

Published in final edited form as:

*Int J Psychophysiol.* 2012 January ; 83(1): 71–78. doi:10.1016/j.ijpsycho.2011.10.006.

## ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback

Oleg Korzyukov<sup>a,\*</sup>, Laura Karvelis<sup>a</sup>, Roozbeh Behroozmand<sup>a</sup>, and Charles R Larson<sup>a</sup>

<sup>a</sup>Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Dr., 2240 Evanston, IL 60208, USA. Tel.: +1-847-491-2428; FAX: +1-847-491-4975

### Abstract

Auditory sensory processing is an important element of the neural mechanisms controlling human vocalization. We evaluated which components of Event Related Potentials (ERP) elicited by the unexpected shift of fundamental frequency in a subject's own voice might correlate with his/her ability to process auditory information. A significant negative correlation between the latency of the N1 component of the ERP and the Montreal Battery of Evaluation of Amusia scores for Melodic organization was found. A possible functional role of neuronal activity underlying the N1 component in voice control mechanisms is discussed.

### Keywords

Voice control mechanisms; Voice perturbation paradigm; Event Related Potentials; Auditory processing; N1 response

## 1. Introduction

### 1.1 Voice perturbation paradigm

Speech production requires phonation for all vowels, most consonants and highly coordinated movements of respiratory, laryngeal and speech articulator muscles. This process must rely on efficient, semi-automatic neuromuscular mechanisms for coordinating vocalization with movement of the upper vocal tract articulators. Since the goal of voicing for speech and singing is the accurate production of the acoustical properties of the output sound, auditory sensory processing of self-vocalization seems to be critical for accurate moment-to-moment control of phonation. Brain mechanisms underlying these audio-vocal interactions can be empirically evaluated with the voice perturbation paradigm (Burnett et al., 1998; Kawahara, 1994). The essence of this paradigm is the elicitation of involuntary motor control commands to the vocal system as a compensatory response to unexpected shifts in pitch feedback in a subject's own voice during self-vocalization. It is reasonable to hypothesize that such compensatory reactions to unexpected changes in self-vocalization include several sequential (probably overlapping to some extent) neural processing stages.

© 2011 Elsevier B.V. All rights reserved.

\*Corresponding author: o-korzyukov@northwestern.edu.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

We hypothesize that early stages should include detection of deviation from the anticipated vocal output. The following phases should involve transfer of information about the detected deviation from auditory processing brain areas to neuronal networks of vocal motor control. One of the initial steps in an effort to elucidate the temporal and spatial brain organization underlying the audio-vocal interactions triggering compensatory responses should be the separation of brain activity associated with auditory processing from brain activity orchestrating compensatory vocal responses.

## 1.2 Aims and scope

Since behavioral vocal compensation occurs within the time range of tens of milliseconds, one of the most appropriate approaches for separation of auditory from motor functions of brain activity is the recording of Event Related Potentials (ERP) elicited by the unexpected shift of fundamental frequency (F0) in a subject's own voice. It is reasonable to suggest that some components of these ERPs should be associated with auditory processing and others might represent brain activity related to compensatory motor actions. In this context the purpose of the present study was to evaluate which components of ERPs elicited by the unexpected shift of F0 in a subject's own voice might correlate with his/her ability to process auditory information. We assume that the ability to process such complex auditory information as speech or music should involve (at least partially) the same neuronal networks that are used in processing auditory input during self-vocalization. Thus, even within the normal population, a subject's ability to process musical organization should correlate with individual variations in parameters of ERP components that reflect processing of auditory input during self-vocalization. One validated (Peretz et al., 2003) method for assessment of an individual's ability to process complex acoustic features of an environment such as music is the Montreal Battery of Evaluation of Amusia (MBEA). This tool is especially suitable for the present study since it was designed for evaluation of individual abilities to process melodic and temporal components of musical organization, and one of the most notable biological functions of voicing is modulation of melodic and rhythmic aspects of speech (termed "intonation"). Thus it is very likely that among the multiple components of auditory information processing cumulatively tested by the MBEA, the elements involved in auditory processing of self-vocalization would be probed.

Previous ERP studies using the voice perturbation paradigm (Behroozmand et al., 2009; Liu et al., 2010) showed that unexpected shifts in the F0 of a person's own voice feedback during vocalization elicited several ERP components. These components can be labeled as the P1-N1-P2 complex, reflecting the sequential order and polarity of components measured from the vertex. The most prominent amplitudes in this complex are the N1 (peak latency around 100ms) and P2 (peak latency within 200–300ms) components. Therefore in the present study, the N1 and P2 were chosen as the ERP components of primary interest. As a test of whether results of our analyses are specific to unexpected shifts of F0 in voice auditory feedback or can be explained by more general cognitive processing of auditory change detection, additional auditory ERP data acquisition was carried out. All subjects were presented with a conventional auditory oddball test that allowed measurement of the N1 elicited by the presentation of pure tone in the absence of vocalization.

## 2. Materials and Methods

### 2.1. Participants and procedure

14 speakers of American English (8 females and 6 males, mean age: 21.5 years std. dev.: 2.7) participated in the study. All subjects passed a bilateral pure-tone hearing screening test at 20 dB SPL (octave frequencies between 250–8000 Hz) and reported no history of neurological disorders. All study procedures including recruitment, data acquisition and

informed consent were approved by the Northwestern University institutional review board, and subjects were monetarily compensated for their participation.

