Effects of Vowel Auditory Training on Concurrent Speech Segregation in Hearing Impaired Children

Hossein Talebi, PhD¹, Abdollah Moossavi, MD², Yones Lotfi, MD¹, and Soghrat Faghihzadeh, PhD³

Abstract

Objective: This clinical trial investigated the ability of concurrent speech segregation in hearing impaired children. The auditory behavioral responses and auditory late responses (ALRs) were compared between test and control groups prior to vowel auditory training and after 3 and 6 months of vowel auditory training to find the effects of bottom-up training on concurrent speech segregation in hearing impaired children.

Methods: Auditory behavioral responses for 5 vowels and ALRs for double synthetic vowels, with special physical properties, were recorded in a timetable in 30 hearing impaired children (test group = 15 and control group = 15).

Results: Identification score and reaction time data showed that the test group was approximately proficient for some vowels ($P < .05$ for vowels /æ/, /e/, and /u:/) and took less time to process after 6 months of training. N1-P2 amplitude indexing of the vowel change detection and reflecting central auditory speech representation without active client participation has been increased in the test group ($P < .05$).

Conclusion: The present study showed training-related improvements in concurrent speech segregation. This information provided evidence for bottom-up training based on F0, its differences in auditory scene analysis, and neural underpinnings.

Keywords

auditory scene analysis, ASA, concurrent speech segregation, hearing impaired children, vowel auditory training

Introduction

We often experience a complex acoustic environment with auditory information originating from several simultaneously active sources that often overlap in many acoustic parameters. However, we are able to identify auditory events and hear distinct auditory objects. Auditory scene analysis (ASA) is the process involving the ability to segregate those sound inputs that originate from different sound sources and integrate those that belong together.¹ Accordingly, segregation and integration processes are 2 fundamental aspects of ASA.

Bregman¹ believes that formation of auditory streams is the result of processes of sequential and simultaneous segregation. Sequential segregation separates and connects sense data over time, whereas simultaneous segregation selects, from the data arriving at the same time, those components that are probably parts of the same sound.

Simultaneous segregation uses the properties of the incoming mixture that tends to be true whenever a subset of its components comes from a common source. For example, there is a broad class of sounds called “periodic,” such as human voice, in which all the component frequencies (harmonic relations) are integer multiples of a common fundamental (fundamental frequency or F0). The auditory system takes advantage of this fact. If it detects, in the input, a subset of frequencies that are all multiples of a common fundamental, it strengthens its tendency to treat this subset as a single distinct sound.¹

Much evidence supporting this view shows that the bottom-up sensory processes handle the segregation of sounds to distinct sources at a pre-attentive level of acoustic processing.²-⁴ This indicates that the auditory streams are sorted prior to stimulus selection. If automatic, low level systems

¹Department of Audiology, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran
²Department of Audiology, Iran University of Medical Sciences, Tehran, Iran
³Department of Social Medicine, School of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran

Corresponding Author:
Abdollah Moossavi, MD, Department of Audiology, Iran University of Medical Sciences, 134875459, Tehran, Iran.
Email: amoossavi@gmail.com
handle the sorting of the incoming mixture of sounds, this could facilitate the ability to select external sources for further processing such as understanding of speech.

Demany\textsuperscript{7} supposed that stream segregation processes are operative very early in life. In addition, Sussman et al\textsuperscript{6} demonstrated that the mechanisms for simultaneous auditory segregation operate similarly in school-age children and adults when frequency proximity is the cue for segregation.

Winkler et al\textsuperscript{7} showed that newborn infants, like adults, are able to segregate concurrent streams of sound and organize the auditory input. They pointed out that neonates have potential capabilities required for shaping auditory conceptual objects.

