



HAL
open science

Speech recovery and language plasticity can be facilitated by Sensori-Motor Fusion (SMF) training in chronic non-fluent aphasia. A case report study

Céline Haldin, Audrey Acher, Louise Kauffmann, Thomas Hueber, Emilie Cousin, Pierre Badin, Pascal Perrier, Diandra Fabre, Dominic Pérennou, Olivier Detante, et al.

► **To cite this version:**

Céline Haldin, Audrey Acher, Louise Kauffmann, Thomas Hueber, Emilie Cousin, et al.. Speech recovery and language plasticity can be facilitated by Sensori-Motor Fusion (SMF) training in chronic non-fluent aphasia. A case report study. *Clinical Linguistics & Phonetics*, 2018, 32 (7), pp.595-621. 10.1080/02699206.2017.1402090 . hal-01651198

HAL Id: hal-01651198

<https://hal.science/hal-01651198v1>

Submitted on 3 Dec 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Speech recovery and language plasticity facilitated by Sensori-Motor Fusion (SMF) training in chronic non-fluent aphasia. A case report study

Céline Haldin¹, Audrey Acher², Louise Kauffmann³, Thomas Hueber⁴, Emilie Cousin^{1,5}, Pierre Badin⁴,
Pascal Perrier⁶, Diandra Fabre⁴, Dominic Pérennou^{1,7}, Olivier Detante⁸, Assia Jaillard⁵, Hélène
Loevenbruck¹, Monica Baciú¹

¹ Laboratoire de Psychologie et NeuroCognition, UMR CNRS 5105, Université Grenoble Alpes, Grenoble, France

² Unité neuro-vasculaire, Pavillon de Neurologie, CHU Grenoble Alpes, Grenoble, France

³ IRMaGE, Plate-forme IRM 3T, CHU Grenoble Alpes, Université Grenoble Alpes, CNRS, INSERM, UMS3552, Grenoble, France

⁴ GIPSA-lab, UMR CNRS 5216/Université Grenoble Alpes, Grenoble, France

⁵ Dept of NeuroRehabilitation, CHU Grenoble Alpes, Université Grenoble Alpes, Université Grenoble Alpes, Grenoble, France

⁶ Neural Mechanisms of Human Communication Research Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Corresponding Author

Monica Baciú

LPNC, UMR CNRS 5105

BSHM, Université Grenoble Alpes

BP 47 38040 Grenoble Cedex 09, France

Email: monica.baciú@univ-grenoble-alpes.fr

Abstract

The rehabilitation of speech disorders benefits from providing visual information which may improve speech motor plans in patients. We tested the proof of concept of a rehabilitation method (Sensori-Motor Fusion, SMF; Ultraspeech player) in one post-stroke patient presenting chronic non-fluent aphasia. SMF allows visualisation by the patient of target tongue and lips movements using high-speed ultrasound and video imaging. This can improve the patient's awareness of his/her own lingual and labial movements, which can, in turn, improve the representation of articulatory movements and increase the ability to coordinate and combine articulatory gestures. The auditory and oro-sensory feedback received by the patient as a result of his/her own pronunciation can be integrated with the target articulatory movements they watch. Thus, this method is founded on sensorimotor integration during speech. The SMF effect on this patient was assessed through qualitative comparison of language scores and quantitative analysis of acoustic parameters measured in a speech production task, before and after rehabilitation. We also investigated cerebral patterns of language reorganisation for rhyme detection and syllable repetition, to evaluate the influence of SMF on phonological-phonetic processes. Our results showed that SMF had a beneficial effect on this patient who qualitatively improved in naming, reading, word repetition and rhyme judgment tasks. Quantitative measurements of acoustic parameters indicate that the patient's production of vowels and syllables also improved. Compared with pre-SMF, the fMRI data in the post-SMF session revealed the activation of cerebral regions related to articulatory, auditory and somatosensory processes, which were expected to be recruited by SMF. We discuss neurocognitive and linguistic mechanisms which may explain speech improvement after SMF, as well as the advantages of using this speech rehabilitation method.

Keywords: Sensori-motor fusion; therapy; non-fluent aphasia; speech disorder

1. Introduction

Aphasia occurs in about 38% of stroke cases (Cortese, Riganello, Arcuri, Pignataro, & Buglione, 2015) and results from the damage of specific cortico-subcortical regions in the left hemisphere predominant for language (Bookheimer, 2002; Démonet, Thierry, & Cardebat, 2005). Non-fluent (NF) or expressive (Broca's) aphasia impacts mainly the oral production of words and sentences (Ballard, Granier, & Robin, 2000) and is induced by lesions located within and around the left inferior frontal gyrus (Ardila, Bernal, & Rosselli, 2016). NFA is characterised by scarce production of speech, poor articulation, short sentences with only a few words and agrammatism (Benson & Ardila, 1996). Commonly used interventions for phonetic disorders in non-fluent aphasia rely on auditory skills. Patients compare their own productions with those of the speech therapist (ST) and try to correct for errors using auditory cues. Supplementing traditional audiovisual information (hearing the voice and seeing the face of a therapist), with visual information on speech articulators that are usually hidden and not visible (such as the internal movements of the tongue) has recently been introduced in speech therapy. It has been shown that the visual modality, which typically provides information on visible articulators such as the lips, jaw, face, tongue tip and teeth, plays a key role in speech perception, even in participants without speech or hearing impairments (see the McGurk effect, McGurk & Macdonald, 1976; see also the role of vision during speech perception in noise, Sumbly & Pollack, 1954). Although visible information on the lips, mouth and face are informative, they are phonetically limited. The internal tongue shapes and movements, which can only be partially seen, convey crucial phonetic information. Recent studies suggest that full vision of internal tongue movements can be natively used by human participants to facilitate speech perception (Badin, Tarabalka, Elisei, & Bailly, 2010). In addition to classical lip reading, tongue reading may therefore contribute to improved speech representations and it could be beneficial in speech therapy. A few recent studies on speech therapy in hearing impaired populations suggested that providing visual information on the movement of speech articulators indeed adds another modality which may supplement auditory information and may contribute to more adapted speech motor plans (Bacsfalvi, Bernhardt, & Gick, 2007; Bernhardt, Gick, Bacsfalvi, & Adler-Bock, 2005; Bernhardt, Gick, Bacsfalvi, & Ashdown, 2003). Multisensory training programs that combine visual, auditory and kinesthetic information, have also been suggested to improve reading abilities, typically in dyslexia children (see the Orton-Gillingham approach; Sparks, Ganschow, Kenneweg, & Miller, 1991; Vickery, Reynolds, & Cochran, 1987).

In this context, two remediation paradigms that include visual representation of the speech articulators can be used, namely visual biofeedback and visual illustration. Visual illustration is usually built from pre-recorded articulatory data. It displays sagittal views of the entire vocal tract, with the lips, teeth, jaw,

palate, tongue, and even sometimes the velum, pharynx and larynx. These illustrations can either be computed from an articulatory model, itself controlled by articulatory data usually obtained with articulography, or be directly drawn from actual tongue ultrasound imaging. Tongue movement recordings are usually corrected for head movement, and are transformed to fit in a fixed display of the vocal tract that includes the palate, jaw, teeth and pharynx. Both methods provide typical lingual movements set in a realistic and interpretable vocal tract display. Visual biofeedback is another relevant paradigm. It provides online visual information about the patient's articulators, typically the patient's own tongue. The easiest way to obtain online articulatory information about the tongue is by using ultrasound recording. Other methods, such as electromagnetic articulography (EMA), require sticking electromagnetic receptor coils to the tongue and wearing a heavy helmet, which is invasive and impractical in speech therapy (although see Katz, McNeil, & Garst, 2010). With ultrasound, patients can simply hold a probe under their own chin and directly visualise their own tongue shape. The ultrasound recording can be processed online and can even be displayed within a vocal tract drawing to facilitate the interpretation of the tongue movements. Using ultrasound feedback, even without a surrounding vocal tract display, Cleland, Scobbie, and Wrench (2015) showed improved pronunciation of targeted phonemes in children with speech sound disorders. The beneficial effect of visual feedback has also been shown in patients with NFA (Fridriksson et al., 2012), in children and adolescents suffering from developmental speech disorders (Bernhardt et al., 2005) or in adults with apraxia of speech (Preston & Leaman, 2014). Cleland et al. (2015) suggested that real-time visual feedback somehow boosted auditory and somatosensory feedback (patients could hear their sound productions and could feel their articulators in motion), leading to the acquisition of articulatory patterns which were not previously possible through traditional speech therapy. However, because it is displayed online, and because head movements are usually not controlled for in such contexts (unless a helmet is worn, like in the Cleland et al.'s (2015) study, which makes speech therapy demanding), it is difficult to adapt online tongue data to a fixed vocal tract display. Visual biofeedback is therefore sometimes difficult to interpret, because tongue movements may occur outside of the vocal tract drawing. This is particularly critical with patients with speech disorders, whose tongue movement may be erratic.

In this study, we have therefore opted for a visual illustration instead of a visual biofeedback paradigm. Chen et al. (2016) conducted a literature review of studies on computer-based speech therapy systems, many of them based on visual illustration. Their review suggests that such systems provide more engaging intervention than traditional explanation, and that the enriched information provided with the visual illustration paradigms is beneficial in various populations of patients with speech disorders, including hearing-impairments, phonological disorders, or aphasia. Among the several augmented-

