Title: Motor cortex compensates for lack of sensory and motor experience during auditory speech perception

Authors: Judith Schmitz[a], Eleonora Bartoli[b], Laura Maffongelli[c], Luciano Fadiga[d,e], Nuria Sebastian-Galles[a], Alessandro D'Ausilio[d,e]

Affiliations:

[a] Center for Brain and Cognition, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain

[b] Neurosurgery Department, University of Texas Health Science Center at Houston, 6431 Fannin Street, 77030 Houston, TX

[c] Department of Psychology, University of Zürich, Binzmuehlestrasse 14, Box 21, CH-8050 Zürich, Switzerland

[d] Dipartimento di Scienze Biomediche e Chirurgico Specialistiche, Section of Human Physiology, Università di Ferrara, via Fossato di Mortara 17/19, 44121 Ferrara, Italy

[e] Center for Translational Neurophysiology for Speech and Communication, Fondazione Istituto Italiano di Tecnologia, via Fossato di Mortara 17/19, 44121 Ferrara, Italy

Corresponding authors:

Judith Schmitz, Center for Brain and Cognition, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain, E-mail: judith-schmitz@gmx.net

Alessandro D'Ausilio, Dipartimento di Scienze Biomediche e Chirurgico Specialistiche, Section of Human Physiology, Università di Ferrara & Center for Translational Neurophysiology for Speech and Communication, Fondazione Istituto Italiano di Tecnologia, via Fossato di Mortara 17/19, 44121 Ferrara, Italy, E-mail: alessandro.dausilio@unife.it

Abstract

Listening to speech has been shown to activate motor regions, as measured by corticobulbar excitability. In this experiment, we explored if motor regions are also recruited during listening to non-native speech, for which we lack both sensory and motor experience. By administering Transcranial Magnetic Stimulation (TMS) over the left motor cortex we recorded corticobulbar excitability of the lip muscles when Italian participants listened to native-like and non-native German vowels. Results showed that lip corticobulbar excitability increased for a combination of lip use during articulation and non-nativeness of the vowels. Lip corticobulbar excitability was further related to measures obtained in perception and production tasks showing a negative relationship with nativeness ratings and a positive relationship with the uncertainty of lip movement during production of the vowels. These results suggest an active and compensatory role of the motor system during listening to perceptually/articulatory unfamiliar phonemes.

Keywords: Speech perception, Speech production, Native language, Non-native language, Motor evoked potentials, Transcranial magnetic stimulation

Introduction

Listening to speech activates temporo-parietal brain regions, as well as the motor system. Activations of the motor regions, including the representation of articulatory muscles of the primary motor cortex, has been tested via corticobulbar excitability modulations (D'Ausilio, Jarmolowska, Busan, Bufalari, & Craighero, 2011; Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Murakami, Restle, & Ziemann, 2011; Nuttall, Kennedy-Higgins, Hogan, Devlin, & Adank, 2016; Rogers, Möttönen, Boyles, & Watkins, 2014; Roy, Craighero, Fabbri-Destro, & Fadiga, 2008; Sato, Buccino, Gentilucci, & Cattaneo, 2010; Sundara, Namasivayam, & Chen, 2001; Watkins, Strafella, & Paus, 2003). Furthermore, the application of (repetitive) Transcranial Magnetic Stimulation (rTMS) to the premotor (Grabski, Tremblay, Gracco, Girin, & Sato, 2013; Meister, Wilson, Deblieck, Wu, & Iacoboni, 2007; Sato, Tremblay, & Gracco, 2009) or the primary motor cortex (Bartoli et al., 2015; D'Ausilio, Bufalari, Salmas, Busan, & Fadiga, 2011; D'Ausilio et al., 2009; D'Ausilio, Bufalari, Salmas, & Fadiga, 2012; Möttönen, Dutton, & Watkins, 2013; Möttönen, van de Ven, & Watkins, 2014; Möttönen & Watkins, 2009; Rogers et al., 2014) showed that the motor system may exert a causal modulatory role in both the discrimination and identification of speech sounds. These results are usually interpreted in line with theories for which knowledge of articulatory gestures may be re-used during speech identification and discrimination tasks (Galantucci, Fowler, & Turvey, 2006; Pickering & Garrod, 2013; Skipper, Nusbaum, & Small, 2006; Pulvermüller & Fadiga, 2010).

