Sensory-motor relationships in speech production in post-lingually deaf cochlear-implanted adults and normal-hearing seniors: Evidence from phonetic convergence and speech imitation

Lucie Scarbela, Denis Beautemps, Jean-Luc Schwartz, Marc Sato

GIPSA-LAB, Département Parole & Cognition, CNRS & Grenoble Université, Grenoble, France
Laboratoire Parole & Langage, CNRS & Aix-Marseille Université, Aix-en-Provence, France

ARTICLE INFO

Keywords:
Phonetic convergence
Imitation
Cochlear implant
Elderly
Sensory-motor interactions
Speech production

ABSTRACT

Speech communication can be viewed as an interactive process involving a functional coupling between sensory and motor systems. One striking example comes from phonetic convergence, when speakers automatically tend to mimic their interlocutor's speech during communicative interaction. The goal of this study was to investigate sensory-motor linkage in speech production in postlingually deaf cochlear implanted participants and normal hearing elderly adults through phonetic convergence and imitation. To this aim, two vowel production tasks, with or without instruction to imitate an acoustic vowel, were proposed to three groups of young adults with normal hearing, elderly adults with normal hearing and post-lingually deaf cochlear-implanted patients. Measure of the deviation of each participant's $f_0$ from their own mean $f_0$ was measured to evaluate the ability to converge to each acoustic target.

Results: showed that cochlear-implanted participants have the ability to converge to an acoustic target, both intentionally and unintentionally, albeit with a lower degree than young and elderly participants with normal hearing. By providing evidence for phonetic convergence and speech imitation, these results suggest that, as in young adults, perceptuo-motor relationships are efficient in elderly adults with normal hearing and that cochlear-implanted adults recovered significant perceptuo-motor abilities following cochlear implantation.

1. Introduction

1.1. Sensory-motor interactions in speech perception and speech production

In the speech communication domain, an old and fundamental debate concerns the nature of processes and representations involved in speech perception and production. Concerning speech perception, auditory theories assume that speech perceptual processing and categorization are based on acoustic features and auditory representations (Stevens and Blumstein, 1978; Blumstein and Stevens, 1979; Lindblom and Maddieson, 1988; Lindblom, 1990). Conversely, the motor theory of speech perception (Liberman and Mattingly, 1985) or the direct realist theory (Fowler, 1986) respectively claim that speech perception involves the recovery of the speaker's motor intentions or articulatory gestures. More recently, various perceptuo-motor theories introduced syntheses of arguments by tenants of both auditory and motor theories, and proposed that implicit motor knowledge and motor representations are used in relationship with auditory representations and processes to elaborate phonetic decisions (Skipper et al., 2007; Schwartz et al., 2012). These theories capitalize on the increasing amount of evidence for the role of motor representations and processes in speech perception (see a recent review in Skipper et al. (2016)).

Concerning speech production, various theories have also been introduced to characterize the speaker's goals. Motor or articulatory theories, like Task dynamics (Saltzman, 1986) and the associated Articulatory Phonology (Brown and Goldstein, 1992), consider that speech targets are defined in the articulatory space. On the contrary, auditory theories like Stevens' Quantal Theory (Stevens, 1972, 1989) and Perkell's speech production control model (Perkell et al., 1995, 2000), suggest that targets are specified in auditory terms. Finally, sensory-motor models claim that speech production control is multimodal and combines auditory and somatosensory information (see Perrier, 2005; and the DIVA model, Guenther et al., 1998; Guenther and Vladusich, 2012). A number of data about the effect of auditory (e.g. Lametti et al., 2014; Shiller and Rochon, 2014) or somatosensory (Tremblay et al., 2003) perturbations applied to speech production in adults or children are in agreement with this sensory-motor framework.

From a number of these theories, sensory-motor relationships seem
to play an important role in both speech perception and speech production in normal conditions. The present study deals with subjects for which some degradation of the sensory-motor link could be expected, because of sensory deficits. The first and primary situation is the case of post-lingually deaf subjects equipped with cochlear implants. Indeed, post-lingually deaf subjects are expected to have acquired efficient sensory-motor relationships before deafness. Their ability to achieve intelligible speech production is interpreted as evidence that they have maintained a stable and rather accurate internal model of these sensory-motor relationships all along their deafness life (Perkell et al., 2000). Cochlear implantation then results in providing them with a new kind of auditory input, quite different from the one they had acquired before deafness considering the very special nature of cochlear-implant coding of acoustic information. The question is to know how the internal relationships between these “new” auditory inputs, the “old” ones acquired at the first stages of development, and the motor representations are established and organized, and how efficient they are.

