Neuroimaging studies of speech
An overview of techniques and methodological approaches

Julie A. Fiez*

Departments of Psychology and Neuroscience, and Center for the Neural Basis of Cognition,
605 LRDC, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received 4 May 2001; received in revised form 25 June 2001; accepted 25 June 2001

Abstract

Over the past two decades, there has been an explosion in the use of imaging technology to study the structure and function of the human brain. The purpose of this article is to explore how functional neuroimaging has been applied to the study of speech production. This article begins with a brief review of neuroimaging methods and limitations. Then, two approaches that can be used to study the brain areas that support speech production are illustrated. The first approach is based upon comparisons across different types of language production tasks; the second approach is based upon comparing the effects of different types of stimuli within a single task. Results obtained using these approaches will be used to dissociate the contributions of different brain regions involved in speech production. For example, evidence will be presented that Broca’s area contributes to phonological encoding, whereas motor cortex, the supplementary motor area (SMA), and the cerebellum support phonetic encoding and articulation.

Learning outcomes: As a result of this activity, the participant will be able to describe basic methods for conducting a positron emission tomography study and a functional magnetic resonance imaging study. The participant will also be introduced to two approaches for fractionating the brain regions involved in speech production: (1) comparisons between tasks, and (2) manipulations of stimulus materials. Finally, the participant will be able to summarize a cognitive model of the components of speech production that will be introduced, and potential mappings between these...
components and particular brain regions will be discussed. © 2001 Elsevier Science Inc. All rights reserved.

Keywords: PET; fMRI; Functional neuroimaging; Speech production; Phonology; Articulation

1. Introduction

The basic rationale for the use of functional neuroimaging studies is that the performance of any task places specific information processing demands on the brain. These demands are met through changes in neural activity in various functional areas of the brain (Posner, Petersen, Fox, & Raichle, 1988). Changes in neuronal activity produce changes in local blood flow (Raichle, 1989), which can be measured directly with PET, and indirectly with fMRI. The design and interpretation of many neuroimaging experiments is based on a particular view of the localization of function in the brain (Posner et al., 1988). The main idea of this framework is that elementary operations, defined on the basis of information processing analyses of task performance, are localized in different regions of the brain. Since many such elementary operations are involved in any cognitive task, a set of distributed functional areas must be orchestrated in the performance of even simple cognitive tasks. In the sections below, the methods used to acquire and analyze PET and fMRI data are reviewed, and then the use of neuroimaging for understanding the neural basis of speech production is illustrated through a discussion of literature review and analysis reported by Indefrey and Levelt (2000), and a study of word reading by Fiez, Balota, Raichle, and Petersen (1999).

2. PET basics

PET is a technology that creates pictures of the distribution of radiation within the central opening of a doughnut-shaped PET scanner. Radioactive substances, called tracers, are employed to “image” different physiological processes, such as brain blood flow or metabolism. Areas of higher blood flow will have a larger amount of radioactive tracer, and thus emit a stronger signal. By comparing images of the distribution of radioactivity, information relating to localization of certain functions can be obtained (Raichle, 1989).

Because 10–15 PET scans can be made in a scan session, different tasks can be performed in each subject. Because within-subjects designs are used, comparison images between different task conditions can be created, with the idea of imaging activity changes related to specific elementary operations of a task. Different techniques have been developed to allow the creation of inter- or intrasubject averaged images, or data sets that increase the signal-to-noise ratio.
(Fox, Mintun, Reiman, & Raichle, 1988). Because these images have been cast into a standard atlas space (Talairach & Tournoux, 1988), the results of experiments from different groups can be compared to one another in a relatively direct way.

Different types of statistical analyses have been applied to identify areas of significant change. The most common approach is to survey images without a priori assumptions about the locations of significant changes. This presents more difficult statistical problems, because of the large number of spatial locations that are not statistically independent from one another and on which there are small numbers of observations (Worsley, Evans, Marrett, & Neelin, 1992). Several theoretical corrections for such multiple comparisons have been developed; in practice, it means that a region of blood-flow change is generally not considered to be significant unless it meets a relatively conservative statistical threshold (e.g., $P < .001$).

