Hemispheric laterality of the motor cortex involved in imitation of lip movement

Akira Toyomura¹,²,*  Tetsunoshin Fujii³,⁴ and Shinya Kuriki⁵

¹Graduate School of Health Sciences, Gunma University, 3—39—22 Showa-machi, Maebashi, 371—8514 Japan
²Research and Education Center for Brain Science, Hokkaido University, Kita 15, Nishi 7, Kita-ku, Sapporo, 060—8638 Japan
³Graduate School of Letters, Hokkaido University, Kita 10, Nishi 7, Kita-ku, Sapporo, 060—0810 Japan
⁴Center for Language Studies, Otaru University of Commerce, Midori 3—5—21, Otaru, 047—8501 Japan
⁵Faculty of Health Sciences, Hokkaido University, Kita 12, Nishi 5, Kita-ku, Sapporo, 060—0812 Japan

(Received 6 April 2018, Accepted for publication 3 July 2018)

Keywords: Imitation, Visual speech, Lip movement, MRI
PACS number: 43.64.Vm  [doi:10.1250/ast.39.432]

1. Introduction

Imitation is a crucial function for learning skills of speech motor production. Infants acquire language by observing and imitating their parents’ mouth movements and sounds. Imitation ability is also necessary for adults for acquiring a second language. In daily life, we unconsciously imitate aspects of the interlocutor’s speech, such as vocal intensity, fundamental frequency, or syntactic structure, as well as behaviors, such as posture, facial expressions, and breathing [1—4].

As the auditory target serves as an explicit target for the pronunciation of the new language, the somatosensory target is visually presented as a model of mouth movement. This target is presented as visual speech, i.e., the talker’s mouth, face, and head movement during articulation [5], which increases the intelligibility of speech [6]. Information obtained from visual speech is crucial in language acquisition in congenitally hearing-impaired persons. Adults who stutter enhance their speech fluency when watching visual speech representing the corresponding sounds [7]. Behavioral research has shown that watching a talking mother’s mouth at the age of 6 months boosts the child’s language acquisition score at the age of 24 months [8].

Recent neuroimaging studies suggest that the auditory cortex, supramarginal gyrus, and Wernicke’s area are involved in the neural substrates of speech imitation [1,9,10]. Ability of speech imitation is associated with activity in the left supramarginal gyrus and postcentral areas [10] and with grey matter volume in the left premotor area, Broca’s area, and the left inferior parietal lobe [9]. Observing visual speech activates brain regions that are typically activated during speech production [e.g., 11], suggesting that visual speech perception is also involved in speech production.

However, it still remains unclear how imitation of visual speech of different familiarity levels is represented in the brain. It is possible that we use different neural processes to reproduce familiar models, which already exist as somatosensory targets, and unfamiliar models during imitation. This functional magnetic resonance imaging (fMRI) study was designed to gain insights into the neural mechanism of speech imitation in the somatosensory domain. Participants were instructed to imitate three types of silent movie clips depicting mouth movement: 1) a somatosensory target already possessed by the speaker as a model (i.e. in their own native language), 2) a somatosensory target that the speaker does not possess as a model (i.e. in a language not spoken by the participant), and 3) nonsense mouth movement for which the speaker lacks a model.

2. Methods

2.1. Participants

Sixteen male participants (age range: 19—34 years, mean age: 23.8 years, SD = 3.5; all right-handed, native Japanese speakers) participated in this study. The data analysis described in Sect. 2.3 revealed that the head movement of one participant during the experiment was substantial. Thus, we excluded this participant’s data from the results. All participants signed an informed consent form approved by the Committee of Medical Ethics, Graduate School of Medicine, Hokkaido University.