A number of data on speech production after cochlear implantation display an increase in acoustic dispersion of vowels after a few months of implantation (e.g. Langereis et al., 1998; Lane et al., 2005; Ménard et al., 2007), suggesting that the internal model has indeed benefited from the new auditory input provided by cochlear implantation. Interestingly however, some studies explored the response to articulatory perturbation using a bite-block task (Lane et al., 2005), a robotic device displacing the jaw during speech (Nasir and Ostry, 2008) or a lip-tube task (Turgeon et al., 2015), with or without auditory feedback. Results showed that post-lingual cochlear-implanted participants (CI) are able to adapt their articulatory trajectory when it is perturbed in order to reach their auditory goals and make their production intelligible, even when the implant is turned off. Such compensatory strategies show that perceptuo-motor abilities acquired during speech acquisition in CI subjects are still at work after deafness, though the addition of auditory information provided by the cochlear-implant does result in enhanced precision and efficiency of the sensory-motor internal model. Finally, a PET-scanning study on visual speech perception in post-lingually deaf CI patients (Rouger et al., 2012) showed that after a short adaptation period with the implant, there was a decrease of the initially abnormal activity in the superior temporal sulcus, a cross-modal brain area, accompanied by a progressive reactivation of frontal premotor speech areas. This suggests that sensorimotor neuroplasticity after cochlear implantation provides a progressive reactivation of the audio-visuo-motor linkage in CI subjects.

The second situation considered in the present study is the case of elderly subjects. They are of interest for us for two reasons. The first one is the interest of assessing the consequences of the potential decline of cognitive and language abilities in relation with decline in sensory and motor accuracy on the efficiency of sensory-motor relationships. The second reason is that a number of the cochlear-implanted post-lingually deaf participants are rather old, and we considered important to include senior subjects with no severe auditory deficit as a control population. As a matter of fact, very few studies investigated perceptuo-motor relationships in elderly population. One of them shows that elderly people adapt their production in case of degraded auditory feedback (Liu et al., 2010, 2011a, 2011b). Sensorimotor neuroplasticity was also observed in elderly people, linked with age-dependent intelligibility effects mainly found in auditory and motor cortical areas ( Tremblay et al., 2013; Bilodeau-Mercure et al., 2015). Taken together, these behavioral and neuro-imaging studies suggest that sensory-motor relationships are well preserved in aging.

1.2. Phonetic convergence and imitation, a paradigm for studying sensory-motor relationships in speech

Imitation is a quite widespread phenomenon in speech communication and can be viewed as a key mechanism in the acquisition of human language. In interactive situations, adult speakers tend to continuously adapt their productions to those of their interlocutor, in order to facilitate communicative exchanges. These adaptive changes in speech production can be voluntary, that is when the speaker consciously imitates his interlocutor, but also unintentional. Indeed, unintentional imitation, or phonetic convergence, is the tendency to automatically imitate a number of acoustic-phonetic characteristics in the productions of an interacting speaker, has been displayed in various studies (see a recent collection of papers on this topic in Nguyen et al., 2013, and for recent reviews see Babel, 2009; Aubanel, 2011; Lelong, 2012). In these studies, convergence effects have been observed both in paralinguistic features, such as gestures, and in speech acoustic parameters, such as intensity, fundamental frequency \(f_0\), or formants. These phonetic convergence effects may be related to the social component of communication, assuming that they would contribute to setting a common ground between speakers (Giles et al., 1991) and that they could be associated to the human desire of affiliation to a social group.

Interestingly however, while most reports on phonetic convergence are based on conversational exchanges in natural conditions, a few studies showed that phonetic convergence can also be observed using laboratory settings involving non-interactive situations of communication (Goldinger and Azuma, 2004; Gentilucci and Cattaneo, 2005; Delvaux and Soquet, 2007; Garnier et al., 2013; Sato et al., 2013). For example, Delvaux and Soquet (2007) obtained phonetic convergence effects during the production of auditorily presented words without interaction, and they also reported offline adaptation to the auditory targets in post-tests following stimulus exposure. These studies hence suggest that phonetic convergence is not only a matter of social attunement, but could also involve a more basic stage of continuous automatic adaptation of the speech production system to the external speech sounds environment (for a recent review, see Sato et al. (2013)).

Such an automatic adaptation mechanism requires the existence of a sensory-motor coupling process in which variations in the external environment provide sensory targets that drive motor control procedures to adapt and produce stimuli closer to these external targets. More in detail, the speaker would program motor commands to achieve an articulatory and ultimately an auditory (or more generally sensory) goal according to the linguistic message to be conveyed to the listener. Then, in the course of speech production, the speaker would compare sensory feedback to the initial sensory objective. Variations in the external environment would shift the sensory targets and accordingly result in modifications of motor commands to converge towards speech sounds in the environment. Hence, the convergence/imitation paradigm is a natural tool to study sensory-motor coupling in speech communication.

Phonetic convergence and imitation in laboratory settings can involve various kinds of acoustic features, though fundamental frequency effects seem to be larger and easier to obtain than variations in e.g. formant values (Sato et al., 2013). The present study therefore focus on convergence and imitation on \(f_0\) variations.