Each subject visited the lab twice, one day for MBEA data acquisition and another for the ERP recordings. MBEA performance data were collected by having subjects listen to approximately 53 minutes of melodies, as per the MBEA protocol (Peretz et al., 2003). A total of 6 musical subtests were administered to subjects in order to measure three areas of musical perception: melodic and temporal perception and memory. Although the full battery was administered to each subject, only scores from the melodic and temporal subtests were used in the present study. Subjects were instructed to listen to each series of melodies, required by the MBEA protocol, and to identify whether two melodies sounded the same or different. For the duration of the MBEA, subjects were seated in a sound-treated booth and listened to the tasks through insert earphones. Before each subtest, subjects were coached how to respond and instructed to write their answers on the provided answer sheet.

During ERP data acquisition, subjects were asked to sustain the vowel sound/a/for approximately 2–3 seconds at their conversational pitch and loudness. Subjects were instructed to vocalize whenever they felt comfortable, i.e., without a cue. This vocal task was repeated 250 times as subjects took short breaks (2–3 seconds) between successive utterances. One upward shift in pitch of the subject's voice feedback (50 or 200 cents magnitude, randomized sequence, 50% probability of occurrence) was delivered during each vocalization within a 500–700 ms time window after vocal onset (Figure 1A). We use the value “cents” for the change in feedback and the subject's response, because this is a logarithmic value related to the 12-tone musical scale. By using this measure, it is easy to equate the same relative change in frequency between any two notes. The original measures in Hertz are converted to cents using the formula  $\text{Cents} = 1200 \times \text{Log}_2(F_2/F_1)$  in which  $F_1$  and  $F_2$  are the pre-stimulus and post-stimulus pitch frequencies. The rise time of the pitch shift was 10–15ms.

During the conventional auditory “novel” oddball ERP data acquisition test, a sequence of 1100 sounds (600 Hz pure tone sine wave, 100 ms duration, 10ms rise and fall times, inter-stimulus (onset to onset) interval 800 ms) was randomly replaced (10% of trials) by a tone of 150 ms duration or by a tone of 700 Hz (10% of trials, 100 ms duration) or by a novel environmental sound (10% of trials, 100 ms duration). The environmental sounds were drawn from a pool of 110 different environmental sounds, such as those produced by a hammer, rain, door, car horn, and so forth. Each novel sound was presented only once during the experiment.

During the ERP data acquisition, subjects were seated in a sound-attenuated room and watched a silent movie of their choice while vocalizing. Subjects' voices were picked up with an AKG boomset microphone (model C420), amplified with Mackie mixer (model 1202-VLZ3) and pitch-shifted through an Eventide Eclipse Harmonizer. The time delay from vocal onset, duration, direction, and magnitude of pitch shifts were controlled by MIDI software (Max/MSP v.5.0 Cycling 74). Voice and auditory feedback were sampled at 10 kHz using PowerLab A/D Converter (Model ML880, AD Instruments) and recorded on a laboratory computer utilizing Chart software (AD Instruments). Subjects maintained their conversational F0 levels with a voice loudness of about 70–75 dB, and the feedback (subject's voice) was delivered through Etymotic earphones (model ER1-14A) at about 80–85 dB. The 10 dB gain between voice and feedback channels (made by Crown amplifier D75) was used to partially mask air-borne and bone-conducted voice feedback. A Brüel & Kjær sound level meter (model 2250) along with a Brüel & Kjær prepolarized free-field microphone (model 4189) and a Zwislocki coupler were used to calibrate the gain between voice and feedback channels.

The electroencephalogram (EEG) signals were recorded from 64 sites on the subject's scalp using an Ag-AgCl electrode cap (EasyCap GmbH, Germany) in accordance with the extended international 10–20 system (Oostenveld and Praamstra, 2001) including left and right mastoids. Recordings were made using the average reference montage in which outputs of all of the amplifiers are summed and averaged, and this averaged signal is used as the common reference for each channel. Scalp-recorded brain potentials were low-pass filtered with 400 Hz cut-off frequency (anti-aliasing filter) and then digitized at 2 KHz and recorded using a BrainVision QuickAmp amplifier (Brain Products GmbH, Germany). Electrode impedances were kept below 5 k for all channels. The electro-oculogram (EOG) signals were recorded using two pairs of bipolar electrodes placed above and below the right eye to monitor vertical and at the canthus of each eye to monitor horizontal eye movements.

## 2.2 Data Analysis

The recorded EEG signals were filtered offline using a band-pass filter with cutoff frequencies set to 1.0Hz and 50.0 Hz (48dB/oct) and then segmented into epochs ranging from 100 ms before and 450 ms after the onset of the pitch shift. Epochs with EEG or EOG amplitudes exceeding 100  $\mu$ V were removed from data analysis. At least 100 epochs were averaged for ERP calculation for all subjects. The ERPs for each electrode site were first grand averaged across the 14 subjects. Then, for each subject, the latency and amplitude of the N1 and P2 ERP components were extracted by finding the most prominent negative and positive peaks within the time windows selected, based upon time-alignment with the grand-averaged responses.

EEG data recorded during the auditory oddball test were filtered offline using a band-pass filter with cut-off frequencies set to 1.0Hz and 55.0 Hz (48dB/oct) and then segmented into epochs ranging from 100 ms before and 400 ms after the onset of each sound. Epochs with EEG or EOG amplitudes exceeding 70  $\mu$ V, and all standard tones that were presented immediately after a deviating tone were removed from data analysis. Only standard tones were submitted for the analyses. Similar to our primary experimental data analysis algorithm, for each subject, the latency and amplitude of the N1 components were extracted by finding the most prominent negative peak within the window selected based upon time-alignment with the grand-averaged responses.

All MBEA and ERP data were tested with the Shapiro-Wilk test for normality of distribution before correlation analyses were done. Only those data sets that were normally distributed were submitted for correlation analyses. The correlations were estimated with Pearson ( $r$ ) product moment correlation coefficient and calculated for each EEG electrode separately.