speech visual illustration paradigms that have been developed, Ultraspeech player (Hueber, 2013), Talking Head (Badin, Elisei, Bailly, & Tarabalka, 2008; Badin & Serrurier, 2006), Virtual Head (Fagel & Madany, 2008) or Diadolab (Menin-Sicard & Sicard, 2012) are promising. The Talking Head (TH) is both comprehensive, accurate and realistic (Badin et al., 2007, 2008). It allows the movements of many articulators (tongue, velum, lips, jaw, face) to be visualised, and these movements are derived from existing articulatory data on one speaker. It is based on 3D models of the speech organs (jaw, tongue, lips, velum and face), built from magnetic resonance imaging, computer tomography and video data acquired from one speaker and aligned on a skull-related reference coordinate system. The animation of the TH is based on motion capture obtained from one speaker (with EMA and video), so that the articulatory dynamics of the different organs is natural. Badin et al. (2010) showed that participants can derive phonetic information (typically on consonants) from the augmented visual display offered by the TH. They suggest that the TH could be used in speech therapy or second language acquisition. The virtual head developed by Fagel (Fagel & Clemens, 2004) includes a visual articulation module based on a simplified dominance model adapted from Cohen and Massaro (1993) and an audio synthesis module based on the MBROLA (Dutoit, Pagel, Pierret, Bataille, & Van Der Vrecken, 1996) speech synthesis engine. It provides a variety of synthesised utterances. It is therefore less accurate and naturalistic than the TH. Yet its integration in a tool designed for speech therapy ("Vivian") has proven to improve sibilant production ([s, z]) in a few children who trained with it (Fagel & Madany, 2008). The authors suggest that it may be applicable to speech therapy. The Diadolab software developed by Menin-Sicard, Sicard, & Bézard, (2016) displays simplified children-size articulatory contours that are inspired from articulatory data. A test with two hearing-impaired children showed that their speech intelligibility improved after training with Diadolab. Ultraspeech-player (Hueber, 2013) is based on a large database of tongue ultrasound movies that were synchronously captured with the audio signal and video recordings of the lips. It includes productions of phonemes and syllables, as well as words and short sentences uttered by several speakers. Since is based on real data, and not reconstructed from articulatory modelling, the articulatory dynamics, and crucially coarticulation patterns, is well preserved and realistic. It allows audio, ultrasound and video data to be displayed with varying rates, so that the tongue gesture can be slowed down to be better observed. By combining visual and auditory perceptual information during speech production training, this software presumably targets crucial speech processes. Speech is grounded or embodied in sensorimotor experiences, and involves intertwined sensory and motor processes (Guenther & Vladusich, 2012; Hickok, 2012; Indefrey & Levelt, 2004). A consequence of this embodiment is that speech abilities can be improved by using rehabilitation strategies targeting both action execution and observation (Marangolo et al., 2010). Improving the representation of the patient's own articulatory movements could induce better selection, production and

combination of speech gestures. Moreover, complementing articulatory training with visual perception of the articulators could be beneficial as visual information would be integrated by the central nervous system and used within the speech production network. Fabre et al. (2016) showed that using Ultraspeech-player software boosted articulatory awareness and improved the performance of children with phonological disorders (substitution of /tʃ/ with [kʃ]) after two rehabilitation sessions. This rehabilitation method seems to be beneficial to patients with speech disorders. In this study, we have therefore chosen to use the Ultraspeech-player instead of other visual illustration systems, given that it is based on real data: the observable tongue contour movies are not derived from an articulatory model but were recorded on actual speakers. The dynamic tongue movements displayed are therefore natural and realistic. In addition, Ultraspeech-player offers to display any types of speech sounds, from isolated phonemes to short sentences. Another therapeutic advantage is that the movies (and sound) can be slowed down to be watched at a more comfortable rate by the patient. Finally, the software is easy to use, does not require any cumbersome apparatus (helmet, probe) and can therefore be used at home, autonomously by the patient. This means that the patient can practice at home, daily, in addition to weekly or monthly training with a clinician.

In this study, we evaluated the effect of Sensory-Motor Fusion (SMF) training, using Ultraspeech-player software, on the speech recovery of one patient with NF chronic aphasia. Our method, referred to as SMF, is similar to that used in other visual illustration studies (with e.g. Ultraspeech player, Talking Head) which rely on visual and auditory information. We opted for the term “Sensory-Motor Fusion” as it evokes two important properties of this kind of approach: (a) it requires interaction and integration between sensory and motor aspects of speech; (b) it is backed up by a theoretical model of speech production (Hickok’s model) based on sensory and motor components. We thus speculate that SMF training could improve the interactions between auditory and somatosensory targets and motor programs. The SMF effect was assessed with qualitative comparison of scores during several language tests and quantitative analysis of the acoustic speech signal recorded in a speech production task, before and after rehabilitation. In addition, fMRI data were acquired for two tasks (syllable repetition and rhyme detection) to assess the SMF effect on the activation pattern for these tasks, pre- and post-rehabilitation.

Overall, through this case report we aim to illustrate the usefulness of a novel speech rehabilitation approach, to be used in non-fluent aphasic patients, targeting sensorimotor integration. The focus of this study is the detailed presentation of the SMF method and the multi-modal assessment of its effects. Specifically, the aim of this study was to evaluate the effect of SMF training on speech recovery in one patient with NF chronic aphasia, by using the Ultraspeech-player software. The SMF effect was

assessed with two methods used before and after training: (a) the qualitative comparison of language scores (language assessment), and (b) the quantitative analysis of the acoustic speech signal based on several parameters related to vowel and consonant production. It is important to underline that this patient did not receive any direct feedback (i.e. via mirrors or ultrasound imaging) of her own productions. In addition, collecting fMRI data during two language tasks (Rhyme Detection, RD and Syllable Repetition, SR), we evaluated the effect of SMF on language networks after training. These tasks were chosen because they require cognitive and linguistic processes that bear on our hypothesis on the potential role of SMF, namely, the improvement of phonological, phonetic and articulatory mechanisms, due to the sensorimotor integration. Sensorimotor integration in Hickok's model (Hickok, 2012) is described in terms of two loops, somatosensory(articulatory)-motor and auditory-motor. These two loops could be activated by using a rhyme detection task and a syllable repetition task, which both strongly rely on somatosensory and auditory representations as well as motor programs. We speculate that SMF training could improve interactions between auditory and somatosensory targets resulting in more accurate motor programs which would improve RD and SR tasks resulting in a modulation of their cerebral substrates. A group of control participants was included in the fMRI experiment to obtain reference maps for the language tasks.

2. Materials and methods

2.1. Patient 1 (P1)

The patient's clinical picture and demographic information is presented in Supplementary Material A (Figure A.1 and Table A.1). This patient shows a typical NFA induced by a lesion restricted to frontal regions after ischemic stroke induced by the obstruction of anterior ramifications of the left medial cerebral artery, and associated with right hemiparesis. This study was approved by the Ethics Committee (CPP-ISIS 07PHR04-N°DCIC/06/25).

2.2. SMF rehabilitation procedure

SMF training was carried out using the Ultraspeech-player software (Hueber, 2013). Ultraspeech-player played sounds simultaneously with the videos of tongue movements recorded on a reference non-pathological speaker (typically the ST) using ultrasound and video imaging. P1 was seated in front of a computer screen and was presented with the production of various isolated vowels, isolated consonants, and vowel-consonant-vowel (VCV) sequences. The sounds and tongue movements were presented synchronously. If necessary, the movements were presented in slow-motion. After each presentation, P1 was invited to produce the sound. The targeted sounds were: isolated French vowels, V: /i, e, ε, a, y, ø, œ, u, o, ɔ, ā, ã, ã̃, ẽ̃/; isolated French consonants, C: /p, t, k, b, d, g, f, v, s, z, ʃ, ʒ, l, m, n,

ʋ, ɲ, j/ and VCV sequences. The rehabilitation started with an initial session in which the ST provided several explanations. Prototypical tongue images were visualised by P1 while the ST described the movements, so that P1 could train to better place the tip, the dorsum and the back of the tongue and could better pronounce the targeted phonemes and syllables. After this guided session, the SMF training included 13 training sessions at home, autonomously, during which P1 watched the target items and repeated them. P1 could hear herself and thus received auditory feedback from her own productions. Then P1 performed 3 supplementary sessions guided by the ST. Each training session lasted 45 minutes. Figure 1 illustrates the Ultraspeech-player software (www.ultraspeech.com).

--- Insert Figure 1 ---

2.3. Assessment of the SMF effect

2.3.1. Qualitative language assessment

Language abilities of P1 were evaluated by a ST during two evaluation sessions, before and after SMF training. Language production was evaluated with several tests generally used in our clinical environment, to access stages of word production (picture naming, DO80; Metz-Lutz et al., 1991), phonological output, buffer and phonetic encoding (syllables and words repetition), naming of written items (letters, syllables and words reading), phonological input (rhyme judgment; Baudo & Vernisse, 2001) and lexical availability (phonemic and categorical fluency). In addition, ST explored verbal working memory (sentence repetition, rated regardless of the articulatory precision) as well as conceptual access and language comprehension (picture designation and morphosyntactical comprehension task, MT-86; Nespoulous, Lecours, & Lafond, 1986). Tests and scores obtained before and after rehabilitation with SMF, are presented in Table 1.

--- Insert Table 1 ---

2.3.2. Quantitative speech assessment

We performed analysis of the acoustic parameters obtained in a speech production task before and after SMF. As detailed in Table A.2, the audio signal was recorded during several repetitions of various monosyllabic or bisyllabic words (or pseudo-words when no words could be found) that included all French vowels and consonants (e.g. “ou” /u/, “where”, “gars” /gal/, “guy”). This allowed the production of all French phonemes inserted in real words (or in regular pseudo-words) to be tested before and after the SMF training. Although it would have been interesting to assess the patient’s ability on longer and more complex words, it was chosen to limit the recordings to simpler words, to make the task feasible (see the patient’s severe deficits in the word repetition task in the language test, Table 1). Three

repetitions were performed during each evaluation. The first one consisted in the production of isolated vowels and of consonants in /a/ vowel context in different sequences, with the consonants (C) placed in initial (/Ca/, /CaCa/), medial (/aCa/) or final (/aC/) positions (latter analysis not described in this article). The other two repetitions only consisted of the production of isolated vowels and of /Ca/ syllables. Table A.2 (Supplementary Material A) provides the corpus with written alphabetic forms as well as phonetic transcriptions (using IPA, the International Phonetic Alphabet).

2.4. Functional MRI assessment

P1 performed two language tasks (two separate runs), Syllable repetition (SR) and Rhyme detection (RD), using visual (RD) and auditory (SR) presentation of stimuli. Runs were block-designed and alternated task and control periods with similar features (5 periods of task/ control, 24 sec/period, 4 stimuli/period). P1 gave verbal (oral) responses during SR and manual responses during the RD task. Oral responses given during SR were recorded via an MRI compatible optical microphone (FOMRITM III, version 1.2). The total duration of the two runs was about 10 min.

2.5. Syllable repetition

During task periods, P1 repeated French mono-syllabic sequences, pronounced by a female voice and transmitted to P1 via MRI-compatible headphones. Syllables were composed of the 10 French oral vowels /i, e, ε, a, y, ø, œ, u, o, ɔ/ and of the 10 lingual consonants /t, d, s, z, ʃ, ʒ, k, g, l, ʁ/ followed by vowel /a/. Each stimulus lasted the time of the production of the syllable and participants were required to repeat the syllable only once. During control periods, white noise was presented, without any response required from P1.

2.6. Rhyme detection

During task periods, P1 was required to judge whether written words presented in pairs rhymed or not. Half of them were phonologically and orthographically congruent (R + O+), the other half being phonologically and orthographically incongruent (R-O-) (Baudo & Vernisse, 2001). Words were written in black, one above the other. During control periods, two horizontal lines were presented, P1 being instructed to simply watch them.

