standards between deviants. In each trial, the stimuli lasted 180 ms with a 500 ms inter-trial interval. Participants were seated in front of a monitor at a distance of approximately 60 cm in an acoustically and electromagnetically shielded chamber. During EEG recording, they were viewing movies without sound or subtitles.

2.4. Data recording, preprocessing, and analysis

The EEG was recorded from 64 Ag/AgCl electrodes (QuickCap, Neuromedical Supplies, Sterling, USA). The electrodes were online referenced to a common vertex reference and were re-referenced to the average of the left and right mastoids for off-line

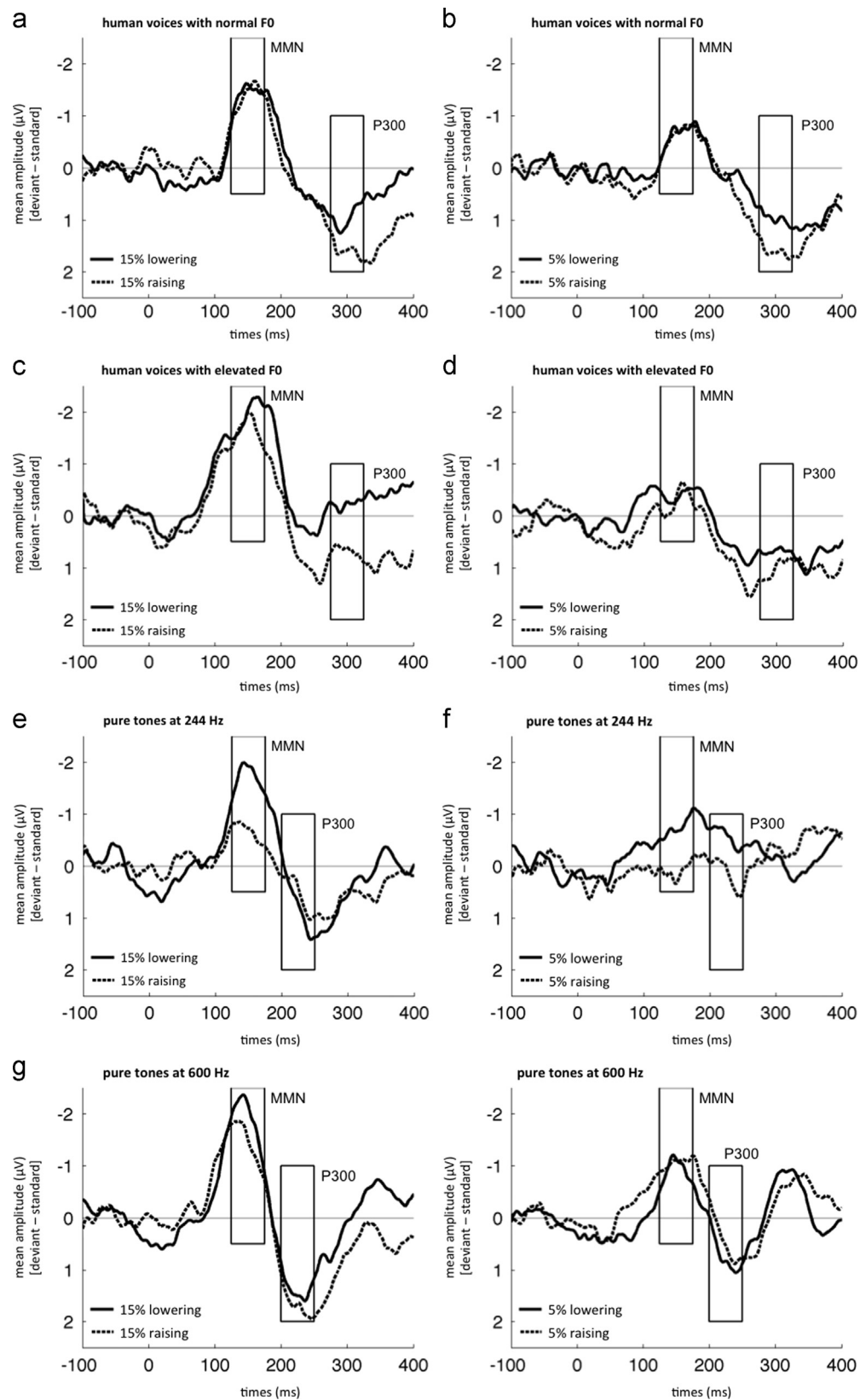


Fig. 1. Grand average of difference waves obtained in FZ electrode. Each wave was estimated by subtracting a wave of standard from that of each deviant. These plots show difference waves of decrement deviants (solid lines) and increment deviants (dash lines) separated by experimental sessions and size of frequency changes.

analysis. The EEG was continuously recorded and digitized at a rate of 1000 Hz and the signal was amplified by SynAmps2[®] (Neuroscan, Inc.) amplifiers with the band-pass at DC–200 Hz. Electrode impedances were kept below 5 k Ω . Eye movements and blinks were monitored with supra- and infra-orbital electrodes and with electrodes in the external canthi.