## 3. Results

### 3.1. MBEA scores

The MBEA data analyses revealed a mean score of 0.89 for the Melodic organization. The minimum was 0.709, maximum 0.989, and the standard deviation was 0.085. The Shapiro-Wilk test for normality of distribution showed that Melodic score data were normally distributed ( $W=0.914$ ;  $p=0.183$ ). The MBEA Temporal organization mean score was 0.908 with a standard deviation 0.081; ranging from 0.737 to 0.983. The Shapiro-Wilk test for normality showed that temporal scores data were not normally distributed ( $W=0.855$ ;  $p=0.026$ ). Since the temporal scores were not normally distributed, only Melodic scores were used for correlation analyses. Since a score of 0.78 was suggested as the approximate division between healthy individuals and persons with congenital amusia (Peretz et al.,

2003), it can be stated that all our subjects had music processing abilities that were within the limits of variation for people who are not amusic.

### 3.2. Behavioral Data

Figure 1B displays a typical response from one of the participants. Response magnitudes reached a maximal decrease of 20 – 30 cents. The time of the maximal response magnitude was quite variable, no doubt due to the long stimulus duration, which likely caused voluntary responses following the immediate reflexive component (Hain et al., 2000). The onset times of the responses were typically about 100 ms.

### 3.3. ERP components

Responses to 50 cents stimuli were not analyzed in the present study since no detectable ERP potentials could be recorded from three of the 14 subjects. These same subjects produced normal ERP responses to 200 cent stimuli. Figure 2 illustrates grand averaged ERPs from three subjects with a response at Cz for a 200 cent stimulus and the lack of a response for the same subjects with a 50 cent stimulus. Inclusion of the 50 cent data from these three subjects would have severely affected the ERP correlations with the behavioral measures. These data also suggest that at least in some people, a 50 cent pitch-shift is not very perceptible, especially compared with a 200 cent pitch-shift. This fact raises a concern that by comparing easily detectable and barely detectable stimuli we might be analyzing different types of neural processes. In order to make analyses between the easily detectable MBEA stimuli and ERP measures as consistent as possible, we did not include the ERP data from the 50 cent pitch-shift condition.

As can be seen from the scalp distribution of the grand average data (Figure 1E), N1 had a maximal amplitude over the central midline electrode Cz. Thus, we selected the eight additional electrodes surrounding Cz, which also had large amplitudes, for ERP latency and amplitude measurements and correlation analyses. Based on comparison with the grand average N1 waveforms (Figure 1C), the latency window from 67 ms to 167 ms was chosen for the ERP latency and amplitude measurements in individual subjects. The Shapiro-Wilk test for normality of distribution showed that data from all nine electrodes across all subjects had a normal distribution. The P2 response was distributed over central-frontal midline electrodes and had a maximal amplitude at the FCz electrode. Therefore measurements of the P2 ERP latency and amplitude from nine electrodes centered over FCz were submitted for correlation analyses. Based on comparison with the grand average P2 waveforms (Figure 1C), the latency window from 167ms to 282ms was chosen for ERP latency and amplitude measurements.

In the oddball experiment, the N1 amplitude was maximal at central electrodes (Figure 3). Therefore the latency and amplitude measurements from C3; C1; Cz; C2; C4 electrodes were submitted for correlation analyses. Based on grand averaged N1 waveforms (Figure 3), the latency window from 80 ms to 100 ms was chosen for the ERP latency and amplitude measurements in individual subjects. The Shapiro-Wilk test for normality of distribution showed that data from all electrodes had a normal distribution.

### 3.4 Correlations

There were no significant correlations between the MBEA scores for Melodic organization and amplitude of the N1 or P2 responses or with the P2 latency. However, the latency of the N1 response was negatively correlated (Figure 1D) with the MBEA score for the Melodic organization at CPz ( $r = -0.562$ ;  $p = 0.036$ ) and C2 ( $r = -0.532$ ;  $p = 0.050$ ). On two other neighboring electrodes, the coefficient of correlation was close but did not reach a level of significance CP1 ( $r = -0.528$ ;  $p = 0.052$ ) and Cz ( $r = -0.493$ ;  $p = 0.073$ ).

In the oddball experiment there were no significant correlations between the MBEA scores for Melodic organization and latency of the N1. However, the peak amplitude of the N1 response was positively correlated (Figure 3) with the MBEA score for Melodic organization at C3 ( $r= 0.611$ ;  $p=0.020$ ); C1 ( $r=0.675$ ;  $p=0.007$ ) and Cz ( $r= 0.542$ ;  $p=0.045$ ).

## 4. Discussion

The present study addressed the general issue of the relation between neural mechanisms of voice control and music perceptual abilities. Specifically, we sought to determine which ERP components triggered by pitch-shifted voice auditory feedback were correlated with scores on the MBEA. Previous studies have shown that the pitch-shift paradigm provides quantitative measures of voice F0 and ERP responses that relate directly to perturbations in the vocal acoustical signal (Burnett et al., 1998; Behroozmand et al., 2009). It has also been shown that ERP responses to pitch-shifted voice feedback vary directly with the magnitude of the pitch-shift signal, and the responses reveal mechanisms related to identification of self-vocalization vs. alien vocalizations (Behroozmand et al., 2009; Heinks-Maldonado et al. 2005). Our present results of the significant negative correlations between the N1 latency and MBEA scores revealed that people with greater music perceptual ability apparently process changes in voice auditory feedback more rapidly than people who are unskilled in these abilities. An auditory oddball paradigm was used as a test of whether or not the ERPs related to pitch-shifted auditory voice feedback reflected a general property of the testing situation or were related more specifically to the vocal task. The negative findings from this experiment support the contention that neural processing of auditory feedback processing occurs more rapidly in people who score highly on a test of musical ability.

