ELSEVIER
NeuroImage
www.elsevier.com/locate/ynimg
NeuroImage 26 (2005) 294–301

Technical Note

Imaging speech production using fMRI

Vincent L. Gracco,[a,b,*] Pascale Tremblay,[a] and Bruce Pike[c]

[a]*School of Communication Sciences and Disorders, McGill University, Faculty of Medicine, 1266 Pine Avenue West, Montreal, Quebec, Canada H3G 1A8*
[b]*Haskins Laboratories, New Haven, CT 06511-6695, USA*
[c]*McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Quebec, Canada*

Received 16 September 2004; revised 13 January 2005; accepted 21 January 2005
Available online 4 March 2005

Human speech is a well-learned, sensorimotor, and ecological behavior ideal for the study of neural processes and brain-behavior relations. With the advent of modern neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), the potential for investigating neural mechanisms of speech motor control, speech motor disorders, and speech motor development has increased. However, a practical issue has limited the application of fMRI to issues in spoken language production and other related behaviors (singing, swallowing). Producing these behaviors during volume acquisition introduces motion-induced signal changes that confound the activation signals of interest. A number of approaches, ranging from signal processing to using silent or covert speech, have attempted to remove or prevent the effects of motion-induced artefact. However, these approaches are flawed for a variety of reasons. An alternative approach, that has only recently been applied to study single-word production, uses pauses in volume acquisition during the production of natural speech motion. Here we present some representative data illustrating the problems associated with motion artefacts and some qualitative results acquired from subjects producing short sentences and orofacial nonspeech movements in the scanner. Using pauses or silent intervals in volume acquisition and block designs, results from individual subjects result in robust activation without motion-induced signal artefact. This approach is an efficient method for studying the neural basis of spoken language production and the effects of speech and language disorders using fMRI.

Keywords: fMRI; Speech production; Neuroimaging

Introduction

Although human speech is one of our most developed functional behaviors and the most common feature of our everyday lives, we still have an incomplete understanding of how the speech production process operates at a fundamental, neurobiological level. For the most part, much of what is assumed about the neuroanatomical and neurophysiological substrate comes from a synthesis of anatomical data on nonhumans, behavioral observations on intact humans and various clinical populations, and a small number of functional neuroimaging studies (see Jurgens, 2002; Kent et al., 2000 for recent summaries). In contrast to the volume of literature using modern functional neuroimaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) for studies of language processing and other cognitive behaviors, "…there are a limited number of PET and fMRI studies that have concentrated on speech production per se…" (Fiez, 2001). For studies employing PET, the most obvious reasons are the limitations in experimental design and the consequences of radioactivity exposure. For studies employing fMRI, the primary reason is motion artefact.

For fMRI studies of speech production, swallowing, and orofacial movements in general, movement-induced artefact has been seen as a significantly limiting factor (cf. Fiez, 2001; Kent et al., 2001). The motion-induced artefact associated with fMRI studies of human speech production comes from two related sources and introduces both direct and indirect signal changes independent of any signal change related to neuronal activation (Birn et al., 1999). The direct source of signal artefact comes from head movement, a fundamental problem for all functional imaging studies. For overt speech and orofacial nonspeech tasks, head motion is always a concern since movements of the mandible are always accompanied by some degree of correlated head motion. The application of any solution to eliminate or minimize head motion during overt speech requires care and attention to head immobilization and the application of a robust motion correction post-processing algorithm. The indirect source of signal artefact, specific to studies of speech, swallowing, and orofacial movements, is more problematic and leads to image warping due to magnetic field variations resulting from motion close to but outside of the field of view (Birn et al., 1998). Producing even single words during scanning results in magnetic field distortions causing voxels in echo planar images (EPI) to shift in the phase-encoding direction by an amount related to the amount of offset in the magnetic field (Birn et al., 1998). The resultant signal changes can either mask or mimic the signal changes due to neural activation.

---

[*] Corresponding author. Fax: +1 514 398 8123.
E-mail address: vincent.gracco@mcgill.ca (V.L. Gracco).
Available online on ScienceDirect (www.sciencedirect.com).

1053-8119/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2005.01.033

A number of approaches have been used to overcome the potential for false activations associated with overt speech. One solution to the motion-induced artefact problem, proposed by Birn et al. (1999), relies on the use of brief (less than 2 s) stimuli and an event-related design to identify different temporal profiles of signal variations associated with the BOLD and motion-induced signals. These time course differences are used to remove the signal change resulting from motion prior to determining the BOLD response. A related, multi-step image processing approach was proposed by Huang et al. (2001) to separate motion-induced signals from activation-induced signals. An alternative and/or complement to signal processing techniques is to discard the images that are acquired during the motion or behavior of interest (overt speech) (Barch et al., 1999; Birn et al., 1999; Riecker et al., 2002; Wilson et al., 2004), thereby eliminating the functional data that may be contaminated by the motion.