For offline analysis, analysis period was 800 ms including 100 ms pre-stimulus intervals and 700 ms of post-stimulus intervals. The pre-stimulus interval was used for baseline correction. To control for effects of number of repeated standard trials (Haenschel, Vernon, Dwivedi, Gruzelić, & Baldeweg, 2005), only standard trials that were followed by at least two standards were analyzed. Before artifact rejection, the data were band-pass filtered at 0.1–30 Hz (zero-phase shift mode). Two stages of artifact removal were performed. The first stage was eye-movement rejection in which trials with voltage variations larger than $\pm 60 \mu\text{V}$ in either vertical or horizontal electrooculography were rejected. In the second stage, trials with voltage variations larger than $\pm 60 \mu\text{V}$ in at least one of the rest of the channels were rejected. After artifact rejection, the number of trials for calculating averaged waveforms ranged from 197 to 100 (mean: 152) for standards and ranged from 99 to 38 (mean: 76) for deviants.

2.4.1. Analytic procedures for measuring MMN and P300

The grand averaged ERPs for each stimulus in each experiment session were computed for each participant and each electrode. To evaluate the MMN and P300 responses, difference waves were calculated by subtracting the ERPs of standard stimuli from those of deviants. Time windows of interest were identified around corresponding peaks showing in the difference waves averaged from all participants. Fig. 1 shows difference waves at the electrode FZ. In each experiment session, there were MMN responses over the frontal scalp peaking at around 150 ms, which was consistent with the time windows and distribution demonstrated in previous studies (Cheng et al., 2013; Lee et al., 2012). Therefore, for MMN responses, mean amplitudes of each stimulus type were measured in the 125–175 ms interval. After the MMN peaks, there were positive waves peaking at around 225 ms for pure tone sessions and at 300 ms for human voice sessions. Accordingly, P300 responses of human voice sessions were estimated by computing the mean amplitude during 275–325 ms, and those of pure tone sessions were estimated by computing the mean amplitude during 200–250 ms.

Based on previous findings (Cheng et al., 2013; Lee et al., 2012), mean amplitudes of MMN and P300 were analyzed at nine electrodes, F3, FZ, F4, FC3, FCZ, FC4, C3, CZ, and C4. The mean amplitudes were submitted to repeated-measure analysis of variance (RM-ANOVA) separated by experiment sessions. Each ANOVA analysis had two within-subject factors, including stimuli (standard, 5% raised deviant, 15% raised deviant, 5% lowered deviant, and 15% lowered deviant) and electrodes. To determine whether each type of deviant elicited MMN or P300 responses, four paired-comparisons, between the standard and each of the deviants, were performed using Dunn–Šidák method for multiple comparisons.

2.4.2. Analysis procedures for time–frequency responses

To further explore whether pre-attentive mechanisms support sensitivity to rises in pitch, time–frequency (TF) analyses were performed with the Fieldtrip software package, a Matlab-based toolbox dedicated to the analysis of electromagnetic and electrophysiological data (www.ru.nl/fcdonders/fieldtrip/). TF representations of the single trial data were obtained by computing the squared norm of the result of the convolution of complex Morlet wavelets with the EEG data. We used wavelets with a 7-cycle width, with frequencies ranging from 1 to 30 Hz, in 1 Hz steps. TF representations were then averaged over trials for each stimulus for each participant. The average power values thus obtained were expressed as a percentage change relative to the power in a baseline interval from 300 to 100 ms prior to stimuli onset. Based on previous studies of MMN and attention orientation (Bishop, 2007; Bishop et al., 2011; Engel & Fries, 2010), we focused on the TF responses in the theta band (from 5 to 7 Hz) and the beta band (from 18 to 22 Hz). To evaluate differences between the standard and each deviant across electrodes and time, a cluster-based random permutation analysis (Maris & Oostenveld, 2007) was conducted in the mean amplitudes of six successive epochs of 50 ms each, from 100 ms to 400 ms. This procedure allowed for identification of the temporal courses at which the mismatch effect appeared and could effectively handle the multiple-comparisons problem. First, a simple dependent-samples *t* test was performed at each electrode. All electrodes that exceeded a significance level ($\alpha = 0.05$) were identified and formed clusters. For each cluster, a cluster-level test statistic was calculated by taking the sum of all the individual *t* statistics within that cluster. Next, a null distribution was created by computing 1000 randomized cluster-level test statistics. Finally, the actually observed cluster-level test statistics were compared against the null distribution, and clusters falling in the highest or lowest 2.5th percentile were considered significant.

3. Results

3.1. MMN and P300

EEG is considered to exhibit a MMN response when there is significant negative activity during the relevant time period, cf. Section 2.4.1. P300 response is observed when there is a positive peak in the EEG difference curve at about 300 ms after stimulus onset. In the following analysis, only planned comparisons were examined. For clarity of exposition, in the following discussion, 5% and 15% *F0* changes are identified as small and large changes.

3.1.1. Human voice with normal *F0* (244 Hz)

There is a significant main effect of stimuli in MMN ($F(4, 44) = 3.822, p < .01$). Fig. 2a shows means and confidence intervals for MMN amplitudes, calculated by subtracting amplitudes of standards from that of deviants. Large deviants significantly elicited more negative activity relative to the standard during the interval of MMN ($ps < .05$), and no significant effect was found for small raises and small lowering ($ps > .2$) in the same time window. In the P300 time window, the main effect of stimuli was significant ($F(4, 44) = 3.357$,

$p < .05$). Fig. 2b shows that both sizes of raised deviants elicited significantly more positive activity than the standard did ($ps < .05$), while there was no significant difference between the standard and lowered deviants in P300 ($ps > .1$).

3.1.2. Human voice with elevated F_0 (600 Hz)

There was a significant main effect of stimuli in MMN ($F(4, 44) = 11.698$, $p < .001$). Specifically, both large deviants elicited significantly more negative activity relative to the standard during the MMN interval ($ps < .001$) (Fig. 2c). On the other hand, during the MMN interval, there was no significant difference between the standard and the small deviants ($ps > .3$). There was no significant P300 response to human voice with elevated F_0 ($F(4, 44) = 1.321$, $p = .277$) (Fig. 2d).