### 4.1. ERP latency and processing efficiency

While there were no significant correlations between musical perception and the amplitude of the ERPs elicited by pitch-shifted feedback, there was a negative correlation between musical perception and the latency of the N1 response. This finding suggests that musical perception imparts more rapid neural processing of unexpected changes in the auditory feedback of a person's vocal performance. The subjects who had shorter latencies of N1 had a higher MBEA score for the Melodic organization. Inter-individual differences in MBEA scores for the Melodic organization suggests that subjects who have higher MBEA scores might have better abilities to detect disparities in melodic structure. Although the MBEA scores might be affected by very complex issues related to the processing of music such as musical background, cultural and genetic heritage, etc. ultimately the scores reflect neural efficiencies in processing such complex auditory information as music.

One of the most obvious indices of efficient auditory processing is speed of change detection in the auditory environment. For example, a study of the mismatch negativity (MMN) component of auditory ERPs (elicited when a change in the auditory stimuli is detected) found that musicians had shorter MMN latencies (for pure and complex tones as well as for speech syllable stimuli) than non-musicians, without an increase in brain response amplitude, suggesting more efficient auditory processing (Nikjeh et al., 2009a). In light of these findings our results can be interpreted as a correlation between auditory processing efficiency (reflected by MBEA scores) and speed of auditory processing (measured by N1 latency) of unexpected shifts in the F0 of a person's own voice during vocalization. Thus our results suggest that the N1 response elicited with the voice perturbation paradigm might be associated with processing of auditory information. The absence of correlations with the P2 response might indicate that the P2 response is related more to motor processing or other aspects of sensory processing elicited by the voice pitch perturbation paradigm. The specific cortical location(s) involved in the pitch-shift paradigm

cannot be determined from ERP electrodes from a restricted montage of 64 sites, but the midline locations of the significant correlation suggest the events are bilateral.

Although these correlations suggest the relation between music perception and the measure of neural activity are weak, it should be remembered that in our subject group, the lowest score on the MBEA was .709 on a scale of 0 to 1.0. Thus, all scores were in the range that is considered to reflect music perception that is good to very good. Had we been able to test subjects whose scores fell across a broader range, i.e., 0 to 1.0, the correlations would have been much stronger across a broader array of electrodes. Therefore, the fact that we were able to observe correlations across electrode sites from a very restricted range of MBEA scores is quite significant.

#### 4.2. ERP components in voice perturbation paradigm

The pattern of ERP responses elicited by the voice pitch perturbation paradigm looks very much like P1-N1-P2 obligatory ERP components usually recorded in response to any auditory stimulus (Figure 1C and 3A). It is very likely that neuronal mechanisms underlying conventionally recorded auditory ERPs are involved in generation of P1-N1-P2 responses elicited by the voice pitch perturbation paradigm. However it is important to note that a simple analogy between the P1-N1-P2 complex routinely recorded in numerous auditory ERP studies and the P1-N1-P2 complex elicited in the voice perturbation paradigm might be misleading. In contrast to conventional auditory studies, P1-N1-P2 components in the voice perturbation paradigm are recorded when a subject is vocalizing, i.e. performing a complex, goal oriented motor act. Thus, the dynamic contribution of neuronal activity underlying this motor activity to the ERP responses elicited by the voice pitch perturbation paradigm has to be considered in the present discussion.

The influence of the vocal motor system on auditory processing was investigated in several studies of brain responses to the onset of voice (Curio et al., 2000; Heinks-Maldonado et al., 2005; Houde et al., 2002). Results of these studies showed that ERP or MEG responses from the auditory cortex are suppressed when the vocal system is active. These findings were interpreted to mean that an efference copy of the intended vocal output suppresses neural areas related to auditory processing of the feedback signal. If the feedback is altered such that it does not match the intended vocal output, the ERP suppression is reduced, and it is almost entirely eliminated for an alien voice or a person's own voice if the feedback is shifted by 400 cents. Thus a match between intended and actual auditory feedback is thought to be part of the neural mechanisms that are used to identify one's own voice. On the other hand, pitch-shifted voice feedback introduced during an ongoing vocalization causes an increase in the amplitude of ERP responses (Behroozmand et al., 2009; Liu et al., 2010). In addition direct recordings from neurons in monkey brains show that individual neurons that are normally suppressed during vocalization become more active in the presence of pitch-shifted voice feedback (Eliades and Wang, 2008). Therefore, although the data reported here might reflect both motor and sensory processing, the techniques used in this study do not allow the specific contributions from each modality to be considered independently of the other.

In light of our hypotheses about sequential stages of information processing that underlie compensatory reactions to unexpected changes in self-vocalization, the degree of the motor system's contribution to the ERPs might change as a function of time following the pitch-shift stimulus. In this context, our present results suggest that auditory processing of a deviation from an anticipated vocal output might last at least to the end of the N1 response time range (i.e., >100 ms). There do not appear to be exact measures of the time range of the N1 response, but based on estimates from visual inspection of averaged waveforms, we estimate the N1 begins about 70 ms following a stimulus and ends about 150 ms after a

stimulus. Although there are no data at present to indicate the timing of cortical activity related to the motor component of the responses to pitch-shift stimuli, estimates of this timing can be made from other measures.

