Composing auditory ERPs: Cross-linguistic comparison of auditory change complex for Japanese fricative consonants

Makiko Sadakata¹, Loukianos Spyrou¹, Mizuki Shingai², & Kaoru Sekiyama²

¹ Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands
² Division of Cognitive Psychology, Kumamoto University, Japan

m.sadakata@donders.ru.nl, l.spyrou@donders.ru.nl, m1_tsu_5@yahoo.co.jp, sekiyama@kumamoto-u.ac.jp

Abstract

Two series of Electroencephalogram (EEG) measurements indicated that average auditory event-related potentials (ERPs) elicited by the disyllabic sound /asu/ were different between Japanese and English native listeners. Significant differences were observed in the time window where the P1-N1-P2 complex for /a/ is expected. This difference may be due to the absence/presence of the Auditory Change Complex (ACC) elicited by /s/. Furthermore, by combining the P1-N1-P2 complex elicited by each component of /asu/ (/a/s/u/) recorded individually, it was possible to compose ERPs similar to those elicited by /asu/. Interestingly, the optimized weights of /s/ were significantly lower for Japanese than for native-English listeners, suggesting that the trace of the ACC associated with /s/ was less visible in the actual ERP response to /asu/ for Japanese native listeners. These results may together suggest that Japanese and English native individuals process the /s/ in /asu/ differently and that the ACC is sensitive to language-specific perceptual categories.

Index Terms: Auditory Change Complex (ACC), P1-N1-P2 complex, fricated consonant, cross-linguistic study

1. Introduction

When hearing speech sounds, individuals apply different perceptual strategies depending on their own linguistic background [1, 2, 3]. Studying patterns of such language-specific listening has contributed considerably to our understanding of speech processing, e.g., how our native phonetic categories are formed [4, 5, 6]. The current study applied an explorative approach to elucidate a cross-linguistic difference in speech perception: we studied how cortical event related potentials (ERPs) to a disyllabic sequence can be estimated from ERPs to each phonetic component, and how this compares with the data collected from individuals from two different linguistic backgrounds. The success of this approach would offer more insight into a potential difference in ERP responses elicited by /asu/ between Japanese and English native listeners. We investigated the following two research questions: First, whether there is a difference in ERP responses elicited by /asu/ between Japanese and English native listeners. Second, whether the composition approach could offer more insight into a potential difference in ERP responses elicited by /asu/. Figure 1 presents a schematic diagram of this approach. We combined the P1-N1-P2 responses to each phonetic component (/a/, /s/, /u/) in order to compose ERPs similar to those elicited by /asu/. In order to minimize the difference between the composed and the actual ERPs, different weights are assigned to each of the P1-N1-P2 complexes.

If the perceptual importance of the onsets of /a/, /s/, /u/ are different between Japanese than English listeners, they should show different ERP responses in the time window of 0-180ms after each phonetic component was presented. Furthermore, the patterns of the weights of /a/, /s/, and /u/ may be different. For example, if the onset of /s/ is perceptually less important for Japanese native speakers, the weights for /s/ are expected to be lower for Japanese than English listeners.
were in total 11 blocks per measurement, where each stimulus was repeated 27 times within a block. Each block lasted about 5 minutes. Within each block, the order of stimuli presentation was randomized. The sessions lasted approximately 2.5 hours including short breaks in between. EEG was measured at a sample rate of 2048 Hz using a BioSemi Active Two amplifier (Biosemi, B.V., Amsterdam, The Netherlands) with 32 Ag/AgCl electrodes according to the international 10/20 system with 6 external electrodes (2 mastoids and 4 EOG). The recording took place in a sound attenuated cabin. During measurements, participants were presented with auditory stimuli with a loudness of approximately 70dB with a pair of canal earphones (Audio Technica CK400h) while watching a self-selected silent movie on a display. The EEG data were resampled to 128 Hz with a reference to the averaged mastoid leads and band-pass filtered between 1-30 Hz. Epochs with amplitudes larger than 2.5 standard deviations from their mean were discarded from the analyses. Individual grand average ERPs were calculated for each stimulus type using responses at electrode site Cz, where the largest responses were observed. Additionally, averaged waveforms for each language group were calculated.