2.2. Experimental procedure

Silent 3-s videos that showed lip movement were prepared in advance. There were three types of videos showed lip movement related to their first language (Japanese), an unfamiliar language (Swahili), or nonsense mouth movement. A single male Japanese speaker appeared in all of the videos. In each silent video, the speaker pronounced a short word, such as a greeting (‘Hello’ in Japanese or Swahili). Prior to making them, he practiced pronunciation of each Swahili word. In the videos showing the nonsense lip movement, he merely moved his mouth randomly. Snapshots of the videos of the conditions are shown in Fig. 1(a). Ten videos for each imitation condition (total of 30) were prepared and were randomly presented during the experiment.

Each participant lay face-up on the MRI bed, and the head was fixed with several belts and small cushions. The videos were presented on the screen in the MRI scanner room, and participants were required to imitate the speech in the videos soon after presentation (Fig. 1(b)). In the rest condition, participants merely looked at a fixation point.

To minimize any movement-related artifacts on MR images, a sparse image acquisition protocol with a 6.5-s delay

*e-mail: ak.toyomura@gmail.com
was used. During the non-scanning periods, the 3-s video was presented first, followed by the participant’s imitation of the movement (Fig. 1(b)). A total of four experimental runs were performed. Each run included 12 blocks, in which two blocks were assigned to each of the imitation task conditions (first language, unfamiliar language, and nonsense movement), which were randomly presented. A single task-block included three MR image volumes of the same condition. The remaining six blocks were assigned to the rest condition, including two MR image volumes.

A whole-body 1.5-T Signa Echospeed scanner (General Electric, Boston, MA, USA) was used to acquire MR images with blood oxygen level-dependent (BOLD) contrast. Seventeen parallel axial slices (thickness: 5 mm; gap: 2 mm) were acquired using echo planar imaging (64 × 64 matrix; field-of-view: 24 × 24 cm; repetition time [TR]: 3 s; echo time [TE]: 40 ms; flip angle: 90°), using a sparse image acquisition protocol with a 6.5-s delay. T1-weighted images (corresponding to 17 axial slices; thickness: 5 mm; gap: 2 mm; 256 × 256 matrix; field-of-view: 24 × 24 cm; TR: 500 ms; TE: 14 ms) were acquired from all participants as an anatomical reference for the functional images.

After completing the fMRI experiment, participants were asked about the videos presented during the experiment. All of the videos were randomly presented again, and each participant was asked whether they felt as 1) Japanese, 2) foreign language, or language) and ‘ ‘1’ ‘ defined that they did not feel the videos as Japanese, a one-way repeated measures ANOVA with stimuli (familiar, unfamiliar languages, and nonsense lip movement) showed significant differences between stimuli ($p < 0.001$) (Fig. 2(a)). The same significance ($p < 0.001$) was obtained from the ANOVA analyses in other questionnaires (Fig. 2(b)(c)). The results of post hoc pairwise comparisons using t-test with Bonferroni correction were shown in Fig. 2. These results suggest that the presented stimuli were perceived by the participants as the study intended; i.e., the video clips of familiar words were perceived more strongly as Japanese, and the clips of unfamiliar words were perceived more strongly as a foreign language than those of other stimuli. In all comparisons, the scores of the video clips of nonsense mouth movement were the lowest, suggesting that these stimuli were not perceived as real language.

Imaging data showed that wide brain areas, including bilateral motor and auditory areas, were activated in all conditions under the $p < 0.05$ FDR correction at the voxel level, with the extent threshold of 10 voxels (Fig. 3(a), Table 1). The significant activations in areas related to speech suggest that the participants performed mouth movement during the experiment. Under the stricter threshold of $p < 0.05$ FWE correction, however, distinct hemispheric laterality was seen in the speech motor cortex (Fig. 3(b)); Imitation of familiar language induced significant activity only in the right middle frontal gyrus (1 in Fig. 3(b)). In contrast, the imitation of unfamiliar language induced significant activity in the left middle frontal gyrus (2) and the left precentral gyrus (3).
Although the imitation of nonsense lip movement induced activity in wider areas under the FDR-corrected threshold, no significant activity was detected under the FWE-corrected threshold. Comparison between conditions did not induce any significant activity under the FDR- or FWE-corrected thresholds.