While the studies described so far show motor activities during the analysis of native speech, less is known about non-native phonemes, which are biomechanically possible oral configurations for which we lack both sensory and motor experience. Neuroimaging studies show that, besides temporo-parietal regions, part of the motor system (premotor cortex) is activated during both listening and production of native phonemes (Wilson, Saygin, Sereno, & Iacoboni, 2004) and listening to non-native phonemes (Wilson & Iacoboni, 2006). Similar results were observed with African clicks, that are judged as non-speech sounds in other languages (Agnew, McGettigan, & Scott, 2011). Regarding the direct measurement of the motor system activity, a TMS study by Swaminathan et al. (2013) measured lips corticobulbar excitability while native and second language learners of English were viewing known speech (English), unknown speech (Hebrew), non-speech movements (gurns) or a static face. Viewing silent faces producing known speech induced larger cortico-bulbar excitability.

Speech perception however, can greatly be affected by the integration of both auditory and visual information (MGurk, McDonald, 1976). More importantly, visual cues are characterized by less discriminative value than the auditory ones, and such a difference may also interact with language proficiency. In this regard, discriminating if two sentences are spoken in the same language, in the visual modality, can only be performed if at least one of the languages is either native (Soto-Faraco et al., 2007) or a high proficiency has been achieved with it (Swaminathan et al., 2013; Weikum et al., 2013). Therefore, measuring motor activities in native versus non-native speech, by using visual speech material, may introduce a partial confound. In fact, it would be difficult to understand if corticobulbar modulations are driven by a general preference for native speech or it is due to a better discriminability of native speech when presented in the visual modality.

Here we investigated if motor activities, during auditory speech listening, depend on sensorimotor experience with the phonemes or otherwise if it maps the amount of motor recruitment present in the auditory stimuli. We tested this hypothesis by measuring lip corticobulbar excitability during passive listening to native and non-native vowels. Stimuli consisted in German vowels from which some had a native counterpart in the language of the participants (Italian), while others were non-native to the participants. Furthermore, by means of linear regression we tested if the corticobulbar excitability was related to subjective ratings regarding vowels nativeness (for each vowel), as well as participant's lip muscle electromyographic activity (EMG) during vowel production. The exploration of motor activities by measuring corticobulbar excitability let us formulate different hypotheses in this regard. In fact, corticobulbar excitability could scale for the amount of sensorimotor experience (hypothesis 1). In this case, listening to over-trained (native) speech sound should elicit stronger motor responses, in agreement with the previous report on visual speech perception (Swaminathan et al., 2013). Otherwise, corticobulbar excitability could instead be larger while listening to unfamiliar speech sounds. The lack of an acoustic-motor model for non-native speech sounds (hypothesis 2) might instead promote additional motor compensatory activities.

Methods

Participants

17 native Italian speakers (9 female, mean age 23.59±4.81 years) took part in this study after giving informed consent, according to the Declaration of Helsinki and to the recommendations of the local Ethical Committee ASL-3 ("Azienda Sanitaria Locale"- Local Health Unit, Genoa, Italy) authorizing the protocol. All participants had normal hearing, were right-handed (Oldfield, 1971) and did not report any neurological/psychiatric disease. They were Italian university students, had no professional training in phonetics and were not proficient in any foreign language whose phonological repertoire includes the non-native vowels used in the experiment, (as accessed by a language questionnaire). One participant was excluded due to technical problems during recordings. Two further par-

ticipants were removed in relation to outlier behavior in one of the tasks (i.e., the participants rated the foreign German vowel /y/ higher in nativeness with respect to the German vowel /u/, that is present in the Italian phonological repertoire). The final sample included 14 participants.

General procedure

The study consisted in a first day where we recorded the stimuli from one German speaker. After stimuli selection, we then had Italian subjects participate in the experiment. They were first asked to fill-in a language questionnaire to test their degree of exposure to the non-native vowels (self-reports about comprehension and production skills in foreign languages). The experimental session was carried out in the same day and consisted of three parts: a TMS experiment, a speech nativeness task and a speech production task. The TMS part was always run as first, while the order of the two remaining parts were counterbalanced across participants. The whole experiment lasted about 2.5 hours. All experiments were programmed using Psychtoolbox functions (Brainard, 1997; Pelli, 1997), running on MATLAB ® (The MathWorks Inc., Natick, MA).

Stimuli recording

Seven German vowels: /a/, /e/, /i/, /o/, /u/ (having a counterpart in Italian) and /ö/ and /y/ (not standard in Italian) were recorded by a male native German speaker using a microphone (AKG c1000s) and surface electrodes placed on the lower and upper right side of the lips to record the electromyography (EMG) of the *orbicularis oris* (OO) muscle. Each vowel was recorded twelve times and the best exemplar for each vowel, matching in average pitch (127 Hz) and intensity (75 dB), was selected. To equal the length for the stimuli, three-hundred ms were selected from the steady middle part of the vowel and 25 ms cosines onset and offset were applied using Praat software (Boersma & Weenink, 2010). The EMG data was band-pass filtered between 20-250 Hz and then low pass filtered with a cut-off frequency of 3 Hz. EMG for each trial was rectified and integrated, obtaining in this way a measure of the area under the curve.

