BRAIN
A JOURNAL OF NEUROLOGY

Speech entrainment enables patients with Broca's aphasia to produce fluent speech

Julius Fridriksson,[1] H. Isabel Hubbard,[1] Sarah Grace Hudspeth,[1] Audrey L. Holland,[2] Leonardo Bonilha,[3] Davida Fromm[4] and Chris Rorden[5]

1 Department of Communication Sciences and Disorders, University of South Carolina, Columbia, SC 29208, USA
2 Department of Speech and Hearing Sciences, University of Arizona, Tucson, AZ, USA
3 Department of Neurosciences, Medical University of South Carolina, Charleston, SC, USA
4 Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA
5 Department of Psychology, University of South Carolina, Columbia, SC, USA

Correspondence to: Julius Fridriksson,
Department of Communication Sciences and Disorders,
University of South Carolina,
Columbia, SC 29208, USA
E-mail: jfridrik@sc.edu

A distinguishing feature of Broca's aphasia is non-fluent halting speech typically involving one to three words per utterance. Yet, despite such profound impairments, some patients can mimic audio-visual speech stimuli enabling them to produce fluent speech in real time. We call this effect 'speech entrainment' and reveal its neural mechanism as well as explore its usefulness as a treatment for speech production in Broca's aphasia. In Experiment 1, 13 patients with Broca's aphasia were tested in three conditions: (i) speech entrainment with audio-visual feedback where they attempted to mimic a speaker whose mouth was seen on an iPod screen; (ii) speech entrainment with audio-only feedback where patients mimicked heard speech; and (iii) spontaneous speech where patients spoke freely about assigned topics. The patients produced a greater variety of words using audio-visual feedback compared with audio-only feedback and spontaneous speech. No difference was found between audio-only feedback and spontaneous speech. In Experiment 2, 10 of the 13 patients included in Experiment 1 and 20 control subjects underwent functional magnetic resonance imaging to determine the neural mechanism that supports speech entrainment. Group results with patients and controls revealed greater bilateral cortical activation for speech produced during speech entrainment compared with spontaneous speech at the junction of the anterior insula and Brodmann area 47, in Brodmann area 37, and unilaterally in the left middle temporal gyrus and the dorsal portion of Broca's area. Probabilistic white matter tracts constructed for these regions in the normal subjects revealed a structural network connected via the corpus callosum and ventral fibres through the extreme capsule. Unilateral areas were connected via the arcuate fasciculus. In Experiment 3, all patients included in Experiment 1 participated in a 6-week treatment phase using speech entrainment to improve speech production. Behavioural and functional magnetic resonance imaging data were collected before and after the treatment phase. Patients were able to produce a greater variety of words with and without speech entrainment at 1 and 6 weeks after training. Treatment-related decrease in cortical activation associated with speech entrainment was found in areas of the left posterior-inferior parietal lobe. We conclude that speech entrainment allows patients with Broca's aphasia to double their speech output compared with spontaneous speech. Neuroimaging results suggest that speech entrainment allows patients to produce fluent speech by providing an external gating mechanism that yokes a ventral language network that encodes conceptual aspects of speech. Preliminary results suggest that training with speech entrainment improves speech production in Broca's aphasia providing a potential therapeutic method for a disorder that has been shown to be particularly resistant to treatment.

Introduction

Broca's aphasia is a common condition resulting from damage to the anterior speech areas in the left hemisphere (Broca, 1861; Geschwind, 1965; Dronkers, 1996). It is characterized by non-fluent spontaneous speech comprising short utterances involving mostly substantive words (Brookshire, 2003). Although speech production in patients with Broca's aphasia may improve following behavioural treatment (Kendall *et al.*, 2008; Edmonds *et al.*, 2009; Fridriksson *et al.*, 2009, 2012; Kiran *et al.*, 2011), most chronic patients never fully recover from this condition and remain non-fluent. Several treatment options have been suggested to improve non-fluent speech in Broca's aphasia (Albert *et al.*, 1973; Rosenbek *et al.*, 1973; Sparks and Holland, 1976; Wambaugh, 2006*a*, *b*; Schlaug *et al.*, 2008; Stahl *et al.*, 2011). Typically, treatment of non-fluent speech focuses on speech production—a task that is inherently difficult for these patients. Based on studies in normal subjects that suggest increased activity in the anterior cortical speech areas associated with speech perception, especially when it involves audio-visual speech stimuli (Hall *et al.*, 2005; Bernstein *et al.*, 2008; Fridriksson *et al.*, 2008; Kotz *et al.*, 2010; Vander Wyk *et al.*, 2010), we examined whether a treatment that focused on speech perception tasks involving hearing speech and seeing the mouth of the speaker would result in greater improvement in speech production (picture naming) compared with auditory-only speech stimuli in a group of non-fluent speakers (Fridriksson *et al.*, 2009). The rationale for this work was as follows: audio-visual speech perception is associated with greater Broca's area activity compared with auditory-only speech (Fridriksson *et al.*, 2008; Vander Wyk *et al.*, 2010). Further, brain stimulation in healthy adults suggests that this region is involved in the perception of auditory (Meister *et al.*, 2007) and visual aspects of speech (Rorden *et al.*, 2008). A treatment that focuses on audio-visual speech perception may stimulate the residual left frontal speech areas in patients with non-fluent speech, and thereby improve their speech production. In short, our study revealed treatment involving audio-visual speech perception resulted in greater improvement in speech production compared with treatment that only relied on auditory speech stimulation (Fridriksson *et al.*, 2009). Although this study did not examine changes in speech fluency, it did motivate us to further examine the use of audio-visual speech feedback as a means of treating non-fluent aphasia. Others have used similar methods that emphasize audio-visual speech perception in aphasic patients (Lee *et al.*, 2010). One interesting aspect of audio-visual speech perception in subjects with non-fluent aphasia is that some patients can actually mimic the speech of others in real-time even though their ability to repeat speech is severely impaired (Rosenbek *et al.*, 1973). A similar effect, usually referred to as choral reading or choral speech, has been shown to induce fluent speech in stuttering, although this effect does not seem to generalize to connected speech (Max *et al.*, 1997; Kiefte and Armson, 2008). For demonstration of this effect, we have included a video of a patient with severe Broca's aphasia, G.B., speaking with and without audio-visual speech feedback (Supplementary Video 1). As demonstrated, G.B. struggles to produce spontaneous speech but is able to mimic fluent speech using audio-visual feedback. We describe this effect as 'speech entrainment', meaning 'drawing or pulling along'. G.B.'s articulation is not

perfect, but he can produce only two spontaneous phrases ('I know' and 'yeah sir/boy') and has been aphasic for 22 years. Although in G.B.'s case the effect is powerful, it is unclear whether speech entrainment is something that would work for most patients with Broca's aphasia.

We suggest that speech entrainment has potential importance for two reasons. First, it could be used as a mechanism for speech rehabilitation in patients with non-fluent speech by practicing speech production, even without the aid of a trained clinician. Second, even if speech entrainment is not effective for improving spontaneous speech, it is possible that patients could rely on speech entrainment in conversation with others. For example, patients could store typical utterances and communication scripts on a portable electronic device (e.g. an iPod) and use it to communicate in appropriate situations. We used one such iPod application, Visually Assisted Speech Technology (VAST)™ (http://www.speakinmotion.com/), in the research described here.

Whereas speech entrainment may represent a novel treatment approach for patients with non-fluent aphasia, it should also contribute to understanding the normal mechanism that supports speech production. Patients with non-fluent aphasia typically have damage to Broca's area and the left anterior insula (Broca, 1861; Dronkers, 1996; Richardson *et al.*, 2012). Accordingly, these areas are thought to play a crucial role in speech fluency. If patients with non-fluent speech can speak more fluently with the aid of speech entrainment, it suggests that the speech entrainment is replacing or enhancing some natural mechanism that is necessary for normal speech fluency. One way to investigate this issue would be to compare brain activation during typical speech production with that of speech entrainment in both normal and aphasic participants. Such an investigation could not only shed light on the mechanism that supports fluent speech production using speech entrainment but might also inform theories of normal speech production. The findings could then be related to current neurocognitive models of speech processing—the 'Directions into Velocities of the Articulators' model and the 'Hierarchical State Feedback Control' model (Guenther *et al.*, 2006; Hickok, 2012), which emphasize the role of distributed feed-forward and feedback mechanisms in fluent speech production.

The purpose of this research was to conduct three experiments: the first examined the effect of speech entrainment on speech fluency in patients with Broca's aphasia. A group of patients with chronic Broca's aphasia completed spoken discourse tasks in three conditions: (i) speech entrainment with audio-visual feedback; (ii) speech entrainment with audio-only feedback; and (iii) spontaneous speech. The second experiment used functional MRI to examine the neural mechanism that supports fluent speech production with speech entrainment in patients and age-matched control subjects. Unlike most neuroimaging studies that emphasize differences in cortical activation among patients and normal subjects (Warren *et al.*, 2009; Fridriksson *et al.*, 2010; Buchsbaum *et al.*, 2011; Kim *et al.*, 2011; Turkeltaub *et al.*, 2011), we attempted to identify a neural mechanism that is common among normal subjects and patients with Broca's aphasia. Moreover, we contrasted brain activation in patients with Broca's aphasia and their normal counterparts to understand how speech entrainment enables the patients to produce fluent

speech in spite of brain damage involving the cortical speech and language regions. Diffusion tensor imaging (DTI) MRI data were also collected to highlight the structure of a potential cortical network that supports speech entrainment. The final experiment tested whether training with speech entrainment improves speech production among patients with Broca's aphasia. For this purpose, a group of patients with Broca's aphasia completed a 6-week training regimen that emphasized daily speech exercises with speech entrainment. Functional MRI was used to examine functional brain changes associated with treatment.