For example, previous studies have shown that the latencies of the F0 compensatory responses can be as short as 100 ms (Larson et al., 2007). Since the laryngeal muscle contraction times preceding a change in voice F0 are about 30 ms (Alipour-Haghighi et al., 1991; Larson et al., 1987; Liu et al., 2011), and the latency from transcranial magnetic stimulation of the laryngeal motor cortex to a laryngeal EMG responses is about 12 ms (Ludlow and Lou, 1996; Rodel et al., 2004), it can be assumed that the earliest the laryngeal cortex could be activated would be about 58 ms following a pitch-shift stimulus. This means that auditory processing reflected by N1 in our study might not be attributed to the detection of change in the auditory feedback signal per se but rather to more advanced (than simply detecting changes in auditory input) sequential stages of auditory scene analyses. Indeed, there is at least one ERP component (labeled P1 in the present study) that could be a suitable candidate for the role of auditory input change detector. It is also notable that in the auditory domain of research on ERPs, it was suggested that P1 (in literature also labeled as P50) might reflect the processes of change detection and N1 might reflect assignment of pitch or phonemic quality to auditory objects in higher auditory centers (Chait et al., 2004).

This role of the auditory N1 in auditory processing is supported by other studies that showed association of the N1 response with object discrimination (Murray et al., 2006), percept formation (Roberts et al., 2000) and auditory memory (Conley et al., 1999; Lu et al., 1992). Making an analogy with the above studies and the N1 recorded in the voice pitch perturbation paradigm, we can speculate that auditory processing of a deviation from an anticipated vocal output reflected by N1 might be associated with higher cognitive aspects of vocal output monitoring. For instance, one of these aspects might be monitoring the results of vocal compensation. Since the vocal compensatory response may begin within the N1 time range, it may be that neuronal activity underlying N1 reflects the monitoring of the compensatory response. This hypothetical, higher cognitive level role of N1 in the pitch perturbation paradigm, suggests that the N1 response might reflect brain activity outside the primary and secondary auditory cortex. It is known that neuronal generators outside auditory cortex are necessary to account for evoked cerebral current fields within the auditory N1 time range (Alcaini et al., 1994; Naatanen and Picton, 1987). Moreover, current density studies suggested that in parallel to auditory cortex contributions to the auditory N1 response, additional frontal currents originating from the motor cortex or the supplementary motor area might contribute to this response (Giard et al., 1994). The authors hypothesize that these separate, simultaneously active temporal and frontal neural systems could be associated with auditory-motor linkages that short-circuit the highest levels of the auditory system (Giard et al., 1994).

If the suggested wide distribution of N1 (in the voice pitch perturbation paradigm) generators can be validated in the future, it will support the vision of N1 as a more integrative index of auditory processing in the pitch perturbation paradigm. Thus the significant correlations found in the present study between N1 latency and the MBEA measures may reflect interactions between auditory processing neuronal networks (N1 measures) and MBEA scores that measure the cumulative output of the same networks. It is interesting that the MBEA, as a cumulative measure of perceptual abilities, covary with inter-individual gray matter volume variations in the left superior temporal sulcus and the left inferior frontal gyrus (Mandell et al., 2007). These results allowed Mandell et al. (2007) to suggest that these regions could also be part of a network that enable subjects to map motor actions to sounds, including a feedback loop that allows for correction of motor actions (i.e., singing) based on perceptual feedback.



Taken together, the above-mentioned auditory ERP and Voxel-Based-Morphometry results are in agreement with our primary finding that the N1 response elicited in the voice perturbation paradigm is associated with processing of auditory information. Nevertheless additional experiments are needed to clarify the functional properties of this N1 response as well other ERP components elicited in the voice perturbation paradigm.

#### 4.3. N1 response in auditory oddball experiment

The significant positive correlation that was found in the present study between the amplitude of auditory N1 elicited by a repetitive pure tone and the MBEA score for Melodic organization indicate that better ability (larger MBEA scores) to detect disparities in melodic structure, the smaller (less negative) auditory N1 amplitudes. The smaller N1 amplitude might be interpreted as a reduction of processing demands occurring during the N1 time window. Thus our result might indicate that increases in processing efficiency (reflected by MBEA score) is accompanied by a reduction in processing demands occurring during the N1 time window. A similar observation was made for the P1 amplitude (Nikjeh et al., 2009a). It was also shown that a decrease in N1 amplitude is correlated with speech identification improvement after practice (Ben-David et al. 2011).

Since we did not find significant latency correlations between the auditory N1 and MBEA scores, we can speculate that our primary result might be specific to the N1 elicited by voice pitch-shifted feedback. This specificity might be associated with active brain processes of laryngeal pitch control that occurred in our primary experiment and not in the oddball experiment. Mechanisms of interactions between neuronal networks underlying laryngeal pitch control and auditory processing occurring within the N1 latency window are not well understood. There are several studies suggesting that auditory processing and laryngeal pitch control may be inherently related systems (Nikjeh et al., 2009b; Amir et al., 2003). However relationships between these two systems might have a complex structure. On one hand, laryngeal pitch control might successfully operate without superior auditory discrimination abilities (Amir et al., 2003) on the other hand auditory discrimination and laryngeal control both benefit from music training (Nikjeh et al., 2009b). Future research is required in order to elucidate neuronal mechanisms of auditory-motor interactions during vocal control.

#### 4.4. Limitations of the study and conclusions

There are several limitations in our study that need to be considered with respect to the conclusions. It is highly likely that the N1 response recorded in the voice perturbation paradigm of the present study is result of overlapping activity from several neuronal generators. Each generator may have different temporal and spatial properties that can vary between and within subjects. It is also possible that only one of the generators is strongly correlated with musical processing. These possibilities would lead to considerable variability that would in turn lead to a decrement in the correlation coefficient. This methodological limitation could to some extent be overcome in future studies by the addition of an additional “baseline” experimental condition related to motor or sensory processing of musical performance. Ideally, such a condition would reduce the number of generators and thereby strengthen the correlation between N1 and musical ability. This “baseline” experimental condition might be designed analogous the ERP subtraction approach that is typical in many studies of the MMN. In the present study, we attempted to correlate the N1 magnitude directly with the MBEA scores, which probably resulted in large variability in N1 measures and hence reduced the correlations with the MBEA scores. Nevertheless, the results of the present study suggest that neuronal activity within the time range of the N1 response might reflect some of auditory processing components of auditory-motor integration, which is an important step in understanding neural mechanisms related to musical ability and audio-motor control of vocal production.

