ELSEVIER

Acta Psychologica 107 (2001) 95–117

acta psychologica

www.elsevier.com/locate/actpsy

Functional imaging during speech production

K.G. Munhall *

Departments of Psychology and Otolaryngology, Queen's University, Kingston, Ont., Canada

Received 20 September 2000; received in revised form 14 December 2000; accepted 18 December 2000

Abstract

Physiological studies of speech production have demonstrated that even simple articulation involves a range of specialized motor and cognitive processes and the neural mechanisms responsible for speech reflect this complexity. Recently, a number of functional imaging techniques have contributed to our knowledge of the neuroanatomical and neurophysiological correlates of speech production. These new imaging approaches have the advantage of permitting study of large numbers of normal and disordered subjects but they bring with them a host of new methodological concerns. One of the challenges for understanding language production is the recording of articulation itself. The problems associated with measuring the vocal tract and measuring the neural activity during overt speech are reviewed. It is argued that advances in understanding fundamental questions such as what are the planning units of speech, what is the role of feedback during speech and what is the influence of learning, await the development of better methods for assessing task performance. © 2001 Elsevier Science B.V. All rights reserved.

PsycINFO classification: 2330; 2720

Keywords: Linguistics; Language; Speech; Motor processes

---

* Tel.: +1-613-533-6012; fax: +1-613-533-2499.
E-mail address: munhallk@psyc.queensu.ca (K.G. Munhall).

PII: S 0 0 0 1 - 6 9 1 8 (0 1) 0 0 0 2 6 - 9

1. Introduction

The production of spoken words is the end product of a complex network of linguistic and cognitive processes. Thoughts and intentions are remarkably transformed into a sequence of movements and sounds in the order of hundreds of milliseconds. One of the great achievements of psycholinguistic research in the past 30 years has been the preliminary system identification of this language output system. Distinct phases of planning and control in this process have been identified through the study of errors (e.g., Fromkin, 1973) and studies of the timing of production (Levelt, 1989) and formal models of language production have been proposed to account for these data.

Levelt's model is representative of these frameworks and it is summarized in Fig. 1. As can be seen in this figure, the progression from concept to articulation is thought to involve separate processes that impart syntactic, morphological, phonological and phonetic organization. At the end of this flow of information processing lies the final and often forgotten stage of the language sequence, articulation itself. Most psycholinguistic models of this kind give minimal attention to the final stage of actual speech production. Like a last minute addition to a guest list, articulation does not feature in early phases of planning. In fact, in many models of language production, articulation is portrayed as a passive and modular final filter for language production. For a variety of technical reasons, this de-emphasis of articulation has been mirrored in functional imaging studies of language production. Many neural imaging studies have used a range of "silent" tasks in which words are not even spoken. [1] This is unfortunate because it de-emphasizes the medium in which the evolution of language presumably took place and it makes an unwarranted "pure insertion" assumption about speech motor control (Jennings, McIntosh, Kapur, Tulving, & Houle, 1997).

There are many reasons for giving the articulation component of language production more attention in functional imaging studies. Foremost amongst these is the fact that the speech motor system is not a passive channel for the transmission of linguistic signals from prior planning stages. Rather, it transforms those signals in a number of ways. At the most peripheral level, the nonlinear mechanics of tissue and muscle, the inertial forces of the moving articulators and the complexities of force generation in muscles each contribute to the form of articulation and thus to the final form of the speech output. In order to accommodate this degree of nonlinearity, the nervous system is thought to have internal models of the vocal tract and the acoustic consequences of articulation (Guenther, 1994; Hirayama, Vatikiotis-Bateson, & Kawato, 1994; Jones & Munhall, 2000; Jordan, 1990, 1996; Kawato, Furukawa, & Suzuki, 1987). An internal model is a representation of the forces and kinematics of movements as well as the feedback that results from those movements (see Miall &

---

[1] To be fair, some of the studies being referred to were not intended to study articulation per se. Their focus was on other aspects of language production. However, the criticisms related to the use of silent tasks are still applicable.

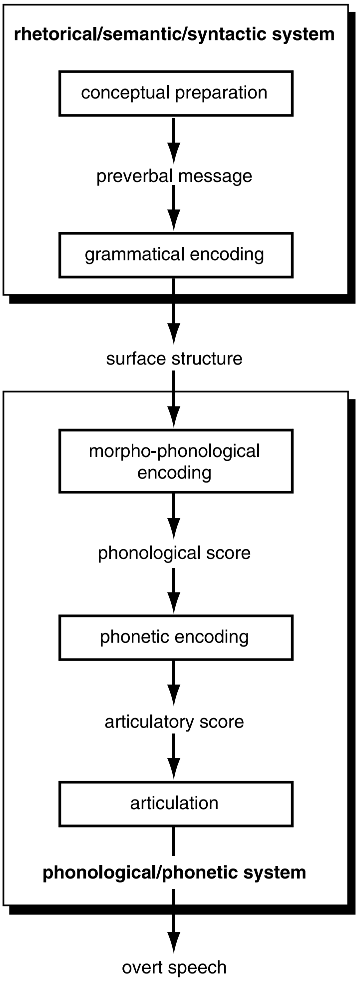

Fig. 1. Framework of processing components in language production (adapted from Indefrey & Levelt, 2000).

Wolpert, 1996; Kawato, 1999, for discussions of internal models in movement control). In this depiction of the act of speaking, articulation involves significant information processing in order to manage the motor system in the vocal tract during phonetic sequencing. Timing, force, and trajectory control of a large number of articulators have to be programmed in rapid succession. This computation engages a neurally distributed planning and control system.