Materials and methods

Experiment 1: Examining the effect of speech entrainment on speech fluency in patients with Broca's aphasia

Participants

Thirteen patients (four females) with stroke-induced Broca's aphasia participated in this research. The mean age was 57 years [standard deviation (SD) = 9.15, range = 45–75], and all participants, except one, were premorbidly right-handed. Each patient was at least 6 months post-stroke, and all had incurred a single event stroke (Table 1). High-resolution imaging (lesion demarcation was done on T_2 MRI; see MRI parameters below) revealed left hemisphere damage in all patients, and greatest lesion overlap was found in the middle portion of the left insula where all 13 patients had damage (Fig. 1). A diagnosis of Broca's aphasia was made based on aphasia assessment using the Western Aphasia Battery-Revised (Kertesz, 2007). The mean aphasia quotient, a measure of aphasia severity where a score >93.8 indicates language performance within normal limits, was 51.5 (SD = 18.5), with a range of 25.8–74.8. The Boston Naming Test was administered to assess overall naming ability (Kaplan et al., 2001) and subtest 6 of the second edition of the Apraxia Battery for Adults (Dabul, 2000) was used to rate severity

of apraxia of speech. The mean number of correctly named pictures on the Boston Naming Test was 23.23 (SD = 19.69; range = 1–50), whereas the mean severity rating on the Apraxia Battery for Adults (second edition) was 11.38 (SD = 2.36; range = 6–15). Each participant included in this research provided an informed consent for study inclusion. The institutional review board at the University of South Carolina approved the study protocol, and all study procedures conformed to principles set forth in the Declaration of Helsinki.

Behavioural tasks

All 13 patients completed three separate behavioural tasks that were used to understand the effects of speech entrainment on speech production in non-fluent aphasia. During the 'speech entrainment–audio visual' condition, participants attempted to mimic a speaker in real-time whose mouth was made visible on the 3.5-inch screen of an iPod Touch and whose speech was heard via headphones. The 'speech entrainment–audio only' condition involved real-time mimicking speech presented through headphones with the screen of the iPod being blank. Finally, the spontaneous speech condition required patients to discuss a given script topic for 1 min. The patients were not trained to produce the scripts before the actual testing session.
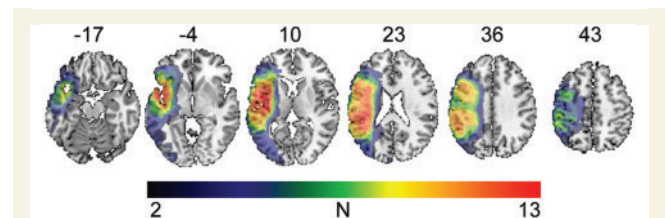

Figure 1 Lesion overlay map showing the distribution of damage among the patients. The greatest lesion overlap (n = 13) was found in the middle portion of the left insula. The numbers above each axial slice designate slice label in standard anatomical space.

Table 1 Biographical and behavioural testing data including subscores from the Western Aphasia Battery-Revised, Apraxia Battery for Adults (second edition) and Boston Naming Test

Patient number	Gender	Age	Months post stroke	AQ	Fluency	Auditory comprehension	Speech repetition	Naming	AOS	BNT
1	M	63	87	68.1	4	9.6	6.6	8.2	13	45
2	M	55	46	70.7	5	7.6	6.8	7.6	9	48
3	F	47	85	43.4	4	6.35	2.8	4.1	10	14
4	M	45	18	25.7	1	6.3	2	2.1	14	2
5	M	75	261	20.9	1	4.25	3.6	2.5	12	3
6	F	57	47	44.8	2	8.5	1.9	4.2	13	17
7	M	60	56	47.6	4	5.4	6.2	2.7	9	9
8	M	53	24	72.7	4	8.25	8.1	9	6	50
9	M	69	31	25.7	0	10	0.9	2	15	1
10	M	66	64	73.5	4	9.05	8.2	8.5	11	44
11	M	50	10	32.7	2	8.6	2	4.2	14	14
12	F	50	28	31.3	2	8.25	4.4	5.1	13	10
13	F	50	72	73.5	4	9.25	5.8	8.7	13	45
Range		45–75	10–261	20.9–73.5	0–5	4.25–10	1.9–8.2	2–9	6–15	1–50
Mean		56.92	63.77	48.51	2.85	7.80	4.56	5.30	11.38	23.23
SD		9.15	64.25	20.62	1.57	1.73	2.53	2.72	2.36	19.69

AOS = apraxia of speech; AQ = aphasia quotient; BNT = Boston Naming Test.

Three different discourse scripts were used and included the following topics: (i) stroke; (ii) Thanksgiving; and (iii) making a peanut butter and jelly sandwich. The scripts used in this research varied in length between 48 and 58 words. Each script took a normal speaker ~1 min to read out loud using a comfortable speaking rate. The content of the scripts was controlled for number of words, number of different words, word class and word frequency. All patients first completed the speech entrainment–audio visual condition, followed by the spontaneous speech condition, and finished with the speech entrainment–audio only condition. This order was maintained to minimize the potential priming effect of spontaneous speech and speech entrainment–audio only for speech entrainment–audio visual; the condition we proposed would result in the most fluent speech. Therefore, any learning effects across conditions should diminish our predicted benefits for speech entrainment. Script topics were randomized among the three conditions so that only one topic was addressed per condition by each participant. For the speech entrainment–audio visual and speech entrainment–audio only conditions, participants were instructed to mimic the speaker as closely as possible, but during the spontaneous speech condition, they were asked to speak in an open-ended manner about a given script topic for 1 min ('tell me about how you make a peanut butter and jelly sandwich'). Patients were videotaped while completing each of the three tasks, and the speech samples were transcribed and scored offline by two naïve listeners. Transcription was done in the CHAT format and linked to the digitized audio and video (MacWhinney, 2000). The CHAT format is designed to operate closely with the CLAN programs (MacWhinney, 2000), which allow for the analysis of a wide range of linguistic and discourse structures. Each transcript was coded by an experienced transcriber and then reviewed in its entirety by a speech-language pathologist with extensive clinical research experience in aphasia and transcription. The two reached forced choice agreement on all features of the transcription.

To assess speech fluency, the number of different words was tallied and recorded as the per cent of correctly produced words per script. Note that for the spontaneous speech condition, this meant that >100% of a given script could be produced (i.e. if a script included 50 words but the participant produced 75 words while discussing a given topic, the recorded percentage for that attempt was 150%). However, the percentage of words produced per script in the speech entrainment conditions exceeded 100% only once during the study. If a target word could be construed from what the patient said (e.g. 'troke' for 'stroke', 'tukey' for 'turkey'), the production was transcribed as spoken, followed by the intended word and an error code. In all conditions, the number of different words produced included these words. The total did not include unintelligible words and segments, neologisms, repetitions and revisions of words, or extraneous comments (e.g. 'no', 'okay', 'oh', 'let's see'). We chose to focus on the number of different words produced per topic instead of the total number of words because some patients with Broca's aphasia repeat the same words without necessarily adding communicative value. It is worth noting, however, that the number of total words and different words produced per script was strongly correlated.

Broca's aphasia frequently co-occurs with some degree of apraxia of speech, an impairment in planning the articulatory movements of speech. In fact, all of the patients involved here had been diagnosed with some degree of apraxia of speech. Although it is not clear what behavioural factors contributed to speech entrainment–audio visual success, it would seem that apraxia of speech severity could be especially important. To understand if the severity of apraxia of speech influences patients' ability to speak with speech entrainment, we calculated the proportional benefit of speech entrainment in enhancing

speech production beyond that observed in the spontaneous speech condition. Specifically, for speech entrainment–audio visual, speech entrainment–audio only and spontaneous speech, the proportion of different words spoken for a given script topic was subtracted from 1.00 (assuming that 1.00 represented the use of all different words included in a given script). For a given patient, these calculations yielded scores representing how much room for improvement was available in each condition. Then, the proportional improvement in the production of different words during speech entrainment–audio visual and speech entrainment–audio only was divided by the potential room for improvement in the spontaneous speech condition. The resulting scores were qualified as the proportional increase in the production of different words during speech entrainment–audio visual or speech entrainment–audio only beyond those produced during spontaneous speech. For example, if a patient was able to produce 60% of the different words for a given script during spontaneous speech but improved to 70% with speech entrainment–audio visual, then the benefit of audio-visual speech entrainment was calculated in the following manner: for spontaneous speech $(1.00 - 0.60 = 0.40)$ and speech entrainment–audio visual $(1.00 - 0.70 = 0.30)$. Then, $1.00 - (0.30/0.40) = 0.25$, suggesting that the proportional benefit of speech entrainment–audio visual beyond spontaneous speech was 25%. The same method was used to assess the proportional increase in word production for speech entrainment–audio only. We chose to specifically examine the effect of apraxia of speech (rather than other impairments commonly seen in Broca's aphasic patients) because of a long-standing controversy regarding the distinction between Broca's aphasia and apraxia of speech (Darley *et al.*, 1975; Johns and LaPointe, 1976; McNeil, 1984; McNeil *et al.*, 1990; Code, 1998; Katz *et al.*, 2001).

Experiment 2: Using functional magnetic resonance imaging to examine the neural mechanism that supports fluent speech production with speech entrainment

Participants

Ten of the 13 patients included in Experiment 1 also completed this portion of the research. Among the three aphasic patients who completed the behavioural aspects of this research, but not its MRI portion, one was excluded, as she had a metal implant that precludes MRI scanning. The two remaining participants had such severely non-fluent speech that we anticipated that neither could contribute much behavioural data while in the scanner (e.g. during the spontaneous speech condition). A control group including 20 normal subjects (14 females) was included to examine cortical activation associated with normal speech entrainment. The mean age of the control group was 53 years (SD = 7, range = 40–76).

Neuroimaging

MRI scanning was accomplished using a Siemens 3T scanner equipped with a 12-channel head coil and audio–video presentation system. Different functional MRI runs were collected for each of three conditions: (i) speech entrainment–audio visual; (ii) spontaneous speech; and (iii) audio/visual speech perception (without an overt response). During the speech entrainment–audio visual condition, patients mimicked a speaker producing speech entrainment scripts. To emphasize visual speech perception and simulate the speech entrainment–audio visual condition in Experiment 1, only the mouth of the speaker

was visible. Functional MRI data collection relied on sparse imaging where whole brain volumes were acquired with a repetition time of 10 s and acquisition time of 2 s. This allowed for stimulus presentation whereby each script was segmented in such a way that single sentences and phrases were presented during the 8 s of silent intervals between actual functional MRI scanning. Participants mimicked the speech that started immediately after the collection of each volume and ended at least 1 s before the start of the next volume collection. Overt speech was recorded using an MRI compatible microphone for later offline scoring to verify task compliance. During the spontaneous speech condition, participants spoke about the event of the current day while watching a speaker's mouth movements and listening to backwards speech. This controlled for low-level perceptual information while minimizing neural activity that would reflect inhibition of real language activation while patients produced spontaneous speech. Although audio-visual stimuli were presented during the spontaneous speech condition to best match it to the speech entrainment–audio visual condition, it is important to note that brain activation associated with spontaneous speech occurred in the presence of audio-visual speech stimuli. The final condition, audio/visual speech perception, involved the same stimuli used during the speech entrainment–audio visual condition, but no overt response was permitted. Audio recording was used to verify that participants were not speaking during audio/visual speech perception. Specific parameters for the functional MRI sequence were as follows: repetition time = 10 s; echo time = 35 ms, attenuation time = 2 s, 64 × 64 axial matrix with 33 3.2-mm-thick slices (no gap); field of view = 208 × 208 mm; number of volumes = 60; total acquisition time = 10 min.