1.3. \(f_0\) perception and motor control in cochlear-implanted adults and normal-hearing seniors

Past studies on pitch control in speech production by CI subjects reported higher \(f_0\) values for CI adults than for adults with normal hearing and, more importantly, larger variations in \(f_0\) for CI than for adults with normal hearing (e.g. Lane and Webster, 1991). Langereis et al. (1998) and Kishon-Rabin et al. (1999) also reported that while pitch production is generally altered in CI patients soon after implantation, almost normal \(f_0\) values were observed one year post-implantation for two participants in Kishon-Rabin et al. (1999) study. Notice that technological evolutions led to important progress in the coding of pitch in cochlear implants since these studies.

Concerning speech perception, several factors appear to influence
performance in CI patients. In Blamey et al. (2012), 2251 CI patients participated to a battery of auditory tests, where they had to recognize phonemes, words and sentences. The experimenters reported that both duration of implant experience, age at onset of severe to profound hearing loss, age at cochlear implantation and duration of deafness influence speech perception to a certain extent – though inter-subject variability was quite large in this kind of study. Regarding $f_0$, for cochlear-implanted patients, pitch is difficult to estimate because of different reasons. First, as the implant is composed of 12–22 electrodes, the merging of frequencies in a given electrode likely impacts frequency discrimination by patients. Furthermore, because of their proximity, electrodes can stimulate nerve fibers, which do not correspond to the intended frequency, which leads to an imperfect tonotopy restitution.

Regarding elderly people, it is well known that auditory capacities decline with aging. More specifically, even with close-to-normal auditory thresholds, elderly people with normal hearing present difficulties in adverse listening conditions, e.g. in a noisy environment (Gelfand et al., 1985; Ohde and Abou-Khalil, 2001; Füllgrabe, 2013). The role of cognitive functions in this decline of speech perception is still under debate (Cienkowski and Carney, 2002; Cienkowski and Vasil-Dilaj, 2010; Moore and Fullgräbe, 2013; see a review in Füllgrabe and Rosen, 2016). On the motor side, elderly people also show deficits in orofacial movements, which might impact speech production. Indeed, a majority of studies demonstrate a degradation of speech production in elderly people, with more variable and less stable speech utterances than observed in younger adults. An alteration of voice control is also observed in elderly people, notably with higher pitch variability and less stable pitch production at the level of individual subjects (Morgan and Rastatter, 1986; Russell et al., 1995; Lortie et al., 2015).

1.4. Goal of the study

In the present study, voluntary and unintentional imitative changes were tested during speech production in both post-lingually deaf or hearing impaired adults equipped with a cochlear implant and in seniors with normal hearing. Our goal was to explore in these populations the underlying perceptuo-motor mechanisms at work in phonetic convergence and speech imitation. To this aim, we capitalized on two recent studies by Sato et al. (2013) and Garnier et al. (2013) displaying both unintentional and voluntary imitative changes on fundamental frequency of auditorily presented vowel targets during speech production in a non-interactive situation of communication. In these studies, participants were asked to produce different vowels according to acoustic targets with their pitch varying around the averaged pitch of their own voice, with or without instruction to imitate the target. Results showed that participants strongly converged to the vowel target not only in the imitative task, but also, at a lower degree, in the production task even if no instruction to imitate the vowel target was given.

To our knowledge, convergence and imitation abilities have never been studied either in CI subjects or in elderly people with normal hearing. We here report online imitative changes on the fundamental frequency in relation to acoustic vowel targets in a non-interactive situation of communication during both unintentional and voluntary imitative production tasks in the two studied populations. Our goal was to determine to which extent CI and elderly participants display convergence and imitation abilities, and to compare these abilities with those observed in young adults with normal hearing (NHY), with the larger aim to evaluate sensory-motor linkage during speech production in CI and elderly population. Three populations have been contrasted: post-lingually deaf or hearing-impaired cochlear-implanted adults (generally not very young), senior adults with no specific auditory deficit apart from aging, and normal-hearing young adults as a control. The same paradigm as the one developed by Garnier et al. (2013) and Sato et al. (2013) was exploited in this study.

2. Methods

2.1. Participants

Three groups of participants performed the experiment. The first group consisted of fifteen young participants with normal hearing (NHY) (10 females and 5 males, mean age: 30 years old, range: 20–40) who reported no history of speaking, hearing or motor disorders. The second group consisted of ten elderly participants with normal hearing (NHE) (4 females and 6 males, mean age: 69 years old, range: 63–78). The third group consisted of ten post-lingually deaf cochlear-implanted (CI) participants (7 males and 3 females, mean age: 58.9 years old, range: 27–76). As indicated in Table 1, the CI group was heterogeneous in several aspects: age at onset of deafness varied from 7 to 65 years, duration of deafness varied from 1 month to 58 years and duration of cochlear implant experience varied from 1 month to 9 years. In addition, seven of the ten participants wore classical hearing aid in the non-implanted ear, and one participant was bilaterally implanted. The experiment was performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki.