Segregating concurrent streams of sound is a critical moment of sound organization,\textsuperscript{1} a prerequisite of perceiving objects and forming conceptual objects.\textsuperscript{5} Concurrent segregation of sound sequences uses the temporal behavior of various acoustic parameters, of which spectral pitch is the most effective one.\textsuperscript{9}

It is believed that segregation of acoustic components of speech such as vowels is very important for speech perception in children.\textsuperscript{10} These speech sounds convey much information about F0, the lowest frequency, and its high-numbered harmonics (F1, F2, etc). This acoustic information (F0 and its harmonic relations) carrying cues for pitch and timbre perception is a basis for speech perception, especially in natural auditory environments.\textsuperscript{10} Although vowels are not complex sound stimuli, they are periodic signals leading to pitch perception.\textsuperscript{10} F0 and first formant contrasts may be more perceptually salient in segregation of vowels and pitch cues.\textsuperscript{11} Most children with severe hearing loss are more likely to have residual hearing in the low frequencies than the higher ones. As vowels are prerequisites for speech formation and perception,\textsuperscript{10} we hypothesized that improving ability of vowel perception in hearing impaired children results in better speech understanding. The present study is the first investigation that connects concurrent speech segregation (as a basis of speech perception in complex auditory environments) and auditory rehabilitation (vowel auditory training). In other words, we investigated the effects of vowel auditory training on concurrent speech segregation and speech perception in children with hearing loss.

Event related potentials (ERPs) are non-invasive measurements of cortical brain activity in response to sensory events. They give information about the timing of certain cognitive processes evoked by a given sound because of the high temporal resolution (in the order of milliseconds) of the responses that are time-locked to stimulus events. Event related potentials provide distinctive signatures for detection of sound change. Of particular importance are auditory late responses (ALRs), which reflect sound or change detection.

To examine simultaneous speech segregation, automatically in hearing impaired children, we used an electrophysiological index of auditory processing that does not require the participant to actively process the sounds or to indicate his or her perception of them. The N1-P2 complex is an obligatory ERP that can reflect central auditory speech representation without active participation.\textsuperscript{12-16}

The N1 response reaches maximal amplitudes at fronto-central sites.\textsuperscript{17} The N1-P2 complex is thought to reflect synchronous neural activation of structures in the thalamic-cortical segment of the central nervous system in response to auditory stimulation.\textsuperscript{18-20} This complex could be used to monitor neurophysiologic changes during speech-sound acquisition after auditory rehabilitation, hearing aid use, or any other form of auditory learning.\textsuperscript{21-23} The N1-P2 complex proved to be a stable measure when it is recorded from electrode Cz, showing the reliable latency or amplitude results from test to retest.\textsuperscript{24} Some evidence demonstrated an increase in N1-P2 amplitude after auditory training.\textsuperscript{25} Increases in N1-P2 amplitude are thought to reflect increases in neural synchrony.\textsuperscript{24} Changes in neural firing patterns coinciding with learned behaviors are consistent with Hebbian principles of neural plasticity.\textsuperscript{26}

Although the mismatch negativity (MMN) provides insight into physiologic processes underlying speech discrimination and training-related plasticity, this ERP may not be the most efficient response for assessing speech-sound representation in individuals.\textsuperscript{27,28} This response is difficult to extract from background electroencephalic noise and often requires prolonged testing time and offline analyses and may be absent in normal persons.\textsuperscript{27,29} The N1-P2 complex may provide a practical tool for clinicians because this response can be collected using most commercially available systems and requires little testing time and minimal offline data manipulation compared to other ERPs. N1-P2 reflects changes in neural activity that coincide with improved perception, a one-to-one relationship between changes in perception and physiology.\textsuperscript{30-32}

Recent evidence indicated an improved identification of synthetic double vowels with increasing the difference of their F0s. In addition, it showed that learning improves this ability, evidenced by an increase of P2 amplitude.\textsuperscript{33}