All participants also underwent high-resolution (voxels = 1 mm³) T_1-MRI and T_2-MRI (patients only) to improve normalization of functional MRI images and lesion segmentation in patients. Each of the structural MRI sequences consisted of 160 sagittal slices. Lesion demarcations were accomplished using the T_2-MRI in native space, and the T_1-MRI images were used for qualitative reference of lesion boundaries.

A 64-direction DTI scan was acquired to examine structural connectivity among the normal subjects. The analysis of DTI images relied on raw images from 17 control subjects (owing to scanner error and time constraints during actual scanning, three normal subjects did not receive a DTI scan) and were transformed into NIfTI format using the software dcm2nii (http://www.mccauslandcenter.sc.edu/mricro/mricron/) that adjusts diffusion gradient directions based on image angulation.

Magnetic resonance imaging data analyses

All functional MRI data were analysed using FMRIB's Software Library (FSL). The first level analysis of each participant's data was carried out using FEAT (FMRI Expert Analysis Tool) Version 5.98, part of FSL. The following pre-statistics processing was applied: motion correction using MCFLIRT (Jenkinson et al., 2002); non-brain removal using BET (Smith, 2002); spatial smoothing using a Gaussian kernel of full-width at half-maximum 5 mm; grand mean intensity normalization of the entire 4D data set by a single multiplicative factor; high-pass temporal filtering (Gaussian-weighted least-squares straight line fitting, with sigma = 30.0 s). Stimulus functions encoding neuronal responses were convolved with a synthetic haemodynamic response (gamma) function to produce functional MRI regressors. A second level analysis was carried out in which activation in different conditions (speech entrainment–audio visual, spontaneous speech and audio/visual speech perception) was compared within subjects using a fixed effects model. This was done by forcing the random effects variance to zero in FLAME (FMRIB's Local Analysis of Mixed Effects; Beckmann et al., 2003; Woolrich et al., 2004; Woolrich, 2008). We tested for the main effect of conditions (using the above contrasts)

and the condition by group interaction using a mixed effects analysis (within FSL). This produced statistical parametric maps testing for the significance of main effects—activations averaged over (common to) both patients and normal subjects; and interactions—differences in activations between the groups. To ensure that these effects were consistent in relation to inter-subject variability in activations, we performed *post hoc* ANOVAs testing for main effects and interactions using a random effects analysis (in regions of interest identified by the mixed effects analysis using SPSS). This allowed us to generalize our findings to the populations from which our subjects were sampled. Normalization of T_2-MRI images to standard space relied on linear image registration tool included in FSL (Jenkinson and Smith, 2001; Jenkinson et al., 2002). Lesion-masked cost function weighting was used to minimize the effect of frank cortical damage on the normalization procedure (Brett et al., 2001); functional MRI images were then normalized to each patient's T_2-MRI in standard space using 12 degrees of freedom.

The Diffusion Toolkit (FDT), part of the software package FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl), was used to pre-process the diffusion-weighted images and construct DTI data. All images first underwent eddy current correction by way of affine transformation of each diffusion weighted image to the base $b = 0$ T_2-weighted image. This step aimed at removing additional spatial distortion in the diffusion weighted images owing to application of diffusion gradients in various directions. The pixel-wise calculation of the diffusion tensor was performed using FDT, with the application of a binary brain mask that was extracted using FSL's Brain Extraction Tool (BET), with a threshold of 0.3 to prevent erroneous DTI calculation in the noise background outside the brain. Mean diffusivity and fractional anisotropy maps were generated in the same space as the $b = 0$ image volume of the original DTI acquisition. Fractional anisotropy maps were then non-linearly normalized to the fractional anisotropy template provided by FSL, FMRIB58_fractional anisotropy (FMRIB's Linear Image Registration Tool—http://www.fmrib.ox.ac.uk/fsl/), and the normalization matrix was applied to the mean diffusivity maps.

Probabilistic tractography was performed using FSL's DTIFit on the DTI data after the pixel-wise calculation of the diffusion tensor. We used BEDPOST (Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques), which runs Markov Chain Monte Carlo sampling to build up voxel-wise distributions on diffusion parameters. Probabilistic tractography was estimated using FSL's probtrackx. Probabilistic tractography was then performed using a seed mask and a termination mask, in order to generate the probabilistic distribution of the fibres travelling from the voxels in the seed mask to the voxels in the termination mask. The masks used in this study were derived from the functional MRI analysis, as described below. Two separate tracts were mapped: (i) from left Brodmann area (BA) 37 to the anterior insula and BA 47; and (ii) between the left middle temporal gyrus and Broca's area. The resulting probabilistic tractography images from each subject were linearly transformed into standard stereotaxic Montreal Neurological Institute (MNI) space, where average connectivity maps were constructed.

Experiment 3: Testing whether training with speech entrainment improves speech production among patients with Broca's aphasia

Participants and functional magnetic resonance imaging

Each patient ($n = 13$) included in Experiment 1 also completed a 6-week training regimen during which time they practiced producing

speech with the aid of speech entrainment–audio visual for 30 min per weekday. The functional MRI sessions described for Experiment 2 were used to establish a pre treatment baseline. Once the speech entrainment treatment phase was complete, all 10 patients underwent functional MRI scanning again using the same sequences that were acquired at baseline. The post-treatment scanning occurred during the week after the 6-week training period was completed. To analyse changes in brain activation associated with training, a third level functional MRI analysis was conducted using the same analysis parameters described in Experiment 2.

Training

Three scripts were targeted for treatment, and one was used to measure generalization from trained to untrained scripts: (i) aphasia; (ii) weather in the South-Eastern USA; (iii) making scrambled eggs; and (iv) explaining speech entrainment. The three scripts used in Experiment 1 (stroke, Thanksgiving, and making a peanut butter and jelly sandwich) were used to establish treatment baseline. Different scripts were selected to establish the baseline and for treatment to minimize the effect of training in the baseline session.

As in the baseline sessions, the training scripts were presented on an iPod equipped with monitoring software that made it possible to verify task compliance (e.g. length of the training session, how many times the patient completed the trained script during the session, and at what time of day the training session occurred). Patients were instructed to complete the training at the same time each day and attempt to get through each script as often as possible during the 30-min session. Training occurred at the patients' place of residence, and a clinician monitored training progress (including downloading compliance data from the iPods) at least once a week. These sessions were accomplished 'in person' or using the computerized communication software Skype. Three scripts were practiced during the 6-week training period—each script was targeted for 2 weeks. The order of scripts (trained and untrained) was randomized among the patients. Patients' performance with each script was assessed within 1 week after they had completed the 2-week training with that specific script and again 6 weeks later. Specifically, within a week following the 2-week treatment period for each specific script, performance for that trained script was assessed with speech entrainment–audio visual. This testing was repeated 6 weeks following training completion of each script. To assess training generalization, patients' ability to produce an untrained script with speech entrainment–audio visual, as well as to produce spontaneous speech about a trained script topic was tested before and after the 6-week training period and again at 6 weeks after training completion.

Although all the patients had Broca's aphasia, their baseline performance varied considerably. Whereas some patients could only produce 5–10% of a given script at baseline, others were able to produce as much as 70% of the script correctly. Based on this variability, we suggest that for someone who can only produce 5 words at baseline, an increase in 10 words produced after treatment constitutes greater improvement than someone whose baseline performance yielded 30 words but also improved by 10 words with training. To account for such differences, training outcome was calculated as the number of different words produced after training compared with the number of different words produced at baseline on a patient-by-patient basis. Accordingly, change scores at 1 week and 6 weeks after treatment completion were calculated in the following manner: number of words produced after treatment divided by number of words produced at baseline. In cases where patients' number of words did not change after training, a score of 1 was assigned. Similarly, scores < 1 signified regression, whereas scores > 1 suggested

improvement. A one-sample *t*-test (one-tailed) was used to infer treatment outcome where the expected value was 1 (signifying no change in speech production).

Results

Experiment 1: Examining the effect of speech entrainment on speech fluency in patients with Broca's aphasia

To understand whether speech entrainment improves speech production, the number of different words produced across the three study conditions (speech entrainment–audio visual, speech entrainment–audio only and spontaneous speech) was compared. A one factor repeated-measures ANOVA with three levels revealed a main effect suggesting that speech performance differed among at least two conditions, $F(2,24) = 11.1$, $P < 0.0004$, Greenhouse-Geisser = 0.9557, $P < 0.0005$ (Fig. 2). *Post hoc* analyses showed that speech entrainment–audio visual [mean = 66.04%, standard error (SE) = 4.36%] elicited a larger variety of words compared with both speech entrainment–audio only [mean = 40.79%, SE = 4.22%, $t(12) = 3.62$, $P < 0.004$] and spontaneous speech [mean = 29.99%, SE = 5.00%, $t(12) = 4.30$, $P < 0.001$]. However, the difference between the number of different words produced during speech entrainment–audio only and spontaneous speech was not significant $t(12) = 1.32$, $P < 0.21$. The patients were able to produce more than twice as many words during the speech entrainment–audio visual compared with the spontaneous speech condition.