2.7. MR acquisition

fMRI experiment was performed in a whole-body 3T MR scanner (Philips Achieva TX – Grenoble MRI facility IRMaGE) with a manufacturer-provided gradient-echo/T2*-weighted EPI sequence. Fifty-two adjacent axial slices parallel to the bi-commissural plane were acquired in non-interleaved ascendant

mode. Acquisition parameters were: slice thickness 2.75 mm with a gap of 0.25 mm; in-plane voxel size $2.5 \times 2.5 \times 3$ mm ($220 \times 220 \times 156$ mm field of view encoded with an 88×85 voxels matrix); TR = 3 sec, TE = 30 msec, flip angle = 80° . A T1-weighted high-resolution ($1 \times 1 \times 1$ mm) three-dimensional anatomical volume was also acquired.

2.8. Data processing

2.8.1. Language scores

Before participating in the study, P1 underwent language assessment with various tests, stimuli and speech-therapists, so that we could not apply a multiple measures baseline method (Nickels, 2002). The evaluation of language scores remained qualitative and we only calculated global evolution indices (EI) based on the comparison of two scores, one before and another one after SMF, for each test. To calculate a global EI for each test (Table 1), we analysed each item; we first attributed a score for each item (+1 if incorrect in pre-training but correct in post-training; -1 if correct in pre-training but incorrect in post-training; 0 if correct in pre- and post-training or if incorrect in pre- and post-training).

2.8.2. Acoustic data

The acoustic signal obtained from the audio recordings before and after SMF training was recorded at 44.1 kHz and downsampled at 12 kHz in order to facilitate the extraction of phonetically relevant spectral parameters. Each recording contains the ST's stimulus productions followed by P1's productions. Before analysis, isolated vowels and syllables were manually labelled using the Praat software (Boersma & Weenink, 2010). All the isolated vowels and all the consonants in the 3 contexts (/Ca/, /CaCa/, /aCa/) were rated for accuracy (78 oral vowels and 216 consonants for the ST and 78 oral vowels and 216 consonants for the patient). The phonemes were labelled as either "correct", "incorrect", or "emergent". In our context we call emergent a phoneme that is preceded by a non-target sound before being reached. It often occurs with voiced sibilants: for a target /v/ phoneme, the patient may start with an initial [f] that gradually transforms into [v]. When the place and manner of articulation were judged correct and only voicing was incorrect, the label was "voicing error". In addition, typical acoustic parameters were examined for each repetition of the 10 oral vowels, of the 6 fricative (/f, v, s, z, ʃ, ʒ/) and of the 6 occlusives (/p, t, k, b, d, g/) consonants. Namely, we measured the first two formant frequencies (F1 and F2) for vowels, voice onset time (VOT) for occlusives and the first two spectral moments (Centre of Gravity, CG and Standard Deviation, SD) for fricatives (the transcription and coding were carried out by a trained transcriber and reliability of the scoring was ensured by having a second rater judge part of the data, see details in Supplementary Material A).

2.8.3. MR data processing

Data analysis was performed using the general linear model (Friston et al., 1994) with SPM12 (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB (Mathworks Inc., Sherborn, MA, USA).

2.9. Pre-processing of data

Data analysis was performed by using the general linear model (Friston et al., 1994) with SPM12 (Wellcome Department of Imaging Neuroscience, London, UK, www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB (Mathworks Inc., Sherborn, MA, USA). Images were initially preprocessed. First, the functional volumes were time-corrected with the mean image as the reference slice (the acquired brain volume was composed of 52 slices) in order to correct artefacts caused by the delay of time acquisition between slices. Subsequently, all volumes were realigned to correct for head motion, by using a rigid body transformation. For healthy participants, T1-weighted anatomical volume was coregistered to mean images created by the realignment procedure and was normalized within the MNI (Montreal Neurological Institute) space. The anatomical normalization parameters were subsequently used for the normalisation of functional volumes. Finally, each functional volume was smoothed by a Gaussian kernel of 8 mm FWHM (Full Width at Half Maximum). For patients, masks covering brain lesion were applied to the T1-weighted anatomical volume which was co-registered to mean images created by the realignment procedure. The masked anatomical volume was then segmented into six different tissues (white matter, grey matter, cerebrospinal fluid, bone, soft tissue and background) and normalised into the MNI space using the tissue probability maps provided in SPM12. The normalisation parameters were subsequently used for the normalization of functional volumes. Each functional volume was then smoothed by a Gaussian kernel of 7 mm FWHM (Full Width at Half Maximum). For both groups of participants, time series for each voxel were high-pass filtered (1/128 Hz cut-off) to remove low-frequency noise and signal drift.

2.10. Statistical analyses of data

Statistical analyses were subsequently performed on the preprocessed data. For each participant, each language paradigm (SR and RD) was declared as a specific fMRI run. It should be noted that since for P1 each paradigm was performed twice (before and after SMF), two runs were declared. For each run, the Task and Control conditions were modelled as two regressors constructed as box-car functions, convolved with a canonical hemodynamic response function in healthy subjects and with a canonical Hemodynamic Response Function (HRF) with its spatial and temporal derivatives (3HRF) for P1.

Movement parameters derived from the realignment corrections (three translations and three rotations) were included into the design matrix as additional factors of no interest. We additionally identified motion-related outlier scans using the artefact detection toolbox (http://www.nitrc.org/projects/artifact_detect/) which were also modelled in the design matrix as regressors of no interest. In P1, the percentage of correct responses (CR) and the reaction time (RT ms) were also included in the design matrix as additional regressors of no interest to account for RT- and accuracy-related variance. The general linear model was then used to generate the parameter estimates of activity for each voxel, each condition and each participant. Statistical parametric maps were generated from the linear contrasts between the HRF parameter estimates for experimental conditions. The spatial resolution of statistical parametric maps was the same as the spatial resolution of functional MR images (3x3x3.5 mm). The statistical analysis was performed at a first level (Individual level analysis). We calculated the main contrast Task vs. Control, for each run (SR, RD). This contrast allowed us to identify language networks involved in syllable repetition (sub-lexical level, audio-motor and sensori-motor interactions) and rhyme detection (phonological processes, phoneme monitoring, phonological buffer and lexico-semantic associated processes). For healthy participants, contrast images were then entered into second-level random effect analyses to test for within group effects. We also computed the contrasts between the two acquisition sessions (before and after SMF) for each task, to examine the effect of SMF on cerebral activation.

3. Results

3.1. Language scores

Table 1 summarises the aphasic and language symptoms in P1 and their evolution after the SMF training. Before rehabilitation, P1 showed naming deficit with articulatory impairment, oral repetition and reading deficits. The level of verbal fluency was lower than the normal values. An analysis of each item indicated that her language production was characterised by phonetic and phonological paraphasias. P1 was not impaired for language comprehension and did not show apraxia of speech. After rehabilitation, based on the global EI for each test, P1 shows general improvement of language production, excepting a decline in phonemic fluency.

3.2. Acoustic results

Table 2 displays accuracy transcriptions for all the segments (all vowels and all consonants), including nasal vowels and the 6 consonants that were not further taken into account in the acoustic analysis. Accuracy transcriptions for each phoneme separately are provided in Supplementary Material A Table A.3.

--- Insert Table 2 ---

Before the SMF training, the productions were correct for 48.98% of all phonemes, with 94.87% for vowels and 32.41% for consonants. After training, the percentage of correct vowels increased to 97.44% and the percentage of correct consonants increased to 43.64%. Voicing errors for consonants decreased (from 24.07% to 17.27%) which suggests that P1 improved her laryngeal control after training. Productions rated as “incorrect” also decreased, which suggests that improvement was not restricted to voicing control but also concerned the place and the manner of articulation. The percentage of productions labelled as “emergent” was also reduced, which argues for an improvement of the dynamic control of phoneme production. Moreover, in order to quantitatively describe the evolution in vowel and consonant production, acoustic measures were carried out on the phonemes rated as correct or emergent. Figure 2 provides vowel distributions in the (F1, F2) plan before and after training. Importantly, P1 and speech-therapist were female, allowing the comparison of their vocalic spaces. The ST’s vowel (F1, F2) plan shows typical small intra-vowel variability (good clustering of productions of each vowel across several repetitions) and large inter-vowel distances (no overlap between vowels). For P1, results for correctly judged vowels (Figure 2, top panel) show an increased similarity for the pronunciation of each of the /i, y/ vowels with a decreased intra-vowel variability after SMF, as well as an increased distance between these two vowels.

--- Insert Figure 2 ---

Results for occlusives are presented in Table 3 with Voice Onset Time (VOT) values, before and after SMF and for the ST as a reference. VOT values for ST are stable in the 2 sessions and show the typical distinction between voiced and unvoiced occlusive. For P1, the VOT values of unvoiced occlusives are stable, positive and in the range of normal production. For voiced occlusives, the VOT values decrease after training and become more negative and closer to the reference. Finally, for fricative consonants (Figure 3), P1 shows inability to produce alveolar fricatives before rehabilitation (absence of the alveolar category in Figure 3a and c). However, after rehabilitation, P1 correctly produced two of the six repetitions of the /s/ consonant (which explains the presence of the alveolar category in Figure 3b and d).

--- Insert Table 3 and Figure 3 ---

Examining the pattern of production errors, not just the accuracy scores, is also informative. For stops, errors mainly consisted in voicing both before and after therapy, the place of articulation being quite well preserved. Before therapy, the alveolar fricative /s/ could not be produced at all, whereas after therapy,

when it became produced, its place of articulation was mostly affected. Instead of the alveolar /s/, the patient sometimes produced the post-alveolar fricative [ʃ] (with correct manner of articulation but backed tongue position) or the labiodental [f] (correct manner but fronted place). This reveals a difficult maintenance of tongue position during friction, when palate contact is not possible. The fact that SMF training improved fricative production suggests that it is effective in re-creating accurate somatosensory representations.

Overall, partial improvement for vowels as well as for consonants was observed for P1 after rehabilitation.

3.3. fMRI results

fMRI result obtained before and after SMF are shown in Table 4 and Figure 4. SMF training modulated brain activity within the RD network in left superior and medial frontal and precentral gyri, right frontal middle and superior, right superior temporal gyri, as well as right cerebellum (Crus I). SMF training modulated also brain activity within the SR network within the left inferior frontal gyrus and left insula.