In this study, hearing impaired children were presented with 5 vowels and 20 double vowels (pairs of synthetic vowels) with different F0 to identify them during auditory behavioral recording (identification score and reaction times) and ALRs before and after 3 and 6 months of vowel auditory training. We hypothesized that hearing impaired children would have more difficulty identifying vowels assessed by identification score and reaction times prior to training. In addition, we hypothesized that changes in processing would also be paralleled with ERPs.
Materials and Methods

Thirty hearing impaired children with symmetrically binaural moderate to severe sensorineural hearing loss participated in this study. All of the participants used binaural behind the ear (BTE) hearing aids. They were 4 to 6 years old and included 15 female and 15 male participants, who were divided equally to control and test groups. Their age and hearing loss levels were the same. They were all right-handed and Farsi speaking (Persian language/mother language). Both groups received traditional rehabilitation programs for their disability; in addition, the test group received the vowel auditory training. Participant recruitment and follow-up occurred between August 22, 2012, and April 20, 2013.

Ethics Statement

The parents of the participants were given informed written consent after the testing procedure was explained to them. The treatment procedure was performed in accordance with the Declaration of Helsinki. Our study was registered with the Iranian Registry of Clinical Trials (registration code IRCT2013011912174N1). The ethics committee of the University of Social Welfare and Rehabilitation Sciences approved our research.

Stimuli and Procedure

The stimuli were 5 recorded steady-state Persian vowels: “AE” as in /æ/, “ER” as in /e/, “AH” as in /ɑ/:/,”EE” as in /i:/, and “OO” as in /u:/ / . A male speaker produced these vowels in an acoustic chamber. Each vowel was 200 milliseconds in duration (16 bits, and 48-kHz sample rate). F0 and formant frequencies were held constant. The source signal was the same in all 5 vowels, simulating “equal vocal effort.”

To create the double vowels, the researchers summed up the digital waveforms of 2 phonetically different vowels. Each pair contained 1 vowel with F0 set at 100 Hz; the other vowel’s F0 was 0.5 semitones higher (1 semitone = 1/12 octave). Snyder and Alain34 showed that performance of young and older adults improved gradually with increasing ΔF0 up to 0.5 semitones, after which there was no significant improvement in performance from 0.5 semitones to 4 semitones ΔF0. Each vowel was paired with every other vowel, giving a total of 20 different pairs. The stimuli were presented binaurally through TDH 49 headphones at 80 dB nHL. Stimuli were randomized in blocks of 20 trials. It took 1 hour. They were designed as stimuli of ALRs. Prior to data collection, participants were presented with each stimulus to become familiarized with the task and the response.

Vowel Auditory Training

Vowel auditory training objectives typically are designed to contrast different vowels.11 We trained 5 vowels—/æ/, /e/, /o/, /i:/, and /u:/—for 15 hearing impaired children (test group) using nonsense syllables. These syllables included unvoiced consonants and vowels as follows: / Pæ/, Shæ/, /Sæ/, /Hæ/, and /Kæ/. These combinations were repeated for the other vowels. The syllables were presented orally behind the children. After presentation, we asked children to identify and produce the syllables verbally. The training was conducted for 2 sessions (duration of each ~2 hours) per week. We recorded correct identification scores and reaction times for 5 vowels at the first time (prior to vowel auditory training) and classified all data after 3 and 6 months of vowel auditory training. Auditory late responses were recorded in parallel to auditory behavioral responses (with test duration of ~ 90 minutes).

Auditory Late Responses

Prior to vowel auditory training, we recorded electrophysiological and behavioral responses of the test group, which received vowel auditory training. Electrophysiological responses were collected (rate: 1 per 2 seconds, ie, 0.5/s rate, artifact reject: ± 100 µV, gain: 10 000-20 000, trials per average: 25-50), digitized, and filtered (band-pass 1-30 Hz) from an array of 5 electrodes (2 inverting: Cz, 2 noninverting: mastoids, and ground: forehead) using BioLogic Auditory Evoked Potential System version 7.0.0 software. Event related potentials were stored for offline analysis. Eye movements were monitored with electrodes placed at the outer canthi and at the superior and inferior margins of orbit. The analysis epoch included 100 milliseconds of pre-stimulus activity and 500 milliseconds of poststimulus activity. For each participant, ERPs were then averaged separately for each double vowel. We measured the amplitude of the N1-P2 component of auditory late responses. After 3 and 6 months of vowel auditory training, the procedures were repeated.