Severity of apraxia of speech was related to how much patients benefited from speech entrainment. Specifically, higher ratings on the Apraxia Battery for Adults (second edition) (suggesting more severe apraxia of speech) were inversely related to how many more words patients were able to produce during the speech entrainment–audio visual condition compared with the spontaneous speech condition, $F(1,11) = 7.97$, $P = 0.017$, $R^2 = 0.42$ (Fig. 3). In contrast, no relationship was revealed between apraxia of speech severity and proportional benefit of speech entrainment–audio only, $F(1,11) = 0.49$, $P = 0.50$, $R^2 = 0.04$. One outlier that was > 2 SD below the regression line was noted for the relationship between apraxia of speech severity and speech entrainment–audio only benefit. This outlier was removed and the same analysis was rerun, $F(1,10) = 0.20$, $P = 0.66$, $R^2 = 0.02$.

Experiment 2: Using functional magnetic resonance imaging to examine the neural mechanism that supports fluent speech production with speech entrainment

Ten patients and 20 normal age-matched control subjects underwent functional MRI scanning while completing three conditions: speech entrainment–audio visual, spontaneous speech and audio/
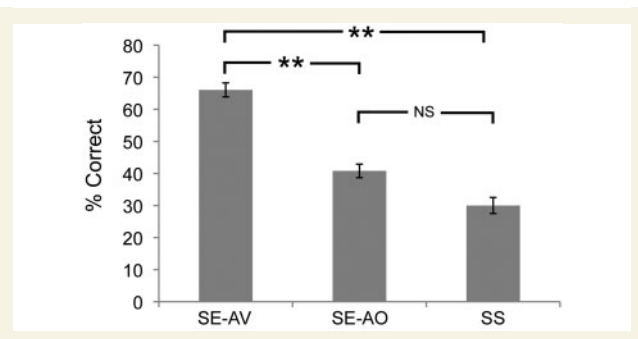
Figure 2 Bar graph showing the percentage of correctly produced different words for each of three conditions: speech entrainment with audio-visual feedback (SE-AV), speech entrainment with audio only feedback (SE-AO) and spontaneous speech (SS) focused on a given script topic. The error bars denote standard error. Statistically significant differences at the $P < 0.01$ level are denoted by **. NS = not significant.
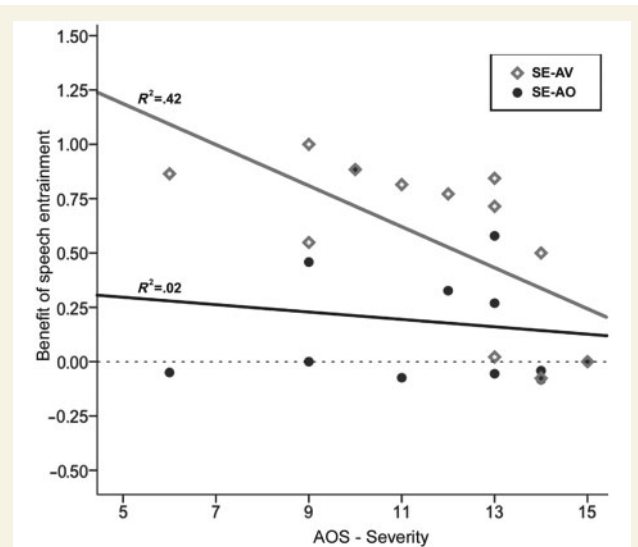

Figure 3 The relationship between proportional benefit of speech entrainment–audio visual (SE-AV) and speech entrainment–audio only (SE-AO) compared with spontaneous speech (y-axis) in relation to severity of apraxia of speech (AOS, x-axis) as rated using subtest 6 of the Apraxia Battery for Adults (second edition). Note that higher scores on the rating scale denote more severe apraxia of speech. Datapoints for speech entrainment–audio visual are shown as grey diamonds and speech entrainment–audio only as black circles (with one speech entrainment–audio only outlier removed). The linear fit for apraxia of speech–speech entrainment–audio visual is shown as a solid grey line whereas the fit for apraxia of speech–speech entrainment–audio only is shown as a solid black line. The dotted line that intersects the y-axis at 0.00 represents a level where no proportional improvement in speech production with speech entrainment was noted compared with the spontaneous speech condition.

visual speech perception. We chose not to include the speech entrainment–audio only condition used in Experiment 1 in the functional MRI experiment, as the patients' speech output did not improve in this condition (in Experiment 1) and to also

minimize the length of the MRI session in patients for whom we typically try to keep scanning time to a minimum. All three functional MRI conditions elicited robust bilateral activation in patients and their normal counterparts. To appreciate cortical activation associated with speech production, we contrasted spontaneous speech and speech entrainment–audio visual with audio/visual speech perception (perception only) for each participant (patients and normal control subjects) and then combined these individual statistical parametric maps to create two mean group maps ('spontaneous speech > audio/visual speech perception' and 'speech entrainment–audio visual > audio/visual speech perception'). Note that these maps included data from all 30 participants, patients and control subjects. The spatial overlap between the contrasts 'spontaneous speech > audio/visual speech perception' and 'speech entrainment–audio visual > audio/visual speech perception' occurred mostly in the bilateral basal ganglia, motor cortex and anterior cingulate gyrus (Fig. 4A). Greater activation for the spontaneous speech condition in comparison with audio/visual speech perception was found mostly in bilateral medial areas including the basal ganglia, precentral gyrus and cingulate gyrus (Table 2, A). The contrast 'speech entrainment–audio visual > audio/visual speech perception' revealed greater activity involving mostly the bilateral anterior insula, posterior and inferior BA 37, and middle temporal gyrus. For each of the aforementioned contrasts, no greater activation was found for the audio/visual speech perception condition. To assess neural activity that supports speech entrainment, we contrasted speech entrainment–audio visual with spontaneous speech. This comparison revealed greater bilateral activation for speech entrainment–audio visual at the junction of the anterior insula and BA 47, in BA 37, as well as unilaterally in the left middle temporal gyrus, and the most dorsal portion of Broca's area (Fig. 4B). The opposite contrast (spontaneous speech > speech entrainment–audio visual) yielded activation mostly in the bilateral orbitofrontal lobe and precuneus (Table 2, D).

Although common cortical activation for the contrast 'speech entrainment–audio visual > spontaneous speech' was found among the normal and aphasic participants (Fig. 4B), it is possible that one group contributed disproportionately to this result. Therefore, a region of interest analysis was conducted where activation (qualified as voxels with 90th percentile of per cent signal change) in each cluster seen in Fig. 4B was compared between the two conditions (speech entrainment–audio visual versus spontaneous speech) for both groups (patients versus normal controls). Using a 2×2 ANOVA, this analysis focused on the main effect of group in each region of interest, as the main effect of condition was already established based on the whole brain analysis. Out of six regions of interest, main effects of group were found in two areas: the right anterior insula/BA 47 (this cluster was contiguous including the anterior insula and extending into BA 47) where patients showed overall greater activation compared with the normal control group ($P = 0.01$), and in the left middle temporal gyrus where patients showed less activation compared with the normal group ($P = 0.005$). No interactions of group by condition were revealed (Fig. 5).

The normal subjects included here were able to mimic speech in real-time, without much effort. Similarly, the patients produced

relatively fluent speech during speech entrainment–audio visual compared with spontaneous speech. Notably, the patients were able to accomplish this despite having non-fluent propositional speech. To better understand the neural mechanism that allows these patients to produce relatively fluent speech while mimicking others, we contrasted the main comparison listed above ('speech entrainment–audio visual > spontaneous speech') between patients and control participants. This analysis showed that patients have relatively higher activation than their normal counterparts in posterior regions, primarily focused around the notch of the sylvian fissure extending into the supra-marginal gyrus and the angular gyrus (Fig. 4C). The opposite contrast showed that normal subjects had greater activation in the left anterior temporal pole, a region that was damaged in many of the aphasic patients (Table 2, F).

As shown in Fig. 4B, the contrast 'speech entrainment–audio visual > spontaneous speech' revealed robust activation most notably in the bilateral anterior insula and BA 47 as well as in bilateral BA 37. To understand the relationship among these areas at a cortical network level, we created probabilistic maps of the white matter connections between each of these areas using DTI data from 17 normal subjects. This analysis revealed white matter connections between the left and right anterior insula/BA 47 via the body of the corpus callosum and between the left and right BA 37 through the splenium of the corpus callosum. The intrahemispheric connections between the anterior insula/BA 47 and BA 37 almost exclusively followed a ventral pathway connecting the anterior portion of the temporal lobe and inferior dorsolateral portion of the frontal lobe via the extreme capsule. In contrast, the two areas that did not show bilateral activation, the middle temporal gyrus and

dorsal portion of Broca's area (only active in the left hemisphere), were connected via a dorsal pathway that extended from the temporal lobe onto the frontal lobe via the anatomical location of the arcuate fasciculus (Fig. 6).

Experiment 3: Testing whether training with speech entrainment improves speech production among patients with Broca's aphasia

All 13 patients completed the 6-week training phase in which they practiced producing each of three scripts for 2 weeks using speech entrainment. Their performance on each script was assessed immediately before the training phase for that specific script started, during the week after training completion of the script, and again 6 weeks later. We applied one-tailed statistical tests because we predicted that training would lead to improved performance. A one-sample *t*-test revealed a statistically significant increase in the number of different words produced with speech entrainment at 1 week after treatment completion but not at 6-weeks (results from all pre- and post-treatment comparisons are included in Table 3 and Fig. 7). To test generalization across scripts, patients' ability to produce an untrained script with the aid of speech entrainment was tested before and after treatment. A significant improvement in producing untrained scripts was found at 1-week and at 6 weeks after completion of the training phase. A significant increase in the number of different words was found at 1-week after training completion for the spontaneous speech condition (where patients were instructed to speak about
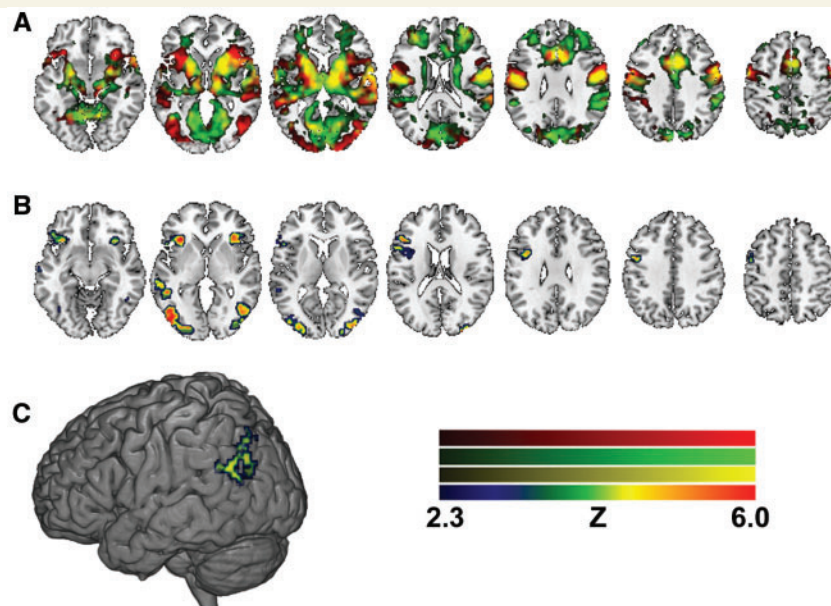

Figure 4 (**A**) Separate mean cortical activation for the contrasts 'speech entrainment–audio visual > audio/visual speech perception' (red colour scale) and 'spontaneous speech > audio/visual speech perception' (green colour scale) including both patients and normal control subjects. (**B**) Results for the contrast 'speech entrainment–audio visual > spontaneous speech' (multi-coloured scale). (**C**) Greater cortical activation was revealed for the aphasic patients in comparison with the normal control group for the contrast 'speech entrainment–audio visual > audio/visual speech perception.' The colour scales denote Z-scores.