2.2. Stimuli

A vowel database was created from /e/, /œ/, /o/ French vowels produced by one male and one female speaker. For each vowel, one clearly articulated occurrence was selected and digitized at 44.1 kHz. From these stimuli, $f_0$ was artificially shifted by steps of ± 5 Hz (from 80 Hz to 180 Hz for the male vowels, and from 150 to 350 Hz for the female vowels) using the PSOLA module integrated in the Praat software (Boersma, 2001). These stimuli allowed us to present to each participant nine distinct stimuli per vowel ranging from −20% to +20% from the mean participant’s average pitch of their own voice, with or without instruction to imitate the target. Results showed that participants strongly converged to the vowel target not only in the imitative task, but also, at a lower degree, in the production task even if no instruction to imitate the vowel target was given.

To our knowledge, convergence and imitation abilities have never been studied either in CI subjects or in elderly people with normal hearing. We here report online imitative changes on the fundamental frequency in relation to acoustic vowel targets in a non-interactive situation of communication during both unintentional and voluntary imitative production tasks in the two studied populations. Our goal was to determine to which extent CI and elderly participants display convergence and imitation abilities, and to compare these abilities with those observed in young adults with normal hearing (NHY), with the larger aim to evaluate sensory-motor linkage during speech production in CI and elderly population. Three populations have been contrasted: post-lingually deaf or hearing-impaired cochlear-implanted adults (generally not very young), senior adults with no specific auditory deficit apart from aging, and normal-hearing young adults as a control. The same paradigm as the one developed by Garnier et al. (2013) and Sato et al. (2013) was exploited in this study.

### Table 1

<table>
<thead>
<tr>
<th>Gender</th>
<th>Age (y.)</th>
<th>Age at onset of deafness (y.)</th>
<th>Hearing aid</th>
<th>Duration of deafness</th>
<th>Duration of CI experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI1</td>
<td>M</td>
<td>65</td>
<td>7</td>
<td>No</td>
<td>58 years 1 month</td>
</tr>
<tr>
<td>CI2</td>
<td>M</td>
<td>56</td>
<td>20</td>
<td>Yes</td>
<td>35 years 3 months</td>
</tr>
<tr>
<td>CI3</td>
<td>F</td>
<td>66</td>
<td>32</td>
<td>Yes</td>
<td>25 years 9 years</td>
</tr>
<tr>
<td>CI4</td>
<td>M</td>
<td>60</td>
<td>59</td>
<td>Yes</td>
<td>1 month 1 years 4 months</td>
</tr>
<tr>
<td>CI5</td>
<td>F</td>
<td>43</td>
<td>20</td>
<td>Yes</td>
<td>13 years 2 months</td>
</tr>
<tr>
<td>CI6</td>
<td>M</td>
<td>27</td>
<td>25</td>
<td>Yes</td>
<td>2 months 2 years 6 months</td>
</tr>
<tr>
<td>CI7</td>
<td>F</td>
<td>67</td>
<td>65</td>
<td>Yes</td>
<td>2 years 7 months</td>
</tr>
<tr>
<td>CI8</td>
<td>M</td>
<td>72</td>
<td>40</td>
<td>Yes</td>
<td>30 years 5 months</td>
</tr>
<tr>
<td>CI9</td>
<td>M</td>
<td>76</td>
<td>48</td>
<td>No</td>
<td>27 years 3 years 4 months</td>
</tr>
<tr>
<td>CI10</td>
<td>M</td>
<td>57</td>
<td>48</td>
<td>Yes</td>
<td>8 years 10 months</td>
</tr>
</tbody>
</table>

2.3. Experimental procedure

The experiment was carried out in a sound-proof room. Participants sat in front of a computer monitor at a distance of approximately 50 cm. The acoustic stimuli were presented at a comfortable sound level through a loudspeaker, with the same sound level set for all participants. The Presentation software (Neurobehavioural Systems, Albany, CA) was used to control the stimulus presentation during all experiments. All participants’ productions were recorded for offline analyses.

The experiment consisted of three vowel production tasks. First, participants had to individually produce /e/, /œ/ and /o/ vowels, according to a visual orthographic target. This allowed the experimen-
ter to measure the participant’s \( f_0 \). In the subsequent task (which will be referred to as “convergence task”), participants were asked to produce the three vowels according to an acoustic target. Importantly, no instruction to “repeat” or to “imitate” the acoustic targets was given to the participants. Finally, the third task (which will be referred to as “imitative task”) was the same as the second task except that participants were explicitly asked to imitate the acoustic targets. The only indication given to participants was to imitate the voice characteristics of the perceived speaker. Importantly, we hence kept a fixed order of the “convergence” and “imitative” tasks so as to avoid cases in which prior explicit imitation instructions could modify further convergence effects and add conscious imitation strategies to automatic convergence processes.