The amount of information that must be conveyed and therefore programmed by the motor system is remarkable. Talkers modify their speech style and precision in real time to fit social and environmental context and to respond to conversational demands. Mood, intent, attention, as well as the conceptual and emotional meaning of a message are all transmitted in parallel with phonetic information by subtle differences in the way in which words are spoken. These subtle differences are produced by systematic modifications to the movement timing and trajectories of the oral articulators. For example, talkers can voluntarily modify the clarity with which they speak (Gagne, Masterson, Munhall, Bilida, & Querengesser, 1995) and speech is perceptually distinct if it is read, recited, voluntarily produced, attentionally engaged, etc. (e.g., Remez, Lipton, & Fellowes, 1996). In addition, there is variability in articulation from a number of other sources. Speaking rate, lexical and emphatic stress and syllabification can vary from repetition to repetition.

How this motor planning is accomplished is not well understood but an interaction between levels of planning is indicated. For example, Saltzman, Lofqvist, and Mitra (2000) have used a phase resetting paradigm to study the relationship between central "clock" mechanisms and the timing of peripheral motor events in speech. When randomly timed mechanical perturbations are applied to the lower lip during repetitive syllable production, the patterns of timing adjustments are consistent with a model in which central timing is modulated by the peripheral motor system (Gracco & Abbs, 1989). Kawato (1997) has suggested that this arrangement is preferred for computational reasons as well. During trajectory planning, the solution space can be reduced more effectively if constraints (e.g., smoothness) are specified at other planning levels (Kawato, 1997). To do this however, the trajectory planning (along with other information processing problems in motor control such as coordinate transformations and motor command generation) must be computed simultaneously rather than sequentially.

Understanding articulation, then, is not simply a matter of understanding how the jaw, for example, moves for a canonical phoneme target. Rather, such an understanding must include a description of coordination in the context of the range of performance conditions occurring during communication. Trajectory formation for the articulators would thus involve interaction with many levels of the language system as well as involving complex motor planning in order to achieve these goals in the presence of the biomechanical and physiological complexity of the vocal tract (Munhall, Kawato, & Vatikiotis-Bateson, 2000).

The nonadditivity of cognitive processing and type of response has recently been demonstrated in a semantic judgment task (Jennings et al., 1997). PET activation patterns [2] varied for the same semantic task depending on the response mode (mouse click, overt speech, silent thought). Jennings et al. suggest that the way in which subjects organized their responses induced activation in different areas of the brain for the semantic processing. This kind of interaction is consistent with a system with many reciprocal connections and feedback projections such as the language system.

Factors such as these argue strongly for careful study of speech production in functional imaging studies. From a practical point of view, the manner in which

---

[2] In this paper I will use the term "activation" to refer to the measures from different imaging techniques. I acknowledge that this is technically wrong (and may even be more than just technically wrong) but it is convenient for expository purposes in this manuscript.

words are spoken defines part of the "task" in any language production study and the presence of real articulation permits more precise task monitoring in imaging studies. Response timing, accuracy and task compliance can only be measured when there actually is a response! There is good evidence that this should be a concern in language production studies. Untrained talkers are not very precise at controlling many speech variables in laboratory studies of language output. For example, speaking rate, in spite of its popularity as a manipulation in speech production studies, is notoriously variable (Miller & Baer, 1983). Rate of articulation has also been reported to influence regional blood flow in PET experiments (e.g., Paus, Perry, Zatorre, Worsley, & Evans, 1996).

Indefrey and Levelt (2000) have recently argued that a range of poorly understood "lead-in" processes used in neuroimaging studies of word production (e.g., picture naming, generating words from beginning letter, word reading, repetition) have to be more carefully considered since they influence neuroimaging results. In a similar vein, details of articulation define a range of "lead-out" processes that will influence imaging data. Variability in speech obviously can result from different neural control strategies at many levels of the production system and this will be reflected in different neural activation patterns.

In addition to these methodological reasons for studying the functional representation of overt speech there are more theoretical reasons as well. Speech can be viewed as a complex *linguistic* process (Liberman & Whalen, 2000), and thus the neural representation of speech gestures is seen as an inextricable part of the language system. In Liberman's view, the phonetic gestures are the "common currency" of two-way communication and thus are the primitives of human language.

More pragmatically, there is a set of classic problems in the understanding of articulation that can profitably be addressed using functional imaging. These include (a) specifying the fundamental units of speech coordination, understanding the implementation of these units in a range of contexts that influence precision and timing, (b) understanding how central representation of units interact with feedback in real time and (c) understanding how these units are acquired and modified by learning. For each of these problems imaging could aid in both the neural mapping of the behavior and potentially the specification of the processing components of the behavior (Kosslyn, 1999). In the final section of the paper I will return to these problems.

To date, functional imaging has played two distinct roles in the study of speaking. First, a form of functional structural imaging has long been important for the precise description of the actual behavior of talking. Secondly, functional neural imaging has recently contributed to our understanding of the cognitive and linguistic processes responsible for articulation and for mapping these putative processes onto the neural architecture. In this paper I will give an overview of these two roles played by imaging, particularly magnetic resonance imaging (MRI), and comment on how functional imaging can help resolve enduring problems in the field. I will begin by commenting on the state of functional structural and functional neural imaging using MR.