## Acknowledgments

This research was supported by a grant from NIH, Grant No. 1R01DC006243.

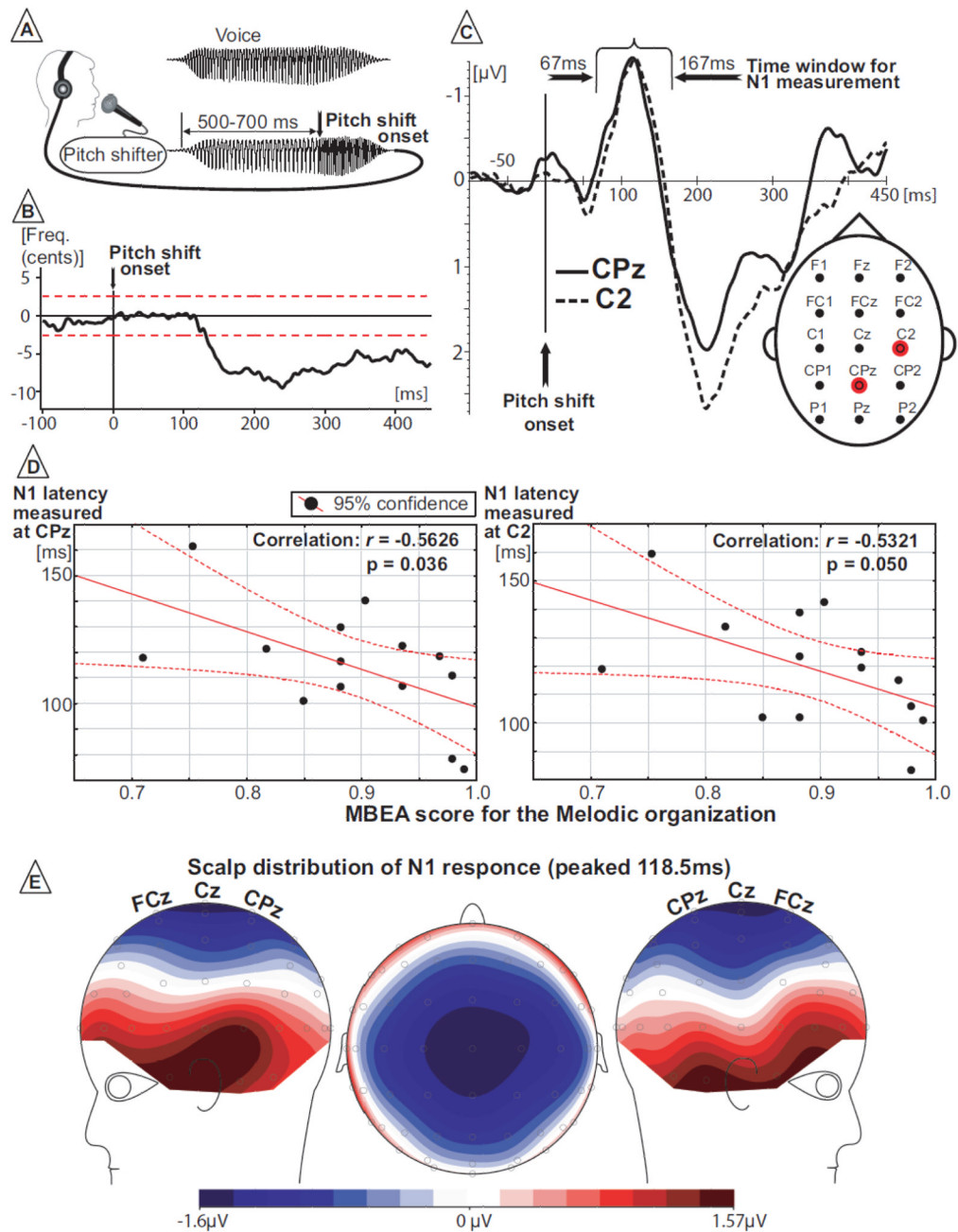
## References

- Alcaini M, Giard MH, Thevenet M, Pernier J. Two separate frontal components in the N1 wave of the human auditory evoked response. *Psychophysiology*. 1994; 31:611–615. [PubMed: 7846222]
- Alipour-Haghighi F, Perlman AL, Titze IR. Tetanic response of the cricothyroid muscle. *The Annals of otology, rhinology, and laryngology*. 1991; 100:626–631.
- Amir O, Amir N, Kishon-Rabin L. The effect of superior auditory skills on vocal accuracy. *The Journal of the Acoustical Society of America*. 2003; 113:1102–1108. [PubMed: 12597203]
- Behroozmand R, Karvelis L, Liu H, Larson CR. Vocalization-induced enhancement of the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin Neurophysiol*. 2009; 120:1303–1312. [PubMed: 19520602]
- Ben-David BM, Campeanu S, Tremblay KL, Alain C. Auditory evoked potentials dissociate rapid perceptual learning from task repetition without learning. *Psychophysiology*. 2011; 48:797–807. [PubMed: 21054432]
- Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America*. 1998; 103:3153–3161. [PubMed: 9637026]
- Chait M, Simon JZ, Poeppel D. Auditory M50 and M100 responses to broadband noise: functional implications. *Neuroreport*. 2004; 15:2455–2458. [PubMed: 15538173]
- Conley EM, Michalewski HJ, Starr A. The N100 auditory cortical evoked potential indexes scanning of auditory short-term memory. *Clin Neurophysiol*. 1999; 110:2086–2093. [PubMed: 10616113]
- Curio G, Neuloh G, Numminen J, Jousmaki V, Hari R. Speaking modifies voice-evoked activity in the human auditory cortex. *Hum Brain Mapp*. 2000; 9:183–191. [PubMed: 10770228]
- Eliades SJ, Wang X. Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature*. 2008; 453:1102–1106. [PubMed: 18454135]
- Giard MH, Perrin F, Echallier JF, Thevenet M, Froment JC, Pernier J. Dissociation of temporal and frontal components in the human auditory N1 wave: a scalp current density and dipole model analysis. *Electroencephalography and clinical neurophysiology*. 1994; 92:238–252. [PubMed: 7514993]
- Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Exp Brain Res*. 2000; 130:133–141. [PubMed: 10672466]
- Heinks-Maldonado TH, Mathalon DH, Gray M, Ford JM. Fine-tuning of auditory cortex during speech production. *Psychophysiology*. 2005; 42:180–190. [PubMed: 15787855]
- Houde JF, Nagarajan SS, Sekihara K, Merzenich MM. Modulation of the auditory cortex during speech: an MEG study. *J Cogn Neurosci*. 2002; 14:1125–1138. [PubMed: 12495520]
- Kawahara H. Interactions between speech production and perception under auditory feedback perturbations on fundamental frequencies. *J Acoust Soc Jpn*. 1994; 15:201–202.