Acoustic targets in the “convergence” and “imitative” tasks for each participant consisted in the 27 stimuli selected from the vowel database (9 per vowel category), repeated three times each so as to obtain 81 trials altogether. Each participant was presented with one model talker (9 per vowel category), repeated three times each so as to obtain 81 trials altogether. Each participant was presented with one model talker of the same sex, with the 9 quantified \( f_0 \) frequencies varying from \(-20\%\) to \(+20\%\) by steps of 5\% around his/her own pitch, as measured in the first task.

At the end of the experiment, participants were asked to perform a frequency discrimination test to estimate their pitch just noticeable difference (JND), based on an experimental paradigm proposed by Vinay and Moore (2010). Two groups of four sounds were presented to participants in a random order, composed either of identical sounds (AAAA) or of two sounds A and B differing in pitch (ABAB), and participants had to determine in which group sounds were different among each other. We used synthesized stimuli, obtained by addition of harmonics of a given \( f_0 \) value, with 0 phase and amplitude based on the spectrum of a vowel /e/ produced by either a male or a female, depending of the participant’s gender. For sound A, the \( f_0 \) value was set at the participant’s mean production value. For sound B, the selected \( f_0 \) value was systematically higher than \( f_0 \) for sound A, beginning with a 5\% deviation further varying along the test. Moreover, small random intensity variations were applied to the two groups of sounds from one trial to the other, in order to force the participant to focus on pitch rather than timbre or intensity perception (Vinay and Moore, 2010).

Variations of the \( f_0 \) values for sound B were driven by an adaptive two-alternative forced choice procedure: after two consecutive correct answers, the frequency difference between sounds A and B decreased, whereas one mistake drove an increase of this difference. The test result corresponded to the mean of the difference between \( f_0 \) values for sound A and sound B for the last eight trials.

### 2.4. Data analysis

All acoustic analyses of participants’ productions were performed using the Praat software (Boersma, 2001). In the second and third tasks which both involved an acoustic target (/e/, /æ/ or /o/) that could possibly be misunderstood by the participant, we annotated the participant productions in order to estimate the percentage of errors in production in relation to the target.

Analysis began by a semi-automatic segmentation procedure aiming at segmenting each individual vowel produced by a given participant in a given task. Using intensity and duration criteria, the algorithm automatically identified pauses between each vowel and segmented individual productions accordingly. Segmentation was hand-corrected when appropriate, checking waveform and spectrogram. Only correctly produced vowels (that is, checked by the experimenter as corresponding to the phonemic target) were further analyzed. For each correctly produced vowel, \( f_0 \) was determined within a time frame around \( \pm 25 \) ms of the maximum intensity of the sound file.

### 2.5. Statistical analysis

For each subject and task, there were altogether 9 values of \( f_0 \) production per target \( f_0 \) value between \(-20\%\) and \(+20\%\) (with 3 vowel types and 3 utterances per type). The mean of the 9 produced \( f_0 \) values was computed, and linear regression between mean produced \( f_0 \) values and \( f_0 \) targets was computed for each participant separately in the convergence and imitation tasks. The slope and correlation coefficients were then determined for each task and each participant. Individual \( t \)-tests were done on these correlation coefficients to determine if they were significant.

For each group, one-tailed \( t \)-tests were then performed on individual slope and correlation coefficients compared to zero in order to determine significant imitative changes in each task. To test for possible differences in imitative changes from one task to the other, additional \( t \)-tests were performed on slope and correlation coefficients between the two tasks. However, it is important to remind that the order between the two tasks being not counterbalanced, any difference between them can be interpreted both as a consequence of difference in tasks and order effect.

In addition, in order to test whether imitative changes in the convergence and imitation tasks may correlate among speakers, a Pearson’s correlation analysis was performed between slope coefficients. Finally, Pearson’s correlation analyses were performed between slope coefficients and JND values within each group of participants and for each task, as well as between slope coefficients and age, duration of deafness, age at implantation, age at the beginning of deafness and duration of implant experience for CI participants in each task.

To compare the results of the three groups, a repeated-measures ANOVA was finally performed on slope coefficients with the group (NHY vs. NHE vs. CI) as between-subject variable and the task (convergence vs. imitation) as within-subject variable. Additionally, a one-factor ANOVA was performed on JND values, with group as categorical factor. The sphericity assumption was tested using a Mauchly test and, when necessary, Greenhouse-Geisser corrections were applied (Greenhouse and Geisser, 1959).