The goal of the study was to investigate corticobulbar excitability in native vs. non-native speech perception. However, regional variations of Standard Italian contain important vowels variability, especially for what regards the mid-vowels (Bertinetto, Loporcaro, 2005). In order to effectively tackle this point, all analyses were limited to vowels that were clearly and consistently evaluated as native or non-native by the participants. Since it is difficult to know beforehand which vowels will be perceived as Standard Italian, all recording sessions were run including the 7 recorded vowels. Later, in the analyses of TMS data, we selected and analyzed only the vowels clearly defined as native or non-native by all the subjects.

Speech nativeness tasks

In the nativeness tasks each vowel (/a/, /e/, /i/, /o/, /u/, /ö/, /y/) was rated three times (total: 21 trials) in a random order on a visual analog scale ranging from “poco” (meaning “less”) on the left side of the screen to “tanto” (meaning “more”) on the right side of the screen. Answers were given by clicking with the mouse on any location along the continuum. Figure 1c shows the average responses produced by subjects.

The nativeness ratings were averaged for each vowel for each participant. A 1-way ANOVA with a within-subject factor “vowel” (/a/, /e/, /i/, /o/, /u/, /ö/, /y/), and as dependent variable the average nativeness rating, showed a significant main effect of vowel ($F(6,84)=36.610$, $p<0.001$). Follow-up Bonferroni corrected t-tests showed that vowels

/ö/ (21.667±3.670%) and /y/ (14.000±1.966%) were rated significantly less native compared to all other vowels (/a/: 92.578±1.777%, /e/: 62.089±4.427%, /i/: 87.156±2.336%, /o/: 59.200±4.712%, /u/: 75.044±4.330%, all p<0.01). The mid-vowels /e/ and /o/, generated the largest ambiguity and were rated less native compared to the most native rated vowel /a/ (/e/-/a/: t(14)=4.243, p=0.017, /o/-/a/: t(14)=4.568, p=0.009). Therefore, all mid vowels were excluded from further analyses (/e/, /o/ and /ö/), and we focused on the regions of the vowel space that were clearly perceived as native or non-native (/a/, /i/, /u/ and /y/). Figure 1 shows the F1-F2 formant space and the lip EMG activation for the 4 selected vowels stimuli.

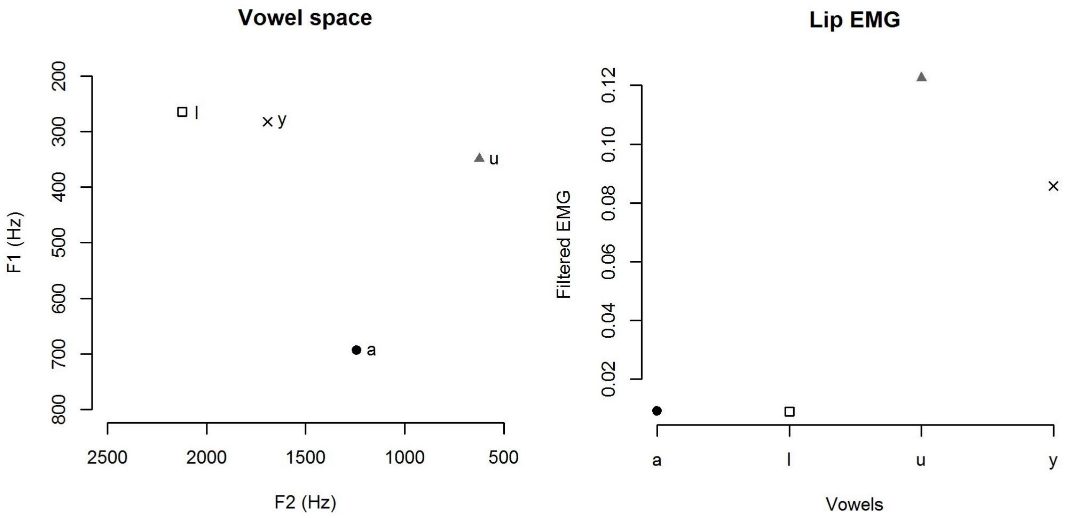

Figure 1: Stimuli. Vowel space (left) and lip EMG (right) of the stimuli.