- Larson CR, Kempster GB, Kistler MK. Changes in voice fundamental frequency following discharge of single motor units in cricothyroid and thyroarytenoid muscles. *Journal of speech and hearing research*. 1987; 30:552–558. [PubMed: 3695447]
- Larson CR, Sun J, Hain TC. Effects of simultaneous perturbations of voice pitch and loudness feedback on voice F0 and amplitude control. *The Journal of the Acoustical Society of America*. 2007; 121:2862–2872. [PubMed: 17550185]
- Liu H, Behroozmand R, Larson CR. Enhanced neural responses to self-triggered voice pitch feedback perturbations. *Neuroreport*. 2010; 21:527–531. [PubMed: 20386347]
- Liu H, Meshman M, Behroozmand R, Larson CR. Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clin Neurophysiol*. 2011; 122:951–957. [PubMed: 20869305]

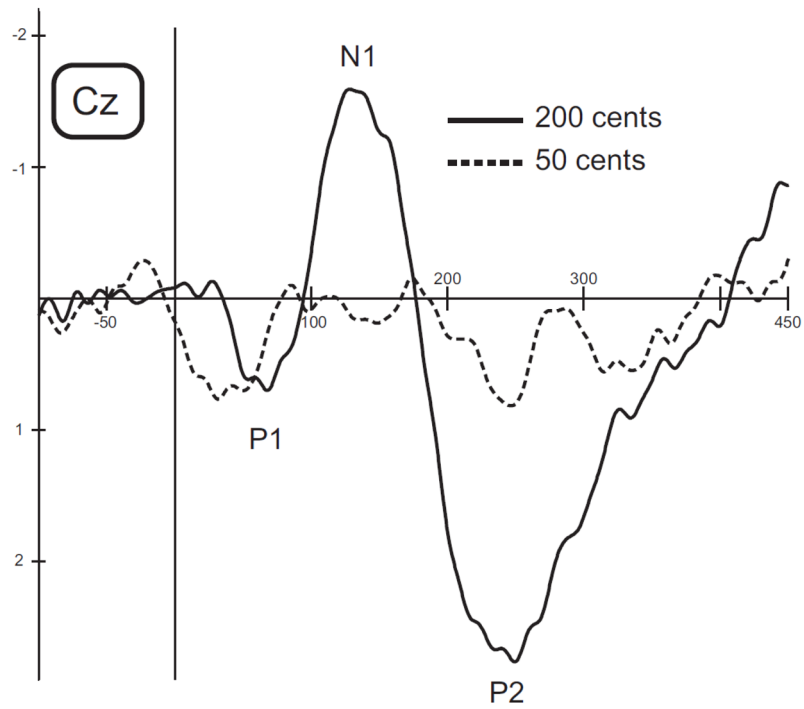
- Lu ZL, Williamson SJ, Kaufman L. Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science (New York, NY)*. 1992; 258:1668–1670.
- Ludlow, CL.; Lou, G. Observations on human laryngeal muscle control. In: Davis, PJ.; Fletcher, NH., editors. *Vocal Fold Physiology: Conrolling Complexity and Chaos*. Singular; San Diego: 1996. p. 201-218.
- Mandell J, Schulze K, Schlaug G. Congenital amusia: an auditory-motor feedback disorder? *Restorative neurology and neuroscience*. 2007; 25:323–334. [PubMed: 17943009]
- Murray MM, Camen C, Gonzalez Andino SL, Bovet P, Clarke S. Rapid brain discrimination of sounds of objects. *J Neurosci*. 2006; 26:1293–1302. [PubMed: 16436617]
- Naatanen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*. 1987; 24:375–425. [PubMed: 3615753]
- Nikjeh DA, Lister JJ, Frisch SA. Preattentive cortical-evoked responses to pure tones, harmonic tones, and speech: influence of music training. *Ear and hearing*. 2009a; 30:432–446. [PubMed: 19494778]
- Nikjeh DA, Lister JJ, Frisch SA. The relationship between pitch discrimination and vocal production: comparison of vocal and instrumental musicians. *The Journal of the Acoustical Society of America*. 2009b; 125:328–338. [PubMed: 19173420]
- Oostenveld R, Praamstra P. The five percent electrode system for high-resolution EEG and ERP measurements. *Clin Neurophysiol*. 2001; 112:713–719. [PubMed: 11275545]
- Peretz I, Champod AS, Hyde K. Varieties of musical disorders. The Montreal Battery of Evaluation of Amusia. *Annals of the New York Academy of Sciences*. 2003; 999:58–75. [PubMed: 14681118]
- Roberts TP, Ferrari P, Stufflebeam SM, Poeppel D. Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J Clin Neurophysiol*. 2000; 17:114–129. [PubMed: 10831104]
- Rodel RM, Olthoff A, Tergau F, Simonyan K, Kraemer D, Markus H, Kruse E. Human cortical motor representation of the larynx as assessed by transcranial magnetic stimulation (TMS). *The Laryngoscope*. 2004; 114:918–922. [PubMed: 15126757]