4. Discussion

Hemispheric laterality in the motor cortex during imitation of familiar and unfamiliar languages was observed. The experimental design of this study was related to neuroimaging studies regarding first (L1) and second language (L2) acquisition. Berken et al. (2015) investigated the neural activation of bilingual individuals during a sentence reading task and showed that the left premotor area and supplementary motor area were more strongly activated when the bilinguals spoke in L2 than in L1. These activations were only observed in participants who had learned their L2 after first acquiring L1 in their learning phase [12]. Similarly, Rüschemeyer et al. (2006) compared the neural activation of native and non-native reading of sentences and found greater activation in the supplementary motor area and the left pre- and postcentral gyrus in non-native speakers than in native speakers [13]. The findings of these and the current studies suggest that the left motor and supplementary motor areas are involved in speech imitation of relatively unfamiliar languages.

Alternatively, because the participants in this experiment were all right-handed native-Japanese speakers, their left speech motor areas may have been optimally tuned to produce speech in their native language, thereby eliciting minimal activity in the left motor cortex. In contrast, imitation of unfamiliar language was more complex than that of familiar language, requiring recruitment of more complex neural

![Fig. 3 Results of whole brain analysis for each imitation condition in the two statistical thresholds. (a) $p < 0.05$ FDR correction at voxel level with the extent threshold of 10 voxels. (b) $p < 0.05$ FWE correction at voxel level with the extent threshold of 10 voxels. Imitation of nonsense movement did not show significant activity under the FWE-corrected threshold. Each number corresponds to the number shown in Table 1.](image)

Table 1 Regions that showed significant activation in each condition.

<table>
<thead>
<tr>
<th>Region</th>
<th>Side</th>
<th>(x, y, z)</th>
<th>Cluster size</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Familiar language</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle frontal gyrus</td>
<td>R</td>
<td>52 6 44</td>
<td>1112</td>
<td>5.20</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>L</td>
<td>−50 0 40</td>
<td>2131</td>
<td>4.60</td>
</tr>
<tr>
<td>Medial Frontal Gyrus</td>
<td>L</td>
<td>−8 0 64</td>
<td>270</td>
<td>4.46</td>
</tr>
<tr>
<td>Putamen</td>
<td>L</td>
<td>−24 2 −4</td>
<td>434</td>
<td>4.35</td>
</tr>
<tr>
<td>Putamen</td>
<td>R</td>
<td>22 10 6</td>
<td>212</td>
<td>3.94</td>
</tr>
<tr>
<td>Postcentral Gyrus</td>
<td>R</td>
<td>56 −30 40</td>
<td>21</td>
<td>3.64</td>
</tr>
<tr>
<td><strong>Unfamiliar language</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Medial Frontal Gyrus</td>
<td>L</td>
<td>−6 −2 60</td>
<td>974</td>
<td>5.86</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>L</td>
<td>−56 0 34</td>
<td>5637</td>
<td>5.33</td>
</tr>
<tr>
<td>Middle Frontal Gyrus</td>
<td>R</td>
<td>54 6 40</td>
<td>1771</td>
<td>5.13</td>
</tr>
<tr>
<td>Putamen</td>
<td>R</td>
<td>30 2 −4</td>
<td>546</td>
<td>4.46</td>
</tr>
<tr>
<td>Superior Temporal Gyrus</td>
<td>R</td>
<td>62 −34 18</td>
<td>335</td>
<td>3.81</td>
</tr>
<tr>
<td>Inferior Temporal Gyrus</td>
<td>R</td>
<td>50 −66 9</td>
<td>91</td>
<td>3.40</td>
</tr>
<tr>
<td>Middle Frontal Gyrus</td>
<td>L</td>
<td>−38 26 22</td>
<td>64</td>
<td>3.15</td>
</tr>
<tr>
<td>Superior Frontal Gyrus</td>
<td>R</td>
<td>22 −10 68</td>
<td>10</td>
<td>2.94</td>
</tr>
<tr>
<td><strong>Nonsense movement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postcentral Gyrus</td>
<td>L</td>
<td>−56 −18 34</td>
<td>15752</td>
<td>4.96</td>
</tr>
<tr>
<td>Precentral Gyrus</td>
<td>R</td>
<td>60 10 12</td>
<td>4.92</td>
<td></td>
</tr>
<tr>
<td>Inferior Frontal Gyrus</td>
<td>L</td>
<td>−46 40 10</td>
<td>63</td>
<td>3.35</td>
</tr>
<tr>
<td>Inferior Frontal Gyrus</td>
<td>R</td>
<td>42 36 4</td>
<td>17</td>
<td>3.10</td>
</tr>
</tbody>
</table>