<u>Speech production task</u>

Audio and EMG activity from the OO muscle were recorded from each participant during the speech production task. Participants produced each vowel after seeing the German grapheme (<a>, <e>, <i>, <o>, <u>, <ö>, <ü>) corresponding to the vowel (/a/, /e/, /i/, /o/, /u/, /ö/, /y/) on the screen. Previous to the production trials, participants were familiarized with the writing of the German graphemes. They were presented three times with each

German grapheme followed by the playback of each vowel. This was necessary since <ö> and <ü> symbols are not used in standard Italian. We recorded 8 repetitions of each vowel and the last five were subsequently used for the analysis. The first three example were removed to avoid fast adaptation effects and thus use only individual stable productions. Individual subject formant extraction (Figure 2a) and EMG data was processed with the same procedure described earlier (*Stimuli* section), but this time the area under the curve was also transformed within subjects into z-scores, to offer a normalized comparison between the participants (Figure 2b).

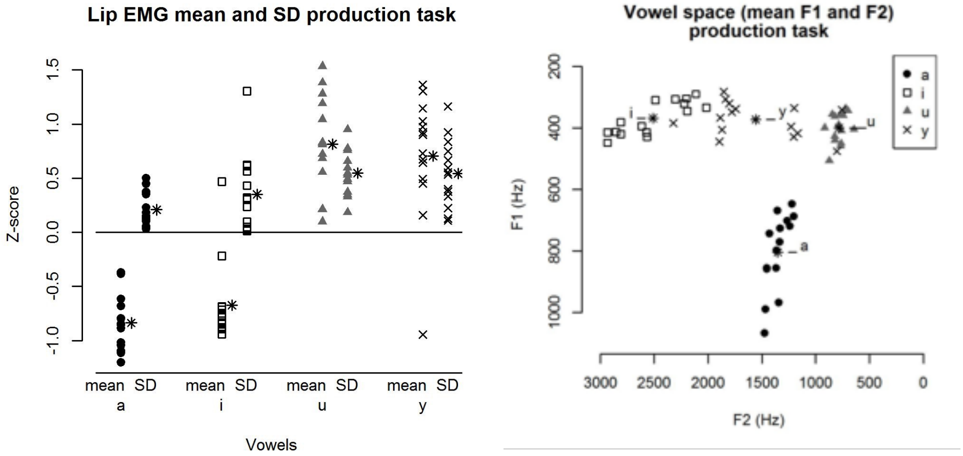

Figure 2: Production task. Left: Lip EMG mean and SD for each participant and vowel. Stars mark the mean values over all participants. Right: vowel space representing the mean value for F1 and F2 for each participant and vowel. Stars mark the mean values for all participants.

TMS experiment

Two surface EMG electrodes were placed on participants' upper and lower side of the right lip (*orbicularis oris*, OO). MEPs were recorded by means of a wireless EMG system (Aurion, ZeroWire EMG; 2KHz sampling). TMS was delivered through a figure-eight coil (70 mm) and a Magstim 200 monophasic stimulator (Magstim Co., Whitland, UK).

The TMS recording session was divided into training phase, cortical mapping and experimental part. During the training phase, participants learned how to maintain a 1.5-2 seconds sustained contraction of the lips, corresponding to 30% of the maximal EMG activity. A yellow dot appeared on the screen to signal the beginning of the contraction, followed. The disappearance of the yellow dot (1.2-1.7 sec duration) signaled the end of the trial. Lips were contracted by rounding and protruding them. Participants could see their EMG activity on the screen and received feedback from the experimenter if necessary. When a satisfactory contraction was achieved, the cortical mapping part started. The hot spot was identified during muscle contraction. Scalp position, coil orientation and intensity of stimulation were adjusted to obtain the lowest possible intensity to elicit a reliable and repeatable MEPs. The criterion was to produce a MEP of at least 200 microvolts on 5 consecutive trials. Once the criterion was met, location was marked on the scalp and the coil position was fixed by a mechanical support and was continuously monitored by the experimenter. Average TMS intensity was 48.786 (SD= 2.833).

At the beginning of each experimental trial a yellow dot appeared on the screen to signal the contraction of the lip and tongue muscle. After a random interval of 1-1.5 seconds one of the seven vowels was played. The TMS pulse was triggered 150 ms after stimulus onset. The delay between trials onset ranged between 5-5.5 seconds, allowing at least 6.35 seconds between two consecutive magnetic pulses. The presentation of auditory stimuli was pseudo-randomized so that the presentation of the same vowel was equally spread throughout the experiment and could never appear twice in a row. Each vowel type was repeated 14 times, 10 times with TMS and 4 times without TMS. TMS and no-TMS trials were randomized. The whole experiment consisted in 98 trials and it was divided in 2 blocks of 49 trials each. An additional task was added in order to keep participants engaged throughout the experiment. Randomly we asked participants if a certain native vowel, printed on the screen, was the same as the one they just listened to (one-back task). Responses were given by pressing one of two buttons on a keyboard, with their right hand. The task was repeated 12 times and equally spread during the experiment and it was never presented twice in a row.