### Highlights

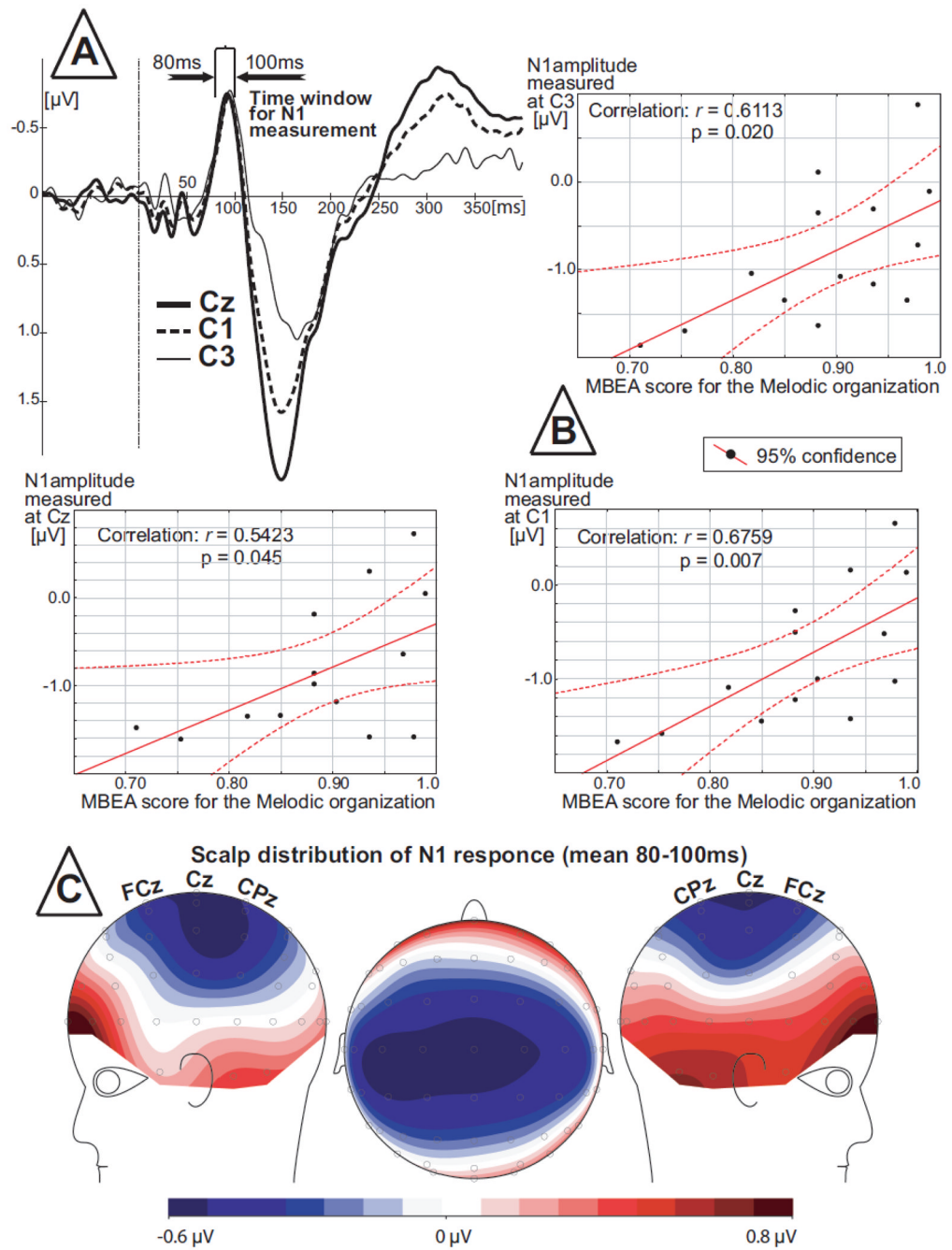
Auditory processing of vocal production is imperative element of mechanisms controlling vocalization. The N1 response to unexpected shifts of fundamental frequency in a subject's own voice is associated with auditory processing. Efficiency of auditory processing is correlated with N1 latency.



**Figure 1.** Voice perturbation paradigm and study results. (A) Schematic illustration of voice perturbation paradigm. (B) Behavioral (voice) data from one representative subject illustrates compensatory response to pitch shift changes in subject’s F0. Horizontal red dashed lines represent two standard deviations of F0 during baseline time window. (C) Grand-averaged ERP for two electrode sites where significant correlations were found. Positions of these electrodes on the scalp are marked by red color. (D) Results of correlation analyses. (E) Scalp distribution of grand-averaged N1 response.



**Figure 2.** Grand averaged ERPs for 3 subjects from the Cz electrode for a 200 cent pitch shift stimulus (solid line) and for a 50 cent stimulus (dashed line).



**Figure 3.** N1 ERP component recorded on the oddball paradigm. (A) Grand-averaged ERP for three electrode sites where significant correlations were found. (B) Results of correlation analyses between N1 peak amplitude and MBEA score. (C) Scalp distribution of grand-averaged N1 response.