3.1.3. Pure tones at 244 Hz

There was a significant main effect of stimuli in MMN ($F(4, 44) = 9.348$, $p < .001$). Fig. 3a shows that large pitch lowering elicited significantly more negative activity than the standard did during the interval of MMN ($p < .001$), and the activities in this interval were marginally different between the large raised deviant and the standard ($p = .08$). There was no significant difference between standard and the smaller deviants ($ps > .1$). Activity in the interval of P300 did not show a significant effect of stimuli ($F(4, 44) = 1.624$, $p = .185$) (Fig. 3b).

3.1.4. Pure tones at 600 Hz

These stimuli showed significant main effects in MMN ($F(4, 44) = 6.018$, $p < .001$) and P300 ($F(4, 44) = 3.636$, $p < .05$). Both sizes of rise, along with large lowering significantly elicited more negative activity than did the standard during the interval of MMN ($ps < .05$) (Fig. 3c), while the activities in this interval were marginally different between small lowered deviant and the standard ($p = .08$). Regarding activities in the P300 time window, both large deviants elicited significantly more positive activities than did the standard ($ps < .05$) (Fig. 3d), but there were no significant differences between small deviants and the standard in P300 ($ps > .2$).

In summary, the results indicated that large deviants yielded significant MMN responses across experimental sessions regardless of the direction of change, which was expected, based on prior studies. In addition to MMN, the P300 response also showed the effect of deviance size in response to pure tones at high frequency. Regarding the effect of deviant direction, the P300 response was elicited for raised deviants regardless of the size of deviant when participants heard a human voice at normal F_0 . Other experimental sessions did not reveal such sensitivity to raised deviants. That is, the P300 response showed sensitivity to F_0 raising when the stimuli matched human speech sounds.

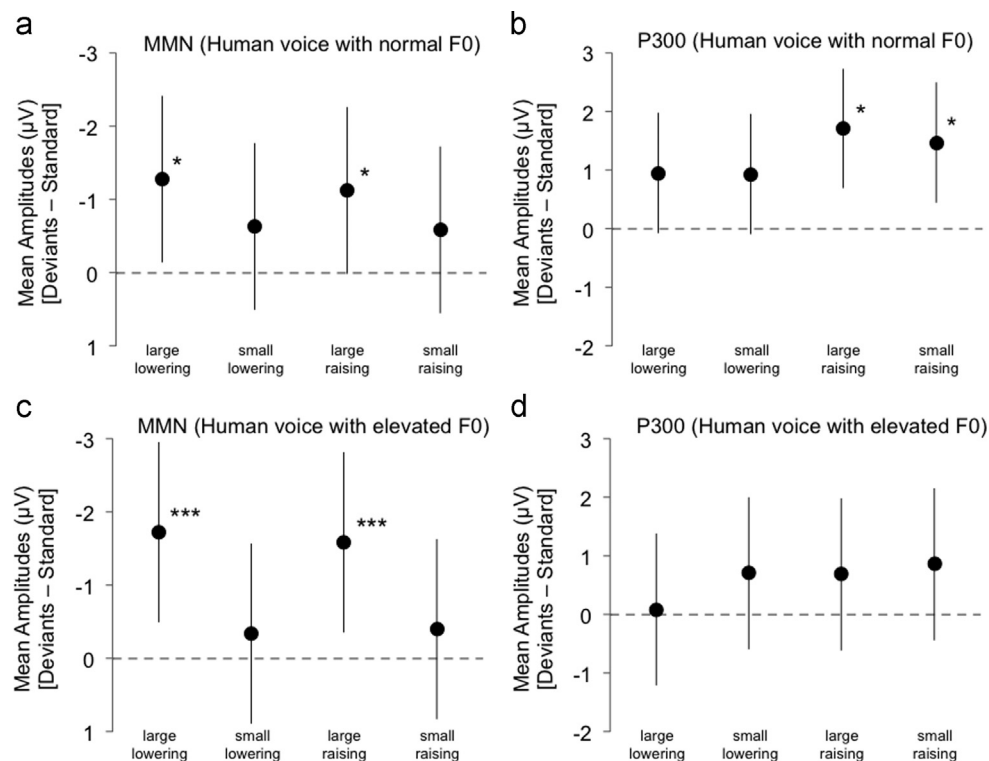


Fig. 2. Black dots represent means from the ANOVA analysis (see Section 2.4.1) by subtracting amplitudes of standards from that of deviants in the time window of MMN. Solid lines represent 95% confidence intervals estimated by the Dunn-Sidak method. $p < .05$ *; $p < .01$ **; $p < .001$ ***.