Table 2 Functional MRI coordinates for cortical activation

Z-score	x	y	z	Hemisphere	Location	BA
A. 'Spontaneous speech > watching only'						
7.23	−52	−10	14	Left	Precentral gyrus	43
7.14	20	14	6	Right	Putamen	NA
6.62	−20	6	14	Left	Putamen	NA
6.53	0	16	36	Left	Cingulate gyrus	32
B. 'Speech entrainment > watching only'						
6.67	−54	−10	22	Left	Insula	43
6.61	34	16	4	Right	Insula	13
6.57	6	20	28	Right	Cingulate gyrus	32
6.56	12	−10	8	Right	Thalamus	NA
3.07	−12	−30	72	Left	Precentral gyrus	4
C. 'Speech entrainment > spontaneous speech'						
5.02	50	−66	−6	Right	Middle occipital gyrus	37
4.47	−34	24	−4	Left	Extranuclear	47
3.95	−50	−68	0	Left	Inferior temporal gyrus	37
3.55	−52	26	22	Left	Inferior frontal gyrus	45
3.79	34	26	−6	Right	Inferior frontal gyrus	47
3.58	34	22	2	Right	Insula	13
3.58	−52	−44	0	Left	Middle temporal gyrus	22
D. 'Spontaneous speech > speech entrainment'						
5.26	−12	−64	20	Left	Precuneus	31
4.87	−10	60	14	Left	Superior frontal gyrus	10
4.71	2	50	−10	Right	Medial frontal gyrus	10
4.88	50	−44	32	Right	Supramarginal gyrus	40
E. 'Aphasic patients > normal control subjects' for 'SE-AV > AV-SP'						
3.02	−48	−60	30	Left	Superior temporal gyrus	39
2.93	−60	−66	18	Left	Middle temporal gyrus	19
2.85	−48	−62	36	Left	Inferior parietal lobule	39
F. 'Normal control subjects > aphasic patients' for 'SE-AV > AV-SP'						
2.98	−62	8	−8	Left	Superior temporal gyrus	38
2.9	−62	−2	−4	Left	Middle temporal gyrus	21
G. Aphasic patients—speech entrainment: 'pretraining > post-training'						
3.56	−66	−34	26	Left	Inferior parietal lobule	40
3.46	−62	−52	28	Left	Supramarginal gyrus	40
H. Aphasic patients—speech entrainment: 'pretraining < post-training'						
4.09	16	−62	22	Right	Precuneus	31
3.81	20	−58	4	Right	Posterior cingulate	30
3.75	0	50	8	Left	Medial frontal gyrus	10
3.52	−4	46	4	Left	Anterior cingulate	32
3.67	32	−32	−16	Right	Parahippocampal gyrus	36
3.52	38	−76	42	Right	Precuneus	19

The first four contrasts listed represent group results involving both patients and normal control participants.
AV-SP = audio-visual–speech perception; SE-AV = speech entrainment–audio-visual.

a trained script topic without the aid of speech entrainment). There was also an increase in the number of different words for the spontaneous speech condition at 6 weeks post-training; however, this change was not statistically significant. Although the magnitude of increase in different words was the greatest for the spontaneous speech condition, far greater variance in task performance was found for this condition compared with the two conditions where patients produced scripts with the aid of speech entrainment. Note that the degrees of freedom for trained scripts with speech entrainment were reduced, as one patient was unable to produce a single word at baseline making it impossible

to estimate proportional increase in spoken words. This was also the case for the two remaining conditions where an additional patient was unable to speak a single word at baseline. To understand whether patients who were able to complete more script runs in each practice session showed greater improvement in word production, Spearman correlation coefficients (one-tailed) were calculated for the relationship between the outcome data (increase in words produced at 1 and 6 weeks after training completion for trained scripts with speech entrainment–audio visual, untrained scripts with speech entrainment–audio visual and spontaneous speech) and the average number of script runs per day. Out of
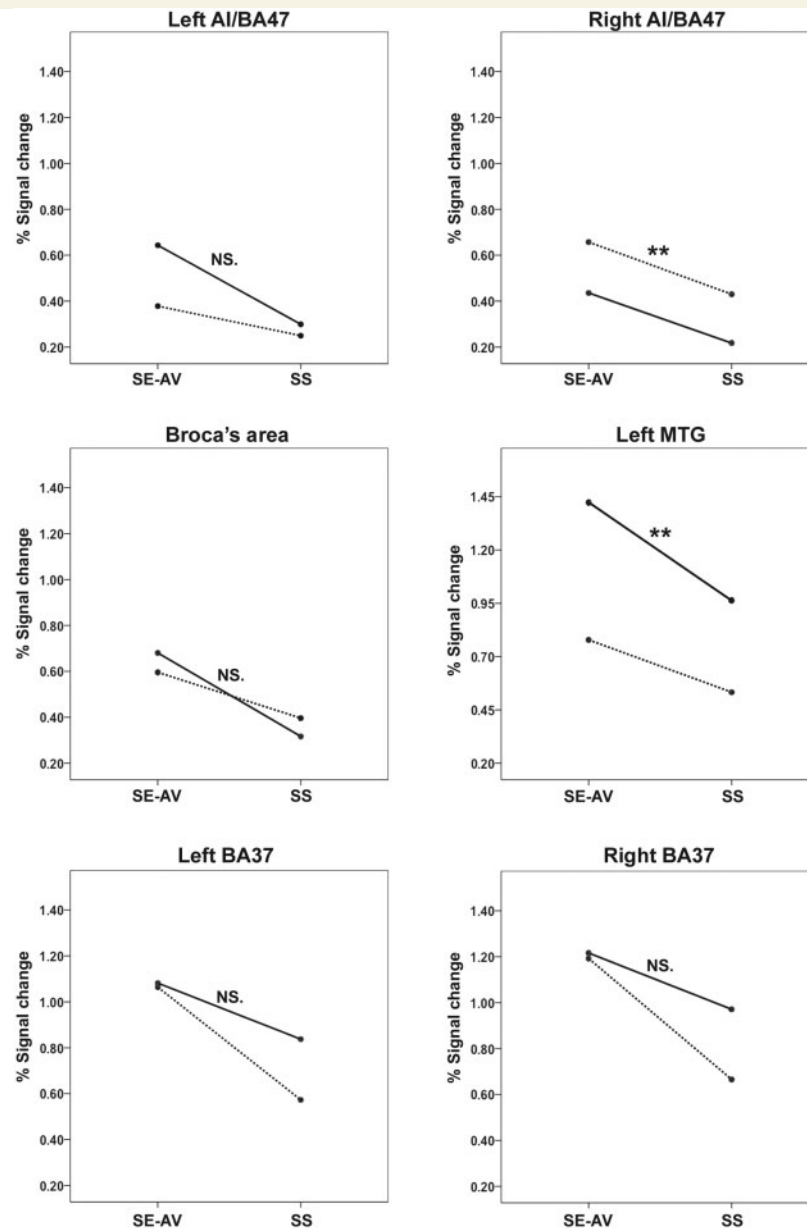
Figure 5 Results from 2 × 2 ANOVA for condition [speech entrainment–audio visual (SE-AV) and spontaneous speech (SS)] by group (patients and normal control subjects) comparing cortical activation in six different regions of interest: left anterior insula (AI)/BA 47, right anterior insula/BA 47, Broca's area, left middle temporal gyrus (MTG), left BA 37 and right BA 37. The solid line denotes data from normal subjects whereas the dotted line denotes patients' data. The scale on the *y*-axis presents signal change in the functional MRI signal in each condition. NS = not significant **$P < 0.01$.
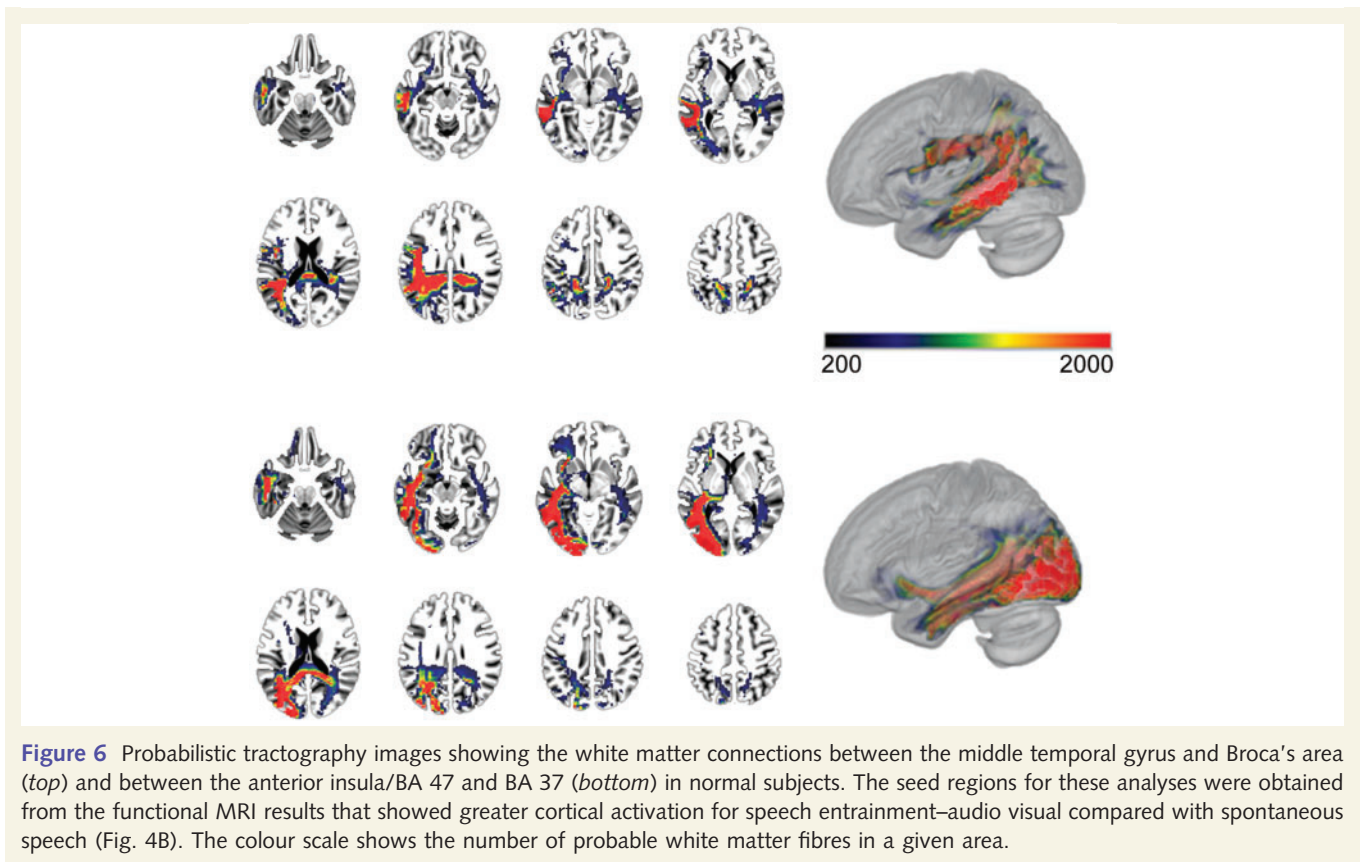
six correlation coefficients, none reached statistical significance (range $r = 0.05$ to $r = 0.32$).