Auditory Behavioral Responses

Auditory behavioral responses including identification scores (in percentages) and reaction times (in milliseconds) were recorded before and after 3 and 6 months of vowel auditory training for the test group and without training for the control group. All of the data were collected at the University of Social Welfare and Rehabilitation Sciences.

Statistical Analysis

As this study was a parallel clinical trial between 2 groups, we analyzed statistical data between test and control groups in repeated measures analysis of variance (ANOVA). In addition, Pearson’s correlation test was used to measure both the auditory behavioral and ERP responses.
Results and Analysis

Behavioral Data

To demonstrate the ability to perceive the vowels used in this study, we analyzed auditory behavioral responses of both groups. This measure reflects identification scores and reaction times before and after 3 and 6 months of vowel auditory training. The test group was nearly proficient to identify some vowels correctly (Table 1).

There was a significant difference between test and control groups with regard to correctly identifying 2 vowels (\( F = 5.320 & P = .029 \) for /æ/, \( F = 2.023 & P = .166 \) for /e/, \( F = 1.891 & P = .180 \) for /ɒː/, \( F = 0.573 & P = .455 \) for /i:/, and \( F = 6.669 & P = .015 \) for /u:/, respectively), suggesting that training affected the identification scores of vowels /æ/ and /u:/.

Overall, participants in the control group tended to take more time than those in the test group to produce their responses.

The RT data presented in Figure 1 suggest that identifying the vowels improved in the test group with vowel auditory training.

Electrophysiological Data

The effect of vowel auditory training on N1-P2 amplitude was significant for 10 double-vowel stimuli (\( F = 7.233 & P = .008 \) for AE,EE, \( F = 7.649 & P = .006 \) for AE,ER, \( F = 13.473 & P = .001 \) for AH,AE, \( F = 16.448 & P < .001 \) for AH,ER, \( F = 7.403 & P = .007 \) for EE,AE, \( F = 29.575 & P < .001 \) for EE,AH, \( F = 15.100 & P < .001 \) for EE,ER, \( F = 25.791 & P < .001 \) for EE,OO, \( F = 9.699 & P = .003 \) for ER,OO, \( F = 8.340 & P = .005 \) for OO,ER) and approximately significant for AE,AH (\( F = 3.424 & P = .060 \)). These results demonstrated that N1-P2 amplitude increased at least in response to half of the double-vowel stimuli after 3 and 6 months of vowel auditory training. Table 3 shows the group mean of N1-P2 amplitudes for the control and test groups. Bonferroni test indicated a statistically significant

---

Table 1. Group Mean Identification Scores of 5 Vowels Used in Our Study.\(^a\)