2. Structural imaging of speaking

Sound production in human speech involves the generation of sound (e.g., by the vocal folds) and its filtering by the acoustic properties of the vocal tract (i.e., shape, size, wall characteristics, etc.). This division of explanation into source and filter has been the working model in speech research for the last 50 years (Fant, 1950) but speech bioacoustics and speech motor control are still far from completely understood. In part, this is due to the relative inaccessibility of the speech articulators. The external face and jaw contribute to the sound character in speech (Lindblom & Sundberg, 1971) and can be easily measured. However, the lips and jaw are not the primary determinants of the overall speech spectrum. A number of other "internal" articulators jointly determine the shape of the oral airway. Measuring the contribution of the tongue and pharyngeal walls on the acoustics of speech, however, requires invasive recording techniques.

This precise measurement of the vocal tract is necessary in order to understand the coordination demands on the speech motor control system and in order to specify the primitives of articulation. Early static X-ray images revealed the complex shape of the vocal tract during speech and significantly advanced our understanding of speech acoustics. However, the advent of high-speed X-ray films revolutionized the field revealing the remarkable nature of the skill of speech motor control for the first time. (See Moll, 1960, for a review of the early research. See Munhall, Vatikiotis-Bateson, & Tohkura, 1994, 1995, for an archival database of high-quality films.) This first functional imaging of speech revealed a rapidly changing "acoustic tube" in which the position and shape of a number of articulators were coordinated to produce a stream of consonants and vowels. For example, the X-ray films revealed that the tongue tip in speech can produce distinct movements within approximately 50 ms of each other (Kent & Moll, 1975). In recent years, the use of this X-ray cineradiography has drastically decreased, primarily due to ethical concerns about the radiation dosage.

Other imaging techniques have been developed including the use of ultrasound (e.g., Stone, Morrish, Sonies, & Shawker, 1987) and electromagnetometry (e.g., Hoole, Nguyen-Trong, & Hardcastle, 1993) but the development of MRI and computed tomography (CT) gave the community opportunity for the first time to record 3D images of the complete vocal tract shape (e.g., Baer, Gore, Gracco, & Nye, 1991; Dang, Honda, & Suzuki, 1994; Moore, 1992; Story, Titze, & Hoffman, 1996; Tiede, 1996). These new technologies have a number of advantages for speech research. They afford remarkable reconstructions of the complete vocal tract airway which have been valuable for tests of aeroacoustic models and for synthesis research. Fig. 2 shows two views of the vocal tract during production of the vowel /a/ from Story, Titze, and Hoffman (1998). As you can see in the image, a complex airway shape produces the pattern of resonances for a vowel.

In recent years, the increasing speed of MR image acquisition has led to the use of MRI as a direct substitute for X-ray imaging (e.g., Narayanan, Alwan, & Haker, 1995; Rokkaku, Hashimoto, Imaizumi, Niimi, & Kiritani, 1986; Yang & Kasuya, 1994). Fig. 3 shows midsagittal images of the three vowels /i, u, a/ collected using a

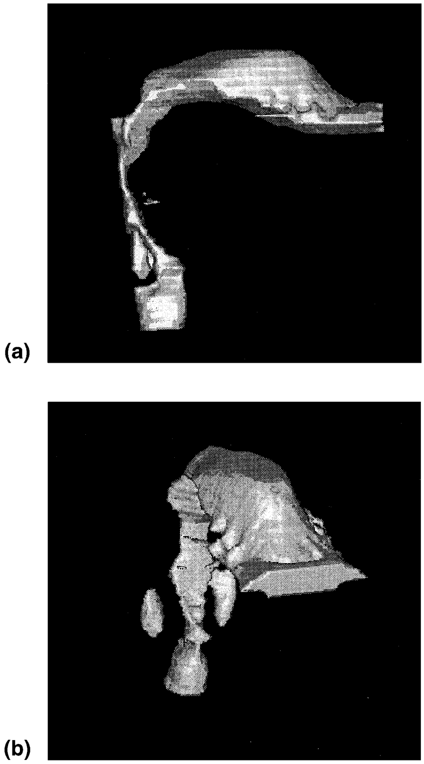

Fig. 2. Volumetric reconstruction of the vocal tract airway of an adult female producing the vowel /a/.

turbo flash technique with data acquisition time in the order of 1.5 s per image. The false color overlay of the three vowels in Fig. 3(d) highlights the different vocal tract configurations required for vowel production.

The most promising development is the possibility of using MRI for dynamic articulation measurement. Stroboscopic MRI has been used to measure the dynamic changes in the vocal tract by building a composite utterance from images sampled from a series of utterances. The subjects repeat a target syllable repetitively and images are recorded from different phases of the production (Mathiak et al., 2000; Shadle, Mohammad, Carter, & Jackson, 1999).

The goal of functional structural imaging is to characterize the shape of the vocal tract for different sound categories and to be able to describe changes in shape over time. The field's dream is to achieve high resolution, 3D imaging at rates that are acceptable for measuring articulation. While this aim no longer seems impossible, it is still far from being achieved.