To understand how changes in speech production abilities with speech entrainment training related to modulation of brain activity, we compared brain activation in each of three conditions—speech entrainment–audio visual, spontaneous speech and audio/visual speech perception—before and after treatment in 10 patients. No significant differences in cortical activity were found for the spontaneous speech and audio/visual speech perception conditions. In contrast, a significant increase in activation was found after treatment completion for the speech

entrainment–audio visual condition primarily involving the following regions: bilateral cingulate gyrus, precuneus and right hippocampus (Table 2, H). A significant training-related decrease in cortical activation was found in areas of the posterior-inferior parietal lobe, including the supra-marginal gyrus (Fig. 7).

Discussion

We believe this is the first study to show that on-line mimicking of audio-visual speech allows patients with Broca's aphasia to

Figure 6 Probabilistic tractography images showing the white matter connections between the middle temporal gyrus and Broca's area (*top*) and between the anterior insula/BA 47 and BA 37 (*bottom*) in normal subjects. The seed regions for these analyses were obtained from the functional MRI results that showed greater cortical activation for speech entrainment–audio visual compared with spontaneous speech (Fig. 4B). The colour scale shows the number of probable white matter fibres in a given area.

increase their speech output by a factor of >2. Importantly, speech entrainment that relies only on audition does not have the same beneficial effect. This suggests that seeing the mouth of the speech model provides crucial information that allows patients to increase their own speech output. Speech entrainment that relies only on visual (and not auditory) speech yields almost no speech production, even for normal subjects. Therefore, the role of auditory speech in speech entrainment cannot be understated.

Although the primary outcome factor here was the number of different words patients could produce in a given condition, our observations suggested that speech entrainment elicited much improved speech fluency among the patients (as seen in Supplementary Video 1). Speech entrainment facilitates speech production in a way that permits patients with Broca's aphasia to produce a greater variety of words and speak at a rate that may better approximate normal fluency compared with their spontaneous speech. This suggests that the motor commands for speech are relatively intact. Unless a mechanism that encodes the motor commands of speech is at least partially preserved, it is difficult to imagine that someone like G.B., who has moderate apraxia of speech, could produce on-line speech with such clarity. Although his speech articulation is somewhat slurred, it must be remembered that his stroke occurred more than two decades ago and that his speech mechanism has been almost quiescent since then.

Can we identify a mechanism that facilitates speech entrainment? We suggest that the neuroimaging results provide some insight. Greater cortical activity associated with spontaneous speech compared with speech entrainment was found mostly in medial regions commonly thought to support reminiscing and recall of events (Burgess et al., 2001; LaBar and Cabeza, 2006). The opposite contrast (speech entrainment–audio visual > spontaneous speech) revealed the greatest activation in regions that support lexical retrieval (left BA 37; Hillis et al., 2001, 2002; Cloutman et al., 2009) and, perhaps, a mechanism that modifies visceral functions (mostly respiration) for speech production (anterior insula/BA 47; Augustine, 1996; Gracco et al., 2005; Shum et al., 2011; Tremblay et al., 2012). Our probabilistic tractography analysis demonstrates that these regions are connected via ventral fibres and not through the arcuate fasciculus, a tract that is crucial for speech articulation (Hickok and Poeppel, 2007; Hickok, 2012). We suggest that this ventral network encodes the conceptual aspects of speech, at a minimum, the lexicon and aspects of homeostatic functions during speech production (Craig, 2009), and that it is enslaved by a dorsal network involving middle temporal gyrus, superior temporal gyrus and dorsal regions of Broca's area (connected via the arcuate fasciculus). It is a common complaint of patients with Broca's aphasia that they know what they want to say but just cannot get the words out. Although Broca's aphasia has been amply described elsewhere (e.g. Geschwind, 1965; Goodglass et al., 1976; Goodglass, 1993), it appears that these individuals suffer from a 'language inertia' manifesting as inability to translate and execute a language code into speech production. Given that at least some of them can speak using speech entrainment, it would seem that lexical retrieval and speech articulation are relatively intact, although, in some cases, not within normal limits. One treatment method that has garnered considerable attention as means to enable patients with Broca's aphasia to speak is melodic

Table 3 One-sample *t*-tests for training outcomes at 1 week and 6 weeks after training completion

	Week 1	Week 6
Speech entrainment—trained	$t(11) = 2.76$, $P = 0.009$ mean = 13%, SD = 1.7%	$t(11) = 1.4$, $P = 0.09$ mean = 8%, SD = 2.2%
Speech entrainment—untrained	$t(10) = 2.65$, $P = 0.012$ mean = 17%, SD = 2.1%	$t(10) = 2.27$, $P = 0.02$ mean = 10%, SD = 1.6%
Spontaneous speech	$t(10) = 1.95$, $P = 0.04$ mean = 65%, SD = 11.1%	$t(10) = 1.5$, $P = 0.08$ mean = 57%, SD = 12.6%

Means and standard deviations denote per cent change in the number of different words spoken before and after treatment. Note that these results correspond to the bar graph in Fig. 7.
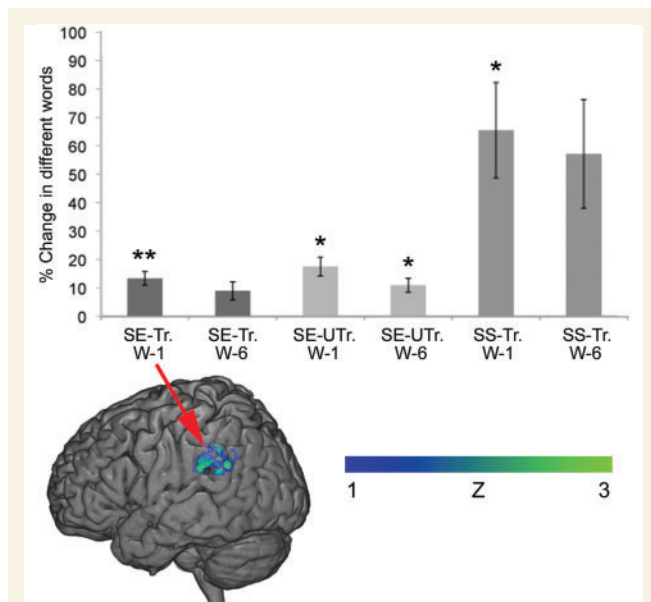

Figure 7 Speech entrainment training results showing per cent increase in different words at 1 week (W-1) and 6 weeks (W-6) after training completion compared with baseline. The first two bars on the left pertain to increase in different words produced with the aid of speech entrainment for trained scripts (SE-Tr.), the two bars in the middle denote increase in different words produced for untrained scripts with the aid of speech entrainment (SE-UTr.), and the two bars on the right refer to increase in different spontaneous words produced while speaking about a given script topic without speech entrainment (SS). The error bars denote standard error. Statistically significant increase in word production is denoted by *$P < 0.05$ or **$P < 0.01$. The brain map shows where significant decrease in cortical activation was found for the speech entrainment condition (red arrow) after treatment completion. The colour scale denotes changes in cortical activation measured as differences in *Z*-scores.

intonation therapy (Albert *et al*., 1973; Sparks and Holland, 1976; Helm-Estabrooks *et al*., 1989). Primarily, this method focuses on patients intoning speech (singing) but also emphasizes speech rhythm. Recently, Stahl *et al*. (2011) demonstrated that providing an external rhythm (using a metronome) promotes greater speech output compared with actual intoning of speech in patients with Broca's aphasia. Based thereon, this group suggested that damage to the basal ganglia was a crucial predictor of speech production. The basal ganglia are thought to play a crucial role in the timing

of motor activity and, when damaged, could affect the timing of motor speech (Fridriksson *et al*., 2005; Giraud *et al*., 2008; Bohland *et al*., 2010; Lu *et al*., 2010a, b; Stahl *et al*., 2011). As can be seen in Fig. 4A, the basal ganglia were particularly active in both the speech entrainment and spontaneous speech conditions, suggesting that speech entrainment may provide patients with crucial speech rhythm that was affected by stroke. However, the basal ganglia were at least partially preserved in all of the patients included in this research (Fig. 1), suggesting that speech entrainment probably does not compensate for an impaired basal ganglia function.