Coordinates (x, y, z) are given in MNI space. L: Left; R: Right. $p < 0.05$ FDR correction at voxel level with the extent threshold of 10 voxels. $p < 0.05$ FWE correction at voxel level with the extent threshold of 10 voxels. Each number corresponds to the number shown in Fig. 3. The maximum peak of each cluster is reported except for the right precentral gyrus (*) in the imitation of nonsense movement condition.

Although the imitation of nonsense lip movement induced activity in wider areas under the FDR-corrected threshold, no significant activity was detected under the FWE-corrected threshold. Comparison between conditions did not induce any significant activity under the FDR- or FWE-corrected thresholds.
activity patterns in the motor areas, which, in turn, induced greater activation. This may also be because the supplementary motor area is related to learning new motor skills [14]. The imitation of nonsense lip movement, however, did not elicit any significant activity, suggesting that the motor cortex areas observed in this study are tuned to contribute only to meaningful speech movement.

The difference might also be associated with differences in articulation of the words. In Swahili, there are five vowels (a, e, i, o, and u) which are similar in pronunciation to their Japanese counterparts. The “u” in Swahili is pronounced by making the lips more round than is required to pronounce “u” in Japanese. The consonants in Swahili are also similar to those used of Roman letters. However, because a Japanese native speaker pronounced all of the words in the video clips presented as targets for imitation, the speaker might not necessarily have pronounced the Swahili words with correct articulations but have pronounced them by combining Japanese vowel/consonant articulations. Therefore, we suggest that the difference in activity might also reflect the imitation of different combination of articulations rather than differences in articulation in both languages. The use of stimuli of unfamiliar language pronounced by a non-native speaker is a limitation of this study, and this point should be considered in further studies.

Activation was observed in wider areas under the liberal statistical threshold using an FDR correction than under the FWE-corrected threshold. The observed pattern was similar to that of brain regions associated with speech found in previous studies (e.g., [15]). This may be because the imitation task is involved in various aspects of speech processing, such as planning of movement, articulation, and sensorimotor feedback.

The speech imitation of unfamiliar language used in this study may be comparable to situations in which people who stutter imitate fluent speech in their native languages. Speech fluency is temporarily improved when speaking while watching visual speech representing the sound [7]. They also show temporary fluency in shadowing or choral reading with others [16]. These fluency-enhancing methods are similar to speech imitation, because the fluent model is explicitly presented for speech production. Therefore, people who stutter may be able to speak fluently under these conditions by using neural networks that differ from the usual networks. Moreover, speaking with different fundamental frequency than usual enhances fluency, possibly suggesting the effect of imitating the voice of another person.

5. Conclusion

Although neural systems for the imitation of visual speech are not yet fully understood, this preliminary study suggests the existence of hemispheric laterality in the speech motor area during speech imitation. Further neuroscience studies of speech imitation are necessary for understanding L1 and L2 learning and for the application of therapy methods for people who stutter.

Acknowledgments

The authors would like to thank Dr. Masahide Yuasa for helpful discussion. We are also grateful to Mr. Shuhei Bannae for providing technical assistance. This study was supported by JSPS KAKENHI (Grant Numbers 21700232, 24680025, 24650138 and 16K00366) and the Strategic Research Project (07H012) for private universities. There are no conflicts of interest to be declared.

References