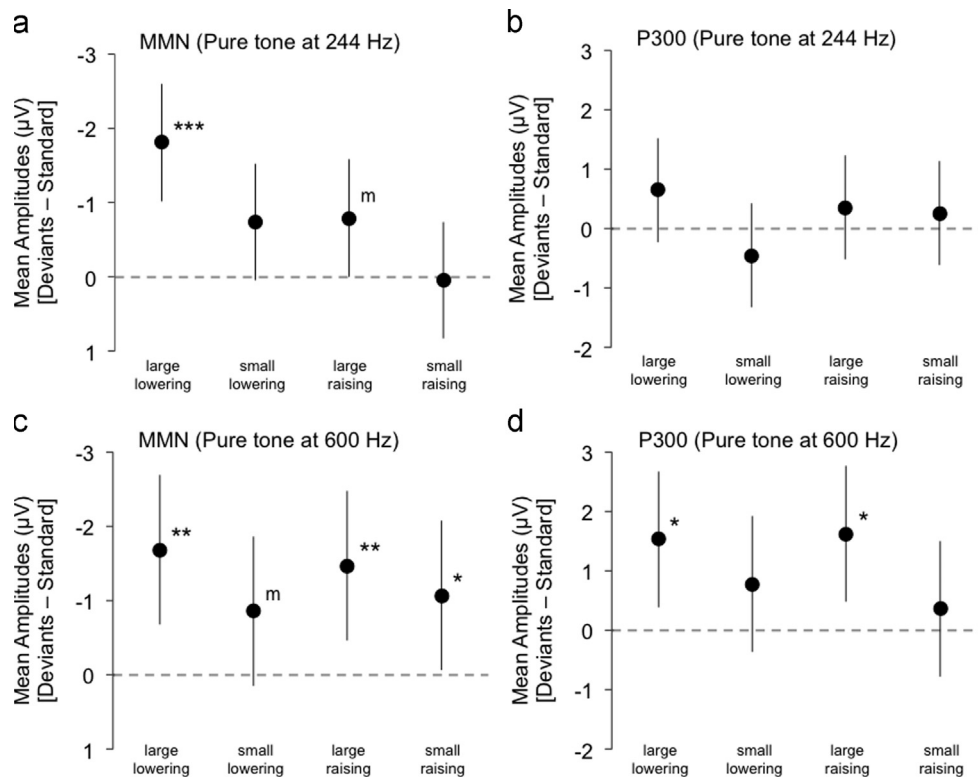


Fig. 3. Black dots represent means from the ANOVA analysis (see Section 2.4.1) by subtracting amplitudes of standards from that of deviants in the time window of P300. Solid lines represent 95% confidence intervals estimated by the Dunn-Sidak method. $p < .1$ ^m; $p < .05$ ^{*}; $p < .01$ ^{**}; $p < .001$ ^{***}.

3.1.5. Differential waveforms

An alternative approach for analyzing MMN responses is to estimate differences. Therefore, based on the time windows and electrodes described in Section 2.4.1, differential waveforms were obtained by subtracting the responses to standards from responses to deviant stimuli. The mean amplitudes were submitted to RM-ANOVA analyses separated by experiment sessions. Each ANOVA analysis had three within-subject factors, including deviant direction (raised deviants and lowered deviants), deviant size (large and small changes), and electrodes.

In the session with normal human voice, the MMN activity showed a significant effect of deviant size ($F(1, 11) = 5.78$, $p < .05$), where large deviants elicited larger MMN than small deviants did. Neither the main effect of deviant direction ($F(1, 11) = .19$, $p = .67$) nor the interaction of deviant size and deviant direction ($F(1, 11) = .02$, $p = .88$) were significant in MMN. In P300 activity, there was a significant main effect of deviant directions ($F(1, 11) = 7.56$, $p < .05$), which showed that raised deviants elicited larger P300 than lowered deviants did. On the other hand, neither the main effect of deviant size ($F(1, 11) = .08$, $p = .77$) nor the interaction of deviant size and deviant direction ($F(1, 11) = .07$, $p = .79$) were significant in P300.

In the session with human voice with elevated F_0 , the MMN activity showed a significant effect of deviant sizes ($F(1, 11) = 22.14$, $p < .001$), and large deviants elicited larger MMN than small deviants did. Neither the main effect of deviant direction ($F(1, 11) = .03$, $p = .87$) nor the interaction of deviant size and deviant direction ($F(1, 11) = .15$, $p = .71$) were significant in MMN. In P300 activity, the main effects of deviant size ($F(1, 11) = 1.24$, $p = .29$) and deviant direction ($F(1, 11) = .93$, $p < .36$) were not significant. The interaction of deviant size and deviant direction were also not significant in P300 ($F(1, 11) = .43$, $p = .52$).

In the session with pure tones at 244 Hz, the MMN activity showed a significant effect of deviant sizes ($F(1, 11) = 3.15$, $p < .01$), and large deviants elicited larger MMN than small deviants did. Deviant directions also affected MMN activity ($F(1, 11) = 8.1$, $p < .05$), and lowered deviants elicited larger MMN than raised deviants did. The interaction of deviant size and deviant direction ($F(1, 11) = .32$, $p = .58$) were not significant in MMN. In P300 activity, there was a significant main effect of deviant sizes ($F(1, 11) = 10.72$, $p < .01$), which showed that large deviants elicited larger P300 than small deviants did. On the other hand, neither the main effect of deviant direction ($F(1, 11) = .26$, $p = .61$) nor the interaction of deviant size and deviant direction ($F(1, 11) = 1.49$, $p = .25$) were significant in P300.

Finally, in the session with pure tones at 600 Hz, the MMN activity showed a marginally significant effect of deviant sizes ($F(1, 11) = 4.01$, $p = .07$), and large deviants elicited larger MMN than small deviants did. Neither the main effect of deviant direction ($F(1, 11) = .001$, $p = .99$) nor the interaction of deviant size and deviant direction ($F(1, 11) = 1.75$, $p = .21$) were significant in MMN. In P300 activity, there was a significant main effect of deviant sizes ($F(1, 11) = 5.02$, $p < .05$), which showed that large deviants elicited larger P300 than small deviants did. On the other hand, neither the main effect of deviant direction ($F(1, 11) = .19$, $p = .67$) nor the interaction of deviant size and deviant direction ($F(1, 11) = .62$, $p = .45$) were significant in P300.