The 'Directions into Velocities of the Articulators' model suggests that posterior regions in the left hemisphere provide feed-forward information during speech production—the supramarginal gyrus involves somatosensory targets and the superior temporal gyrus and superior temporal sulcus provide auditory targets for speech articulation (Guenther *et al*., 1998; Bohland *et al*., 2010; Golfinopolous *et al*., 2010). In a recent review that integrates the 'Directions into Velocities of the Articulators' model with more conventional psycholinguistic models of speech production, Hickok (2012) suggests coordination of the motor programs of speech in Broca's area (specifically BA 44) and auditory syllable targets in the superior temporal gyrus and superior temporal sulcus are supported by an area in the sylvian region at the parietal–temporal boundary. The crucial lesion location that causes impaired speech is in Broca's area (Hillis *et al*., 2004; Richardson *et al*., 2012). Consistent with Hickok's (2012) account, the patients studied here are non-fluent because the motor programmes for speech are impaired. However, speech entrainment permitted speech fluency in these individuals, raising the possibility that some other mechanism might be impaired. We suggest that this crucial mechanism relates to binding and temporal gating (in Broca's area) that pulls along (entrains) lexical processing (BA 37) and on-line modifications of visceral functions for speech production (anterior insula/BA 47). Although lexical processing and on-line visceral modifications for speech production would be expected in the speech entrainment–audio visual and spontaneous speech conditions, normal participants who completed the functional MRI portion of this study reported having to rely heavily on prediction to execute the speech entrainment–audio visual task, suggesting that greater activity in BA 37 and anterior insula/BA 47 could reflect on-line predictions of the upcoming word and modifications to respiratory state to reflect utterance length (e.g. greater inspiration for longer sentences). Furthermore, we propose that speech entrainment works by providing an external gating

mechanism via a visual route (i.e. watching a speech model) that compensates for damage to Broca's area. However, as demonstrated in Fig. 3, speech entrainment does not work for all patients, especially in some cases of very severe apraxia of speech. This implies that motor speech processing needs to be at least partially preserved. Interestingly, the benefit of speech entrainment that only relies on audition is not related to apraxia of speech severity, suggesting that visual speech perception provides crucial feedback that primarily benefits non-fluent patients with milder forms of apraxia of speech. Although Broca's area has been suggested as a crucial area for binding for syntactical processing (Hagoort, 2003), we suggest it has a role in uniting language and articulatory processing for speech production. Using neurotransmitter receptor mapping, Amunts et al. (2010) suggested that Broca's area and surrounding cortex can be further subdivided beyond the traditional BA 44 and BA 45 into several different regions including dorsal (44d) and ventral (44v) regions of BA 44. Accordingly, it is pertinent to emphasize that the gating mechanism that we propose could potentially be attributed to a sub-region of Broca's area, perhaps in area 44d (Fig. 4), although a precise anatomical substrate of this mechanism cannot be determined based on the current data.

Consistent with the Hierarchical State Feedback Control model, fluent speech production relies on feed-forward information for auditory and somatosensory targets. Although speech entrainment may function to allow patients with Broca's aphasia to successfully gait visceral aspects of speech and lexical processing, it is clear that other regions must compensate for damage to Broca's area. Based on our data that showed greater activation at the left temporal–parietal junction for patients compared with their normal counterparts for 'speech entrainment–audio visual > spontaneous speech', it is possible that this compensation occurs primarily in and around area the sylvian region at the parietal–temporal boundary. Cortical activity in approximately the same region was reduced following speech entrainment training in the patients. This provides further evidence that this region may play a crucial role in compensation for impaired gating at the level of Broca's area. Clearly, more data are needed to verify our account regarding speech entrainment and its relation to the importance of Broca's area in gating for speech production. However, the present data provide reasonable evidence to support such an explanation. Furthermore, we would emphasize that any hypothesis explaining speech entrainment in Broca's aphasia would have to account for the fact that in many such individuals, the motor programmes for speech are at least relatively intact.

Although speech entrainment in patients may be interesting from a theoretical point of view as something that could inform normal speech production, our interest was piqued as we saw this as a potential mechanism to treat impaired speech production in non-fluent aphasia. For many patients with Broca's aphasia, speech output is severely limited. If speech entrainment promotes relatively improved speech fluency among patients, it could potentially be used as a way to practice speech over a prolonged period of time. Our results suggest that speech entrainment training not only promotes greater speech output while patients rely on speech entrainment but their ability to speak without it also improves. During the week after completion of the 6-week training phase, patients were able to produce a greater variety of words for trained and untrained scripts while using speech entrainment as well as during a spontaneous speech condition. Importantly, these findings suggest that speech entrainment training generalizes to spontaneous speech. Overall, patients were able to produce >60% more words upon training completion. If this effect is verified in future studies, speech entrainment will have to be considered as an important treatment option for this patient population. As importantly, more research is needed to understand factors that relate to treatment outcome with speech entrainment.

Many different aphasia treatment approaches have been tested with patients with non-fluent aphasia (Rosenbek et al., 1973; Thompson and McReynolds, 1986; Wambaugh, 2006a, b; Crosson et al., 2009; Lee et al., 2009, 2010; Kiran et al., 2011; Hurkmans et al., 2012). Perhaps the most similar approach to speech entrainment is melodic intonation therapy, a technique that emphasizes melody and rhythm to produce fluent speech in non-fluent patients. As reviewed by van der Meulen et al. (2012), a considerable number of studies have examined the effect of melodic intonation therapy on speech production in non-fluent aphasia. A vast majority of these studies relied on single case examination or small groups of patients. As is often the case with aphasia treatment research, outcome measures varied widely across studies. Nevertheless, at least one group study by Bonakdarpour et al. (2003) did find that melodic intonation therapy training was associated with increased number of correct information units in seven patients with non-fluent aphasia and that this increase generalized to spontaneous speech, a finding that is consistent with our results. Clearly, far more research is needed to verify the effects of melodic intonation therapy and speech entrainment to treat non-fluent aphasia. Given the current results, it would seem that comparisons between these two approaches in a larger number of patients are warranted.

As we suggested earlier, even for patients whose spontaneous speech does not improve, it would seem that improvement in speech production that relies on speech entrainment could also promote greater quality of life if patients are able to practice personal scripts that can be used with simple hand-held devices in highly predictable contexts (e.g. telling one's own story of stroke or what it is like to have aphasia) or for short, communicatively rich personally selected phrases.

Acknowledgements

Funding

Supplementary material

Supplementary material is available at Brain online.

References

Albert M, Sparks R, Helm N. Melodic intonation therapy for aphasia. Arch Neurol 1973; 29: 130–1.

Amunts K, Lenzen M, Friederici A, Schleicher A, Morosan P, Palomero-Gallagher N, et al. Broca's region: noel organization principles and multiple receptor mapping. PLoS Biol 2010; 21: e1000489.

Augustine J. Circuitry and functional aspects of the insular lobe in primates including humans. Brain Res Brain Res Rev 1996; 22: 229–44.

Beckmann C, Jenkinson M, Smith S. General multi-level linear modelling for group analysis in fMRI. Neuroimage 2003; 20: 1052–63.

Bernstein L, Auer E, Wagner M, Ponton C. Spatiotemporal dynamics of audiovisual speech processing. Neuroimage 2008; 39: 423–35.

Bohland J, Bullok D, Guenther F. Neural representations and mechanisms for the performance of simple speech sequences. J Cogn Neurosci 2010; 27: 1504–29.

Bonakdarpour B, Eftekharzadeh A, Ashayeri H. Preliminary report on the effects of melodic intonation therapy in the rehabilitation of Persian aphasic patients. Irn J Med Sci 2000; 2: 156–60.

Bonakdarpour B, Eftekharzadeh A, Ashayeri H. Melodic intonation therapy in Persian aphasic patients. 2003; 17: 75–95.

Brett M, Leff A, Rorden C, Ashburner J. Spatial normalization of brain images with focal lesions using cost function masking. Neuroimage 2001; 14: 486–500.

Broca P. Loss of speech, chronic softening and partial destruction of the anterior left lobe of the brain. Bulletin de la Société Anthropologique 1861; 2: 235–8.

Brookshire R. Introduction to neurogenic communication disorders. 6th edn. Mosby: St. Louis; 2003.

Buchsbaum B, Baldo J, Okada K, Berman K, Dronkers N, D'Esposito M, et al. Conduction aphasia, sensory-motor integration, and phonological short-term memory—an aggregate analysis of lesion and fMRI data. Brain Lang 2011; 119: 119–28.

Burgess P, Gonen-Yaacovi G, Volle E. Functional neuroimaging studies of prospective memory: what have we learnt so far? Neuropsychologia 2001; 49: 2246–57.

Cloutman L, Gottesman R, Chaudry P, Davis C, Kleinman J, Pawlak M, et al. Where (in the brain) do semantic errors come from? Cortex 2009; 45: 641–9.

Code C. Models, theories and heuristics in apraxia of speech. Clin Ling Phon 1998; 12: 47–65.

Craig A. How do you feel—now? The anterior insula and human awareness. Nat Rev Neurosci 2009; 10: 59–70.

Crosson B, Moore A, McGregor K, Chang Y, Benjamin B, Goptinath K, Sherod M, Wierenga C, et al. Regional changes in word-production laterality after a naming treatment designed to produce a rightward shift in frontal activity. Brain and Language 2009; 111: 73–85.

Dabul B. Apraxia battery for adults. 2nd edn. Austin: PRO-ED Inc.; 2000.

Darley F, Aronson A, Brown J. Motor speech disorders. Philadelphia: Saunders; 1975.

Dronkers N. A new brain region for coordinating speech articulation. Nature 1996; 384: 159.

Edmonds L, Nadeau S, Kiran S. Effect of verb network strengthening treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. Aphasiology 2009; 23: 402–24.