### 3. Results

#### 3.1. Young participants with normal hearing (see Fig. 1)

For NHY participants, imitative changes were observed in both tasks, though stronger during voluntary imitation. Slope coefficients differed significantly from zero in both the convergence (t(14) = 5.98; \( p < 0.001 \)) and imitation (t(14) = 55.78; \( p < 0.001 \)) tasks. In addition, slope coefficients were higher in the imitation compared to the convergence tasks (on average: 0.87 vs. 0.45; t(14) = 6.02; \( p < 0.001 \)). No significant correlation was observed between the slope coefficients in the convergence and imitation tasks (\( r^2 = 0.12 \) (Figs. 1–4).
Similarly, correlation coefficients differed significantly from zero in both the convergence \( (t(14)=8.3; p<0.001) \) and imitation \( (t(14)=93.34; p<0.001) \) tasks, and were higher in the imitation compared to the convergence task (on average: 0.94 vs. 0.64; \( t(14)=4.3; p<0.001 \)). Individual analyses showed that fourteen of the sixteen participants converged to the target in the convergence task \( (r^2=0.91) \), and in the imitative task all of the participants converged to the target \( (r^2=0.99) \).

JND values for NHY participants were at 1.06 Hz on average (range: 0.14–3.88 Hz), with no significant correlation between the slope coefficients in the convergence or the imitation tasks and JND values (convergence: \( r^2=0.00 \) \( p=0.82 \), imitation: \( r^2=0.00 \) \( p=0.9 \)).

### 3.2. Elderly participants with normal hearing (see Fig. 2)

For NHE participants, imitative changes were observed in both tasks, though stronger in voluntary imitation. Slope coefficients differed significantly from zero in both the convergence \( (t(9)=3.52; p<0.01) \) and imitation \( (t(9)=8.5; p<0.001) \) tasks. In addition, slope coefficients were higher in the imitation compared to the convergence task (on average: 0.75 vs. 0.33; \( t(9)=4.97; p<0.001 \)). There was no significant correlation between convergence and imitation tasks \( (r^2=0.59) \).

Similarly, correlation coefficients differed significantly from zero in both the convergence \( (t(9)=5.96; p<0.001) \) and imitation \( (t(9)=11.78; p<0.001) \) tasks, and were higher in the imitation compared to the convergence tasks (on average: 0.84 vs. 0.54; \( t(9)=4.12; p<0.001 \)). Individual analyses showed that eight of the ten participants converged to the target in the convergence task \( (r^2=0.35) \) to 0.99, and in the imitative task all of the participants except one converged to the target \( (r^2=0.28) \) to 0.99.

JND values for NHE participants were at 0.95 Hz on average (range: 0.43–1.8 Hz) with no correlation with either the convergence or the imitation slopes (convergence: \( r^2=0.03 \) \( p=0.63 \), imitation: \( r^2=0.00 \) \( p=0.79 \)).

### 3.3. Cochlear-implanted participants (see Fig. 3)

As for NHY and NHE participants, imitative changes were also observed in both tasks for CI participants. Slope coefficients differed significantly from zero in both the convergence \( (t(9)=3.24; p<0.02) \) and imitation \( (t(9)=4.84; p<0.001) \) tasks. In addition, slope coefficients were higher in the imitation compared to the convergence tasks (on average: 0.39 vs. 0.14; \( t(9)=3.53; p<0.001 \)). As for NH participants, there was no significant correlation between convergence and imitation tasks \( (r^2=0.28) \).

Similarly, correlation coefficients differed significantly from zero in both the convergence \( (t(9)=4.08; p<0.01) \) and imitation \( (t(9)=5.53; p<0.001) \) tasks, and were higher in the imitation compared to the convergence tasks (on average: 0.49 vs. 0.24; \( t(9)=2.95; p<0.02 \)).

JND values for CI participants were at 14.06 Hz on average (range: 1.71–45.35 Hz). There was also no correlation between the convergence or imitation tasks and the others factors, that are JND values, age, deafness duration, age at implantation, age at the beginning of deafness and duration of implant experiment (all \( r^2<0.3 \)).

While young and elderly participants with normal hearing made no errors in both tasks, cochlear-implanted participants made a number of errors in both tasks, with similar error percentage in the convergence (15%) and imitation tasks (13%). In order to verify if these errors influence convergence effects, correlations between production and target were measured in both tasks on error trials only, for participants presenting at least 15% of errors (four participants in each task). The correlation values are displayed in Table 2. For those participants, correlations between production and target for error trials appear similar to those obtained for correct trials in both convergence and imitation tasks. As a matter of fact, the correlation values for correct trials and for error trials (when they are sufficient to do computations) are highly correlated among the 8 cases where both can be computed \( (r^2=0.71, t(6)=3.81, p<0.005) \).

To relate convergence effects with etiology variability and JND values, we report determination coefficients in both tasks for each participant (see Table 2). In the convergence task, we obtained significant correlation between target and production for five participants, with determination coefficients from 0.05 and 0.75. In the imitative task, correlation is significant for seven participants, with determination coefficients from 0.09 and 0.94. As shown in Table 2, the participants with no significant convergence in one or the other tasks do...
not display obvious similarity in etiology in e.g. age at deafness onset or deafness duration. Importantly, clear convergence in both tasks is obtained for some subjects after a quite reduced duration of implantation: see the cases of subjects CI1, CI2, CI5 or CI9 with significant convergence or imitation on f0, after 1–5 months of implantation.