3.2. Time–frequency analysis

In order to verify brain behaviors related to the hearing of raised and lowered deviants in spoken *F0*, a time–frequency analysis was performed that focused on the EEG data in response to normal *F0* human speech sounds. Figs. 4 and 5 are topographical plots of the differential activities of theta band activity, and beta band activity in six time windows. Dots show the position of electrodes viewing from the top of the head. The differential activities were calculated by subtracting the responses to the standard stimuli from those to the deviant stimuli. Red regions indicate larger time–frequency activation/response in response to deviants than to standard stimuli. Blue regions showed less time–frequency activation/response in response to deviants than to standard stimuli. White dots indicate clusters of electrodes with significant differences described below.

3.2.1. Theta band (5–7 Hz)

Cluster-based random permutation analysis showed that large lowered deviants yielded theta activities that were significantly larger than the standard in the frontal scalp area during three time periods: 100–150 ms, 150–200 ms, and 200–250 ms ($p < .01$) (Fig. 4). However, hearing large raised deviants yielded smaller theta activities than hearing the standard did in the central–posterior scalp during the time periods 150–200 ms, 200–250 ms, 250–300 ms, 300–350 ms, and 350–400 ms ($p < .05$). The contrast between small lowered deviant and standard yielded a marginally significant cluster during 200–250 ms ($p = .049$). Finally, there was no significant difference between small raised deviants and the standard.

3.2.2. Beta band (18–22 Hz)

The cluster-based random permutation analysis showed that large raised deviants yielded significantly smaller beta activities than the standard did in the right posterior scalp during 250–300 ms ($p < .05$), 300–350 ms ($p < .05$), and 350–400 ms ($p < .01$) (Fig. 5). Hearing small raised deviants also yielded smaller beta activity than hearing the standard did in the right posterior scalp during 200–250 ms ($p < .05$). On the other hand, neither large nor small lowering stimulated any significant difference in the beta power relative to the standard stimuli.

4. Discussion

The present study tested whether human subjects were sensitive to the direction of sudden *F0* changes, as indicated by MMN, P300, and/or time–frequency responses. These measurements were chosen because they aid in exploring the mechanisms that would support differences in sensitivity to *F0* rises and falls. Namely, MMN is taken as an indicator of the pre-attentive mechanism of detecting acoustic changes, P300 has been shown to reflect overt evaluation of sensory inputs, and time–frequency response results can help evaluate the brain behaviors associated with mismatch responses. To reconcile contradicting results in previous ERP studies, we employed experimental sessions with different types of sounds to which we applied the same deviant sound design. By using different types of sounds, this study could evaluate whether subjects have sensitivity to stepwise rises in pitch, and whether pitch change sensitivities differ according to high/normal *F0*, and/or according to pure tones versus human voice. The study employed shifts of 5% and 15% in order to model small and large shifts in *F0*. These sizes of change are easily produced and perceived. In the following discussion we summarize the findings of each type of response; finally, we consider some possible linguistic implications of the findings.

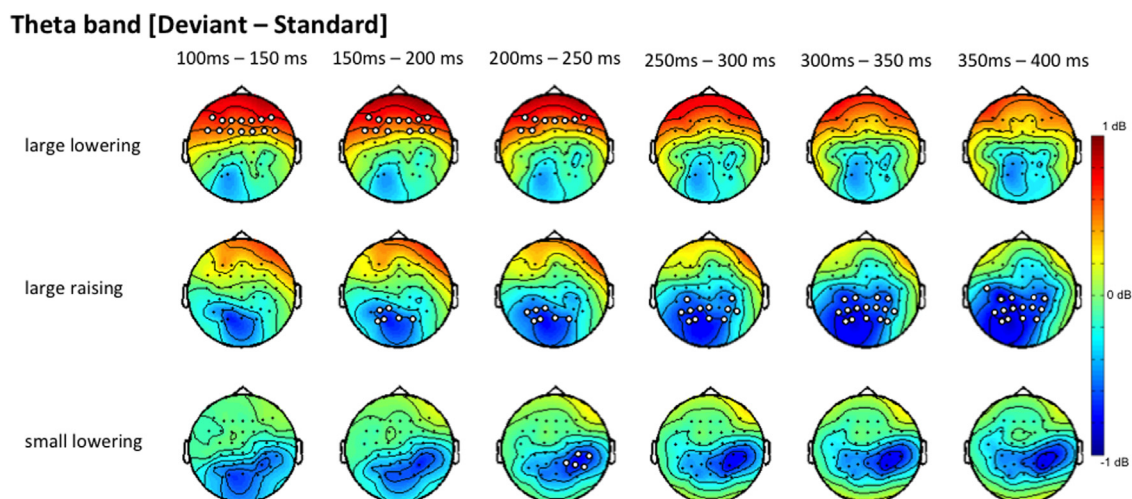


Fig. 4. Topographical plots of differential activities of theta band (5–7 Hz) power in six time windows. White dots indicate clusters of electrodes with significant differences between standard and deviants. (For interpretation of references to color in this figure, the reader is referred to the web version of this article.)