Statistical Analyses

In the analysis of MEP size, we computed the area under the rectified curve for each trial. For each participant, single trials were removed if the MEP area exceeded 2 standard deviations from the average MEPs area or if the muscle contraction was above 2 standard deviations from average muscle contraction in a time window just prior to the TMS pulse delivery (-100 to 0 ms). On average 3.29 (SD=1.27) trials were excluded from the analyses for each subject. The time range of the MEP area was identified manually for each subject by overlaying all trials in one plot and marking the common start and end time of the MEPs. The MEPs area were standardized for each participant using z-scores and then trials corresponding to the same vowel were averaged together. The ANOVA on MEPs tested for differences in MEP size accounted by the within-subject factor “vowel” (/a/, /i/, /u/ and /y/).

In case of significant effect, a linear regression analysis was run to test relationship between the amplitude of MEPs (for each vowel and participant) and the nativeness task or the production task data. For the production data, we used mean and SD of lips EMG data. For the nativeness task we used mean nativeness value. Due to having several measures for each subject, subjects were included in the model as a random effect while the MEPs and measures from nativeness and production tasks were included as fixed factors.

For the analysis, all data were pre-processed using MATLAB ® (The MathWorks Inc., Natick, MA). Statistical analyses (ANOVAs, t-tests and linear regressions) were performed by means of R statistical software (R Core Team, 2013). All t-tests were corrected using the Bonferroni correction.

Results

Corticobulbar excitability

The ANOVA for the lip MEPs showed a significant main effect of "vowel" (F(3,39)=4.074, p=0.0131). Follow-up Bonferroni corrected t-tests revealed a difference in MEP size between the vowels /a/ and /y/ (t(13)=-3.348, p=0.031, see Figure 3). To further explore if the significant differences for the lip MEPs were related to the listening of the different vowels, linear regression analyses were performed with the lip MEPs and the scores obtained in the nativeness and production tasks.

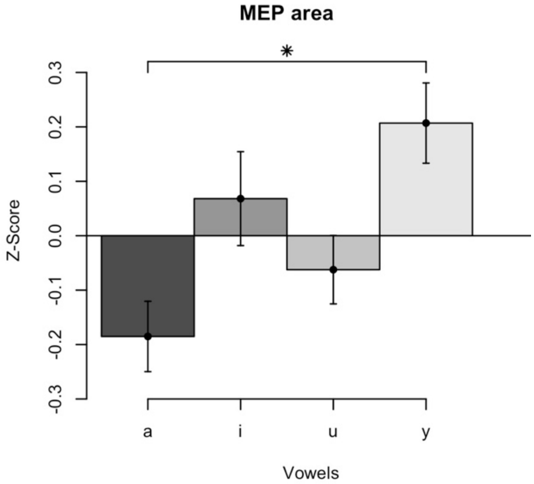

Figure 3: TMS results. Lips MEP area +/- SEM for the German vowels /a/, /i/, /u/ (having a counterpart in Italian) and /y/ (non-native).

The linear regression analyses for the lip MEPs and the perception data revealed a significant negative relationship between the lip MEPs and the nativeness ratings (β=–0.003, t(41)=-2.950, p=0.005, see Figure 4a). Thus, higher MEP size was associated to lower nativeness ratings. A positive relationship was present between the lip MEPs and the SD of lip EMG only (β=0.430, t(41)=3.401, p=0.002, see Figure 4b). Larger MEPs were thus associated with larger EMG variability during speech production. Mean EMG did not yield to significant results (all p>0.05).