While these approaches can and will reduce the motion artefact associated with tasks involving brief motion, they have a number of limitations. The behavior of interest must be short so that the motion-induced signal changes do not overlap with the hemodynamic response function (HRF). This precludes the investigation of long, sequential speech or nonspeech tasks. Moreover, the use of block designs, which generally result in greater functional contrast-to-noise levels compared to event-related designs, is also precluded as there can be no overlap in the motion-induced signal and HRF (Birn et al., 1999). When attempting to eliminate motion artefact by discarding the volumes acquired during the actual motion, some part of the neural activity associated with the motor planning process may be eliminated because it will most likely overlap in time with the behavior of interest, or, if speech motor performance varies, the likelihood is high that the motion-induced activity will not be confined to the discarded volume. Finally, speech-related movement during multislice EPI can cause misalignment of slice selection relative to the brain, potentially disturbing the MRI signal equilibrium. If the motion exceeds the inter-slice gap during sequential slice acquisition, the same region of tissue can be excited more than once within a scan, while other regions are not excited. This can produce differential spin history effects, which decay according to the T1 relaxation time. Thus, even if scans containing artefact are discarded and no motion occurs during the retained volumes, artefact-related effects can still influence subsequent scans. The preferable way to eliminate motion-induced signal change is to avoid its occurrence.

One recent approach to avoiding motion in fMRI studies of speech and spoken language is to use covert or silent speech, in which there is no actual movement of the speech articulators (Ackermann et al., 1998; Riecker et al., 2000; Wildgruber et al., 1996, 2001). Here, a fundamental assumption is that internal speech is similar to, and a valid replacement for, actual speech production. However, this assumption appears to be unsupported or at best, partially supported. A number of studies have demonstrated that covert or silent speech does not activate the same networks that are used during overt speech (Barch et al., 1999; Bookheimer et al., 1995; Huang et al., 2001; Price et al., 1994). The lack of correspondence between brain activation for covert and overt speech may result from two different phenomena. First, it is not possible to monitor what subjects actually do when they engage in covert speech nor is it clear how best to instruct them. Uncontrolled activation, either in terms of level of activity or location of activity, due to the lack of experimental control of the subjects behavior, introduces a potential confound to the data. Second, since any motor behavior involves a network of interacting brain regions modulated by the peripheral (self-generated) aspects of the behavior, there will be no contribution of self-generated feedback to the activated areas and no way to determine what is normally contributed to the network by the areas that are not activated for the covert task.

The requirements, then, for the acquisition of valid and reliable fMRI activations investigating speech or orofacial nonspeech behaviors can be summarized as follows. The functional imaging data need to be directly associated with the production of the behavior of interest, the blood-oxygen level dependent (BOLD) signal needs to be artefact-free, the head needs to be comfortably immobilized, and the experimental design should be optimized for functional contrast-to-noise levels. Recently, we have obtained excellent functional imaging data for overt speech and nonspeech orofacial movement using a technique in which the behavior of interest is produced during pauses in EPI volume time-series acquisitions. Specifically, using a clustered or sparse image acquisition technique (Eden et al., 1999; Edmister et al., 1999; Hall et al., 1999), the gradients are switched off during periods of speech and nonspeech orofacial movement and then switched back on for image acquisition (the behavior interleaved gradients technique of Eden et al. (1999). With appropriate temporal considerations in the design, the technique takes advantage of the physiological delay in the hemodynamic response function and can be used with either event-related or block design acquisition.

A number of recent studies have employed this approach with subjects producing single words (Abrahams et al., 2003; de Zubicaray et al., 2000, 2001). These researchers have presented apparent artefact-free, group-level cortical activations for single-word productions. Only one study of Japanese spoken language has reported use of this technique with speech material longer than single words (Hashimoto and Sakai, 2003). In this report, we present an overview of our results demonstrating the utility of this technique for acquiring functional speech and nonspeech orofacial movement data. We focus on data from individual subjects to illustrate the robustness of the functional activations and the utility of the technique for both longer speech utterances and nonspeech orofacial voluntary movements. We also report results using a customized head-restraining device to minimize slow head drift within and across experimental runs that improves the quality of speech-related functional data.

Materials and methods

Subjects

The data presented in this report were obtained from a total of 20 healthy subjects between the ages of 20 and 30 years who participated in two separate experiments. For both studies, ten subjects (balanced for gender) comprised the experimental groups. All subjects were right-handed.

Tasks

In order to minimize uncontrolled experimental factors that may contribute to various kinds of artefact, a number of pre-experiment routines are used. Prior to each experimental session, the experimental design is explained to the subject, and each subject is allowed a short practice session outside of the magnet. All subjects are provided specific instructions about minimizing head motion within and across experimental runs and regarding breathing. The potential problem with breathing, if uncontrolled, is that respiratory activity associated with speech has been shown to contribute significantly to levels of functional activation in cortical sensorimotor areas (Ramsay et al., 1993). In order to eliminate the potential confound of different amounts and patterns of respiratory activity during speech, speech breathing needs to be considered in experimental designs. In the studies presented here, speaking short sentences or repeating words on a single breath is representative of natural speech behavior.

Study one

The experimental design for study one was blocked (five trials per block) and consisted of two listening conditions and two speaking conditions (three words in sequence or three word sentences). Subjects were fitted with high-quality MR-compatible headphones (Resonance Technology) and were instructed through a back-projected visual display to either listen or listen and then repeat words or sentences. All subject responses were recorded with an MR-compatible microphone attached to the headphones (Resonance Technology). A schematic of the experimental design for study one is shown in Fig. 1.

Study two

The experimental design for study two was also blocked. Subjects performed blocks of 4 different tasks (3 speech and one nonspeech), each repeated 2 or 3 times per block. There were two nonspeech orofacial movements used, consisting of lip pursing followed by lip retracting, and jaw lowering followed by tongue raising.