Concerning JND values, the participants with no significant convergence in one or the other tasks do not present the worst JND values, and the participant with the highest JND value obtains significant convergence effect in both tasks. More strikingly, the corresponding JND value for this subject, at 45 Hz, is actually larger than the largest variation in F0 applied to the stimuli in both convergence tasks. As a matter of fact, F0 for this subject equals 125 Hz, and the maximal variation of F0 at 20% results in a 25 Hz modification. Altogether, these facts converge to suggest that the task used for JND estimation, well adapted for normal-hearing subjects (Vinay and Moore, 2010), is probably too complex for cochlear-implanted subjects.

3.4. Comparison between groups (see Fig. 4)

The main effect of group was significant (F(2,32) = 31.3; p < 0.001) with lower mean slope coefficient for CI participants than for NHY and NHE participants (CI: 0.26; NHY: 0.65; NHE: 0.55). No difference was obtained between NHY and NHE groups. The main effect of task was also significant (F(1,32) = 67.29; p < 0.001), with lower slope coefficients for the convergence task than for the imitation task (convergence: 0.33; imitation: 0.69). No significant interaction was found between the group and the task.

For the JND test, the effect of group was significant (F(2,32) = 11.03; p < 0.001), with larger values for CI participants than for NHY and NHE participants (CI: 14.06; NHY: 1.06; NHE: 0.95) without significant difference between NHY and NHE participants. This shows that CI participants stay largely impaired in f0 discrimination.

4. Discussion

The goal of this study was to investigate sensory-motor linkage in speech production on post-lingually deaf cochlear-implanted patients, in comparison with young and elderly adults with normal hearing, through the abilities of f0 convergence and imitation. To this aim, we used a paradigm of intentional and non-intentional imitation of vowels with modified fundamental frequency.

Table 2

Determination coefficients in both tasks for each participant of the cochlear-implanted group (CI), normal-hearing elderly adults group (NHE) and normal-hearing young adults (NHY).

<table>
<thead>
<tr>
<th>CI</th>
<th>NHE</th>
<th>NHY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergence</td>
<td>Imitation</td>
<td>JND</td>
</tr>
<tr>
<td>0.25 (0.33)</td>
<td>0.51 (0.32)</td>
<td>15.22</td>
</tr>
<tr>
<td>0.51</td>
<td>0.70</td>
<td>45.35</td>
</tr>
<tr>
<td>0.78 (0.51)</td>
<td>0.77 (0.73)</td>
<td>1.71</td>
</tr>
<tr>
<td>0.04 (0.17)</td>
<td>0.39</td>
<td>9.76</td>
</tr>
<tr>
<td>0.27</td>
<td>0.94</td>
<td>1.79</td>
</tr>
<tr>
<td>0.12</td>
<td>0.77 (0.41)</td>
<td>4.55</td>
</tr>
<tr>
<td>0.49</td>
<td>0.62</td>
<td>18.52</td>
</tr>
<tr>
<td>0.05</td>
<td>0.92</td>
<td>10.55</td>
</tr>
<tr>
<td>0.02 (0.03)</td>
<td>0.22</td>
<td>16.21</td>
</tr>
<tr>
<td>0.60</td>
<td>0.09</td>
<td>28.54</td>
</tr>
</tbody>
</table>

4.1. Young adults

We firstly replicated the previous findings by Sato et al. (2013) and Garnier et al. (2013) for young adults with normal hearing. Imitative changes towards the acoustic target were indeed observed in both tasks, with stronger convergence in the task with direct instruction than in the convergence task. It is likely that this is mainly due to the effect of the explicit instruction, though we cannot discard the possibility that order intervened here, with a trend to increase convergence in the second task just becomes it comes after the first one. As in the two previous studies, no correlation was found between the two tasks, that is individual f0 changes in the convergence task were not related to those observed in the imitative task among subjects. This lack of correlation could be due to ceiling effects provided by the very high degree of convergence in the imitative task and the very low variability of these imitative changes for almost all participants. These results confirm that convergence can occur even in a non-interactive situation of communication, in line with previous studies (Goldinger and Azuma, 2004; Gentilucci and Cattaneo, 2005; Delvaux and Soquet, 2007; Garnier et al., 2013; Sato et al., 2013).

Such non-interactive phonetic convergence effects likely rely on low-level sensory-motor mechanisms described in the Introduction section, according to which participants would first analyze the target stimulus to elaborate their motor and sensory goals, resulting in an adaptation of their production to the acoustic target.