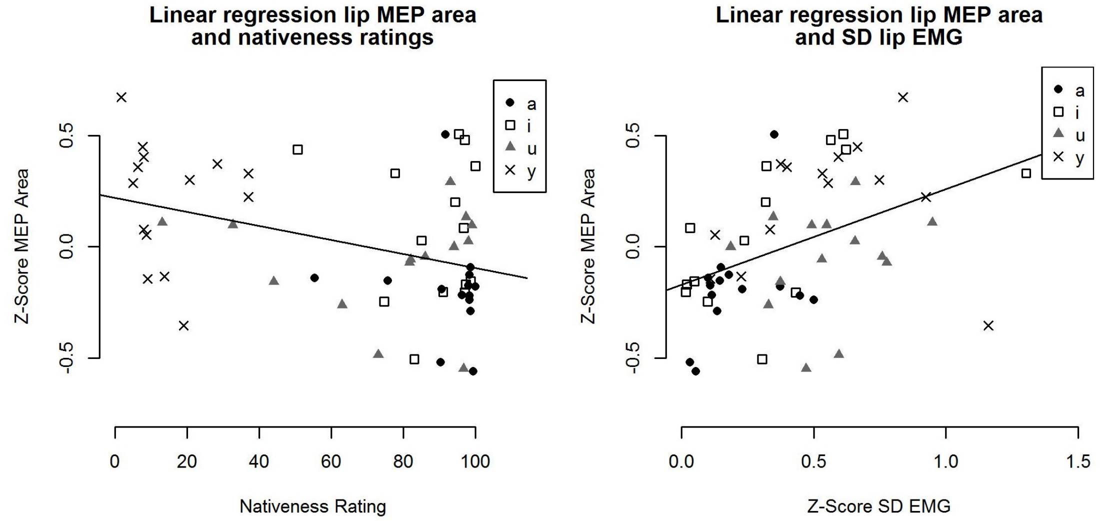

Figure 4: Linear regression analyses for the lip MEP area and the nativeness rating (left) and the lip MEP area and the SD of the lip EMG (right).

Discussion

Here we investigated if activations of the motor system during speech listening depend on sensorimotor experience with the phonemes. In line with hypothesis 2, which stated that the lack of acoustic-motor model for the non-native speech sounds might lead to motor compensatory activities, the strongest lip facilitation was found for the non-native vowel /y/ compared to the native vowel /a/. Values for the other native vowels /i/ and /u/ were in between the native vowel /a/ and the non-native vowel /y/. Differently from to /a/, the native vowels /i/ and /u/ require important lip movements for their production (/i/

lips pulled back and /u/ lip rounding) and this may have led to some residual modulation of corticobulbar excitability.

Further support for hypothesis 2 came from linear regression analyses revealing that lip MEPs amplitude is related to speech perception and production measures. A negative relationship between nativeness ratings and MEPs was found, showing that lip corticobulbar excitability increases when the perceived nativeness decreases. Additionally, a positive relationship between MEPs and standard deviation of EMG during production was found, showing that lip corticobulbar excitability increases when the production of the vowels was more variable. Trial by trial variability in vowel production suggests that participants were more insecure about how much movement of the lip was expected to produce the target vowel. Taken together, these results suggest that corticobulbar excitability of the lip is larger for perceptually and articulatory unfamiliar vowels, in line with the idea that the motor activities might compensate for the lack of an acoustic-motor model for the non-native speech sounds.

This interpretation is in line with a previous neuroimaging study reporting greater activations in the motor system during passive listening to non-native phonemes compared to native ones (Wilson & Iacoboni, 2006). Analogously, the motor system has been shown to increase its activation during the identification of a difficult second language contrast (/l/-/r/ for Japanese) in comparison to both an easy second language contrast (/b/-/g/ for Japanese, Callan et al., 2003) and the same contrast (/l/-/r/) in native speakers (English) for which the contrast is easier to identify (Callan, Jones, Callan, & Akahane-Yamada, 2004). Based on these results, Callan et al. (2003, 2004) proposed that learning a second language not only reorganizes auditory brain regions, but additionally requires stronger functional connectivity between articulatory-auditory and articulatory-orosensory brain regions.

In parallel to these neuroimaging studies showing motor compensatory activations during the perception of non-native phonemes, lip corticobulbar excitability was also found to increase when raising the difficulty level in native speech perception (Murakami et al., 2011). Lips corticobulbar excitability was also increased during perception of distorted speech produced using a tongue depressor, relative to naturally produced speech (Nuttal et al., 2016). These combined results suggest that motor compensatory activities can be found whenever the audio signal alone cannot be easily identified because of external perturbation of native sounds (e.g. acoustic perturbation; Murakami et al., 2011), internal perturbation of native speech sounds (e.g. articulatory perturbation; Nuttall et al., 2016). The present results fit with these previous accounts and suggests that even in absence of any perturbation, internal or external, compensatory motor activities are generated by listening to non-native speech sounds.

The present results need to be discussed in relation to the study by Swaminathan et al. (2013) showing the opposite pattern of corticobulbar excitability for visual speech. In that case, higher lips corticobulbar activity was demonstrated while viewing a known language compared to viewing an unfamiliar language. However, this is likely due to the different nature of the stimuli. Visual continuous speech as in Swaminathan et al. (2013) is more difficult to understand in the native language (Bernstein, Demorest, & Tucker, 2000; Ronquest, Levi, & Pisoni, 2010; Soto-Faraco et al., 2007). Instead, we used auditory presented single vowels that are much easier to identify. Most importantly, one key aspect is related to the intelligibility of non-native stimuli and more specifically if they provide enough cues to the participant. Continuous visual speech may not provide enough sensory cues to start the process of sensory-to-motor coordinate transformation (Soto-Faraco et al., 2007). Our participants could gain articulatory related information from the speech acoustic signals and in fact, they imitated the non-native vowel relatively well in the production task, placing it close to the values of the native speaker. Interestingly, larger lip MEPs to the non-native vowel /y/ shows that participants could extract information about the gesture of the non-native phoneme – a lip movement – probably by assimilating the non-native vowel to the articulatory similar native vowel /u/ that also uses lip rounding and thus actively trying to find a match for the perceived phoneme.