2.2. Stimuli

We carried out two EEG recording sessions. For the first session, a continuum of /asu/ - /assu/ was constructed with different durations of /s/, ranging from 60-240ms with steps of 30ms. This range roughly corresponds with the durational difference between fricative singleton (/s/) and geminate (/ss) consonants. The duration of /u/ and /au/ were 63ms and 90ms, respectively. The pitch accent patterns of two vowels were high-low. Additionally, the following four filler stimuli were constructed: /asu/ (/a/=93ms, /s/=240ms, /u/=90ms), /asu/ with different average loudness levels (+7.5dB/+15dB, /s/=240ms), and /aku/ (/a/=63ms, silent = 200ms, /k/=40ms, and /u/=90ms). For the second session, the five component sounds that constituted /asu/ were separately constructed: /a/=63ms, /s/=60, 150 and 240ms (short, medium and long, respectively), and /u/=90ms. These three durations were used instead of seven in order to record responses for the following six filler composite elements: /as/ (/a/=63ms + /s/=60, 150, and 240ms), /su/ (/s/=60, 150, and 240ms + /u/=90ms). Filler stimuli were included in order to investigate the effect of various speech parameter changes. However, the current paper focuses on responses to the non-filler sounds. A female Japanese native speaker recorded the materials (/asu/). Speech materials were synthesized using Praat [13]. The first and the last 5ms of each sound component were ramped to avoid onset transients.

2.3. Electrophysiological measurements

Both EEG recording sessions included 11 types of stimuli (the 1st session: 7 versions of /asu/ and 4 filler stimuli, 2nd session: 5 components and 6 components). All stimuli were presented 297 times with an inter stimulus interval (ISI) of 900ms. There were in total 11 blocks per measurement, where each stimulus repeated 27 times within a block. Each block lasted about 5 minutes. Within each block, the order of stimuli presentation was randomized. The sessions lasted approximately 2.5 hours including short breaks in between. EEG was measured at a sample rate of 2048 Hz using a BioSemi Active Two amplifier (Biosemi, B.V., Amsterdam, The Netherlands) with 32 Ag/AgCl electrodes according to the international 10/20 system with 6 external electrodes (2 mastoids and 4 EOG). The recording took place in a sound attenuated cabin. During measurements, participants were presented with auditory stimuli with a loudness of approximately 70dB with a pair of canal earphones (Audio Technica CK400h) while watching a self-selected silent movie on a display. The EEG data were resampled to 128 Hz with a reference to the averaged mastoid leads and band-pass filtered between 1-30 Hz. Epochs with amplitudes larger than 2.5 standard deviations from their mean were discarded from the analyses. Individual grand average ERPs were calculated for each stimulus type using responses at electrode site Cz, where the largest responses were observed. Additionally, averaged waveforms for each language group were calculated.

2.4. Composing ERPs

The ERPs elicited by the following 5 phonemes were used for the composition: /a/=63ms, /s/=60, 150, 240ms, and /u/=90ms. The weighted sum of these components was calculated using a least squares fit for estimating the best possible composition of ERPs for the three versions of /asu/ (short, medium and long). We used the following model:

\[
\mathbf{a}_{\text{asdu}}(t) = \mathbf{w}_{\text{asdu}} \mathbf{a}_{\text{d}}(t) \times \mathbf{w}_{\text{ud}} \mathbf{s}_{\text{u}}(t) \]

where the only unknowns are the weights expressed in the variables \(w_{\text{asdu}}\), \(w_{\text{ud}}\), \(w_{\text{d}}\) for /a/, /s/ and /u/, respectively (d denotes duration of s:60, 150, 240ms). The \(a_{\text{asdu}}(t)\) represents the composite ERP and \(a_{\text{d}}(t)\), \(s_{\text{u}}(t)\) and \(u_{\text{d}}(t)\) represent the ERPs of the individual components. We chose nonnegative least squares (nonnegative LS, expressed by the constraint \(w_{\text{asdu}}\), \(w_{\text{ud}}\), \(w_{\text{d}}\geq 0\)) for the determination of the weights to avoid any spurious solutions. We opted for this because it incorporates the assumption that the P1-N1-P2s do not invert their positive and negative signs (i.e. the P1 amplitude is positive and the P2 amplitude is negative).

<table>
<thead>
<tr>
<th>Table 1. Predicted peak timing (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration of /s/ (ms)</td>
</tr>
<tr>
<td>/a/</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>/s/</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>/u/</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

2.1. Participants

Five Japanese native speakers and six native-English speakers reporting normal hearing took part in the experiment. Both Japanese and English native listeners were students of the University of Kumamoto. All English native listeners were visiting students who had lived in Japan for approximately 8 months. All participants provided informed consent.