Recently, faster image collection has permitted the recording of dynamic sequences within a single utterance. For example, Demolin, Metens, and Soquet (2000), using an ultra fast Turbo Spin Echo (TSE) Zoom sequence, recorded real time midsagittal images during the production of the vowel sequence /ieaou/ at a rate

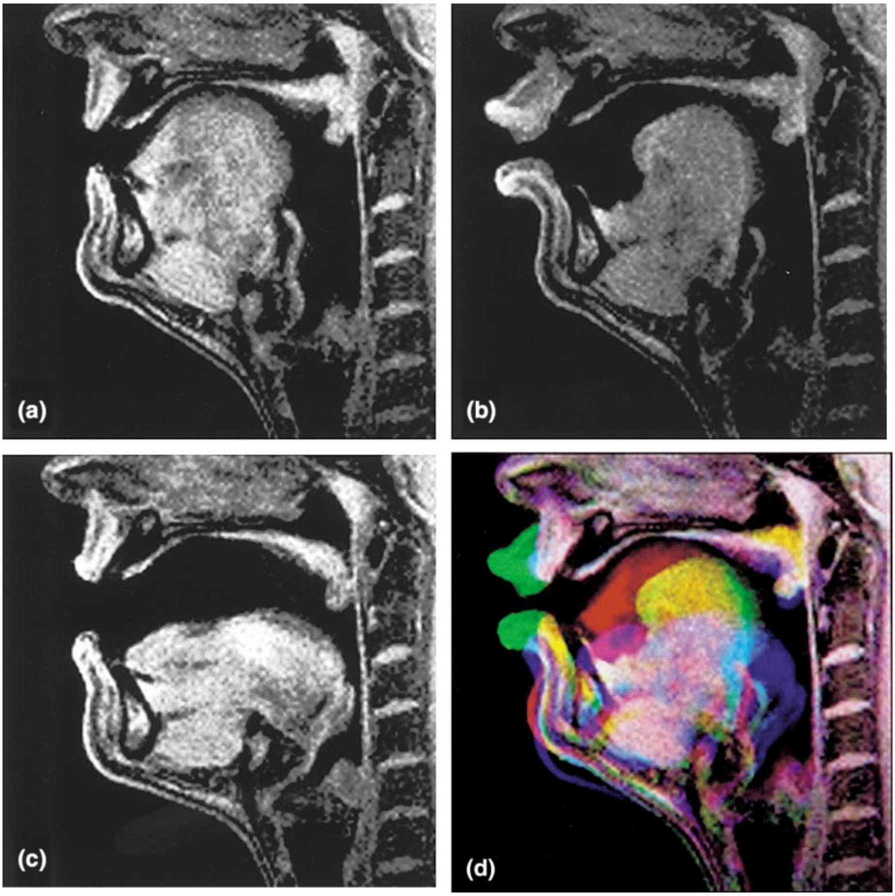

Fig. 3. Mid-saggital images of a talker producing the vowels (a) /i/ (b) /u/ and (c) /a/. Panel (d) shows the three vowels superimposed in false colours.

of 4–5 images per second. In a related development, Stone, Dick, Douglas, Davis, and Ozturk (2000) recorded tagged Cine-MRI to examine the changes in the internal biomechanics of the tongue during speech. A magnetic grid with a decay time of approximately 500 ms was imposed on the tongue tissue and its deformation was recorded over time. In a one second interval, 24 time-phases or image frames were recorded. While Stone et al. required ensemble summing of multiple repetitions to increase signal strength at each time-phase, this is still remarkable sampling speed.

In spite of these successes, MRI presents a number of problems for the study of speech. The recording environment is very noisy leading to changes in the acoustic feedback for subjects (see below) and making acoustic analysis of productions difficult. Data acquisition is still too slow and full volumetric reconstructions require sustained, repetitive productions. Thus, only static postures that can be maintained for a number of seconds can be studied in high resolution volumetric recordings.

Some boundaries of the vocal tract (e.g., the teeth) do not image well due to their protein density and there are even concerns about the changes in shape of the vocal tract when subjects produce speech in a supine position (Shiller, Ostry, & Gribble, 1999; Tiede, Masaki, & Vatikiotis-Bateson, 2000). Finally, because of constraints on eligible utterances, only monosyllabic stimuli produced in a laboratory citation style have been measured. As a result, the older, high-speech X-ray films still remain the best overall view of the moving vocal tract during fluent speech (Munhall et al., 1994, 1995).

3. Functional neural imaging of speaking

It has been known for over a century that multiple regions of the brain are involved in producing spoken language. Our knowledge of these neuroanatomical correlates of speech production primarily comes from deficit/lesion studies (see Caplan, 1992) and electrical stimulation studies (e.g., Ojemann, 1991; Penfield & Roberts, 1959) but recently, a number of imaging techniques have also been used. (See Savoy, 2001, for an extensive historical review.)

The key components of the language production system, first identified in the nineteenth century, continue to shape neural accounts of language. The classical view involves an area in the posterior portion of the inferior frontal gyrus (Broca's area), an area in the posterior superior temporal gyrus (Wernicke's area) and connections between these areas (e.g, the arcuate fasciculus). In the past 30 years, this view has been refined in many ways. First, the areas involved in even simple language behavior such as naming have been shown to be much more extensive than originally proposed. Fig. 4 shows cortical stimulation sites (Ojemann, 1991) that produced disruptions in naming. As can be seen broad regions of the temporal, parietal and frontal cortex can be involved in language production. This figure also shows the individual variability that has come to define the study of the neural representation of language. The numbers in circles represent the percentage of subjects who showed disruption when stimulated at each site. As can been seen, the percentages showing stimulation effects vary considerably with no single location being essential for all subjects. Twenty-one percent of the subjects showed no disruption even in the region that corresponds with Broca's area (see also Dronkers, 1996).