<table>
<thead>
<tr>
<th>Group</th>
<th>Vowel</th>
<th>First Session</th>
<th>3 Months Later</th>
<th>6 Months Later</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (without vowel auditory training)</td>
<td>/æ/</td>
<td>68.00 ± 14.73</td>
<td>73.33 ± 12.34</td>
<td>70.67 ± 10.32</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>65.33 ± 11.87</td>
<td>69.33 ± 14.86</td>
<td>68.00 ± 10.14</td>
</tr>
<tr>
<td></td>
<td>/ɒː/</td>
<td>66.67 ± 14.47</td>
<td>73.33 ± 14.47</td>
<td>73.33 ± 12.34</td>
</tr>
<tr>
<td></td>
<td>/e:/</td>
<td>62.67 ± 14.86</td>
<td>69.33 ± 14.86</td>
<td>69.33 ± 10.32</td>
</tr>
<tr>
<td></td>
<td>/u:/</td>
<td>58.67 ± 14.07</td>
<td>66.67 ± 17.99</td>
<td>64.00 ± 13.52</td>
</tr>
<tr>
<td>Test (with vowel auditory training)</td>
<td>/æ/</td>
<td>64.00 ± 15.49</td>
<td>77.33 ± 10.32</td>
<td>96.00 ± 8.28</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>53.33 ± 12.34</td>
<td>72.00 ± 10.14</td>
<td>90.67 ± 10.32</td>
</tr>
<tr>
<td></td>
<td>/ɒː/</td>
<td>61.33 ± 14.07</td>
<td>76.00 ± 13.52</td>
<td>92.00 ± 10.14</td>
</tr>
<tr>
<td></td>
<td>/e:/</td>
<td>50.67 ± 12.79</td>
<td>66.67 ± 9.75</td>
<td>92.00 ± 10.14</td>
</tr>
<tr>
<td></td>
<td>/u:/</td>
<td>57.33 ± 12.79</td>
<td>72.00 ± 12.64</td>
<td>93.33 ± 9.75</td>
</tr>
</tbody>
</table>

\(^a\)The data show that both groups of participants were approximately proficient in correctly identifying vowels. Values are mean ± standard deviation.

Table 2. Group Mean Reaction Times Needed to Identify the Vowels Used in Our Study.\(^a\)

<table>
<thead>
<tr>
<th>Group</th>
<th>Vowel</th>
<th>First Session</th>
<th>3 Months Later</th>
<th>6 Months Later</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (without vowel auditory training)</td>
<td>/æ/</td>
<td>818.67 ± 110.18</td>
<td>818.00 ± 67.31</td>
<td>818.33 ± 63.32</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>802.67 ± 97.35</td>
<td>788.00 ± 73.60</td>
<td>793.33 ± 79.16</td>
</tr>
<tr>
<td></td>
<td>/ɒː/</td>
<td>802.00 ± 89.61</td>
<td>709.33 ± 76.48</td>
<td>707.33 ± 74.20</td>
</tr>
<tr>
<td></td>
<td>/e:/</td>
<td>805.07 ± 127.87</td>
<td>736.00 ± 92.10</td>
<td>749.33 ± 108.59</td>
</tr>
<tr>
<td></td>
<td>/u:/</td>
<td>767.33 ± 106.33</td>
<td>748.67 ± 121.99</td>
<td>761.33 ± 120.34</td>
</tr>
<tr>
<td>Test (with vowel auditory training)</td>
<td>/æ/</td>
<td>736.00 ± 86.42</td>
<td>645.33 ± 67.28</td>
<td>599.33 ± 66.05</td>
</tr>
<tr>
<td></td>
<td>/e/</td>
<td>741.33 ± 122.17</td>
<td>638.67 ± 84.67</td>
<td>595.33 ± 68.75</td>
</tr>
<tr>
<td></td>
<td>/ɒː/</td>
<td>726.67 ± 112.22</td>
<td>678.67 ± 124.89</td>
<td>628.00 ± 123.82</td>
</tr>
<tr>
<td></td>
<td>/e:/</td>
<td>782.00 ± 180.20</td>
<td>717.33 ± 241.55</td>
<td>655.33 ± 155.60</td>
</tr>
<tr>
<td></td>
<td>/u:/</td>
<td>722.67 ± 130.95</td>
<td>654.67 ± 141.26</td>
<td>611.33 ± 126.76</td>
</tr>
</tbody>
</table>

\(^a\)Values are mean ± standard deviation.
difference in the test group between the first and third sessions (6 months of auditory training) and between the second (3 months of auditory training) and third sessions. In addition, Figure 2 shows the group mean of ERPs elicited by the double-vowel stimuli for the control and test groups, separately. Both groups showed a characteristic N1-P2 response following stimulus onset, with larger responses in the test group. This amplitude difference could in part be due to improvement of N1-P2 responses in the test group as a result of vowel auditory training.