--- Insert Table 4 and Figure 4 ---

4. Discussion

This study is a proof of concept for using a novel speech rehabilitation method, the SMF with the Ultraspeech-player software, in chronic patients with non-fluent aphasia. The main objective of this work was to provide a description of the method and the procedure used as well as a multimodal evaluation of the effects. SMF effects were assessed by a qualitative comparison of language scores and by a quantitative analysis of acoustic characteristics of vowel and consonant production, before and after rehabilitation. Moreover, we examined how the SMF training modulates the cerebral networks elicited by two appropriate language tasks, RD and SR. It should be noted that the patient was included more than 1 year after stroke. It is very unlikely that spontaneous recovery still occurs so late after stroke (Lazar, Speizer, Festa, Krakauer, & Marshall, 2008; Lendrem & Lincoln, 1985). Consequently, it is safe to consider that the observed modulation of cerebral language networks, as assessed with fMRI, is the consequence of the rehabilitation training. RD and SR tasks mainly require phonological, phonetic and articulatory processes targeted by the SMF, such as sensori-motor, audio-motor and visuo-motor interactions. These interactions are supposed to improve internal models and motor plans necessary in speech production.

The patient P1 examined here shows typical NFA induced by a lesion of frontal regions after ischemic stroke induced by the obstruction of anterior ramifications of the left medial cerebral artery, and associated with right hemiparesis. The qualitative comparison of language scores (Table 1) indicates that after SMF training, P1's language production abilities improved, specifically for articulatory and phonological processes. Indeed, excepting a decline of verbal fluency, EI values (Table 1) for all tests were positive, indicating that SMF training might have had a positive effect. However, language improvement was not complete. Verbal fluency score remained low after SMF, suggesting remaining difficulties with word retrieval, attributable to an association between domain-specific (language) and domain-general (executive functions or cognitive control) deficits, frequently observed in patients with aphasia (Brownsett et al., 2014).

In addition, quantitative acoustic analyses revealed that P1 significantly improved pronunciation of several vowels and consonants after rehabilitation. Acoustic results on vowels (see Figure 2) suggest that rehabilitation facilitated lip rounding control (which distinguishes /i/ and /y/ vowels). In addition, the F2 values obtained for /o/ and /u/ are better differentiated (increased inter-vowel distance) after training, suggesting a better control of the tongue height in posterior position (which distinguishes between these two vowels). However, the rehabilitation did not reveal improved reproducibility across repetitions for /a, œ, e/, suggesting that tongue position remains imperfectly controlled (see Figure 2). As illustrated in Tables 2–3 and Figure 3, acoustic results on occlusive consonants suggest that P1 improved her production of voicing after training. For fricative consonants, acoustic results showed that the /s/ and /ʃ/ distinction improved after SMF training, as well as the distinction of labiodentals /f, v/ from the other fricatives /s, z/ and /ʃ, ʒ/. Altogether, these patterns suggest that supplying visual articulatory information improved articulator placement and movement. Since voicing control was also improved, SMF seems to have additionally facilitated the coordination between laryngeal and supra-laryngeal movements. A better mastery of articulatory configurations and dynamics could have allowed her to better concentrate on an additional gesture (laryngeal closing and opening).

At the cerebral level, the comparison before vs. after rehabilitation revealed interesting results. We first discuss the effects related to the RD task. As illustrated in Table 4 and Figure 4, SMF modulated brain activity within the RD network in left superior-medial frontal and left precentral gyrus, right frontal middle and superior and superior temporal gyri, as well as the right cerebellum (Crus I). Left frontal and precentral activation might suggest supplementary audio-motor and sensori-motor interactions, as suggested by recent speech motor control models (Guenther & Vladusich, 2012; Hickok, 2012). Previous studies indicated that speech depends on strong relationships between production and perception networks, two functional systems which are intimately linked (Guenther, Ghosh, & Tourville,

2006; Indefrey & Levelt, 2004; Tomasello, Garagnani, Wennekers, & Pulvermüller, 2016) and whose links we assumed to be reinforced by the SMF rehabilitation. Hence our results provide evidence for the efficacy of the SMF rehabilitation in this patient. In addition, the activation of the right cerebellum and specifically the Crus I region, reflects the involvement of word generation processes (Stoodley, Valera, & Schmahmann, 2012). In a meta-analysis, Stoodley and Schmahmann (2009) suggested that among various processes, language and working memory were dependent on the lobules VI-Crus I and Crus II. This suggests that, after SMF, our patient might more strongly resort to linguistic and verbal memory processes to perform RD more efficiently. Our fMRI results indicating neurocognitive efficiency improvement were coherent with the positive EI (Table 1) obtained for rhyme detection in the language assessment. Finally, the right activation of the cerebellum may also be related to an improvement in speech production-related internal models (Hickok, 2012; Ito, 2005). All these processes could be boosted after SMF training. Furthermore, the SMF effect was also observed in terms of supplementary activation of the right middle and superior frontal gyri after training. These regions have been considered to play multiple roles, such as that of a gateway between top-down and bottom-up control of attention (Japee, Holiday, Satyshur, Mukai, & Ungerleider, 2015). This suggests that, after training, our patient might be able to recruit attentional resources to perform the task. The activation of right fronto-temporal regions has also been observed after applying other methods of rehabilitation in non-fluent aphasia, such as melodic intonation therapy, which engages fronto-temporal regions of the right hemisphere (Schlaug, Norton, Marchina, Zipse, & Wan, 2010).

In sum, the effect of SMF on the RD network seems to reflect improvement of sensorimotor interactions, attention, motor planning and internal models for speech, which is consistent with the expected impact of the SMF rehabilitation method.

We now turn to the effect of SMF on the SR network. After training, increased activation of the left inferior frontal gyrus and left insula was observed (see Table 4 and Figure 4). As already mentioned, the activation of the left inferior frontal gyrus after SMF might be related to the increase in sensori-motor interactions due to the supplied visual information and to a better phonological output and phonetic encoding. This finding is in agreement with the positive EI obtained for the repetition test in language assessment (Table 1). The activation of the left insula suggests an improvement in the intra- and inter-syllabic coordination movements during speech (Baldo, Wilkins, Ogar, Willock, & Dronkers, 2011; Dronkers, 1996). In sum, the effect of SMF on the SR network suggests an improvement in sensori-motor interactions and in coordination of articulation in our patient.

In future works, next to SMF, we will explore the acoustic, linguistic, cognitive and neurophysiological effects of other rehabilitation methods, including traditional speech therapy. Moreover, it will be also important to evaluate the effects of SMF when using a visual feedback paradigm (by means of an ultrasound probe), rather than mere visual illustration. We do not have sufficient data yet to compare the specific effects of all these methods, but we are currently including more patients and we will use randomised controlled trials to robustly assess the respective effects of the different methods. In addition, other questions should be addressed, such as for instance, the most appropriate timing for the use of SMF or other speech rehabilitation methods. Presumably, an efficient rehabilitation should not only use a specific method in isolation, but should associate several methods, which should be applied at different optimal times after stroke, and should be patient-tailored, meaning that they should be customised according to clinical variables specific to each patient. Another important aspect that should be highlighted is how much relevant to language recovery these different methods are, according to the post-stroke phase - acute, subacute or chronic. SLT (Speech and Language Therapy) is known to work in the first months after stroke, but we do not know how effective this method could be if was immediately associated with other rehabilitation approaches. Moreover, neurocognitive plasticity and its efficiency (whether cerebral reorganisation is useful for the patient because his/her language skills improve) in patients with non-fluent aphasia after stroke, are not fully described yet. Even if a majority of findings suggest that an efficient language network reorganisation should involve the left hemisphere (the right hemisphere activation typically measured in the acute phase should shift to the left in the chronic phase, Cornelissen et al., 2003; Saur et al., 2006), the activity of which is correlated with improvement in language performance (Léger et al., 2002; Meinzer et al., 2008; Musso et al., 1999; Saur et al., 2006), supplementary evidence is required. Overall, we should consider these results with caution. Essentially, our present study is a proof of concept in that it assesses the feasibility of using SMF training for speech recovery in one patient with post-stroke aphasia. This study has the merit to show that training in autonomy could be useful, engaging and even playful for the patients. Indeed, P1 improved in her speech abilities after training. Her speech production improvement was associated with the activation of a cerebral network typically underlying motor control, auditory and somatosensory processes. It is precisely these specific networks that we expected to be activated by using this speech recovery method. However, this study has several limitations: (a) it is a single case study, which means that the results and the conclusions should be considered with caution and more patients should be tested in the future; (b) a monitoring of how the patient performs during the autonomous training at home is required; (c) the patient underwent both traditional speech therapy and the SMF training in parallel; thus, the acoustic, linguistic and neuroimaging effects are probably also related to traditional therapy, at least in part; future studies with group designs should compare the relative effects of the

different methods. Moreover, there are still many aspects of the therapy that need to be assessed. In our future studies, we will collect more data for a better understanding of (a) the effect of using SMF autonomously vs. in accompaniment with a SLT; (b) the optimal duration of each training session, the ideal session number and frequency, and (c) the respective advantages of SMF training with visual illustration only vs with online visual feedback (with a tongue imaging ultrasound probe). In addition, although autonomous daily training seems to be feasible for the patient, given that the software is easy to use, non-invasive and engaging, we do not know yet whether autonomous daily training is more or less effective than daily or weekly training with a clinician (ST). Training with continuous support from a clinician could also be advantageous, as the ST can monitor the patient's performance, and can help the patient in adjusting articulatory execution and in avoiding error perseverations. In addition to monitoring errors, the clinician can also encourage and stimulate the patient. We think that mixing autonomous training with clinician-guided training would result in optimal rehabilitation scores. The amount of ST support should also be adapted to the post-stroke phase, patients with subacute post-stroke aphasia needing more significant support than patients with chronic aphasia.

Acknowledgments

We thank Geneviève Meloni and Manon Gabriel for their help with the acoustic data processing. The Ultraspeech-player software is free to download at www.ultraspeech.com.

Declaration of interest

The authors report no conflicts of interest.

Funding

This work was partly funded by the French program "Investissement d'Avenir" run by the 'Agence Nationale pour la Recherche' ; grant 'Infrastructure d'avenir en Biologie Santé -ANR-11-INBS-0006.