This observation is consistent with recent changes to the view of the language areas in the cortex. The definition of classical areas such as Broca's area and to a lesser extent Wernicke's area have undergone significant revision in recent years. Both the boundaries of Broca's area and its role in language processing have been questioned. Broca's area is now thought to involve a much larger area than the classical view (Mohr, 1976). To some extent, the exact area designated "Broca's area" is not that critical and as Damasio and Damasio (1988) suggest its precise location may be a matter of taste. However, there is general acceptance of a larger anterior language area. The functionality of this region has also varied in recent decades with a greater emphasis on the role of Broca's area in syntax and reduced role for Broca's area for actual articulation (e.g., Dronkers, Redfern, & Knight,

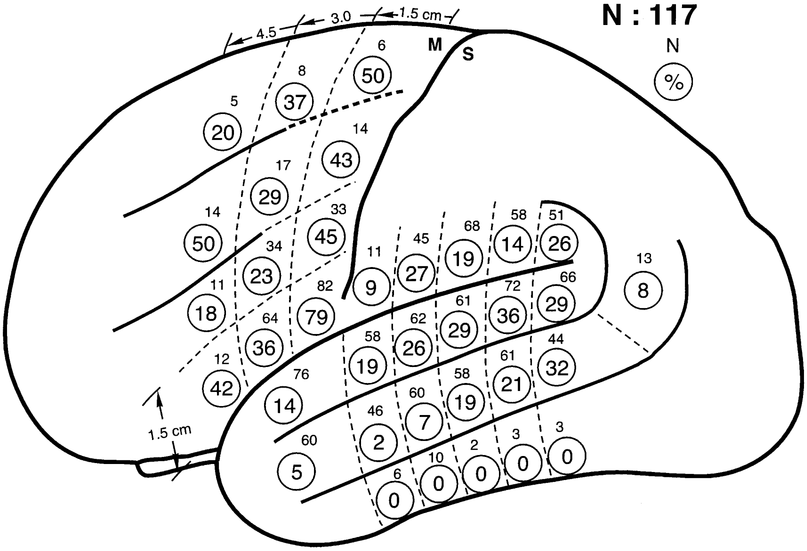

Fig. 4. Variability in localisation of sites deemed essential for naming. The circled numbers indicate the percentage of subjects who produced naming errors when electrically stimulated at that site. The upper numbers indicate the number of subjects stimulated (from Ojemann, 1991).

2000). In addition, the anterior inferior frontal lobe may be more involved in semantic decisions while the posterior inferior frontal lobe is implicated more in phonological processing (Fiez, 1997). Finally, the role played by subcortical structures (e.g., thalamus, cerebellum, basal ganglia) in articulation and other language processes has been revised (Damasio, Damasio, Rizzo, Varney, & Gersh, 1982).

Following Petersen, Fox, Posner, Mintum, and Raichle's (1988) PET study of single word processing, a range of functional imaging studies of language output have contributed to our growing neurological model of language production. These studies in combination with recent studies of lesions have led to the view that language production involves an extended cortical and subcortical network (e.g., Ackermann & Hertrich, 2000; Fiez & Raichle, 1997; Grabowski & Damasio, 2000; Nadeau & Crosson, 1997). Recent studies of oral movement and overt speech indicate that articulation similarly involves a distributed network (Corfield et al., 1999). For example, Rosen, Ojemann, Ollinger, and Petersen (2000) compared silent articulation to speaking aloud in a stem completion task. Expected differences in the activity in the primary motor cortex and the superior temporal gyrus were found. In addition, increased activity was seen in the thalamus and putamen during overt speech.

One of the recurring conclusions from the existing body of functional imaging work on language production is that the patterns of neural activation are extremely sensitive to the task the subject is asked to perform. For example, in Indefrey and Levelt (2000) meta-analysis of 58 word production studies involving functional imaging, a variety of language output tasks were used ranging from picture naming aloud, to silent word reading, to repetition aloud. [3] By Indefrey and Levelt's reckoning, all of these protocols involved a "phonological encoding" component. However, no single area was universally active across studies. Broca's area and the left mid superior temporal gyrus were most common. At least in part, this pattern can be attributed to the range of demands placed on the subjects by the different tasks used in the experiment. For example, the silent tasks in their meta-analysis ranged from "viewing" to "thinking" to silent "mouthing". A similar pattern was found in Poeppel's (1996) review of five PET studies of phonological processing. The "sparse overlap" that Poeppel criticized may be accounted for by subtle differences in task demands between experiments (e.g., segmentation of stimuli into syllables vs. segmentation into onset and rime of the syllables) (Demonet, Fiez, Paulesu, Petersen, & Zatorre, 1996).

In many of the studies on language output, particularly fMRI studies, the language tasks have not involved overt speech but rather various silent tasks such as imagined speech have been used. This is understandable given the problems associated with overt responses (see below). However, there may be costs associated with adopting this strategy. Direct comparisons between overt and covert speech have revealed predictable differences associated with the motor control of the speech articulators but also differences in activation that may relate to task or linguistic/cognitive control (Barch et al., 1999; McCarthy, Rothman, Gruetter, & Shulman, 1993; Price, Moore, & Frackowiak, 1996a; Zelkowicz, Herbster, Nebes, Mintun, & Becker, 1998). Further, Barch et al.'s examination of the response pattern in their study revealed that subjects did not always follow the task instructions. For anyone who carries out behavioral research, this observation should be a very familiar one!

Measuring neural activity with fMRI during overt speech, however, shares many of the problems of functional structural measurements with MRI. The head moves during articulation to signal prosodic distinctions such as intonation, stress, and accent. In studies of the relationship of head motion to pitch of the voice, we have found that the two variables are highly correlated (e.g., Kuratate, Munhall, Rubin, Vatikiotis-Bateson, & Yehia, 1999). As talkers produced sentences, the head motion covaried with the intonation contour of the sentence. The head may also move slightly for mechanical reasons related to jaw rotation. Inhibiting head motion completely without extraordinary and uncomfortable measures to secure the head, is difficult. When the head does move, even slightly, activations from a given cortical location can be spread across different voxels at different points in time. This effectively decreases the activation signal-to-noise ratio and distorts anatomical localization.