Scanning protocol

All data were acquired on a 1.5-T Siemens Sonata MR scanner at the Montreal Neurological Institute. For study one, twenty-six axial slices oriented parallel to the AC–PC line (thickness = 6 mm, no gap) were acquired in 2.6 s using a mulitslice EPI sequence (TE = 40 ms, TR = 10.0 s, delay in TR = 7.4 s., 128 × 128 matrix). The delay in TR occurred following each volume acquisition. For each subject, a total of 266 volumes were acquired. For study two, thirty-nine axial slices oriented to the AC–PC line (thickness 4 mm, no gap) were acquired in 3.3 s using a mulitslice echo planar imaging sequence (TE = 40 ms, TR = 9.5 s, delay in TR = 6.2 s., 64 × 64 matrix). For each run, 120 volumes were acquired (30 per condition) in 20 min; 3 experimental runs were obtained for each subject. A 3D T1-weighted high-resolution scan was acquired as an anatomical reference. All conditions occurred during the silent period.

Head restraint

For study one, the head was restrained using a vacuum-bag filled with polystyrene balls, fitted around the subject's head. The air is removed from the vacuum bag, and the bag and polystyrene hold the head in place during the experimental session. More recently (for study two), we have combined the use of the vacuum bag with a custom-built head-restraining device to further minimize head movement. The subject is outfitted with the MR-compatible headphones and microphone, the head secured in the polystyrene bag, and the head-restraining device is secured in place. The head-restraining device is made of Ultem 1000, a plastic that can be autoclaved. The base is an acrylic plastic with an adjustable pad that rests firmly against the forehead. While not used for these studies, a chin cup, bite bar, or ear cups can also be mounted on the restraint system. This added restraint has not been reported as uncomfortable by subjects and, in our experience, limits head motion more than tape or a velcro strap.

Movement estimation and correction

All functional images were realigned to the 4th frame of the first functional run and corrected for movement using a six-parameter 3D automated algorithm (AFNI; Cox and Jesmanowicz,

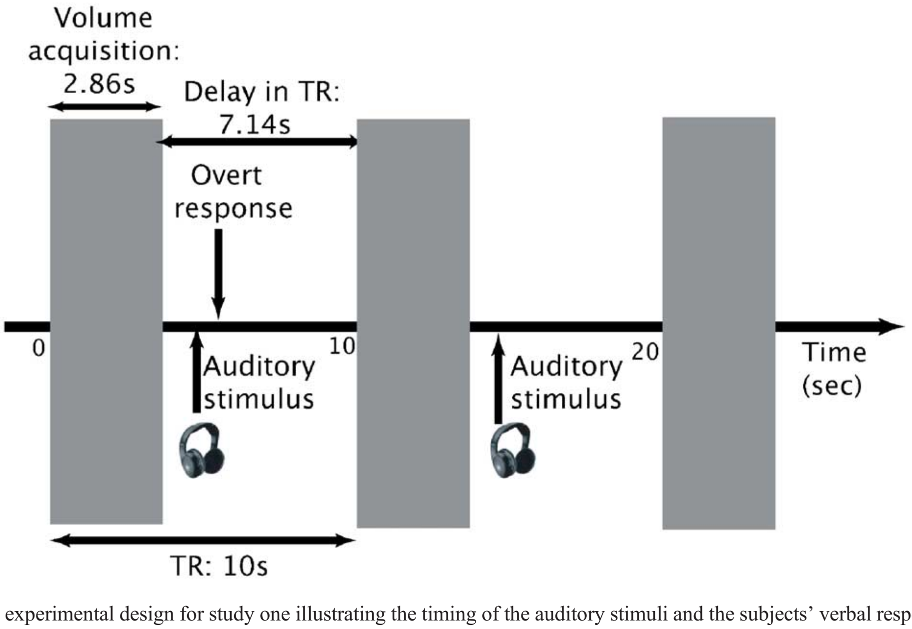

Fig. 1. A schematic of the experimental design for study one illustrating the timing of the auditory stimuli and the subjects' verbal responses with respect to EPI volume acquisition.

1999). Two sets of movement parameters (roll, pitch, yaw in degrees, and linear displacement in millimeters in 3 spatial dimensions) were obtained from AFNI.

Data analysis

The functional images were low pass filtered (6-mm FWHM Gaussian kernel) and transformed into stereotaxic space (Collins et al., 1994). The three first scans, and all trials in which the subjects made a response error, were excluded from the analysis.

The statistical analysis of fMRI data was performed using a linear model with correlated errors (Fmristat, Worsley et al., 2002). The BOLD response for the tasks was compared against a baseline. The *t* statistic images were thresholded using the minimum given by Bonferroni correction and random field theory (Worsley et al., 2002). We used a random effects model for all analyses.

Results

For this report, we present representative data highlighting the quality of the functional data obtained in the absence of motion-induced signal change as well as presenting results in which significant speech motion artefact is present. First, however, we present head motion data illustrating the magnitude of head motion (slow drift) accompanying the two tasks under different degrees of head restraint.