References

- Ardila, A., Bernal, B., & Rosselli, M. (2016). Why Broca's area damage does not result in classical Broca's Aphasia. *Frontiers in Human Neuroscience*, 10, 249.
- Bacsfalvi, P., Bernhardt, B. M., & Gick, B. (2007). Electropalatography and ultrasound in vowel remediation for adolescents with hearing impairment. *Advances in Speech Language Pathology*, 9 (1), 36–45.
- Badin, P., Elisei, F., Bailly, G., Savariaux, C., Serrurier, A., & Tarabalka, Y. (2007). Têtes parlantes audiovisuelles virtuelles: Données et modèles articulatoires: Applications. *Revue De Laryngologie Otoologie Rhinologie*, 128(5), 289–295.
- Badin, P., Elisei, F., Bailly, G., & Tarabalka, Y. (2008). An audiovisual talking head for augmented speech generation: Models and animations based on a real speaker's articulatory data. *Vth Conference on Articulated Motion and Deformable Objects (AMDO 2008, LNCS 5098) (Vol.5098, pp. 132–143)*.
- Badin, P., & Serrurier, A. (2006). Three-dimensional linear modeling of tongue: Articulatory data and models. In H. C. Yehia, D. Demolin, & R. Laboissière (ed.), *7th International Seminar on Speech Production, ISSP7* (pp. 395–402). Ubatuba, SP, Brazil: UFMG, Belo Horizonte, Brazil.
- Badin, P., Tarabalka, Y., Elisei, F., & Bailly, G. (2010). Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding. *Speech Communication*, 52(6), 493–503.
- Baldo, J. V., Wilkins, D. P., Ogar, J., Willock, S., & Dronkers, N. F. (2011). Role of the precentral gyrus of the insula in complex articulation. *Cortex*, 47(7), 800–807.
- Ballard, K. J., Granier, J. P., & Robin, D. A. (2000). Understanding the nature of apraxia of speech: Theory, analysis, and treatment. *Aphasiology*, 14(10), 969–995.
- Baudo, K., & Vernisse, A.-L. (2001). Influence de la mémoire de travail sur l'ordre de réussite aux épreuves métaphonologiques. (Mémoire pour l'obtention du Certificat de capacité d'orthophonie). Lyon: Université Claude Bernard Lyon 1.
- Benson, D. F., & Ardila, A. (1996). *Aphasia: A clinical perspective*. New York, NY: Oxford University Press.
- Bernhardt, B., Gick, B., Bacsfalvi, P., & Adler-Bock, M. (2005). Ultrasound in speech therapy with adolescents and adults. *Clinical Linguistics & Phonetics*, 19(6–7), 605–617.
- Bernhardt, B., Gick, B., Bacsfalvi, P., & Ashdown, J. (2003). Speech habilitation of hard of hearing adolescents using electropalatography and ultrasound as evaluated by trained listeners. *Clinical Linguistics & Phonetics*, 17(3), 199–216.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 6.0.05). Retrieved from <http://www.praat.org/>

- Bookheimer, S. Y. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, 25(1), 151–188.
- Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain*, 137(1), 242–254.
- Chen, Y.-P.-P., Johnson, C., Lalbakhsh, P., Caelli, T., Deng, G., Tay, D., . . . Morris, M. E. (2016). Systematic review of virtual speech therapists for speech disorders. *Computer Speech & Language*, 37, 98–128.
- Cleland, J., Scobbie, J. M., & Wrench, A. A. (2015). Using ultrasound visual biofeedback to treat persistent primary speech sound disorders. *Clinical Linguistics & Phonetics*, 29(8–10), 575–597.
- Cohen, M. M., & Massaro, D. W. (1993). Modeling coarticulation in synthetic visual speech. In N. Magnenat-Thalmann & D. Thalmann (eds.), *Models and techniques in computer animation* (pp. 139–156). Tokyo, Japan: Springer.
- Cornelissen, K., Laine, M., Tarkiainen, A., Järvensivu, T., Martin, N., & Salmelin, R. (2003). Adult brain plasticity elicited by anomia treatment. *Journal of Cognitive Neuroscience*, 15(3), 444–461.
- Cortese, M. D., Riganello, F., Arcuri, F., Pignataro, L. M., & Buglione, I. (2015). Rehabilitation of aphasia: Application of melodic-rhythmic therapy to Italian language. *Frontiers in Human Neuroscience*, 9, 520.
- Démonet, J. F., Thierry, G., & Cardebat, D. (2005). Renewal of the neurophysiology of language: Functional neuroimaging. *Physiological Reviews*, 85(1), 49–95.
- Dronkers, N. F. (1996). A new brain region for coordinating speech articulation. *Nature*, 384(6605), 159–161.
- Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vrecken, O. (1996). The MBROLA project: Towards a set of high-quality speech synthesizers free of use for non-commercial purposes. *Proceedings ICSLP*, 3, 1393–1396 vol.3.
- Fabre, D., Hueber, T., Canault, M., Bedoin, N., Acher, A., Bach, C., . . . Badin, P. (2016). Apport de l'échographie linguale à la rééducation orthophonique. In N. Joyeux & S. Topouzkhianian (eds.), *XVIèmes Rencontres Internationales d'Orthophonie: "Orthophonie et technologies innovantes"* (pp. 199–225). Paris, France: Ortho Edition.
- Fagel, S., & Clemens, C. (2004). An articulation model for audiovisual speech synthesis: Determination, adjustment, evaluation. *Speech Communication*, 44(1), 141–154.
- Fagel, S., & Madany, K. (2008). A 3-D virtual head as a tool for speech therapy for children. *Proceedings of Interspeech*, 2008, 2643–2646.
- Fridriksson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A. L., Bonilha, L., Fromm, D., & Rorden, C. (2012). Speech entrainment enables patients with Broca's aphasia to produce fluent speech. *Brain*, 135(12), 3815–3829.

- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, 96(3), 280–301.
- Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, 25(5), 408–422.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, 13(2), 135–145.
- Hueber, T. (2013). Ultraspeech-player: Intuitive visualization of ultrasound articulatory data for speech therapy and pronunciation training. *Proceedings of Interspeech (show&tell)*, Lyon, France. (pp. 752–753).
- Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1–2), 101–144.
- Ito, M. (2005). Bases and implications of learning in the cerebellum-adaptive control and internal model mechanism. *Progress in Brain Research*, 148, 95–109.
- Japee, S., Holiday, K., Satyshur, M. D., Mukai, I., & Ungerleider, L. G. (2015). A role of right middle frontal gyrus in reorienting of attention: A case study. *Frontiers in Systems Neuroscience*, 9, 23.
- Katz, W. F., McNeil, M. R., & Garst, D. M. (2010). Treating apraxia of speech (aos) with EMA-supplied visual augmented feedback. *Aphasiology*, 24(6–8), 826–837.
- Lazar, R. M., Speizer, A. E., Festa, J. R., Krakauer, J. W., & Marshall, R. S. (2008). Variability in language recovery after first-time stroke. *Journal of Neurology, Neurosurgery & Psychiatry*, 79(5), 530–534.
- Léger, A., Démonet, J.-F., Ruff, S., Aithamon, B., Touyeras, B., Puel, M., . . . Cardebat, D. (2002). Neural substrates of spoken language rehabilitation in an aphasic patient: An fMRI study. *NeuroImage*, 17(1), 174–183.
- Lendrem, W., & Lincoln, N. B. (1985). Spontaneous recovery of language in patients with aphasia between 4 and 34 weeks after stroke. *Journal of Neurology, Neurosurgery, and Psychiatry*, 48(8), 743–748.
- Marangolo, P., Bonifazi, S., Tomaiuolo, F., Craighero, L., Coccia, M., Altoè, G., . . . Cantagallo, A. (2010). Improving language without words: First evidence from aphasia. *Neuropsychologia*, 48 (13), 3824–3833.
- Mcgurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588), 746–748.

- Meinzer, M., Fleisch, T., Breitenstein, C., Wienbruch, C., Elbert, T., & Rockstroh, B. (2008). Functional re-recruitment of dysfunctional brain areas predicts language recovery in chronic aphasia. *NeuroImage*, 39(4), 2038–2046.
- Menin-Sicard, A., & Sicard, E. (2012). *Intérêt de la visualisation de la position et du mouvement des articulateurs dans la prise en charge des troubles phonologiques*. Paris, France: Les entretiens de Bichat.
- Menin-Sicard, A., Sicard, E., & Bézard, M. (2016). Intérêt de la visualisation de la position et du mouvement des articulateurs pour améliorer l'intelligibilité : Plate-forme Diadolab. In N. Joyeux & S. Topouzkhianian (eds.), *XVIèmes Rencontres Internationales d'Orthophonie. Orthophonie et technologies innovantes* (pp. 261–289). Paris, France: Ortho Edition.
- Metz-Lutz, M. N., Kremin, H., Deloche, G., Hannequin, D., Ferrand, R. A., Perrier, N. D., . . . Larroque, C. (1991). Standardisation d'un test de dénomination orale: Contrôle des effets de l'âge, du sexe et du niveau de scolarité chez les sujets adultes normaux. *Reviews Neuropsychology*, 1(1), 73–95.
- Musso, M., Weiller, C., Kiebel, S., Müller, S. P., Bülau, P., & Rijntjes, M. (1999). Training-induced brain plasticity in aphasia. *Brain*, 122, 1781–1790.
- Nespoulous, J.-L., Lecours, A. R., & Lafond, D. (1986). *MT-86-Protocole Montréal-Toulouse d'examen linguistique de l'aphasie*. Isbergues, France: Ortho-Edition.
- Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology*, 16(10–11), 935–979.
- Preston, J. L., & Leaman, M. (2014). Ultrasound visual feedback for acquired apraxia of speech: A case report. *Aphasiology*, 28(3), 278–295.
- Saur, D., Lange, R., Baumgaertner, A., Schraknepper, V., Willmes, K., Rijntjes, M., & Weiller, C. (2006). Dynamics of language reorganization after stroke. *Brain*, 129(6), 1371–1384.
- Schlaug, G., Norton, A., Marchina, S., Zipse, L., & Wan, C. Y. (2010). From singing to speaking: Facilitating recovery from nonfluent aphasia. *Future Neurology*, 5(5), 657–665.
- Sparks, R. L., Ganschow, L., Kenneweg, S., & Miller, K. (1991). Use of an orton-gillingham approach to teach a foreign language to dyslexic/learning-disabled students: Explicit teaching of phonology in a second language. *Annals of Dyslexia*, 41(1), 96–118.
- Stoodley, C. J., & Schmahmann, J. D. (2009). Functional topography in the human cerebellum: A meta-analysis of neuroimaging studies. *NeuroImage*, 44(2), 489–501.
- Stoodley, C. J., Valera, E. M., & Schmahmann, J. D. (2012). Functional topography of the cerebellum for motor and cognitive tasks: An fMRI study. *NeuroImage*, 59(2), 1560–1570.
- Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The Journal of the Acoustical Society of America*, 26(2), 212–215.

Tomasello, R., Garagnani, M., Wennekers, T., & Pulvermüller, F. (2016). Brain connections of words, perceptions and actions: A neurobiological model of spatio-temporal semantic activation in the human cortex. *Neuropsychologia*, 98, 111–129.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M. (2002). Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain. *NeuroImage*, 15(1), 273–289.

Vickery, K. S., Reynolds, V. A., & Cochran, S. W. (1987). Multisensory teaching approach for reading, spelling, and handwriting, Orton-Gillingham based curriculum, in a public school setting. *Annals of Dyslexia*, 37(1), 189–200.