In addition, the acoustic noise in the magnet significantly alters the talking environment and this causes changes in the talker's speech and presumably in their

---

[3] It should be noted that many of the studies reviewed in Indefrey and Levelt used only partial brain acquisition. Thus, activations in a number of important language areas could not be imaged in these studies.

articulation strategy. Talkers rely on hearing their own speech and the planning system calibrates speaking volume, clarity and precision based on the auditory input during talking (Jones & Munhall, 2000).

The potential for magnetic susceptibility artifact also increases during overt speech (Birn, Bandettini, Cox, Jesmanowicz, & Shaker, 1998) but susceptibility problems are a concern for measuring neural speech areas even without overt articulation. The sinuses, the oral cavity and the pharynx border the cortex and brain stem and are the source of variations in the homogeneity of the magnetic field. Because of the large changes in magnetic susceptibility between tissue and these air-filled cavities, there are large field gradients near these regions. These gradients can reduce sensitivity to task-dependent signal changes in nearby brain areas and also can lead to geometric distortions in the image. This is one of the sources of observed differences between fMRI and PET (Devlin et al., 2000; Veltman, Friston, Sanders, & Price, 2000). These static problems are exacerbated in the presence of movement. When the head moves or the vocal tract or nasal cavity change in volume during speech, there are now dynamic, susceptibility-induced changes to the field (see also Savoy, 2001) Sinus volumes change over time and thus even the static susceptibility gradient will not remain constant over repeated experiments with a subject.

The solutions to these movement problems include the use of event-related designs (Birn, Bandettini, Cox, & Shaker, 1999) and the use of pooled group data (Barch et al., 1999). The former takes advantage of the slow hemodynamic response to the initiation of neural processing (3–6 s), acquiring activation data only after articulation is completed. The latter approach assumes that idiosyncratic artifacts may disappear in the averaging process. Neither approach is entirely satisfactory. Event-related designs increase the length of studies compared to block designs while the pooling of data across subjects acts to minimize the individual differences that are the hallmark of the neural representation of language. Given the variability shown between subjects (e.g., Fig. 4), there is a distinct advantage to the event-related designs. (See Savoy, 2001; Op de Beeck, Wagemans, & Vogels, 2001, for more information about event-related designs. See Martin, Goodyear, Gati, & Menon, in press, for the use of even-related designs in the study of swallowing.)

4. Three problems in speech motor control

As indicated above, recording the neural correlates of articulation is a methodologically difficult task. However, the full problem involves more challenges than just the complexity of the data acquisition. As Indefrey and Levelt (2000) suggest, some kind of task decomposition is required before functional imaging of language production can succeed. We need to specify the necessary and sufficient components of speaking aloud. In doing so, we not only must identify the key neural subsystems of speaking and determine their involvement in different speaking tasks but in addition we must understand the processes represented by this neural architecture. In this Section 1 consider three example problems in current speech motor control. Resolution of each of these problems will require innovative behavioral and functional imaging studies. As will be seen, preliminary progress in functional mapping has taken place but little progress has been made in specifying (a) the sensitivity to task and (b) the nature of the processing at these activation sites.

4.1. Units of planning and production

Fig. 1 portrays a multi-staged planning system that coordinates behavior spanning different extents of time (or information). Phrasal structure often spans seconds while phonetic structure is implemented in the order of tens of milliseconds. The speech motor system must plan and produce movements that show pattern across these varied spans of time.

Recently, lesion studies have indicated that the anterior portion of the insula in the left hemisphere may play a special role in speech planning. In a large study of aphasic patients Dronkers (1996) found that all 25 of the patients in this population who were diagnosed with apraxia of speech had lesions in the anterior tip of the precentral gyrus of the insula. Apraxia of speech is viewed as a disorder of motor planning and thus this data suggested strongly that the insula has a special role in this aspect of speech motor control. A series of recent imaging studies have corroborated this finding in normal subjects. Riecker, Ackermann, Wildgruber, Dogil, and Grodd (2000) compared overt speech and singing to covert versions of those tasks using fMRI. While the motor cortex and cerebellum showed activation in both overt and covert tasks the left anterior insula only showed activation during overt speech. Interestingly, the right hemisphere showed activity during overt singing. Wise, Greene, Buchel, and Scott (1999) using PET similarly reported activity in the left anterior insula when subjects spoke. Kuriki, Mori, and Hirata (1999) using MEG observed a source in the same region but slightly superior to the location originally reported by Dronkers.

While these studies are promising, it is not clear what “planning” means in this context. For all of the temporal planning spans suggested by Fig. 1, the production of speech involves the coordinated activity of dozens of muscle groups and articulators ranging from the muscle activity responsible for respiration to the control of the articulators directly involved in sound production (larynx, pharynx, velum, tongue, lips and jaw). The coordination of this complex articulator system requires synergistic behavior between individual articulators in order to produce acoustic targets (Perkell et al., 1997).