Similar assimilation phenomena are observable if we look at first and second language acquisition. In fact, newborns can perceive most phoneme contrasts from any language at birth (Werker & Tees, 1984), this capacity rapidly declines at the end of first year of life, as infants start to focus on the phonemes of their native language (Werker & Tees, 1984). The ability to successfully acquire later in life the sounds of a new language largely depends on the relative structure of the phonological system of our native language and the foreign language (Best & Tyler, 2007; Best, 1995; Flege, 1995, 2003). Phonemes that fall within the same phonological category in the native language but in different ones in the foreign language – such as the /l/-/r/ contrast for Japanese learners of English or /e/-/ɛ/ for Spanish learners of Catalan – are extremely hard to acquire (Miyawaki et al., 1975; Pallier, Bosch, & Sebastián-Gallés, 1997), due to the assimilation of the new sounds to the native ones. Our study shows that even when confronted to an unknown and untrained phoneme, the corticobulbar excitability of the lip muscle – a muscle used during the articulation of the non-native phoneme – increases. This is probably due to an assimilation mechanism which compensates for the lack of an acoustic-motor model for the non-native speech sounds. These results suggest that the motor system plays an active role in speech perception, even when confronted with new and untrained phonemes and this role might be relevant even during the first contact with a new, foreign language.

Acknowledgements

This research was supported by the European Community Grant POETICON ++ (STREP-Project ICT – 288382) awarded to L.F, and grants from the European Commission Seventh Framework Programme (FP7/2007-2013): ERG grant agree-ment number 323961 (UNDER CONTROL); Cooperation grant agreement num-ber 613465 – (AThEME), the Spanish Ministerio de Economía y Competitividad (PSI2015-66918-P) and the Catalan Government (SGR 2014–1210) awarded to N.S.G.. She also received the prize ‘‘ICREA Acadèmia’’ for excellence in research, funded by the Generalitat de Catalunya.

References

Agnew, Z. K., McGettigan, C., Scott, S. K., 2011. Discriminating between auditory and motor cortical responses to speech and nonspeech mouth sounds. J. Cogn. Neurosci. 23(12), 4038–47.

Bartoli, E., D'Ausilio, A., Berry, J., Badino, L., Bever, T., Fadiga, L., 2015. Listener-Speaker Perceived Distance Predicts the Degree of Motor Contribution to Speech Perception. Cereb. Cortex, 25(2), 281–288.

Bernstein, L. E., Demorest, M. E., Tucker, P. E., 2000. Speech perception without hearing. Percept. Psychophys. 62(2), 233–252.

Best, C. T., 1995. A direct realist view of cross-language speech perception. In W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp. 171–204). Baltimore: York Press.

Best, C. T., Tyler, M. D., 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), Language experience in second language speech learning: In honor of James Emil Flege (pp. 13–34). Amsterdam: John Benjamins Publishing Company.

Boersma, P., Weenink, D., 2010. Praat: doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Brainard, D. H., 1997. The Psychophysics Toolbox. Spatial Vision, 10(4), 433–436.

Callan, D. E., Jones, J. A., Callan, A. M., Akahane-Yamada, R., 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. NeuroImage, 22(3), 1182–94.

Callan, D. E., Tajima, K., Callan, A. M., Kubo, R., Masaki, S., Akahane-Yamada, R., 2003. Learning-induced neural plasticity associated with improved identification performance after training of a difficult second-language phonetic contrast. NeuroImage, 19(1), 113–124.

D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P., Fadiga, L., 2011. Vocal pitch discrimination in the motor system. Brain Lang. 118(1-2), 9–14.

D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L., 2012. The role of the motor system in discriminating normal and degraded speech sounds. Cortex, 48(7), 882–887.

D'Ausilio, A., Jarmolowska, J., Busan, P., Bufalari, I., Craighero, L., 2011. Tongue corticospinal modulation during attended verbal stimuli: priming and coarticulation effects. Neuropsychologia, 49(13), 3670–3676.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., Fadiga, L., 2009. The motor somatotopy of speech perception. Current Biology, 19(5), 381–385.

Fadiga, L., Craighero, L., Buccino, G., Rizzolatti, G., 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur.n J. Neurosci. 15(2), 399–402.

Flege, J. E., 1995. Second Language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research (pp. 233–277). Baltimore: York Press.

Flege, J. E., 2003. Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities (pp. 319–355). Berlin: Mouton de Gruyter.