Appendix A: Description of the analysis of acoustic parameters for phonemes

Four tiers were used in the Praat software (Boersma & Weenink, 2010): the first one described the phonetic target, the second one included several acoustic landmarks, the third one described sound accuracy and the last tier included comments from the transcriber. For vowels, the first tier provided the beginning and the end of the vowel and the second tier included a landmark for a stable point at which formant frequencies would be later automatically extracted (most often corresponding to vowel midpoint). For consonants, /Ca/ syllables were extracted from the following sequences: /Ca/, /aCa/, /CaCa/ (2 syllables). For occlusives (/p, b, t, d, k, g/), the landmarks on the first tier were the beginning and the end of the /Ca/ syllable, and on the second tier, the time of the burst peak and the onset of voicing. For occlusives with double or triple burst, the largest burst was considered. The onset of voicing was defined as the first indication of voicing, as evident from the voicing bar in the spectrogram, as well as from visible initiation of regular cycle of periodicity in the waveform. The first zero-crossing of that periodic waveform was marked as voicing onset. For fricatives (/f, v, s, z, ʃ, ʒ/), the landmarks on the first tier were the beginning and the end of the /Ca/ syllable, and on the second tier, the onset of the fricative and the onset of vowel /a/. The onset of the fricative was defined as the first appearance of aperiodic noise on the waveform, simultaneously accompanied with frication noise above 2500 Hz from the spectrogram. The onset of the vowel was defined as the beginning of the periodic waveform after the fricative. For each target vowel or consonant, the accuracy tier coded accuracy in the following way. When the phoneme was judged correct, a "1" was keyed in the accuracy tier; when it was judged incorrect, a "0" was keyed. When the phoneme was emergent, i.e. preceded by a different sound before reaching the correct phoneme, an "E" was keyed. When place and manner of articulation were judged correct and only voicing was incorrect (such as when the voiced consonants /b, d, g, z, ʒ, v/ were pronounced as their unvoiced counterparts [p, t, k, s, ʃ, f] respectively, and vice versa), "V" was keyed. The comment tier allowed the transcriber to provide details about the incorrect production, describing substitutions, emergences, deletions, insertions, intelligibility. When a patient repeated a sequence, only the best produced syllable was kept for analysis. Transcriptions and coding were carried out by a trained native speaker transcriber. Part of the data was re-transcribed by a second trained native speaker, with more than 70% agreement, which is standard in patient audio data transcription. The Praat textgrid files provided by the transcription were fed into custom Praat scripts which derived acoustic parameters for all phonemes that were judged as correct or emergent ("1" or "E" in the accuracy tier).

For vowels, F1 and F2 formant frequencies were obtained using the Burg algorithm for formant detection available in Praat, with a time shift of 10 msec, a 25 msec Gaussian analysis window and a

preemphasis filter with a cut on frequency of 50 Hz. The formant search range was 0–5500 Hz. F1 and F2 values were subsequently all manually checked and corrected by the experimenters when necessary. An R script (R Development Core Team, 2008) was used to plot formant spaces for each speaker, i.e. F1 as a function of F2 for each repetition of each vowel. Interspeaker variability is usually dealt with by normalising data (Adank, Smits, & van Hout, 2004). Normalisation procedures were not used, however, because only 3 repetitions were available for each vowel. Therefore, only visual inspection of the patient' formant spaces compared with the ST's was possible. Two features of vowel spaces were considered: intra-vowel distance, which captures the acoustic stability of the productions of the same vowel across several repetitions (and should be small) and inter-vowel distance, which captures the degree of acoustic distance between different vowels (and should be large). For occlusives, VOT was determined by subtracting the instant of the burst from the instant of voicing onset. In French, VOT values are negative for voiced occlusives (/b, d, g/) and are positive for unvoiced occlusives (/p, t, k/). For fricatives, the spectral moments' analysis used algorithms implemented in Praat. An FFT spectrum was made over a 40 msec Hamming window centred at the midpoint of the fricative noise (where the fricative is most stable). The first two spectral parameters were considered in this study: centre of gravity (CG) and standard deviation (SD). Centre of gravity usually allows one to distinguish between alveolar (/s, z/) and post-alveolar (/ʃ, ʒ/) fricatives, while standard deviation allows one to distinguish between labiodental (/f, v/) and alveolar and post-alveolar fricative consonants (Li, Edwards, & Beckman, 2009). As shown by Nissen & Fox (2005), in normal condition, centre of gravity values are ranked in the following order: Alveolar > labiodental > post-alveolar. In normal condition, standard deviation values are expected to be ordered as follows: labiodental > alveolar > post-alveolar.

References

- Adank, P., Smits, R., & van Hout, R. (2004). A comparison of vowel normalization procedures for language variation research. *The Journal of the Acoustical Society of America*, 116(5), 3099–3107.
- Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (Version 6.0.05). Retrieved from <http://www.praat.org/>
- Li, F., Edwards, J., & Beckman, M. E. (2009). Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers. *Journal of Phonetics*, 37(1), 111–124.
- Nissen, S. L., & Fox, R. A. (2005). Acoustic and spectral characteristics of young children's fricative productions: A developmental perspective. *The Journal of the Acoustical Society of America*, 118 (4), 2570–2578.
- R Development Core Team. (2008). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org>

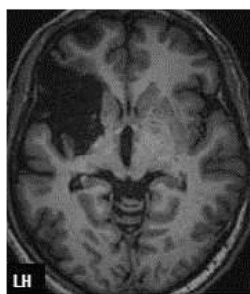


Figure A.1. Anatomical MR (T1) axial slice for the female patient (P1) illustrating the localisation of the post-stroke lesion.

| GENDER, AGE (YEARS) | HANDEDNESS | EDUCATION | STROKE ONSET (YEARS) | ISCHEMIC LESION | LANGUAGE DEFICIT | ASSOCIATED NEUROLOGICAL DEFICIT | SPEECH LANGUAGE THERAPY (DURATION, FREQUENCY) |
|---------------------|------------|-------------------|----------------------|--|--|--|--|
| F, 64 | Right | Bachelor's degree | 60 | Anterior territory of left sylvian artery (superficial and profound) | Broca's aphasia Symptoms: naming deficit, phonetic, phonological and rare semantic paraphasias. | Initial: Right hemiplegia → Right hemiparesis (facio-brachial) | First year after stroke: 5 sessions/week; After 1 Year: periodic sessions |

Table A.1. Demographic and clinical information of patient P1.

| Vowel corpus | | Consonant corpora | | | | |
|----------------------|-----|---------------------------|------|------|------------------------|-----|
| Orthographic | IPA | Conditions (orthographic) | | | | IPA |
| | | Ca | aC | aCa | CaCa | |
| hi (<i>oh</i>) | /i/ | pas | ape | apa | papa (<i>dad</i>) | /p/ |
| hé (<i>he</i>) | /e/ | (<i>step</i>) | abe | aba | baba (<i>small</i>) | /b/ |
| haie | /ɛ/ | bas | ate | ata | cake | /t/ |
| (<i>hedge</i>) | /a/ | (<i>down</i>) | ade | ada | tata (<i> aunty</i>) | /d/ |
| ha (<i>ah</i>) | /y/ | tas | ake | aka | dada (<i>hobby-</i>) | /k/ |
| hue (<i>gee</i> | /ø/ | (<i>heap</i>) | ague | aga | horse) | /g/ |
| up) | /œ/ | da (<i>yes</i>) | afe | afa | kaka (<i>poo</i>) | /f/ |
| eux (<i>them</i>) | /u/ | cas | ave | ava | gaga (<i>doting</i>) | /v/ |
| œuf (<i>egg</i>) | /o/ | (<i>case</i>) | asse | assa | fafa | /s/ |
| où (<i>where</i>) | /ɔ/ | gars | aze | aza | vava | /z/ |
| oh (<i>oh</i>) | /ã/ | (<i>guy</i>) | ache | acha | sasa | /ʃ/ |
| or (<i>gold</i>) | /õ/ | fa (<i>fa</i>) | aje | aja | zaza | /ʒ/ |
| an (<i>year</i>) | /ẽ/ | va (<i>go</i>) | ale | ala | chacha | /l/ |
| on (<i>one</i>) | | sa (<i>her</i>) | ame | ama | jaja (<i>wine</i>) | /m/ |
| hein (<i>what</i>) | | za | ane | ana | lala | /n/ |
| | | chat | are | ara | mama | /ɛ/ |
| | | (<i>cat</i>) | agne | agna | nana (<i>girl</i>) | /ŋ/ |
| | | ja | aye | aya | rara | /j/ |
| | | la (<i>la</i>) | | | gnagna | |
| | | ma (<i>my</i>) | | | yaya | |
| | | na | | | | |
| | | rat (<i>rat</i>) | | | | |
| | | gna | | | | |
| | | ya | | | | |

Table A.2. Vowel and consonant corpora Ca, aC, aCa, CaCa used for the acoustic evaluation; with orthographic and phonetic notations; C=consonant (/p, t, k, b, d, g, f, v, s, z, ʃ, ʒ, l, m, n, ɛ, ŋ, j/).

| | P1 | | | | | | | |
|-----------------------|-------------------------|--------------|------------------|---------------|------------------------|--------------|------------------|---------------|
| | Accuracy before SMF (%) | | | | Accuracy after SMF (%) | | | |
| | Correc t | Emergen t | Voicing error | Incorrec t | Correc t | Emergen t | Voicing error | Incorrec t |
| /i/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /y/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /u/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /e/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /ø/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /o/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /ɛ/ | 66.67 | 33.33 | NA | 0 | 100 | 0 | NA | 0 |
| /œ/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /ɔ/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /a/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /ã/ | 100 | 0 | NA | 0 | 66.67 | 33.33 | NA | 0 |
| /ê/ | 100 | 0 | NA | 0 | 100 | 0 | NA | 0 |
| /õ/ | 66.67 | 33.33 | NA | 0 | 100 | 0 | NA | 0 |
| ALL VOWELS | 94.87 | 5.13 | NA | 0 | 97.44 | 2.56 | NA | 0 |
| /b/ | 0 | 0 | 100 | 0 | 50 | 0 | 33.33 | 16.67 |
| /d/ | 33.33 | 0 | 50 | 16.67 | 66.67 | 0 | 16.67 | 16.67 |
| /g/ | 0 | 0 | 66.67 | 33.33 | 0 | 0 | 83.33 | 16.67 |
| /p/ | 100 | 0 | 0 | 0 | 66.67 | 0 | 16.67 | 16.67 |
| /t/ | 50 | 0 | 33.33 | 16.67 | 100 | 0 | 0 | 0 |
| /k/ | 50 | 0 | 0 | 50 | 50 | 0 | 0 | 50 |
| /f/ | 100 | 0 | 0 | 0 | 100 | 0 | 0 | 0 |
| /s/ | 0 | 0 | 0 | 100 | 28.57 | 0 | 0 | 71.43 |
| /ʃ/ | 66.67 | 0 | 16.67 | 16.67 | 50 | 0 | 0 | 50 |
| /v/ | 0 | 16.67 | 66.67 | 16.67 | 16.67 | 0 | 83.33 | 0 |
| /z/ | 0 | 0 | 0 | 100 | 0 | 0 | 14.29 | 85.71 |
| /ʒ/ | 16.67 | 0 | 83.33 | 0 | 16.67 | 0 | 50 | 33.33 |
| /l/ | 16.67 | 0 | 0 | 83.33 | 33.33 | 0 | 0 | 66.67 |
| /ʁ/ | 16.67 | 0 | 16.67 | 66.67 | 33.33 | 0 | 16.67 | 50 |
| /j/ | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 100 |
| /m/ | 83.33 | 0 | 0 | 16.67 | 100 | 0 | 0 | 0 |
| /n/ | 50 | 0 | 0 | 50 | 66.67 | 0 | 0 | 33.33 |
| /ɲ/ | 0 | 0 | 0 | 100 | 16.67 | 0 | 0 | 83.33 |
| ALL CONSONANTS | 32.41 | 0.93 | 24.07 | 42.59 | 43.64 | 0 | 17.27 | 39.09 |
| ALL PHONEMES | 48.98 | 2.04 | 17.69 | 31.29 | 57.72 | 0.67 | 12.75 | 28.86 |