Head motion

On the left side of Fig. 2 are average angular and translational head motion parameters for the group for each of the two experiments. For this comparison, only the head motion parameters from the first functional runs are included. It can be seen that the magnitude of all rotational and translational head motion parameters was reduced for the second (B) experiment compared to the first (A). As mentioned, the head-restraining device was used for study two only. While the experiments are not directly comparable due to the different tasks, as part of a related experiment three of the same subjects from study one were scanned while producing the sentence portion of study one. In this follow-up study, the same subjects produced the same sentence material using the same scanning parameters and experimental design with the head-restraining device in place. The word condition was not included. The results from this more direct comparison are presented on the right side of Fig. 2. All head motion parameters obtained for the subjects with the head restraint in place (D) are reduced compared to the head motion parameters obtained without additional head restraint (C).

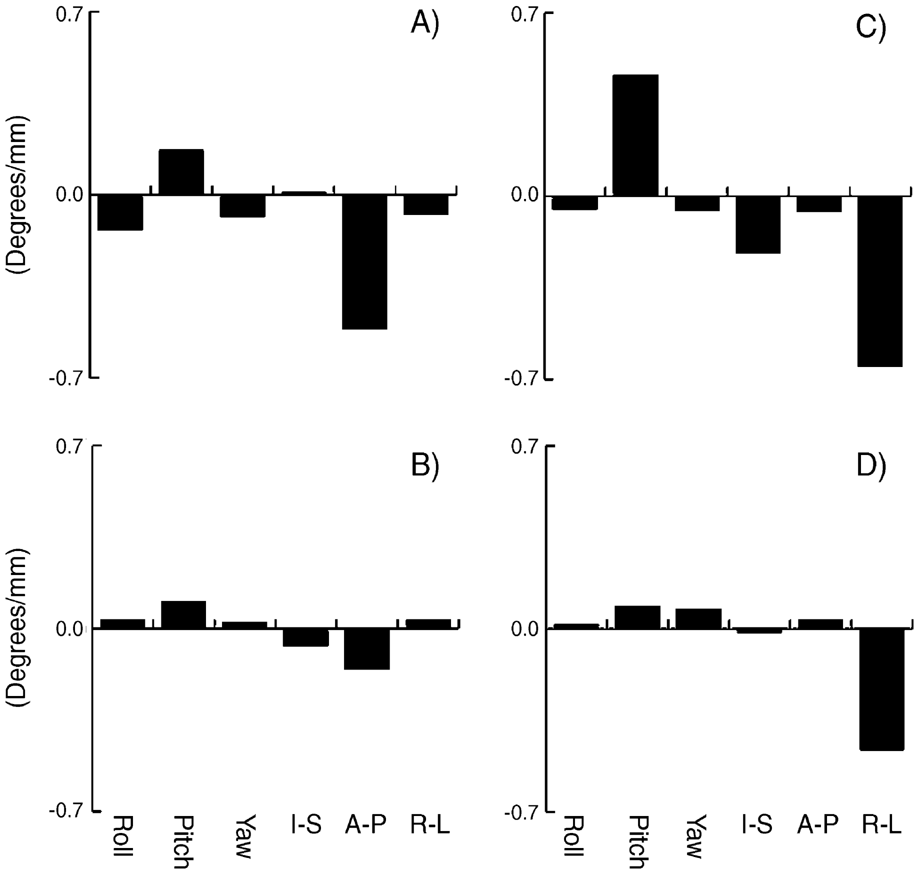

Fig. 2. The average head motion parameters for the group from study one (A) and study two (B). The three rotational parameters (roll, pitch, and yaw) and the three translational parameters [Inferior–Superior (I–S)], [Anterior–Posterior (A–P)], [Right–Left (R–L)] are reported on the same scale with the rotational parameters in degrees and the translational parameters expressed in mm. The top right side of the figure (C) are data from three of the subjects from study one. The bottom right side of the figure (D) are the same subjects re-scanned on a different day using the same scanning protocol and producing the same sentences from study one. This follow-up study, in comparison to study one, used the head-restraining device. The difference in the magnitude of head motion can be seen in a visual comparison of the results in panel C with those presented in panel D.

The values indicate that, on average, head motion was less than a degree in the three rotational dimensions and less than a millimeter in the three translational dimensions. Routinely, head motion parameters are examined and any trial in which the values are greater than 1 mm of translation or 1° of rotation is excluded from analysis. For study one (without the additional head restraint), an average of 4.7 volumes per subject was excluded due to excessive head motion; for study two, only one volume in total was excluded. Overall, while there is head motion accompanying overt speech, the motion was acceptable.

Qualitative observations of speech and nonspeech oral motor activations

In order to illustrate the quality and robustness of the functional data using a delay or pause in volume acquisition (Eden et al., 1999), we examined the magnitude and locations of maximal activations for the two studies and qualitatively examined axial sections looking for any evidence of obvious or subtle motion-induced artefact. We first examined axial sections from study one for any evidence of "halo" artefacts typical of motion-induced signal change. As a point of reference, presented in Fig. 3 is an example from a single subject from an experiment in which an error in timing between speech production and volume acquisition was unintentionally introduced resulting in volume acquisitions in the presence of overt speech movement. As can be seen in the activation maps, there is clear evidence of movement artefact around and within the anterior crown in the axial sections (2-mm steps from $Z = 62$ to $Z = 40$; total 22 mm range). Above and below these levels, the artefact, while still observable, was not as obvious. Based on our recent data (see below) and those in the literature, these spurious regions of statistical significance are undoubtedly mixed with real activations.