3. fMRI basics

fMRI is a technique that is based upon the magnetic properties of the hydrogen atom (a component of water, the most abundant substance in the body) and hemoglobin (the blood’s oxygen carrier) (Howseman & Bowtell, 1999). The main signal is created by applying a magnetic force to artificially align the hydrogen atoms along a main pole, and then transient pulses of radio frequency energy are used to tip the magnetization of the atoms away from the main pole. The energy released as the atoms relax back to the main pole is measured to create images of the distribution of the magnetic signal, which will be related to the distribution of water in the brain tissue. The generation of images that are related to blood flow takes advantage of an additional property of the system: namely, that hemoglobin has different magnetic properties when it is not carrying oxygen (deoxyhemoglobin) than when it is carrying oxygen (hemoglobin). When blood flow to a particular brain area is increased, the number of oxygen-carrying hemoglobin molecules is also increased. In fact, the tissue use of oxygen does not keep pace with the increase in hemoglobin, so as a result areas that are functionally active experience a decrease in the percentage of deoxyhemoglobin. The magnetic properties of deoxyhemoglobin disrupt the strength of the local signal produced by the hydrogen atoms, and so as the amount of deoxyhemoglobin decreases, the local signal created by the hydrogen atoms in the nearby tissue increases. Altogether then, the method permits images of changes in blood flow to be measured indirectly through the effects of changing percentages of deoxyhemoglobin (Howseman & Bowtell, 1999).

Unlike with PET, images of the changes in blood flow are created in seconds, thus a single 2-h imaging session may generate more than 1000 images of the brain (as compared to 10 for PET). However, the signal-to-noise ratio in an individual image is very low, so it is necessary to average many images together...
to generate an image that shows reliable changes. When fMRI was first developed, the predominant paradigms for analyzing task-related changes in blood flow were conceptually similar to PET paradigms: in these “blocked” paradigms, subjects would alternate between performing an “active” and a “control” task for short time periods (e.g., 30 s), and then the images acquired during the active task blocks would be statistically compared to the images acquired during the control task block.

Further development of the technique revealed that task-related changes could also be observed by measuring the blood flow generated during an individual trial, and then averaging the data across trials to increase the signal-to-noise ratio (Rosen, Buckner, & Dale, 1998). This approach is analogous to that used in event-related potential studies of electrical activity. One advantage of this approach is that trials of different types (e.g., high-frequency words versus low-frequency words) can be randomly intermixed, as is done in most behavioral studies. The initial use of so-called event-related designs relied upon the use of long interstimulus intervals (total trial times of 10–12 s), because the blood-flow changes (hemodynamic responses) that accompany neuronal changes are sluggish. They typically begin 2–4 s after the trial onset, peak at about 6–8 s, and decay back to baseline levels by 10–12 s. Data acquired using this approach are often analyzed using ANOVA techniques that assess changes in signal across time, and across different experimental conditions.

Subsequent work demonstrated that hemodynamic responses in many cases add in a linear fashion, and this has made it possible to use rapid event-related designs, in which subjects perform a trial every 1–4 s (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000). The resulting data is a complex waveform, and the analysis is often dependent upon regression techniques in which expected trial events are convolved with the general shape of the hemodynamic response to determine a predictor function that accounts for a significant portion of the variance in the complex waveform.

As fMRI techniques have emerged, it has became the preferred imaging method of most investigators, for several reasons. These include the fact that the hardware is widely available, it is less expensive, it is noninvasive, it imposes minimal risks for the subject, and it has higher temporal resolution (and potentially higher spatial resolution as well, though in most cases the raw data are blurred to help compensate for anatomical differences across subjects). Unfortunately, for studies of speech, hearing, and language, fMRI has some notable disadvantages. One problem is that the scanner is noisy, which can make studies that hinge upon subtle differences in auditorily presented stimuli difficult to implement. A second problem is that speaking leads to movements of the sinus air cavities, and this can lead to disruptions in the magnetic signal, causing speech-related artifacts in the data. A third problem is that the signal generated in response to a single stimulus item is potentially smaller than it is in PET. This may be an important constraint for studies in which only a limited number of well-controlled stimulus items are available, as is often the
case in language studies. Fortunately, recent methodological advances have made all of these problems tractable, and so this issue will be less of a concern in the future (Bandettini, Jesmanowicz, Van Kylen, Birm, & Hyde, 1988; Barch et al., 1999).

4. Understanding speech through task comparisons

For whatever reason, there are a limited number of PET and fMRI studies that have concentrated on speech production per se, though the article by Ingram (this volume) reviews a body of research that is a notable exception. Thus, much of what we know about the neural basis of speech production comes from language and cognitive tasks that have incorporated overt or covert (silent) speech. Although the focus of these tasks is usually upon core linguistic processes, they provide a rich database that can be used to generate insights about speech production. This point will be illustrated by considering two basic approaches to studying the regions involved in speech and language. One approach is based upon comparisons between different types of tasks, and it is one of the most common approaches used in neuroimaging. The second approach relies upon keeping the task the same while varying the types of items that subjects receive. This approach has strong ties to methods used in cognitive psychological behavioral research and theory.