4.2. Elderly population

The results for elderly participants with normal hearing were similar to those of young adults. Indeed, NHE imitated the auditory target in the imitative task and converged towards the acoustic target in a probably subconscious way in the convergence task. In addition, as for young participants with normal hearing, no correlation was observed between the two tasks. The observed convergence effects in both voluntary and unintentional imitation in this elderly population, as high as in young subjects, suggests that sensory-motor relationships during speech production are efficient in these participants, which is in line with previous studies suggesting that sensory-motor relationships are still active and efficient in seniors (Liu et al., 2010, 2011a, 2011b; Tremblay et al., 2013).

4.3. Post-lingually deaf cochlear-implanted participants

In spite of their strongly impaired ability to discriminate frequencies, as displayed by JND values 10–15 times larger than those of
subjects with normal hearing, it is striking that CI participants were also found to be able to imitate and to converge towards the acoustic targets. This suggests that they are able to estimate the pitch of a vowel target to a certain extent and to monitor their own vocal source to attempt to get closer to this target. Even more strikingly, they do it even in the convergence task where no explicit imitation instruction was provided – hence the importance to know that all participants performed this task before the “imitative task”, reducing the risk that they could have considered that imitation mattered in the task.

Although these results demonstrate that cochlear-implanted patients are able to perceive and imitate the fundamental frequency of the vowel targets, no correlation was however found between their imitative abilities and their auditory discrimination scores in the JND test. This suggests that their imitation abilities are not related to their auditory capacity of perceiving frequencies in a simple way. Nevertheless, cochlear-implanted participants showed less convergence and imitation than adults with normal hearing whatever their age, and higher pitch variability between participants. At this stage, it is unclear if pitch variability in production is due to inaccurate estimation of the target pitch or to degraded voice control, but interestingly, the lack of correlation between convergence and JND suggests that apart from sensory ability, the sensory-motor relationships could differ from one subject to the other. Considering the importance and efficiency of somatosensory control for deaf subjects (Nasir and Ostry, 2008), together with the range of inter-individual variations in the relative importance of auditory and somatosensory control (Lametti et al., 2012; Feng et al., 2011), it is likely that variations in somatosensory dependence could be at hand in these data.

Contrary to Blamey et al. (2012) who showed that deafness duration and age at deafness onset are two factors influencing speech perception performance, we did not find any correlation between these factors and CI abilities to imitate or to converge towards an acoustic target. However, it is important to note that in the Blamey et al. (2012) study, residual variability not explained by CI- and deafness-related factors was considerable (representing 90% of variation in their study). It seems therefore likely that the limited sample of our CI participants impeded any clear correlation with CI- and deafness-related characteristics to appear in the present study.

Finally, contrary to NH participants, CI participant’s responses were not always correct (with up to 13–15% errors in each task), which suggests some auditory difficulties in the decoding of isolated vowels. Actually, it is known that CI listeners need formant transition to accurately categorize a vowel, as shown by Hanna (2011). The isolated vowels used in our study are hence probably difficult to categorize efficiently by the CI participants. Moreover, convergence effects tested in error trials appear to be similar to those of correct trials, which suggests that timbre and pitch estimations seem to be relatively independent for CI participants. Nevertheless, the major result of this study is that a number of CI participants have already recovered a good ability to associate auditory with motor parameters even a short time after implantation (with significant convergence after only one, two or three months of implantation for some subjects). Indeed, results displayed both automatic and conscious imitation abilities in postlingually deaf cochlear-implanted patients, displaying their ability to adapt their production to the environment, which is uniquely due to a functional sensory-motor linkage. This should be crucial for the retuning of their speech production system (see Perkell et al., 1992), enabling them to improve their internal model (Perkell et al., 2000).

Interestingly, recovering perceptuo-motor abilities could also be of importance for their speech perception abilities, considering the proposals about the role of the motor system in speech perception (e.g. Skipper et al., 2007; Schwartz et al., 2012).

5. Conclusion

The present study reports convergence and imitation abilities in adults with normal hearing, both young and elderly, and in cochlear-implanted subjects. Indeed, in the three groups, we obtain significant and generally strong imitation of the acoustic target during a task with explicit instruction, and also convergence effects in a task without direct instruction. These results first confirm that sensory-motor relationships during speech production are still efficient in elderly subjects, with basically no difference with younger ones. Moreover, crucially, we also show for the first time that cochlear-implanted participants have the ability to converge to an acoustic target, intentionally and unintentionally. Therefore, they have recovered significant sensory-motor abilities, which could be crucial for improving both their speech production and speech perception abilities.

Competing financial interests

The authors declare no competing financial interests.

Acknowledgments

This work was supported by the French National Research Agency (ANR) through funding from the PlasMody project (Plasticity and Multimodality in Oral Communication for the Deaf, ANR-11-BSH2-0008).

References


Gueneh, F.H., Vladonich, T., 2012. A neural theory of speech acquisition and...