Figure 1: Schematic diagram of the auditory ERP composition from P1-N1-P2 complexes of speech component. The w denotes weights.
always expected to represent the positive deflection). The goodness of fit is indicated by the residual of the following:

\[ \sqrt{\frac{1}{T} \sum_{i=1}^{T} (x(t) - y(t))^2} \]  

(2)

where \( x(t) \) represents the actual ERP elicited by /asu/, \( y(t) \) represents the composed ERP and T represents the total number of samples of the ERP.

### 3. Results

#### 3.1. ERPs of /asu/ with different /s/ durations

Figure 2 presents group mean waveforms (Cz) elicited by /asu/ with different durations of /s/ ranging from 60-240ms in steps of 30ms. Japanese and English ERPs show consistent amplitude differences, especially at the time point where P1-N1-P2 complex associated with /a/ is expected (80-180ms). A 3-way repeated measures ANOVA was performed on peak amplitudes of P1-N1-P2 associated with /a/ with Group (JP/EN) as a between-subjects factor, and Peak (P1/N1/P2 associated with /a/) and Stimuli (/asu/ with 7 /s/ durations) as within-subjects factors. It indicated significant main effects of Peak (\( F(2,18)=16.8, p<.001 \)) and Stimuli (\( F(6,54)=2.4, p<.01 \)) as well as significant interactions between Peak and Group (\( F(2,18)=4.6, p<.05 \)). Simple effect analyses indicated that Japanese P2 (around 180ms) was significantly more positive than P1 and N1 (p<.05) whereas none of English peaks differed significantly. These confirmed that Japanese and English data exhibited different P1-N1-P2 patterns associated with /a/. Interestingly, the Japanese ERP pattern is more similar to the conventional P1-N1-P2 complex than the English one. English data seem to exhibit extra peaks in between 80-180ms time window. A trend was also observed in which English listeners had larger P1 (p<.09) and N1 amplitudes (p<.07) than for Japanese listeners. Simple effect analyses with regard to the significant main effect of Stimuli indicated that amplitude of s=60ms condition was significantly larger than the rest but except for 240ms, and that of s=90ms was larger than s=150ms.

#### 3.2. ERPs of component sounds

Figure 3 presents group mean waveforms (Cz) elicited by /a/, /s/ (short, medium and long) and /u/ for the two groups. Both groups showed clear P1-N1-P2 complexes for most sounds. A 3-way repeated measures ANOVA with Group (JP/EN) as a between-subjects factor and Component (/a/, short /s/, medium /s/, long /s/, /u/) and Peak (P1/N1/P2) as within-subjects factors with Group (JP/EN) as a between-subjects factor indicated strong significant main effects of Component (\( F(4,36)=9.8, p<.001 \)) and Peak (\( F(2,18)=37.9, p<.001 \)). There were significant interactions between Peak and Group (\( F(2,18)=10.4, p<.001 \)) as well as Component and Peak (\( F(2,18)=14.2, p<.001 \)). Further analysis of group differences indicated the following significant effects: Amplitudes of three peaks (P1-N1-P2) in all the Japanese data were significantly different (p<.01), whereas none of English peaks differed significantly.

#### 3.3. Composing auditory ERPs

Figure 4 shows the actual ERPs elicited by /asu/ with three different durations of /s/ (short, medium and long, solid line) as well as ERPs composed using a weighted simulation (dotted line). The goodness of fit measure was analyzed by a 2-way repeated measures ANOVA with Group (JP/EN) as a between-subjects factor and Stimuli (/asu/ with short, medium, long /s/) as a within-subjects factor. It indicated a non-significant main effect of Group (\( F(1,9)=.11, n.s. \)) and a significant main effect of Stimuli (\( F(2,18)=13.3, p<.001 \)) with no significant interactions (\( F(2,18)=3.31, n.s. \)). Further analysis indicated that the fit was significantly better for the medium /asu/ than for the short duration /asu/ (p<.01). Currently, we do not have good explanation for this effect. However, as such, this analysis indicated that the accuracy of composition was equivalent for both groups because the main effect of Group was not significant.
which may have led to the clear P1-N1-P2 complex of /a/. In ACC associated with /s/ may be absent in the Japanese data, presence of ACCs elicited by /s/ in the recorded ERPs: The P2 complex. This difference may be due to the lack or the same time window, in contrast to the conventional P1-N1-P2 complex. These differences were more similar to the conventional P1-N1-P2 complex than were more similar to the conventional P1-N1-P2 complex as with /s/. The Japanese ERP patterns within that time window were more similar to the conventional P1-N1-P2 complex than were more similar to the conventional P1-N1-P2 complex. This difference may be due to the lack or /u/ (p<.01). Although there was no interaction between Component and Group, we performed an additional simple effect analysis in order to test the predicted difference between the weights of /s/ by the two groups. It indicated that weights for /a/ were significantly higher than that for /u/ (p<.01). Although there was no interaction between Component and Group, we performed an additional simple effect analysis in order to test the predicted difference between the weights of /s/ by the two groups. It indicated that weights for /a/ were significantly lower for Japanese than for English native listeners (p<.05). Further analyses indicated trends that Japanese native listeners showed higher weights for /a/ than /u/ (p=.06) and that English native listeners showed significantly higher weights for /u/ than /a/ (p=.07).