In contrast, presented in Fig. 4 are serial axial sections taken at the same location in Talairach space as the sections in Fig. 3, from four of the ten subjects from study one. The activation maps clearly illustrate the lack of any movement-induced artefact and the lack of overlap in the location of the real activations with the spurious ones (Fig. 3). Overall, the data from individual subjects revealed robust levels of activation due in part to the use of a block design. Table 1 is a summary of the maximal t values for each subject with the corresponding stereotaxic coordinates and closest Brodmann areas associated with the maximal t value clusters for all subjects for both studies. In all cases, the maximum activation clusters were found in areas consistent with the sensorimotor representations for speech. Overall, no clusters were found in brain areas outside of those known or suspected to be involved in speech production.

To illustrate further the quality of the functional data for individual subjects acquired with the technique, we examined qualitatively the nonspeech movement data from study two for a representative individual subject. Fig. 5 presents the two non-speech conditions, voluntary lip movement (left) and tongue/jaw movement (right) for a male subject. The figure presents data at the t value threshold corrected for multiple comparisons ($P < 0.05$; top) and at a lower t value threshold ($P < 0.25$; bottom). As can be seen, the quality of the functional data is similar at both threshold levels with no evidence of artefact. In addition, there are clear neuroanatomical differences in the two conditions consistent with different lip and tongue/jaw representation in sensorimotor cortex.

Discussion

The major limiting factor in investigating neural mechanisms of spoken language has been the potential confounding of neural-induced functional activations with motion-induced signal artefact. As a result, the extant literature on the neural mechanisms of speech, spoken language, swallowing, singing, and voluntary orofacial behaviors is small and dominated by studies using PET, silent (covert) speech, or singing, and signal processing approaches whose main focus is to separate true activations from artefact. To date, however, these solutions have seen either limited application or provided results of questionable validity. As mentioned in the Introduction and discussed in numerous publications, movements within the head coil during volume acquisition can result in potential problems that cannot be fully known, therefore cannot be

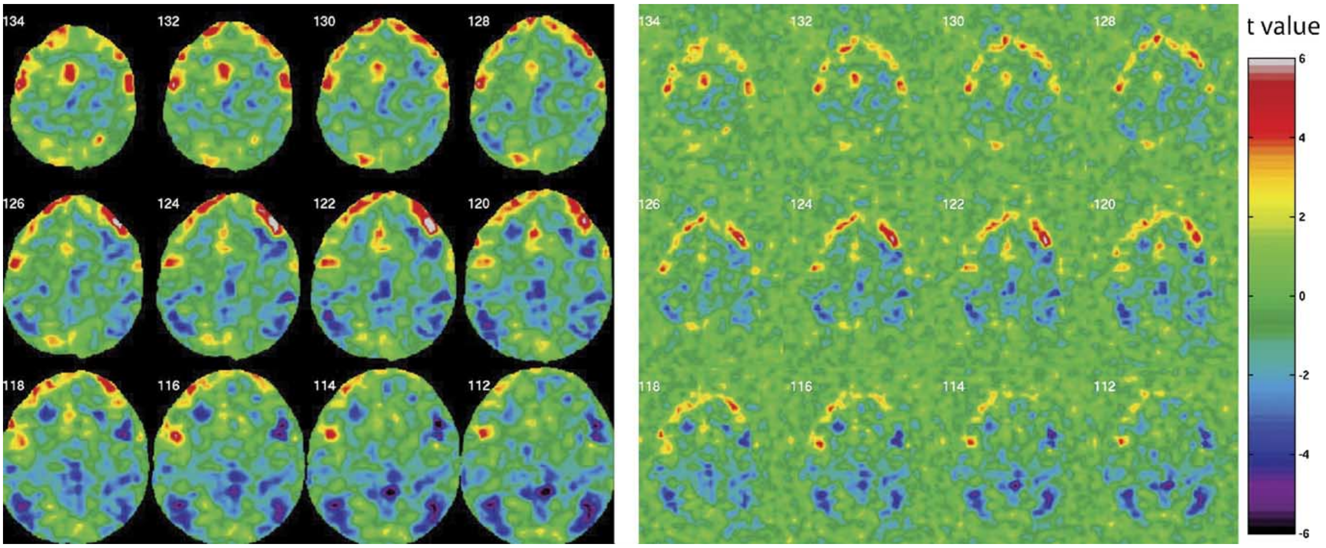

Fig. 3. Serial sections from a single subject illustrating the presence of motion-induced signal artefact. The right side of the figure is the activation map and the left side of the figure is the activation map superimposed on the subjects' anatomical scan. These sections represent slices taken every 2 mm from $Z = 62$ to $Z = 40$. The number in the upper left hand portion of each slice represents the slice number in Talairach space.