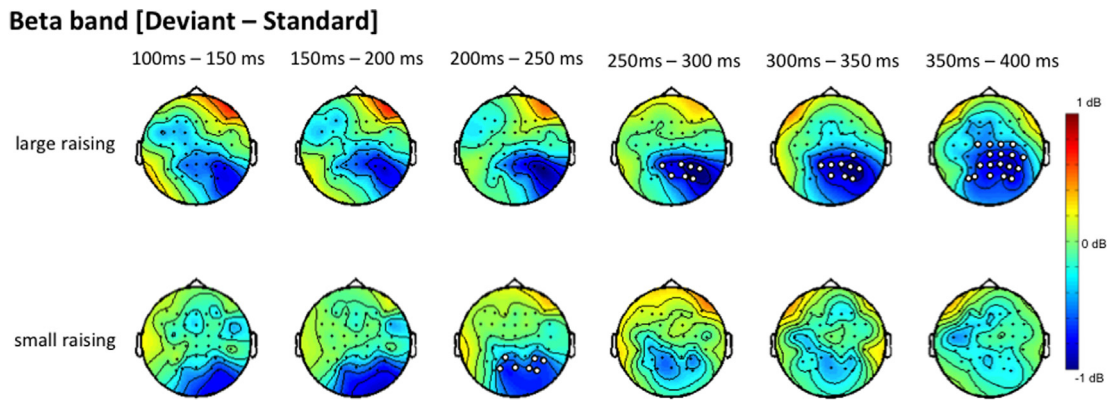


Fig. 5. Topological plots of differential activities of beta band (18–22) power in six time windows. White dots indicate clusters of electrodes with significant differences between standard and deviants. (For interpretation of references to color in this figure, the reader is referred to the web version of this article.)

4.1. MMN discussion

An MMN response is an indication of the pre-attentive mechanism of detecting acoustic changes. Consistent with the ERP literature, the results demonstrated robust effects of size of deviance in MMN across experiment sessions (Näätänen & Alho, 1995). That is, for all four types of stimuli, large changes in F_0 significantly elicited MMN responses. Regarding the manipulation of deviant directions, both raised and lowered deviants could elicit the MMN response, which seems to indicate that MMN amplitude is not sensitive to the direction of pitch changes. Such findings are consistent with the results in previous studies (Jacobsen & Schröger, 2001; Maess et al., 2007). These studies also argued that using high range of pitch (from 500 Hz to 1000 Hz) might underestimate the effect of pitch changes on MMN responses. However, the present study used stimuli of both extra high (600 Hz) and normal (244 Hz) speaker F_0 , and found MMN responses in both ranges.

Although Peter et al. (2010) demonstrated that raised pitch deviants elicited a larger MMN than lowered deviants did, their design of deviants was confounded with the size of deviance, which is a robust predictor of MMN responses. That is, the size of raising was 40% of the standard pitch, and the size of lowering was 28% of the reference pitch. Thus, the consistent finding seems to be that the pre-attentive mechanism of detecting acoustic changes does not appear to be responsible for the sensitivity of detecting elevated pitch in human speech. Rather, pre-attentiveness seems to be more responsive to degrees of acoustic change than to direction of change.

4.2. P300 discussion.

The amplitude of P300 is related to the evaluation of stimuli at the conscious level. For example, in studies of attention, P300 was elicited while participants succeeded to report or to overtly respond to targets embedded in streams of distractors in a rapid serial presentation (Sergent, Baillet, & Dehaene, 2005; Vogel, Luck, & Shapiro, 1998), or to targets in a dual-task paradigm (Kramer, Wickens, & Donchin, 1983). In the present study, the P300 responses to pure tones at high pitch showed an effect of deviant size. These results support the assumption that encountering salient acoustic changes triggers the orientation of attention. The normal human voice stimuli, however, produced a different P300 response than the other stimuli did. While hearing human voices at normal F_0 , the P300 indicated sensitivity to F_0 rises. Specifically, both large and small rises elicited a large P300 response, but their lowered counterparts did not significantly elicit a P300 response. Sensitivity to pitch change direction was not found in response to pure tones nor to artificially high human voice F_0 . These results suggest that hearing human voice elevations, especially in the normal speaking F_0 range, triggers the overt processing of stimuli evaluation, while hearing F_0 lowering does not. For pure tones, large deviants, regardless of the deviant direction, elicited P300 responses. This finding is consistent with results in previous studies (Jacobsen & Schröger, 2001; Maess et al., 2007; Ruusuvirta & Astikainen, 2012), which suggested that attention processing of pure tones might not be sensitive to deviant direction.

4.3. EEG time–frequency responses discussion.

The present results in time–frequency responses also indicate that overt attention to raised deviants was related to sensitivity to elevations in pitch of a normal human voice. The theta band activity generally reflected differences between large deviants and the standard regardless of the direction. This finding was consistent with the assumption that theta activity is related to the pre-attentive mechanism in MMN responses (Bishop et al., 2011). On the other hand, beta activity showed a significant difference between raised deviants and the standard. This finding was related to results in previous studies that indicated that beta-band activity is the signature of a state of attention (Engel & Fries, 2010). That is, beta-band oscillation decreases when the current cognitive set is disrupted by incoming sensory information, an event which triggers attention. For example, Shahin et al. (2009) found decreased beta activity in tasks requiring memory demands. Loo et al. (2009) also found that adults with attention-deficit hyperactivity disorder have abnormally

low beta activity. These previous studies suggest that our finding of decreased beta activity in hearing raised *F0* deviants reflects overt attention to human voice elevation.

Fig. 5 shows the beta-power decrease around the right parietal-posterior scalp. It has been demonstrated that the right hemisphere, particularly the right parietal cortex, dominates the direction of attention to sensory information (Spotorno & Faure, 2011; Tomer, 2008). Minami et al. (2014) also suggested that decreased beta activity in the parietal-posterior scalp was related to transition of cognitive state. Therefore, the present finding of beta-power suggests that sensitivity to rises in human voice *F0* is related to the mechanism of attention modulation.