Grabski, K., Tremblay, P., Gracco, V. L., Girin, L., Sato, M., 2013. A mediating role of the auditory dorsal pathway in selective adaptation to speech: a state-dependent transcranial magnetic stimulation study. Brain Res., 1515, 55–65.

Levy, E. S., 2009. On the assimilation-discrimination relationship in American English adults' French vowel learning. J. Ac. Soc. Am., 126, 2670–2682.

MacDonald, J., McGurk, H., 1978. Visual influences on speech perception processes. Percept. Psychophys. 24(3), 253–257.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., Iacoboni, M., 2007. The essential role of premotor cortex in speech perception. Curr. Biol., 17(19), 1692–6.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., Fujimura, O., 1975. An effect of Linguistic Experience: The discimination of [r] and [l] by native Speakers of Japanes and English. Percept. Psychophys. 18(5), 331–340.

Möttönen, R., Dutton, R., Watkins, K. E., 2013. Auditory-motor processing of speech sounds. Cereb. Cortex, 23(5), 1190–1197.

Möttönen, R., van de Ven, G. M. Watkins, K. E., 2014. Attention fine-tunes auditory-motor processing of speech sounds. J. Neurosci. 34(11), 4064–4069.

Möttönen, R., Watkins, K. E., 2009. Motor representations of articulators contribute to categorical perception of speech sounds. J. Neurosci. 29(31), 9819–9825.

Murakami, T., Restle, J., Ziemann, U., 2011. Observation-execution matching and action inhibition in human primary motor cortex during viewing of speech-related lip movements or listening to speech. Neuropsychologia, 49(7), 2045–54.

Nuttall, H. E., Kennedy-Higgins, D., Hogan, J., Devlin, J. T., Adank, P., 2015. The effect of speech distortion on the excitability of articulatory motor cortex. NeuroImage, 128, 218–226.

Oldfield, R. C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia, 9, 97–113.

Pallier, C., Bosch, L., Sebastián-Gallés, N., 1997. A limit on behavioral plasticity in speech perception. Cognition, 64(3), B9–B17.

Pelli, D. G., 1997. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spatial Vis. 10, 437–442.

Pulvermüller, F., Fadiga, L., 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nat. Rev. Neurosci. 11(5), 351–360.

R., 2013. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

Rogers, J. C., Möttönen, R., Boyles, R., Watkins, K. E., 2014. Discrimination of speech and non-speech sounds following theta-burst stimulation of the motor cortex. Front. Psychol. 5(July), 754.

Ronquest, R. E., Levi, S. V, Pisoni, D. B., 2010. Language identification from visual-only speech signals. Atten. Percept. Psychophys. 72(6), 1601–1613. doi:10.3758/APP.72.6.1601

Roy, A. C., Craighero, L., Fabbri-Destro, M., Fadiga, L., 2008. Phonological and lexical motor facilitation during speech listening: a transcranial magnetic stimulation study. J. Physiol, 102(1-3), 101–5.

Sato, M., Buccino, G., Gentilucci, M., Cattaneo, L., 2010. On the tip of the tongue: Modulation of the primary motor cortex during audiovisual speech perception. Speech Comm. 52(6), 533–541.

Sato, M., Tremblay, P., Gracco, V. L., 2009. A mediating role of the premotor cortex in phoneme segmentation. Brain Lang. 111(1), 1–7.

Soto-Faraco, S., Navarra, J., Weikum, W., Vouloumanos, A., Sebastián-Gallés, N., Werker, J. F., 2007. Discriminating languages by speech-reading. Percept. Psychophys. 69(2), 218–231.

Strange, W., Levy, E. S., Law, F. F., 2009. Cross-language categorization of French and German vowels by naive American listeners. J. Ac. Soc. Am. 126(September), 1461–1476.

Sundara, M., Namasivayam, A. K., Chen, R., 2001. Observation-execution matching system for speech: a magnetic stimulation study. Neuroreport, 12(7), 1341–1344.

Swaminathan, S., MacSweeney, M., Boyles, R., Waters, D., Watkins, K. E., Möttönen, R., 2013. Motor excitability during visual perception of known and unknown spoken languages. Brain Lang. 126(1), 1–7.

Watkins, K. E., Strafella, A. P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia, 41(8), 989–994.

Weikum, W., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., Werker, J., 2013. Age-related sensitive periods influence visual language discrimination in adults. Front. Syst. Neurosci. 7, 1–8.

Werker, J. F., Tees, R. C., 1984. Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life, Infant. Behav. Dev. 7, 49–63.

Wilson, S. M., Iacoboni, M., 2006. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. NeuroImage, 33(1), 316–325.

Wilson, S. M., Saygin, A. P., Sereno, M. I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7(7), 701–702.