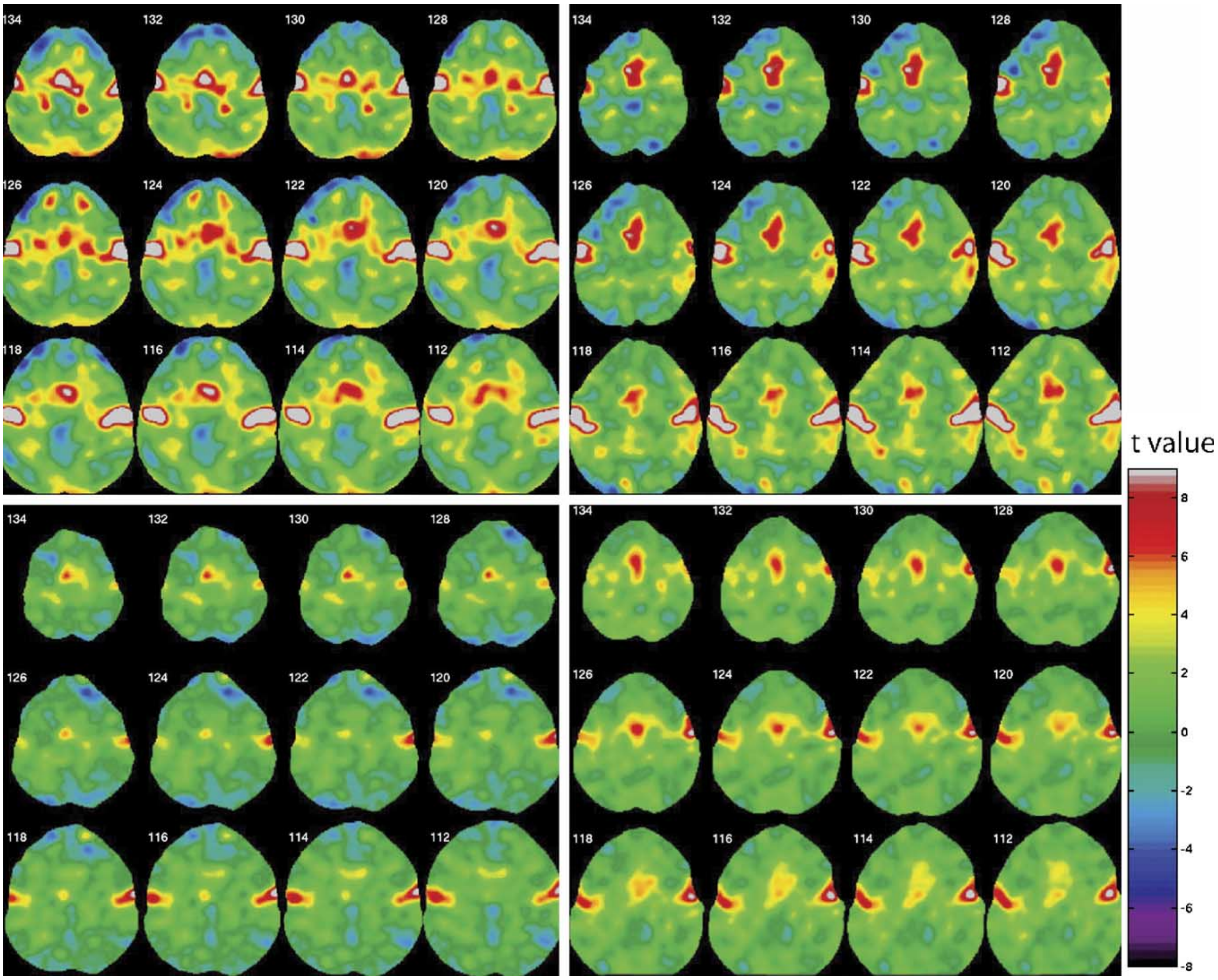

Fig. 4. Serial sections taken from the same locations in Talairach space as in Fig. 3 ranging from Z = 62 to Z = 40 in 2-mm steps. The activation maps are superimposed on the anatomical scans and illustrate the patterns of activation for the sentence condition for 2 male (A) and 2 female (B) subjects.

fully compensated for and lead to false activations and spurious results. In addition, the head movement that is an inherent part of speech (and other orofacial actions such as swallowing) can add to the potential problems by creating drift in head position that, if of sufficient magnitude, can lead to spatial misalignment and introduce another source of false activation. Here we have presented an approach to acquiring BOLD signal changes associated with serial speech and nonspeech orofacial tasks that substantially minimize motion-induced signal artefact in brain activations obtained with fMRI.

Compressed volume acquisition (the behavior interleaved gradients technique) was suggested by Eden et al. (1999) as an

Table 1
Summary of maximum activations for each subject in the two studies indicating the maximum t value, Talairach coordinates of the maximum value, and the corresponding Brodmann areas

	Study one			Study two		
	t value	Coordinate	Area	t value	Coordinate	Area
F1	16.9	(−53, −15, 23)	BA 3/4	9.09	(61, 1, 37)	BA 6
F2	11.9	(64, −11, 9)	BA 42	8.68	(62, −4, 36)	BA 6
F3	16.2	(55, −4, 41)	BA 6	6.19	(−55, −15, 32)	BA 3
F4	8.9	(53, −7, 3)	BA 22	10.15	(−60, −18, 35)	BA 3
F5	10.6	(61, −6, 36)	BA 6	9.13	(−55, −3, 29)	BA 6
M1	17.34	(−52, −22, 47)	BA 2	14.33	(−64, −6, 25)	BA 4
M2	16.31	(44, −12, 40)	BA 4/6	11.94	(59, −8, 23)	BA 4
M3	14.92	(−56, −14, 48)	BA 3	11.11	(−48, −24, 40)	BA 2
M4	13.02	(−42, −18, 40)	BA 4	7.09	(−54, 4, 40)	BA 6
M5	14.71	(−38, −38, 20)	BA 13	12.14	(60, −10, 48)	BA 3