4.4. Conclusions and possible linguistic implications

The present findings in P300 and beta band power indicate that changes in human speech *F0* are processed differently than the same degrees of change in non-speech signals. The difference consists of two parts. First, human speech sound is attended to differently from pure tones. Second, sudden rises in spoken *F0* activate the brain differently than do equivalent drops. Specifically, perception of rises in spoken *F0* activates attention orienting mechanisms.

Nevertheless, present results are not sufficient to determine whether the sensitivity is due to the fact that the sound is of a human voice, or whether it is due to the fact that the sound has harmonic structure. Sensitivity to *F0* raising is not seen in the extra-high human voice, although it has the same complex structure as the human voice at 244 Hz. This difference suggests that the brain treats raising in spoken *F0* differently from lowering, and also from *F0* changes in other signals, even when the spoken sound carries no linguistic content. This hypothesis awaits confirmation by using musical instruments or other complex sounds with the same *F0* as human speech stimuli.

Not surprisingly, sudden pitch rises in spoken *F0* signal a diverse set of linguistic and paralinguistic content. Across many of the world's languages, beginnings of new discourse units, stressed syllables, emotionally laden speech, information-seeking questions, and contrastive focus are all signaled (at least in part) by sudden rises in spoken *F0*. The present study suggests that a set of brain responses to spoken *F0* rises serve to orient the speaker-listener's attention to the associated linguistic content.

It is impossible to say for certain why these brain activities occur when raised spoken *F0* is perceived. Nevertheless, it is possible to be aware of ways that speakers and listeners utilize this response. Listeners, who in most cases are also speakers, know the effort that is needed to raise their own *F0*. This awareness could be a reason that listeners direct attention to perceived speech that contains raised *F0*. Likewise, speakers with intact hearing have experienced their own attention orienting mechanism being triggered by rises in spoken *F0*. Thus, they could choose (at some level of awareness) to initiate *F0* rises to either acquire or maintain the attention of their interlocutors.

Acknowledgments

This work was finished under the support from the thematic research program of Academia Sinica, Taiwan (AS-102-TP-C06, Lee), and Ministry of Science and Technology (Taiwan) grant #100-2628-H-001-008-MY4 (Evans). The authors would like to thank Ching-Fang Wu for help with testing participants and data analyses.

References

- Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, 43.3, 425–436.
- Bishop, D. V. M. (2007). Using mismatch negativity to study central auditory processing in developmental language and literacy impairments: Where are we, and where should we be going?. *Psychological Bulletin*, 133.4, 651–672.
- Bishop, D. V. M., Hardiman, M. J., & Barry, J. G. (2011). Is auditory discrimination mature by middle childhood? A study using time–frequency analysis of mismatch responses from 7 years to adulthood. *Developmental Science*, 14.2, 402–416.
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer.
- Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford University Press.
- Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford University Press.
- Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of intonation*. University Park Press.
- Cheng, Y.-Y., Wu, H.-C., Tzeng, Y.-L., Yang, M.-T., Zhao, L.-L., & Lee, C.-Y. (2013). The development of mismatch responses to mandarin lexical tones in early infancy. *Developmental Neuropsychology*, 38.5, 281–300.
- Chobert, J., Francois, C., Habib, M., & Besson, M. (2012). Deficit in the preattentive processing of syllabic duration and VOT in children with dyslexia. *Neuropsychologia*, 50.8, 2044–2055.
- Collard, P., Corley, M., MacGregor, L. J., & Donaldson, D. I. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34.3, 696–702.
- De Looze, C., & Hirst, D. (2010). L'échelle OME (Octave-Médiane): une échelle naturelle pour la mélodie de la parole. In *Proceedings of the XXIIIème Journées d'Étude sur la Parole (JEP 2010)*. Mons [paper 107].
- Donchin, E. (1981). Surprise!... surprise?. *Psychophysiology*, 18.5, 493–513.
- Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., et al. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, 120.11, 1883–1908.
- Engel, A. K., & Fries, P. (2010). Beta-band oscillations-signalling the status quo?. *Current Opinion in Neurobiology*, 20.2, 156–165.
- Fernald, A., & Mazzei, C. (1991). Prosody and focus in speech to infants and adults. *Developmental psychology*, 27(2), 209.
- Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 12(4), 377–430.
- Fujisaki, H., Ohno, S., & Wang, C. (1998). A command-response model for *F0* contour generation in multilingual speech synthesis. In *Proceedings of the third ESCA/COCOSDA workshop (ETRW) on speech synthesis*.
- Giard, M. H., Lavikainen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & Näätänen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: an event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience*, 7(2), 133–143.
- Goldsmith, J. (1984). Meeussen's rule. In M. Aronoff, & R. T. Oehre (Eds.), *Language Sound Structure* (pp. 245–259). Cambridge, MA: The MIT Press.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.