Table A.3. Accuracy (%) of each phoneme separately judged by the transcriber as correct, incorrect, emergent or with voicing error, before and after rehabilitation for P1.

Tables

Table 1. Scores obtained by P1 before and after rehabilitation during assessment of language production. We calculated global evolution indices (EI) for each test: each item was attributed a score (+1 if incorrect in pre-training but correct in post-training; -1 if correct in pre-training but incorrect in post-training; 0 if correct in pre and post-training or if incorrect in pre and post-training); scores/items were cumulated to provide a global EI/task

| | REPETITION | READING | NAMING /80 | WORKING MEMORY (PHRASES) | VERBAL FLUENCY (2 MIN) | CONCEPTUAL ACCESS AND LANGUAGE COMPREHENSION | RHYME JUDGMENT | LANGUAGE SYMPTOMS SYNTHESIS |
|---------------------------------|---|--|---------------|--------------------------------|-------------------------------|---|-------------------|--|
| PATIENT 1 BEFORE SMF | Syllables 18/30 Simple words 12/20 Complex words 6/20 | Letters, sounds 13/20 Syllables 6/20 Regular words 7/10 Irregular words 6/10 | 67/80 | 10/10 | Categorical 8 Alphabetic 7 | Picture designation 34/36 Morphosyntax (MT 86) 7/8 | 26/40 | <ul style="list-style-type: none"> - Repetition and reading impairment with phonetic, phonological and rare semantic paraphasia (output phonology, phonetic encoding and phonological buffer impairment) - Naming deficit - Significant decreased verbal fluency - Preserved conceptual access and language comprehension - Significant deficit of input phonology - Lack of apraxia |
| PATIENT 1 AFTER SMF | Syllables 21/30 Simple words 13/20 Complex words 9/20 | Letters, sounds 17/20 Syllables 6/20 Regular words 5/10 Irregular words 6/10 | 69/80 | 10/10 | Categorical 9 Alphabetic 3 | Picture designation 36/36 Morphosyntax (MT 86) 8/8 | 33/40 | |
| SMF EFFECT EI | +3; +1; +3 (+7) | +4; 0; -2; 0 (+2) | +2 | Normal | +1; -4 (-3) | +3 | +7 | →Improvement |

Table 2 : Accuracy (%) of the entire set of vowels and consonants judged by the transcriber as correct, incorrect, emergent or with voicing errors; before and after rehabilitation in P1.

| | Patient 1 | | | | | | | |
|-----------------------|--------------------------------|-----------------|----------------------|------------------|-------------------------------|-----------------|----------------------|------------------|
| | Accuracy before SMF (%) | | | | Accuracy after SMF (%) | | | |
| | Correct | Emergent | Voicing error | Incorrect | Correct | Emergent | Voicing error | Incorrect |
| All vowels | 94.87 | 5.13 | NA | 0.00 | 97.44 | 2.56 | NA | 0.00 |
| All consonants | 32.41 | 0.93 | 24.07 | 42.59 | 43.64 | 0.00 | 17.27 | 39.09 |
| All phonemes | 48.98 | 2.04 | 17.69 | 31.29 | 57.72 | 0.67 | 12.75 | 28.86 |

Table 3 : Mean values of VOT (Voice Onset Time) obtained before and after rehabilitation for P1 as well as by the speech therapist (ST) for voiced and unvoiced occlusives (/ p, t, k, b, d, g /) in condition Ca (three repetitions), CaCa and aCa (see main text).

| | VOT (sec) for occlusives | | | |
|-------------------------|---------------------------------|----------------------|----------------------|---------------------|
| | Voiced | | Unvoiced | |
| | Before SMF | After SMF | Before SMF | Before SMF |
| PATIENT 1 | -0.0321 (±0.0116) | -0.0594 (±0.0371) | 0.0311 (±0.0201) | 0.0306 (±0.0156) |
| SPEECH THERAPIST | -0.119 (±0.0346) | -0.106 (±0.0567) | 0.0271 (±0.00907) | 0.0267 (±0.0118) |

Table 4 : Activated regions obtained in Patient 1 in Syllable repetition (SR) and Rhyme detection (RD), before and after Sensori-Motor Fusion (SMF) training. Only the most significant peak within each cluster of activation is reported, with its corresponding size in voxels (k), MNI coordinates (AAL labelling Tzourio-Mazoyer et al., 2002) and t value ($p < 0.001$, uncorrected, $k > 10$).

| AAL Label | K | x | y | z | T | AAL Label | k | x | y | z | T |
|--|------|-----|-----|------|-------|---|------|-----|-----|------|-------|
| | | | | {mm} | value | | | | | {mm} | value |
| RD Task > Control - Before SMF | | | | | | RD Task > Control - After SMF | | | | | |
| Occipital_Inf_L | 925 | -39 | -85 | -7 | 13.57 | Occipital_Inf_R | 2068 | 33 | -85 | -7 | 13.67 |
| Occipital_Inf_R | 557 | 30 | -88 | -7 | 12.11 | Occipital_Inf_L | 2590 | -39 | -82 | -7 | 13.12 |
| Precentral_L | 375 | -36 | 2 | 62 | 8.31 | Precentral_L | 1099 | -36 | 2 | 62 | 10.45 |
| Frontal_Mid_R | 295 | 36 | 38 | 35 | 8.07 | Precentral_R | 22 | 27 | -4 | 50 | 4.26 |
| Frontal_Mid_L | 34 | -33 | 50 | 5 | 5.57 | Precentral_R | 18 | 33 | -22 | 50 | 3.63 |
| Frontal_Mid_L | 40 | -48 | 29 | 29 | 5.51 | Frontal_Mid_R | 123 | 39 | 41 | 35 | 7.71 |
| Frontal_Mid_Orb_R | 206 | 33 | 50 | -1 | 7.97 | Cerebelum_8_L | 60 | -21 | -67 | -49 | 6.74 |
| Occipital_Mid_L | 479 | -27 | -67 | 29 | 7.49 | Frontal_Inf_Orb_R | 165 | 45 | 26 | -7 | 6.49 |
| Parietal_Inf_R | 679 | 33 | -46 | 41 | 7.09 | Frontal_Inf_Orb_R | 15 | 21 | 17 | -22 | 3.91 |
| Frontal_Sup_L | 110 | 36 | -4 | 65 | 6.28 | SupraMarginal_L | 51 | -60 | -40 | 23 | 5.61 |
| Temporal_Mid_L | 39 | -54 | -58 | 17 | 6.20 | SupraMarginal_R | 44 | 60 | -43 | 29 | 4.30 |
| Cerebelum_8_L | 83 | -18 | -64 | -49 | 5.94 | Precuneus_L | 29 | -3 | -43 | 11 | 5.49 |
| SupraMarginal_L | 37 | -60 | -40 | 23 | 5.71 | Temporal_Pole_Sup_R | 27 | 36 | 20 | -31 | 5.00 |
| Supp_Motor_Area_L | 38 | 0 | 2 | 56 | 4.91 | Temporal_Pole_Sup_L | 21 | -39 | 17 | -28 | 4.95 |
| Fusiform_R | 16 | 39 | -37 | -28 | 4.37 | Supp_Motor_Area_L | 26 | -6 | 5 | 53 | 4.78 |
| | | | | | | Poscentral_R | 19 | 15 | -46 | 74 | 4.40 |
| | | | | | | Poscentral_L | 15 | -42 | -22 | 44 | 3.80 |
| | | | | | | Insula_R | 17 | 33 | -22 | 8 | 4.39 |
| RD Before > After SMF | | | | | | RD After > Before SMF | | | | | |
| Cerebelum_Crus1_L | 12 | -42 | -43 | -34 | 4.09 | Frontal_Mid_R | 206 | 39 | 14 | 50 | 4.48 |
| Poscentral_L | 61 | -27 | -22 | 83 | 3.77 | Temporal_Sup_R | 11 | 45 | -28 | 2 | 4.30 |
| Frontal_Sup_L | 25 | -30 | -1 | 71 | 3.72 | Frontal_Sup_Medial_L | 45 | -9 | 35 | 44 | 4.08 |
| Vermis_4_5_R | 44 | 6 | -58 | -13 | 3.63 | Cerebelum_Crus1_R | 31 | 39 | -67 | -25 | 3.98 |
| Parietal_Sup_L | 40 | -21 | -52 | 80 | 3.56 | Precentral_L | 48 | -36 | 2 | 47 | 3.92 |
| Cerebelum_4_5_R | 11 | 24 | -37 | -28 | 3.46 | Frontal_Sup_R | 20 | 24 | 5 | 68 | 3.78 |
| Cerebelum_6_L | 14 | -21 | -52 | -28 | 3.44 | | | | | | |
| SR - Task > Control - Before SMF | | | | | | SR - Task > Control - After SMF | | | | | |
| Frontal_Med_Orb_L | 29 | -9 | 65 | -1 | 6.40 | ParaHippocampal_R | 213 | 21 | -16 | -22 | 6.04 |
| Frontal_Inf_Tri_L | 102 | -42 | 44 | 8 | 6.30 | Temporal_Pole_Sup_L | 1406 | -39 | 17 | -28 | 5.94 |
| Frontal_Inf_Tri_R | 23 | 45 | 35 | 26 | 4.38 | Frontal_Inf_Tri_L | 37 | -45 | 35 | 14 | 5.53 |
| Postcentral_L | 37 | -51 | -7 | 53 | 5.70 | Parietal_Sup_L | 320 | -21 | -67 | 68 | 4.71 |
| Temporal_Sup_R | 100 | 66 | -16 | 8 | 5.63 | Parietal_Sup_R | 124 | 39 | -43 | 71 | 4.02 |
| Temporal_Sup_L | 21 | -60 | -43 | 17 | 5.31 | Cingulum_Mid_R | 108 | 21 | -31 | 44 | 4.50 |
| Temporal_Sup_L | 40 | -54 | -22 | 11 | 4.65 | Frontal_Sup_L | 17 | -24 | 17 | 65 | 4.44 |
| Temporal_Mid_L | 33 | -54 | -58 | 17 | 5.46 | Frontal_Inf_Orb_R | 31 | 33 | 26 | -19 | 4.39 |
| Temporal_Mid_L | 38 | -54 | -37 | -1 | 5.15 | Frontal_Inf_Orb_L | 25 | -30 | 26 | -19 | 3.68 |
| Temporal_Mid_R | 22 | 48 | -34 | 2 | 5.42 | Temporal_Sup_R | 16 | 57 | -10 | -10 | 4.35 |
| Precentral_R | 162 | 63 | 2 | 23 | 5.36 | Temporal_Mid_R | 26 | 54 | -34 | -1 | 4.15 |
| Precentral_L | 51 | -33 | -13 | 71 | 4.54 | Postcentral_L | 34 | -27 | -31 | 41 | 3.97 |
| Supp_Motor_Area_R | 31 | 9 | 8 | 74 | 4.81 | Cerebelum_8_L | 25 | -24 | -58 | -52 | 3.90 |
| Lingual_L | 57 | -21 | -70 | 2 | 4.50 | SupraMarginal_R | 22 | 60 | -46 | 32 | 3.79 |
| Cuneus_R | 36 | 6 | -88 | 20 | 4.07 | Heschl_R | 16 | 54 | -10 | 5 | 3.51 |
| Cerebelum_4_5_L | 21 | -6 | -45 | -7 | 3.98 | | | | | | |
| SR Before > After SMF | | | | | | SR After > Before SMF | | | | | |
| Precentral_L | 1026 | -27 | -4 | 56 | 5.12 | Frontal_Inf_Orb_L | 101 | -51 | 17 | -4 | 4.73 |
| Temporal_Mid_L | 17 | -51 | -22 | -10 | 4.54 | Insula_L | 17 | -36 | -1 | -1 | 3.63 |
| Frontal_Mid_R | 140 | 48 | -4 | 56 | 3.97 | | | | | | |
| Precuneus_L | 13 | -15 | -52 | 74 | 3.94 | | | | | | |