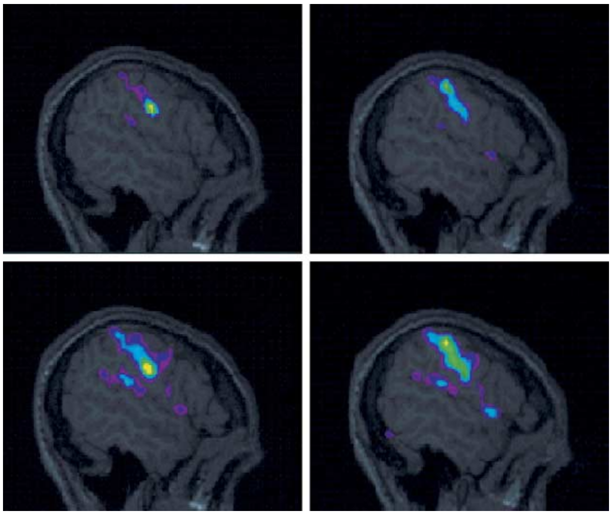

Fig. 5. Functional activations over the left motor cortex for the tongue/jaw (left) and lip (right) movement tasks from study two for a single male subject. Activations are superimposed on the 3D anatomical scan. The top portion of the figure is thresholded at $t = 5.2$ ($P < 0.05$) while the bottom portion of the figure is thresholded at $t = 3.0$ ($P < 0.25$).

approach well suited for studies of speech production. In the results reported here, we have included representative data from individual subjects engaged in more movement-intensive tasks than have been used previously with this technique (Abrahams et al., 2003; de Zubicaray et al., 2000, 2001; Hashimoto and Sakai, 2003). In the first study, subjects produced extensive speech movements (three words in sequence or three word meaningful sentences), while in the second study, subjects produced sequences of novel orofacial movements as well as multisyllable words under different production constraints (whispered speech and mimicked speech). The quality and magnitude of the functional activations obtained from individual subjects illustrate the robustness of the technique.

Head motion

As part of our attempt to maximize signal to noise in our data, we found that adding additional head restraint limited the slow head drift that is a consequence of speech production. While changes in head position can be compensated for in the spatial processing of the functional images, the ability to completely compensate for the large changes in head positioning that may accompany speech is limited (Friston et al., 1996). In order to minimize the effects of any change in head position, head restraint should be a major consideration in studies using overt speech or other orofacial nonspeech tasks.

With regard to the head motion parameters measured for these two studies, a few points need mention. First, while not routinely reported, the head motion parameters for these studies were substantially greater than those reported by Barch et al. (1999) in their study comparing overt and covert speech. This is no doubt due to the more extensive speech motor actions represented in the present studies (producing sentences, voluntary nonspeech movements of the lips, jaw, and tongue). However, the head movements were almost always less than 1 mm and 1° of rotation and were apparently well compensated for by the motion correction algorithm. Second, given the lack of head motion artefacts noted in our studies to date, we suggest that slow head drift (drift within and across experimental runs) due to speaking contributed little to motion-induced artefact.

In contrast, the most significant source of head motion artefact, as has been shown in previous studies (Birn et al., 1998, 1999; Huang et al., 2001), appears to come from the actual movements of the speech organs during volume acquisition. Moreover, the extent of movement during volume acquisition will have variable effects depending on the magnitude and nature of the motion and the slice thickness. It should be noted that artefacts are possible even during isometric contraction of jaw muscles. Recently, Tamura et al. (2002) provide evidence of motion-induced signal artefact during jaw clenching. Moreover, any motor task, not just speech or speech-related movements produced during volume acquisition, may contribute motion-induced artefact. For example, a recent study by Hoeller et al. (2002) reports motion-induced artefact associated with finger tapping and hand clenching. Presumably, these artefacts were associated with head movement as a consequence of the motor tasks. Any movement within the magnet has the potential to move the head during acquisition and introduce artefact to the acquired signal.

In summary, we suggest that future functional imaging studies of speech and other motor behaviors, including studies of clinical populations with movement disorders, should avoid volume acquisition during movement. Previously described solutions do not completely eliminate motion artefact and clearly limit the validity of the data. Assuming that the head is comfortably and securely restrained, the approach reported here is well suited for even more extensive, natural, and ecologically-relevant studies of spoken language production and interpersonal interactions.

Acknowledgments

We thank Mike Ferreira for his valuable assistance on all stages of the work reported here. The studies reported here were supported by grants from the National Institutes of Health (NIDCD DC 00594), Natural Sciences and Engineering Research Council of Canada, and the McGill University Research Development Fund.

References

Abrahams, S., Goldstein, L.H., Simmons, A., Brammer, M.J., Williams, S.C.R., Giampietro, V.P., Andrew, C.M., Leigh, P.N., 2003. Functional magnetic resonance imaging of verbal fluency and confrontation naming using compressed image acquisition to permit overt responses. Hum. Brain Mapp. 20, 29–40.

Ackermann, H., Wildgruber, D., Grodd, W., 1998. Does the cerebellum contribute to cognitive aspects of speech production? A functional MRI study in humans. Neurosci. Lett. 247, 187–190.

Barch, D.M., Sabb, F.W., Carter, C.S., Braver, T.S., Noll, D.C., Cohen, J.D., 1999. Overt verbal responding during fMRI scanning: empirical investigations of problems and potential solutions. NeuroImage 10, 642–657.