The N1-P2 amplitude correlated significantly with reaction times in identification of vowels in the test group (Table 4).

**Discussion**

The ability to correctly identify single and double vowels showed improvement after vowel auditory training. In the present study, overall, the control group was less accurate in identifying vowels compared to the test group, meaning that hearing impaired children would have more difficulty
identifying vowels without vowel auditory training. This problem may in part be related to deficits in parsing concurrent sounds.35-38

**Behavioral Results**

The test group showed statistically significant changes in RT results (for 3 vowels) after 3 and 6 months of training regarding the results of pretraining tests (P < .05). The current results did not show a persistent and significant decrease in RTs for vowels in the control group, although this group exhibited an overall but not significant increase in the identification score of some vowels during the 6 months. The effect of vowel auditory training in the test group showed a substantial decrease in RTs for all of the vowels. The analysis of RT data suggested that hearing impaired children without vowel auditory training required more processing time for correct vowel identification. We also found that data of the identification score (vowels /æ/ and /u:/) showed statistically significant results. It is
apparent that vowel identification improved after 6 months of vowel auditory training. Together, the accuracy (correct identification score) and RT data suggest that vocal auditory training leads to decreased processing time of stimuli. This is consistent with a study by Atienza et al demonstrating the improved time course of perceptual learning as revealed by changes in reaction times after auditory training.

According to our findings, acoustic training with nonsense syllables (bottom-up stimuli) leads to improvement of auditory information reaching the higher auditory system and results in the ultimate betterment of higher order (top-down) processes followed by auditory plasticity, for example, speech perception. This hypothesis could be studied comprehensively in future research.

Electrophysiological Results

The electrophysiological results provide comprehensive time-locked evidence for training-related processes. In the present study, we examined changes of N1-P2 amplitude associated with vowel auditory training. As we know, the N1-P2 complex event related potential holds promise as a clinical tool for assessing changes in neural activity associated with auditory rehabilitation.

Overall, ERP N1-P2 amplitude was larger in the test group, an effect at least attributable to vowel auditory training effects. N1-P2 amplitude showed a significant increase in 3 and 6 month vowel auditory training (at least for half of double vowels) relative to the control group. Furthermore, the N1-P2 amplitude changes showed a significant correlation with vowel identification in the test group.

In this study, the training-related increase in N1-P2 amplitude suggests improvement in the early bottom-up parsing of concurrent vowels without active participation. This interpretation is based on the fact that the N1-P2 amplitude becomes larger after auditory training. In addition, it is shown that the presence of the N1-P2 complex in an auditory stimulus provides physiologic evidence of the arrival at the auditory cortex of sensory information. Hence, this complex reflects the presence and registration of audible stimuli (bottom-up procedure). Although necessary for discrimination (top-down procedure), its presence does not by itself indicate discrimination. In the present study, we showed by this electrophysiologic response that hearing impaired children receiving vowel auditory training (nonsense syllables), which prevents top-down and high-level cognitive processing such as categorical perception, have more ability to detect concurrent speech sounds without active participation. The increased N1-P2 amplitude may be related to improvements in frequency selectivity, which might affect the ability to form accurate F0 representations for concurrent vowels.

Acknowledgments

The authors are especially thankful to Al Bregman, who, after a very beneficial electronic discussion, suggested E. Sussman for further follow-up of the project, and also thankful to Elyse S. Sussman for introducing Joel S. Snyder (Department of Psychology, University of Nevada, Las Vegas), who provided his kindly co-supervision and aid in this study and who humbly rejected being named as co-author. The authors are so thankful to all the children and their families for their participation in this research. This study was conducted as a PhD thesis at the University of Social Welfare and Rehabilitation Sciences.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References