ERPs for Japanese and English native speakers. Figure 5: Weights used to compose the best estimated ERPs for Japanese and English native speakers.

4. Discussion

Two series of EEG measurements indicated that Japanese and English native listeners show differing amplitude patterns in the ERPs elicited by a disyllabic sound (/asu/). Significant differences were found in component amplitudes between 80 and 180ms following stimulus onset. This corresponds with the time window of ERPs associated with /a/, and partially with /s/. The Japanese ERP patterns within that time window were more similar to the conventional P1-N1-P2 complex than the English ones. English data exhibited additional peaks in the same time window, in contrast to the conventional P1-N1-P2 complex. This difference may be due to the lack or presence of ACCs elicited by /s/ in the recorded ERPs. The ACC associated with /s/ may be absent in the Japanese data, which may have led to the clear P1-N1-P2 complex of /a/. In contrast, the ACC may well be present in the English data, leading to additional positive peaks between 80-180ms, as well as decreased P2 amplitudes of /a/, due to a second N1 response to /s/ that partially overlaps with the P2 response elicited by /a/. Our ERP composition approach further suggested that Japanese ERP patterns can be predicted well when assigning smaller weights to the P1-N1-P2 complex of /s/ than for English native listeners. The overall results were in line with our hypothesis that the onset of /s/ is processed differently by Japanese and non-Japanese native listeners. More specifically, it was harder to find a trace of the ACC associated with /s/ in the actual ERP responses to /asu/ measured from Japanese native listeners. This aligns well with the previous finding that Japanese native speakers falsely perceive a silent duration during ongoing friction. Why this is the case remains an interesting but open question.

The current study replicated the previous finding [10] that it is possible to make use of the P1-N1-P2 complexes measured in response to individual phonetic components (/a/, /s/, /u/) in order to compose ERP responses to /asu/ that are similar to the actual ones. Instead of a simple averaging of the individual components used in the previous study, we used a non-negative LS fitting procedure in order to estimate the optimal combination of the individual components. This made it possible to gain insight into how different phonetic components contribute to the formation of ERP responses to disyllabic sounds. It is also possible to make a more precise estimation by taking the variances of individual components into account (e.g., by means of non-negative weighted LS). Similarly, it would be interesting to compare different methods for estimating the goodness of fit to evaluate the results, e.g., the R² test, which provides a measure of the extent to which the model generalizes.

It is important to note that Japanese and English native listeners showed clear differences in cortical responses, even though all English listeners had extensive exposure to Japanese speech. This suggests that it takes a rather long time before one acquires skills related to speech processing in a second language. However, this is not too surprising, as it has been repeatedly demonstrated that adults need a long time to achieve fluency in various skills of second language speech processing [15, 16]. In addition, it has been previously demonstrated that the distinction of fricative geminate and singleton consonants embedded in /asu/ is hard to learn for non-native Japanese listeners [11, 12, 16].

Finally, the results of this study indicated that top-down information processing influences the elicitation of the ACCs. If indeed the ACC is sensitive to acoustic changes that correspond with language specific phonetic categories, this offers new opportunities for studying language-specific listening; it would be interesting to investigate whether it is possible to modulate the ACCs as one learns to perceive new speech segmentation strategies and how long it takes before such modulation occurs. This awaits further research.

5. Acknowledgment

This research was supported by a Grant-in-Aid for Scientific Research (B) 21330169 under the project leadership of Shuji Mori (Japan Society for the Promotion of Science). Authors are thankful for useful comments from Alex Brandmeyer.
6. References