Fridriksson J, Ryalls J, Rorden C, Morgan P, George M, Bayliss G. Brain damage and cortical compensation in foreign accent syndrome. Neurocase 2005; 11: 319–24.

Fridriksson J, Moss J, Davis B, Baylis G, Bonilha L, Rorden C. Motor speech perception modulates the cortical language areas. Neuroimage 2008; 41: 605–13.

Fridriksson J, Baker J, Whiteside J, Eoute D, Moser D, Vesselinov R, et al. Treating visual speech perception to improve speech production in non-fluent aphasia. Stroke 2009; 40: 853–8.

Fridriksson J, Bonilha L, Baker J, Moser D, Rorden C. Activity in preserved left hemisphere regions predicts anomia severity in aphasia. Cereb Cortex 2010; 20: 1013–101.

Fridriksson J, Richardson J, Fillmore P, Cai B. Left hemisphere plasticity and aphasia recovery. Neuroimage 2012; 60: 854–63.

Geschwind N. Disconnexion syndromes in animals and man. Brain 1965; 88: 237–94.

Giraud A, Neumann K, Bachoud-Levi A, von Gudenberg A, Euler H, Lanfermann H, et al. Severity of dysfluency correlates with basal ganglia activity in persistent developmental stuttering. Brain Lang 2008; 104: 190–9.

Golfinopolous E, Tourville J, Guenther F. The integration of large-scale neural network modeling and functional brain imaging in speech motor control. Neuroimage 2010; 52: 862–74.

Goodglass H. Understanding aphasia. Foundations of neuropsychology. San Diego: Academic Press; 1993.

Goodglass H, Kaplan E, Weintraub S, Ackerman N. The "Tip-of-the-tongue" phenomenon in aphasia. Cortex 1976; 12: 145–53.

Gracco V, Tremblay P, Pike B. Imaging speech production using fMRI. Neuroimage 2005; 26: 294–301.

Guenther F, Ghosh S, Tourville J. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang 2006; 96: 280–301.

Guenther F, Hampson M, Johnson D. A theoretical investigation of reference frames for the planning of speech movements. Psychol Rev 1998; 105: 611–33.

Hagoort P. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. Neuroimage 2003; 20: S18–29.

Hall D, Fussell C, Summerfield A. Reading fluent speech from talking faces: Typical brain networks and individual differences. J Cogn Neurosci 2005; 17: 939–53.

Helm-Estabrooks N, Nicholas M, Morgan A. MIT. melodic intonation therapy manual. San Antonio: Special Press; 1989.

Hickok G. Computational neuroanatomy of speech production. Nat Rev Neurosci 2012; 13: 135–45.

Hickok G, Poeppel D. The cortical organization of speech processing. Nat Rev Neurosci 2007; 8: 393–402.

Hillis A, Tuffiash E, Wityk R, Barker P. Regions of neural dysfunction associated with impaired naming of objects and actions in acute stroke. Cogn Neuropsychol 2002; 19: 523–34.

Hillis A, Barker P, Wityk R, Aldrich E, Restreop L, Breese E, et al. Variability in subcortical aphasia is due to variable sites of cortical hypoperfusion. Brain Lang 2004; 89: 524–30.

Hillis A, Kane A, Tuffiash E, Ulatowski J, Barker P, Beauchamp N, et al. Reperfusion of specific brain regions by raising blood pressure restores selective language functions in subacute stroke. Brain Lang 2001; 79: 495–510.

Jenkinson M, Smith S. A global optimisation method for robust affine registration of brain images. Med Image Anal 2001; 5: 143–56.

Jenkinson M, Bannister P, Brady M, Smith S. Improved optimisation for the robust and accurate linear registration and motion correction of brain images. Neuroimage 2002; 17: 825–41.

Johns D, LaPointe L. Neurogenic disorders of output processing: apraxia of speech. In: Whitaker H, Whitaker HA, editors. Studies in neurolinguistics. Vol. I. New York: Academic Press; 1976.

Kaplan E, Goodglass H, Weintraub S. Boston naming test. 2nd edn. Philadelphia: Lippincott: Williams, & Wilkins; 2001.

Katz W, Bharadwaj S, Carstens B. Electromagnetic articulography treatment for an adult with Broca's aphasia and apraxia of speech. J Speech Lang Hear Res 1999; 42: 1355–66.

Kendall D, Rosenbek J, Heilman K, Conway T, Klenberg K, Gonzalez , Rothi L, et al. Phoneme-based rehabilitation of anomia in aphasia. Brain Lang 2008; 105: 1–17.

Kertesz A. The western aphasia battery—revised. New York: Grune & Stratton; 2007.

Kiefte M, Armson J. Dissecting choral speech: properties of the accompanist critical to stuttering reduction. J Commun Disord 2008; 41: 33–48.

Kim S, Lee D, You H, Son S, Cho Y, Chang M, et al. The clinical application of the arcuate fasciculus for stroke patients with aphasia: a diffusion tensor tractography study. NeuroRehabilitation 2011; 29: 305–10.

Kiran S, Sandberg C, Sebastian R. Treatment of category generation and retrieval in aphasia: effect of typicality of category items. J Speech Lang Hear Res 2011; 54: 1101–17.

Kotz S, D'Ausilio A, Raettig T, Begliomini C, Craighero L, Fabbri-Destro M, et al. Lexicality drives audio-motor transformations in Broca's area. Brain Lang 2010; 112: 3–11.

LaBar K, Cabeza R. Cognitive neuroscience of emotional memory. Nat Rev Neurosci 2006; 7: 54–64.

Lee J, Kaye R, Cherney L. Conversational script performance in adults with non-fluent aphasia: Treatment intensity and aphasia severity. Aphasiology 2009; 23: 885–897.

Lee J, Fowler R, Rodney D, Cherney L, Small S. IMITATE: an intensive computer based treatment for aphasia based on action observation and imitation. Aphasiology 2010; 24: 449–65.

Lu C, Chen C, Ning N, Ding G, Guo T, Peng D, et al. The neural substrates for atypical planning and execution of word production in stuttering. Exp Neurol 2010a; 221: 146–56.

Lu C, Peng D, Chen C, Ning N, Ding G, Li K, et al. Altered effective connectivity and anomalous activity in the basal ganglia-thalamocortical circuit of stuttering speakers. Cortex 2010b; 46: 49–67.

MacWhinney B. The CHILDES project: tools for analyzing talk. 3rd edn. Mahwah: Lawrence Erlbaum Associates Inc.; 2000.

Max L, Caruso AJ, Vandevenne A. Decreased stuttering frequency during repeated readings: a motor learning perspective. J Fluency Disord 1997; 22: 17–33.

McNeil M. Current concepts in adult aphasia. Int Rehabil Med 1984; 6: 128–134.

McNeil M, Liss J, Tseng C, Kent R. Effects of speech rate on the absolute and relative timing of apraxic and conduction aphasic sentence production. Brain Lang 1990; 38: 135–158.

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

Richardson J, Fillmore P, Rorden C, LaPointe L, Fridriksson J. Re-establishing broca's original findings. Brain Lang, 2012. In press.

Rorden C, Davis B, George M, Borckardt J, Fridriksson J. Broca's area is crucial for visual discrimination of speech but not non-speech oral movements. Brain Stimul 2008; 1: 383–5.

Rosenbek J, Lemme M, Ahern M, Harris E, Wertz R. A treatment for apraxia of speech in adults. J Speech Lang Hear Res 1973; 38: 462–72.

Schlaug G, Marchina S, Norton A. From singing to speaking: why singing may lead to recovery of expressive language function in patients with Broca's aphasia. Music Percept 2008; 25: 315–23.

Shum M, Shiller D, Baum S, Gracco V. Sensorimotor integration for speech motor learning involves the inferior parietal cortex. Eur J Neurosci 2011; 34: 1817–22.

Smith S. Fast robust automated brain extraction. Hum Brain Mapp 2002; 17: 143–55.

Sparks R, Holland A. Method: melodic intonation therapy for aphasia. J Speech Hear Disord 1976; 41: 287–97.

Stahl B, Kotz S, Henseler I, Turner R, Geyer S. Rhythm in disguise: why singing may not hold the key to recovery from aphasia. Brain 2011; 134: 3083–93.

Thompson C, McReynolds L. Wh-interrogative production agrammatic aphasia: An experimental analysis of auditory-visual stimulation and direct-production treatment. Journal of Speech and Hearing Research 1986; 29: 193–206.

Tremblay P, Deschamps I, Gracco V. Regional heterogeneity in the processing and the production of speech in the human planum temporale. Cortex, 2012. In press.

Turkeltaub P, Messing S, Norise C, Hamilton R. Are networks for residual language function and recovery consistent across aphasic patients? Neurology 2011; 76: 1726–34.

van der Meulen I, van de Sandt-Koenderman M, Riggers G. Melodic intonation therapy: present controversies and future opportunities. Arch Phys Med Rehabil 2012; 93: S46–52.

Vander Wyk B, Ramsay G, Hudac C, Jones W, Lin D, Klin A, et al. Cortical integration of audio-visual speech and non-speech stimuli. Brain Cogn 2010; 74: 97–106.

Wambaugh J. Treatment guidelines for acquired apraxia of speech: a synthesis and evaluation of the evidence. J Med Speech Lang Path 2006a; 14: 15–33.

Wambaugh J. Treatment guidelines for apraxia of speech: lessons for future research. J Med Speech Lang Path 2006b; 14: 317–21.

Warren J, Crinion J, Lambon Ralph M, Wise R. Anterior temporal lobe connectivity correlates with functional outcome after aphasic stroke. Brain 2009; 132: 3428–42.

Woolrich M. Robust group analysis using outlier inference. Neuroimage 2008; 41: 286–301.

Woolrich M, Ripley B, Brady J, Smith S. Temporal autocorrelation in univariate linear modelling of fMRI data. Neuroimage 2001; 14: 1370–86.

Woolrich M, Behrens T, Beckmann C, Jenkinson M, Smith S. Multi-level linear modelling for fMRI group analysis using bayesian inference. Neuroimage 2004; 21: 1732–47.

Worsley K. Ch 14, Statistical analysis of activation images. In: Jezzard P, Matthews P, Smith S, editors. Functional MRI: an introduction to methods. New York: OUP; 2003.