The idea of isolating brain regions involved in speech production through task comparisons is well illustrated by a meta-review of 58 studies of language production conducted by Indefrey and Levelt (2000). This review of the literature was guided by theoretical models of language production that are supported by a large body of behavioral research. The authors propose a number of components that help support language production, as suggested by a number of theoretical models. Some of these components can be considered lead-in processes that are outside the core domain of language production. For instance, for a picture-naming task this would include processes that help support visual object recognition. After these lead-in processes are completed, the task may require a number of core linguistic processes. These include conceptual preparation of a message to be communicated, access to and selection of the lexical items that will communicate the concept, and encoding of the phonological forms of the lexical items, to generate a surface form of the entire message. Next, it is necessary to transform the message into an articulatory output. This is accomplished by a phonetic encoding process that generates a gestural score, which can be used to guide the final process of articulation.

Using this broad theoretical model, the authors (Indefrey & Levelt, 2000) next conducted a task analysis in which they reviewed the different components of language production that would be required for the tasks employed in prior neuroimaging studies. For example, they argued that overt production of a verb in response to a presented noun would require all of the components outlined...
above, while reading aloud a pronounceable nonword would not entail conceptual preparation, lexical selection, or phonological code retrieval. Similarly, production of an overt response, regardless of the specific task, should require phonetic encoding and articulation, whereas covert (silent) production of a response should not.

Based upon the task analysis, which linked different types of tasks to different components of speech production, the authors (Indefrey & Levelt, 2000) were then in a position to analyze whether particular brain regions showed patterns of activity that were consistent with the theoretical model and associated task analysis. They argue that the data supports a number of potential mappings. For instance, they link portions of the middle temporal gyrus to conceptual preparation and lexical selection, based in part upon the fact these processes are shared by picture naming and word generation, but not necessarily word reading. The posterior superior temporal lobe may help support phonological code retrieval, because it is active during tasks that require this process (naming, word generation, and reading), but is not active during pseudoword reading. All of the tasks they examined involved phonological encoding, and thus brain areas that support this process should be active across all tasks. Broca’s area came close to showing this pattern of activity. Finally, areas related to the production of the abstract articulatory program and its execution should be isolated by identifying areas that are active only during overt pronunciation conditions that are contrasted to silent control condition. Several areas involved in motor control and coordination met this criteria, including motor cortex, supplementary motor area (SMA), and the medial cerebellum.

5. Understanding speech through stimulus manipulations

The study by Indefrey and Levelt (2000) demonstrates how broad distinctions can be drawn between the large number of brain regions that contribute to even simple language production tasks. In this section, a second approach that is based upon manipulations of stimulus type is reviewed. One benefit of this approach is that it may offer a more fine-grained view into the types of representations and processes that are supported by different regions.

Behavioral studies of word reading have revealed many different types of manipulations that affect the speed and accuracy of reading. Three of the most commonly studied factors are frequency (how often a word is encountered), consistency (whether the spelling-to-sound correspondence of the word is consistent with other visually similar words), and lexicality (whether an item is a real word, or a pronounceable nonword) (e.g., see Forster & Chambers 1973; Monsell, Patterson, Graham, Hughes, & Milroy 1992). Subjects can read aloud all kinds of high-frequency words relatively quickly (e.g., the consistent word “gave,” and the inconsistent word “from”), and they are nearly as fast to read
aloud low-frequency consistent words (e.g., “hint”). However, subjects typically read aloud low-frequency inconsistent words (e.g., “pint”) much more slowly (Andrews, 1982; Monsell et al., 1992). Several theoretical models have been proposed to account for this effect (Coltheart, Curtis, Atkins, & Haller, 1993; Plaut, McClelland, Seidenberg, & Patterson, 1996), but in all of the models the interaction effect is considered to be a hallmark of the processes involved in the transformation between how a word looks (orthography) and how it sounds (phonology). Subjects are also slow to read aloud nonwords, and again this is typically considered to be a hallmark of the processes involved in orthographic to phonological transformation (e.g., in dual-route models of reading it is thought to reflect the slower operation of a sublexical procedure that converts individual letters and letter clusters to their corresponding sounds).