- Haenschel, C., Vernon, D. J., Dwivedi, P., Gruzelier, J. H., & Baldeweg, T. (2005). Event-related brain potential correlates of human auditory sensory memory-trace formation. *Journal of Neuroscience*, 25.45, 10494–10501.
- Honorof, D. N., & Whalen, D. H. (2005). Perception of pitch location within a speaker's F0 range. *Journal of the Acoustical Society of America*, 117.4, 2193–2200.
- Hsu, C.-h., Lin, S.-k., Hsu, Y.-y., & Lee, C.-y. (2014). The neural generators of the mismatch responses to Mandarin lexical tones: An MEG study. *Brain Research*, 1582, 154–166.
- Hyman, L. M. (2001). Privative tone in Bantu. In Shigeki Kaji (Ed.), *Cross-linguistic studies of tonal phenomena* (pp. 237–257). Tokyo: Institute for the Study of Languages and Cultures.
- Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch?. *Psychophysiology*, 38.4, 723–727.
- Kishon-Rabin, L., Roth, D. A.-E., Dijk, B. V., Yinon, T., & Amir, O. (2004). Frequency discrimination thresholds: The effect of increment versus decrement detection of frequency. *Journal of Basic and Clinical Physiology and Pharmacology*, 15.1–2, 29–40.
- Kochanski, G., Shih, C.-L., & Jing, H.-Y. (2003). Quantitative measurement of prosodic strength in Mandarin. *Speech Communication*, 41(4), 625–645.
- Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication*, 34.4, 391–405.
- Kramer, A. F., Wickens, C. D., & Donchin, E. (1983). An analysis of the processing requirements of a complex perceptual-motor task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 25.6, 597–621.
- Lee, C.-y., Yen, H.-l., Yeh, P.-w., Lin, W.-h., Cheng, Y.-y., Tzeng, Y.-l., et al. (2012). Mismatch responses to lexical tone, initial consonant, and vowel in Mandarin-speaking preschoolers. *Neuropsychologia*, 50.14, 3228–3239.
- Liu, T., Shi, J., Zhao, D., & Yang, J. (2008). The event-related low-frequency activity of highly and average intelligent children. *High Ability Studies*, 19.2, 131–139.
- Loo, S. K., Hale, T., Sigi, Macion, J., Hanada, G., McGough, J. J., McCracken, J. T., et al. (2009). Cortical activity patterns in ADHD during arousal, activation and sustained attention. *Neuropsychologia*, 47.10, 2114–2119.
- Macdonald, M., & Campbell, K. (2011). Effects of a violation of an expected increase or decrease in intensity on detection of change within an auditory pattern. *Brain and Cognition*, 77.3, 438–445.
- Maess, B., Jacobsen, T., Schröger, E., & Friedericia, A. D. (2007). Localizing pre-attentive auditory memory-based comparison: Magnetic mismatch negativity to pitch change. *Neuroimage*, 37.2, 561–571.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164.1, 177–190.
- Minami, T., Noritake, Y., & Nakachi, S. (2014). Decreased beta-band activity is correlated with disambiguation of hidden figures. *Neuropsychologia*, 56, 9–16.
- Moreau, P., Policoeur, P., & Peretz, I. (2013). Pitch discrimination without awareness in congenital amusia: Evidence from event-related potentials. *Brain and Cognition*, 81.3, 337–344.
- Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral and Brain Sciences*, 13.2, 201–233.
- Näätänen, R. (1992). *Attention and brain function*. Hillsdale: Lawrence Erlbaum Associates.
- Näätänen, R., & Alho, K. (1995). Mismatch negativity—A unique measure of sensory processing in audition. *International Journal of Neuroscience*, 80.1–4, 317–337.
- Näätänen, R., Astikainen, P., Ruusuvirta, T., & Huottilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory–cognitive core of cognitive processes. *Brain Research Reviews*, 64.1, 123–136.
- Pell, M. D., Monetta, L., Paulmann, S., & Kotz, S. A. (2009). Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior*, 33.2, 107–120.
- Peter, V., McArthur, G., & Thompson, W. F. (2010). Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neuroscience Letters*, 482.1, 71–75.
- Robinson, D. W., & Dadson, R. S. (1956). A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, 7(5), 166.
- Ruusuvirta, T. T., & Astikainen, P. (2012). Mismatch negativity of higher amplitude for melodic ascendance than descendance. *Neuroreport*, 23.4, 220–223.
- Schröder, M. (2001). Emotional speech synthesis: A review. In *Proceedings of the 7th European conference on speech communication and technology: Vol. 1. Eurospeech 2001* (pp. 561–564). Aalborg, Denmark: International Speech Communication Association, September 3–7, 2001.
- Schröger, E. (1996). The influence of stimulus intensity and inter-stimulus interval on the detection of pitch and loudness changes. *Electroencephalography and Clinical Neurophysiology/ Evoked Potentials Section*, 100.6, 517–526.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience*, 8.10, 1391–1400.
- Shahin, A. J., Picton, T. W., & Miller, L. M. (2009). Brain oscillations during semantic evaluation of speech. *Brain and Cognition*, 70.3, 259–266.
- Shih, C.-L. (2000). *A declination model of Mandarin Chinese. Intonation* (pp. 243–268)Netherlands: Springer243–268.
- Spotorno, S., & Faure, S. (2011). The right hemisphere advantage in visual change detection depends on temporal factors. *Brain and Cognition*, 77.3, 365–371.
- Stevens, S. S., Volkman, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, 8.3, 185–190.
- Sun, J. T.-S. (1997). The typology of tone in Tibetan. *Chinese Languages and Linguistics IV: Typological studies of languages in China*, 485–521.
- Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, 101(1), 514–521.
- Tomer, R. (2008). Attentional bias as trait: Correlations with novelty seeking. *Neuropsychologia*, 46.7, 2064–2070.
- Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception and Performance*, 24.6, 1656–1674.
- Williams, C. E., & Stevens, K. N. (1972). Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America*, 52.4B, 1238–1250.
- Yip, M. (2002). *Tone*. Cambridge University Press.
- Yuan, J.-H. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, 130(6), 4063–4069.