This synergistic behavior can be seen in experiments involving the application of unexpected loads to the lips or jaw. Compensations from other articulators are observed so that goals at the phonetic task level are achieved. For example, perturbations of the lips result in compensatory changes in the lips and jaw (Abbs & Gracco, 1984; Gracco & Abbs, 1985, 1988) and the larynx (Löfqvist & Gracco, 1991; Munhall, Löfqvist, & Kelso, 1994); jaw perturbations result in compensatory changes in the tongue (Kelso, Tuller, Vatikiotis-Bateson, & Fowler, 1984), lips (Folkins & Abbs, 1975; Shaiman, 1989), and velum (Kollia, Gracco, & Harris, 1992). Most remarkable is that task-specific responses are observed in a given articulator when it is actively involved in a sound segment being produced but not when the same articulator is not involved in a sound segment (Kelso et al., 1984; Shaiman, 1989). This behavior derives from a neural system that can flexibly and rapidly plan and implement task-specific couplings of articulators (see also Saltzman & Munhall, 1989).

At a longer time scale, the timing and pitch contours responsible for prosody must also be planned and controlled. From patient data there is some indication that the production of prosody (i.e., sentence-level temporal structure) may be differentially lateralized cortically. While the production of phonemic contrasts are selectively impaired with left hemisphere damage, both linguistic and affective structure that spans a phrase or sentence seems to be less strongly lateralized (Baum & Pell, 1999). Mayer et al. (2000) have shown similar results using fMRI. The activation corresponding to prosodic organization was localized to the superior temporal gyrus but they conclude that degree of lateralization is determined by the length of the linguistic unit.

Neurally distributed control of movement characteristics is well established (e.g., Rao et al., 1997). Beyond the insula and superior temporal cortex other regions are consistently associated with articulation in functional imaging studies. For example, cerebellum and supplementary motor area (SMA) show increased activity during overt speech (e.g., Hirano et al., 1996, 1997; Kroll & de Nil, 1998; Wise et al., 1999). Which areas are necessary and for what component of articulation are unclear at present. The recent report of LR, a comatose woman with severe damage to most of the right hemisphere and major portions of the left hemisphere and subcortical structures demonstrates that isolated neural structures can still produce fluent articulation (Schiff, Ribary, Plum, & Llinas, 1999). In spite of being unconscious for 20 years, she produces spontaneous words that are well formed.

In order to begin to understand what role and which structures are playing in language output, greater task control in articulation must be achieved. Clearly some kind of production monitoring is mandatory in language output studies. As was indicated in Section 2, measurement of speech movement in the vocal tract is technically demanding and not completely solved by the research community even outside of the magnet. However, at a minimum, acoustic recordings are necessary. The extremely poor recording conditions may require noise-cancelation or some other type of active filtering approach.

4.2. Auditory feedback

Talkers are responsive to a range of different feedback sources. The oral cavity is richly populated with cutaneous and kinesthetic sensors and this information is used in controlling movements and vocal learning. Auditory feedback has a number of influences on articulation. Background noise produces an immediate modulation of speaking volume (Lombard effect: [4] Lane & Tranel, 1971). Also, clinical and

---

[4] The Lombard effect refers to the tendency to increase the loudness of your voice in the presence of noise.

experiment evidence indicates that perception of your own speech is an important part of speech motor control. Post-linguistically acquired deafness often leads to a deterioration of many aspects of speech production (e.g., intensity and pitch control, variability in the production of vowels and consonants; Cowie & Douglas-Cowie, 1992). In laboratory studies, when subjects hear the pitch of their own voice suddenly raised or lowered artificially, they compensate by shifting their pitch in the opposite direction (e.g. Burnett, Freedland, Larson, & Hain, 1998; Kawahara, 1995). Adaptation to these modified conditions can occur quickly within an experimental session (Houde & Jordan, 1998; Jones & Munhall, 2000). Finally, dynamic conversational interaction involves the monitoring and adaptation to a range of visual signals from a conversational partner (e.g., Bavelas & Chovil, 1997).

From recent stimulation and imaging studies we have some indications about the neural activity during auditory feedback in speech. A group of recent MEG studies have reported reduced auditory cortical response during speech production (Houde, Nagarajan, & Merzenich, 2000; Numminen & Curio, 1999; Numminen, Salmelin, & Hari, 1999). When subjects spoke, the perception of their voice produced smaller magnetic field recordings from the auditory cortex than when they listened to tape-recorded speech (Houde et al., 2000). This attenuated response during speaking is consistent with stimulation data in patients. Creutzfeldt, Ojemann, and Lettich (1989) observed suppression of activity in neurons in the lateral temporal cortex during speech. Similar data have been reported from animal studies. Muller-Preuss and Ploog (1981) found a population of primary auditory cortex cells in squirrel monkeys that were suppressed during the animal's voicing.

On the other hand, a series of imaging studies have shown increased activation in cortical regions in response to talkers hearing their own voice during speech. Price et al. (1996b) and McGuire, Silbersweig, and Frith (1996) have shown activation in the anterior temporal cortex during overt speech production when the speaker's voice can be heard. The activated area, auditory association cortex, corresponds with areas activated by listening to someone else's speech or recordings of your own speech. A preliminary study of singing (Perry et al., 1999) showed similar activation near the right superior temporal region during overt voicing.

In a series of studies, Hirano and colleagues (1996, 1997) suggest that observations of activation of auditory speech areas may depend on the task demands and the role auditory feedback is playing in a particular utterance. In Hirano et al. (1996), they report infrequent activation of the superior temporal region in overt compared to silent reading of familiar written text. In a follow-up study (Hirano et al., 1997), in which subjects heard modified feedback (low-pass filtered or delayed auditory feedback) they did show increased regional blood flow in the auditory speech areas. McGuire et al. (1996) gave subjects pitch-shifted speech feedback and found trends of increased activity relative to normal auditory feedback in the superior temporal region but not as dramatic as the difference shown in Hirano et al. (see also Houde et al., 2000).