Figure Legends

Figure 1. Illustration of the display provided by Ultraspeech-player (<http://www.ultraspeech.com> and http://www.ultraspeech.com/video/hueber_ultraspeech_player.mp4 Hueber, 2013).

Figure 2. Vowel F1/F2 formant space for Patient P1 (top) before (left) and after (right) rehabilitation, and for the speech therapist (ST, bottom).

Figure 3. Boxplots of the two spectral parameters (center of gravity and standard deviation) obtained before and after rehabilitation for P1 for fricatives (/f, s, ʃ, v, z, ʒ/) in condition Ca (three repetitions), CaCa and aCa (see the main text).

Figure 4. Activated regions projected onto 2D axial in Patient P1 obtained for syllable repetition and rhyme detection ($p < 0.001$, uncorrected, $k > 10$).

Figure 1



Figure 2

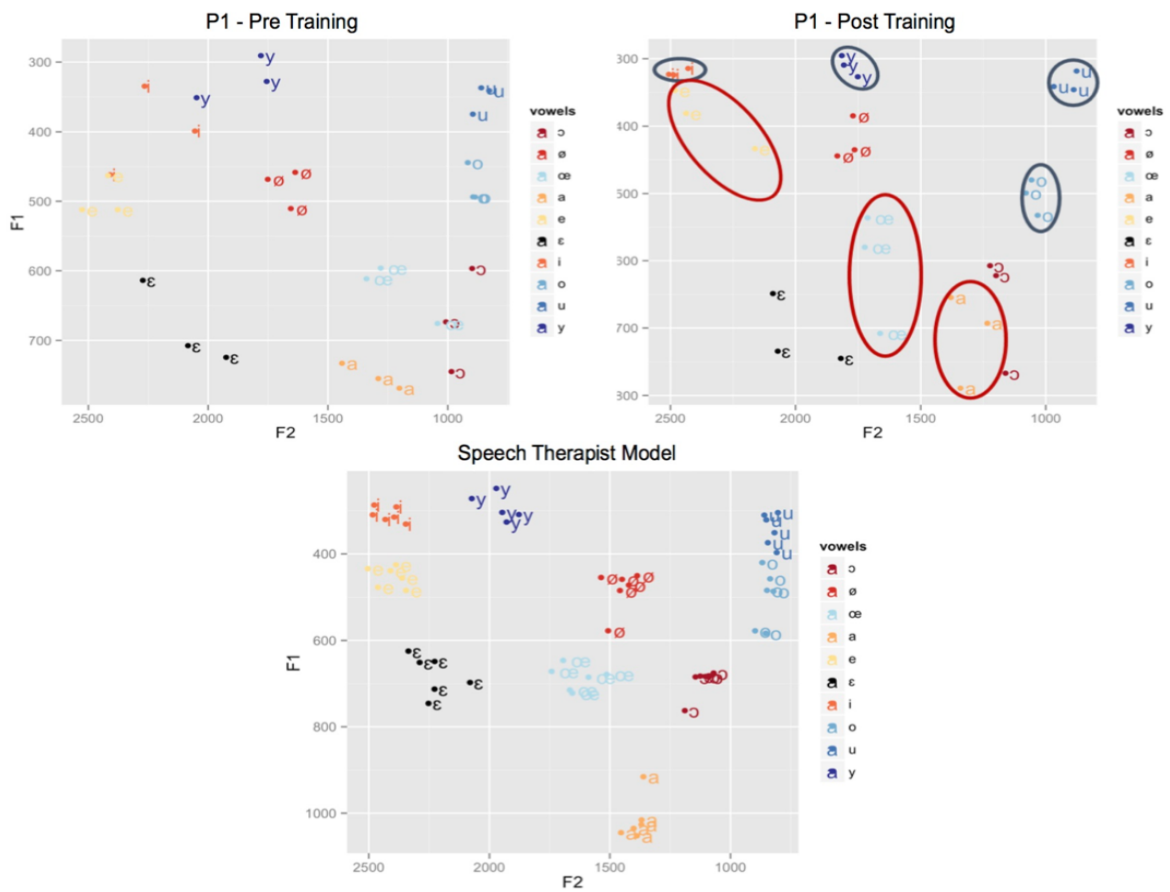


Figure 3

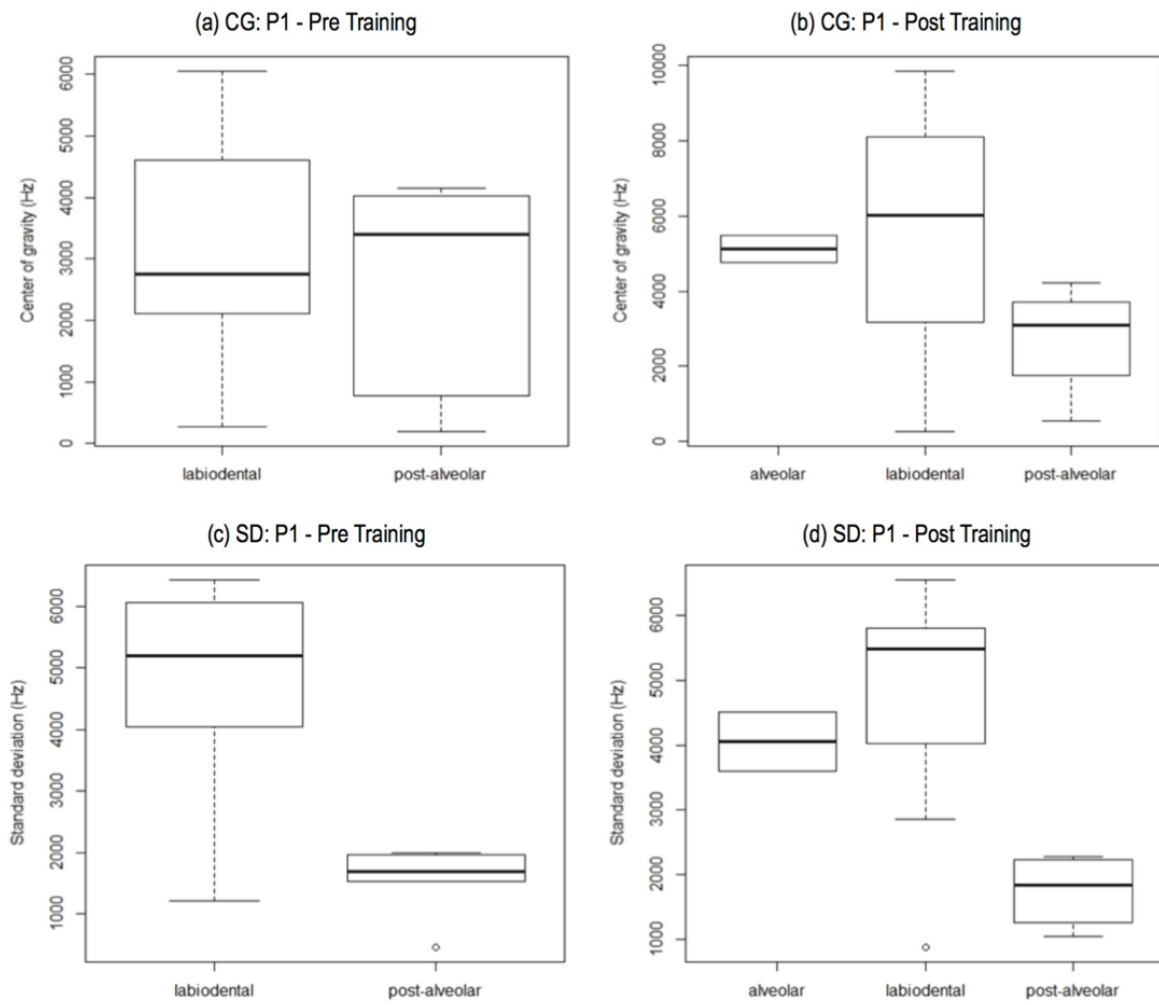


Figure 4