Birn, R.M., Bandettini, P.A., Cox, R.W., Jesmanowicz, A., Shaker, R., 1998. Magnetic field changes in the human brain due to swallowing or speaking. Magn. Reson. Med. 40, 55–60.

Birn, R.M., Bandettini, P.A., Cox, R.W., Shaker, R., 1999. Event-related fMRI of tasks involving brief motion. Hum. Brain Mapp. 7, 106–114.

Bookheimer, S.Y., Zeffiro, T.A., Blaxton, T.A., Gaillard, W., Theodore, W.H., 1995. Regional cerebral blood flow during object naming and word reading. Hum. Brain Mapp. 3, 93–106.

Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D inter-subject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.

Cox, R.W., Jesmanowicz, A., 1999. Real-time 3D image registration for functional MRI. Magn. Reson. Med. 42, 1014–1018.

de Zubicaray, G.I., Zelaya, F.O., Andrew, C., Williams, S.C.R., Bullmore, E.T., 2000. Cerebral regions associated with verbal response initiation, suppression and strategy use. Neuropsychologia 38, 1292–1304.

de Zubicaray, G.I., Wilsom, S.J., McMahon, K.L., Muthiah, S., 2001. The semantic interference effect in the picture-word paradigm: an event-related fMRI study employing overt responses. Hum. Brain Mapp. 14, 218–227.

Eden, G.E., Joseph, J.E., Brown, H.E., Brown, C.P., Zeffiro, T.A., 1999. Utilizing hemodynamic delay and dispersion to detect fMRI signal change without auditory interference: the behavior interleaved gradients technique. Magn. Reson. Med. 41, 13–20.

Edmister, W.B., Talavage, T.M., Ledden, P.J., Weisskoff, R.M., 1999. Improved auditory cortex imaging using clustered volume acquisition. Hum. Brain Mapp. 7, 89–97.

Fiez, J.A., 2001. Neuroimaging studies of speech an overview of techniques and methodological approaches. J. Commun. Dis. 34, 445–454.

Friston, K.J., Williams, S., Howard, R., Frackowiak, R.S., Turner, R., 1996. Movement-related effects in fMRI time-series. Magn. Reson. Med. 35, 346–355.

Hall, D.A., Haggard, M.P., Akeroyd, M.A., Palmer, A.R., Summer field, A.Q., Elliot, M.R., Gurney, E.M., Bowtell, R.W., 1999. Sparse temporal sampling in auditory fMRI. Hum. Brain Mapp. 7, 213–223.

Hashimoto, Y., Sakai, K.L., 2003. Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: an fMRI study. Hum. Brain Mapp. 20, 22–28.

Hoeller, M., Krings, T., Reinges, M.H.T., Hans, F.J., Gilsbach, J.M., Thron, A., 2002. Movement artefacts and MR BOLD signal increase during different paradigms for mapping the sensorimotor cortex. Acta Neurochir. 144, 279–284.

Huang, J., Carr, T.H., Cao, Y., 2001. Comparing cortical activations for silent and overt speech using event-related fMRI. Hum. Brain Mapp. 15, 39–53.

Jurgens, U., 2002. Neural pathways underlying vocal control. Review. Neurosci. Behav. Rev. 26, 235–258.

Kent, R.D., Kent, J.F., Weismer, G., Duffy, J.R., 2000. What dysarthrias can tell us about the neural control of speech. J. Phon. 28, 273–302.

Kent, R.D., Duffy, J.R., Slama, A., Kent, J.F., Clift, A., 2001. Clinicoanatomic studies on dysarthria: review, critique, and directions for research. J. Speech Lang. Hear. Res. 44, 535–551.

Price, C.J., Wise, R.J.S., Watson, J.D.G., Patterson, K., Howard, D., Frackowiak, R.S.J., 1994. Brain activity during reading: the effects of exposure duration and task. Brain 117, 1255–1269.

Ramsay, S.C., Adams, L., Murphy, K., Corfield, D.R., Grootoonk, S., Bailey, D.L., Frackowiak, R.S., Guz, A., 1993. Regional cerebral blood flow during volitional expiration in man: a comparison with volitional inspiration. J. Physiol. 461, 85–101.

Riecker, A., Ackermann, H., Wildgruber, D., Dogil, G., Grodd, W., 2000. Opposite hemispheric lateralization effects during speaking and singing at motor cortex, insula and cerebellum. NeuroReport 11, 1997–2000.

Riecker, A., Wildgruber, D., Dogil, G., Grodd, W., Ackermann, H., 2002. Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. NeuroImage 16, 169–176.

Tamura, T., Kanayama, T., Yoshida, S., Kawasaki, T., 2002. Analysis of brain activity during clenching by fMRI. J. Oral Rehabil. 29, 467–472.

Wildgruber, D., Ackermann, H., Klose, U., Kardatzki, B., Grodd, W., 1996. Functional lateralization of speech production at primary motor cortex: a fMRI study. NeuroReport 7, 2791–2795.

Wildgruber, D., Ackermann, H., Grodd, W., 2001. Differential contributions of motor cortex, basal ganglia, and cerebellum to speech motor control: effects of syllable repetition rate evaluated by fMRI. NeuroImage 13 (1), 101–109.

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

Worsley, K.J., Liao, C., Aston, J.A.D., Petre, V., Duncan, G.H., Morales, F., Evans, A.C., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1–15.