Neuroimaging studies have manipulated frequency, consistency, and lexicality as a means of identifying key areas that are involved in orthographic-to-phonological transformation. For instance, in a study by Fiez et al. (1999), subjects read aloud high-frequency consistent words, high-frequency inconsistent words, low-frequency consistent words, low-frequency inconsistent words, and pronounceable nonwords. One of the findings from this study was that a region near the junction of the anterior insula and the inferior frontal gyrus (Broca’s area) showed a pattern of activation that paralleled the behavioral data: significantly more activation was observed in the low-frequency inconsistent word and nonwords conditions, than in the low-frequency consistent, and high-frequency consistent and inconsistent conditions. The authors concluded that Broca’s area plays a critical role in orthographic to phonological transformation, potentially by supporting effortful sublexical phonological analysis.

An unexpected result from the study by Fiez et al. (1999) was that motor-related areas also showed effects of frequency and consistency. Specifically, SMA showed a weak effect of frequency, and primary motor cortex showed a significant effect of consistency. These findings were surprising, because theoretical models of reading typically place the origin of these effects in the cognitive processing stages that precede phonetic encoding and articulation. A number of interpretations of the results are discussed, but one of the more interesting possibilities is that these effects provide insights into the types of representations used by motor-related areas for speech production. For instance, the frequency effects in SMA may indicate that this region is involved in the generation of motor program for an entire spoken word, and the level of activity observed in SMA is influenced by how often this motor program is implemented. In motor cortex, the effects of consistency may reflect a covarying factor that affects motoric aspects of response initiation or articulation. For instance, if motor cortex represents articulatory gestures at the syllable or phoneme-sequence level (Levelt & Wheeldon, 1994), then the effects of consistency may actually represent an effect of frequency at a sublexical level. Since both of these interpretations posit that the influences of frequency and consistency are related to phonetic encoding and articulation, and not orthographic-to-phonological transformation, similar
patterns of activation would be expected if the items were presented auditorily and repeated aloud.

6. Summary and conclusions

Functional neuroimaging studies are based upon the idea that changes in task demands produce localized changes in neuronal activity that in turn produce localized changes in blood flow. These changes in blood flow can be imaged directly with PET, and indirectly with fMRI. One limitation to understanding the brain regions that contribute to speech production is that neuroimaging studies have tended to concentrate on studies of language and other cognitive processes, rather than speech production itself. However, since many of the tasks used in these studies involve overt or silent speech production, it is possible to use this literature to learn something about the neural basis of speech production. Two illustrations of this point were provided. The first exemplified the use of cross-task comparisons to fractionate the brain regions involved in language production. Based upon a meta-review of the literature, it was concluded that middle temporal regions are primarily involved in conceptual analysis and lexical retrieval, the superior temporal gyrus supports phonological retrieval, Broca’s area contributes to phonological encoding, and motor cortex, SMA, and the cerebellum play critical roles in phonetic encoding and articulation. The use of stimulus manipulations to examine the function of language production areas was then exemplified through a study of word reading. Results from this study provided further evidence that Broca’s area contributes to phonological encoding, possibly at the sublexical level, and suggested that motor cortex and SMA support phonetic encoding and articulation at different levels of representation (syllable versus whole word).

Appendix A. Continuing education

1. Functional neuroimaging is based upon the idea that:
   a. Changes in task demands lead to changes in localized neuronal activity.
   b. Changes in neuronal activity lead to changes in blood flow.
   c. Changes in blood flow can be imaged directly using PET.
   d. Changes in blood flow can be imaged indirectly using PET.
   e. All of the above.

2. An advantage of fMRI is that it:
   a. Is quieter than PET.
   b. Is less susceptible to movement artifacts.
   c. Has potentially greater spatial and temporal resolution.
   d. More directly measures changes in electrical activity.
   e. All of the above.
3. Cross-task comparisons of speech production suggest that:
   a. Lexical selection is accomplished in Broca’s area.
   b. Lead-in processes occur in the superior temporal gyrus.
   c. Conceptual processes are supported by anterior prefrontal cortex.
   d. Phonological retrieval involves the supplementary motor area.
   e. Phonetic encoding involves the cerebellum.
   f. None of the above.

4. A low-frequency inconsistent word:
   a. Is a word that is not experienced very often.
   b. Is a word with an unusual spelling-to-sound correspondence.
   c. Takes longer to read than a low-frequency consistent word.
   d. Produces more activation in Broca’s area than a low-frequency consistent word.
   e. All of the above.

5. Stimulus-related differences in functional brain activation:
   a. May extend to motor-related areas involved in speech production.
   b. Only provide very crude fractionations of the brain regions involved in speech production.
   c. Are impossible to study using current neuroimaging techniques.
   d. Often yield results that contradict those observed using across-task comparisons.
   e. None of the above.

References