Hirano et al. (1997) interpret their findings as evidence for different cortical mechanisms for processing overt speech – one for more automatized or overlearned speech and another for situations that require more feedback. While this suggestion fits with a longstanding dichotomy between open- and closed-loop control systems, current work on internal models in motor control suggests that feedback can play a range of possible roles in learning and control and that multiple feedback systems can be in operation at once (Kawato, 1999; Perkell et al., 1997). Thus, this simple dichotomy may underestimate the range of possible control systems involved in speech control. For example, some kind of internal feedback or corollary discharge could be involved. Paus et al. (1996) have demonstrated that, at least in part, activation patterns may not require that subjects hear themselves. In a PET study, their subjects whispered syllables at a range of speaking rates while listening to a white masking noise that prevented the perception of their own speech. As syllable rate increased cerebral blood flow increased in a range of areas associated with speech production as well as the left planum temporale and the left posterior perisylvian cortex. The authors interpret this as evidence for motor-to-sensory discharge.

These findings are consistent with other neuroimaging and stimulation studies that find support for a linkage between perception and production. Ojemann and Mateer (1979) showed that phoneme identification and sequential lip movements were both disrupted by stimulating the same cortical region (posterior Broca's area). More recently, Wise et al. (1999), in a recent PET study, found that the left anterior insula and lateral premotor area showed activation during both listening and repetition.

Auditory transformation manipulations such as used by McGuire et al. are an important tool for understanding the role played by feedback in controlling articulation. However, these manipulations raise a series of concerns about task sensitivity. Pitch shifts and feedback delays modify speaking rate and the amplification that is required in these experiments to overcome the subjects true airborne speech induces an increase in the talker's speaking effort and loudness (Mayer et al., 2000). In addition, attention to the feedback may vary with the feedback condition (see also Price et al., 1996a,b). All of these speech manner variables can influence neural activation.

4.3. Learning

Learning and development and the corresponding neural plasticity that is associated with these changes are the focus of research across a broad spectrum of cognitive neuroscience. The study of learning can make contributions, as well, to the understanding of spoken language. Learning to speak requires mastery of extensive new knowledge including linguistic categories as well as perceptual-motor skills. An essential part of learning to speak is learning the geometry of the vocal tract and the mapping between speech acoustics and vocal tract shapes (Callan, Kent, Guenther, & Vorperian, 2000). For children this means learning the dimensions of their articulators and learning the unique morphology of their tongue and palate. In addition, it means that the child's motor control must be adapted to the slow but significant changes in shape and size of the vocal tract due to growth (Saltzman & Munhall, 1992). From a linguistic perspective, children must learn the sound system of their native language and manners in which sequences of sound are put together

(Callan et al., 2000). Learning continues beyond childhood in the acquisition of new vocabulary and new lexical forms, adaptation to dental prostheses or oral cavity changes, language and dialect acquisition and changes following speech and language therapy, etc.

Using functional imaging to study learning in speech is important for two reasons. First, it allows us to address fundamental questions about learning and the acquisition of speech skills. Secondly, there are methodological reasons for understanding learning in this context. In functional imaging, a range of short-term changes in performance can be broadly classified as learning. For example, Petersen, van Mier, Fiez, and Raichle (1998) showed a shift in the areas used in performing the same verbal task over the course of 15 min of practice. In studies that are less formally designed to examine practice, similar changes can be seen in activation patterns over the course of the experiment (e.g., McGuire et al., 1996).

One area in which learning paradigms have been used is in the evaluation of treatment effects in people who stutter (e.g., Kroll & de Nil, 1998; see Fox et al., 2000; Salmelin, Schnitzler, Schmitz, & Freud, 2000, for other studies of stuttering). In these studies PET imaging was used pre- and post-therapy in order to examine differences between the patient population and normal talkers and to track changes over time. In Kroll and de Nil's work the stutterers showed changes in the activity level of a number of regions such as the anterior cingulate cortex and the cerebellum. The observation of cerebellar changes fits with a growing literature that suggests that parts of this structure play an important role in information processing during motor learning (e.g., Imamizu et al., 2000; Raymond, Lisberger, & Mauk, 1996; see Parsons, 2001).

However, distinguishing changes in learning from a number of possible confounds may be difficult. Performance differences, fluctuations in vigilance, anxiety, etc., can contribute to changes in neural activity over time (Poldrack, 2000) without there being true differences in the functional anatomy of the task. As Kroll and de Nil (1998) note, the anterior cingulate cortex effects in their studies may be related to the emotional response in their subjects. On the other hand, changes in processing or changes in the efficiency of the circuitry may also be occurring (Petersen et al., 1998). Given that these kinds of short-term changes may be a factor in all imaging studies it can be argued that some type of learning paradigm should always be used.

5. Conclusions

After more than a decade of functional neural imaging of language production, we have seen significant advances in our understanding of language production as well as the methodologies for studying it. A variety of new imaging and neural recording techniques are available to researchers today (e.g., fMRI, PET, MEG, ERP, etc.), each with its own strengths and weaknesses. The study of articulation, the end product of language production, is particularly challenging and may require the use of multiple techniques. One of the major challenges facing this field is task specification. As I have suggested above, articulation involves significant information processing for its sensorimotor aspects and in addition it is highly interactive with other levels of communication planning. Understanding what neural structures are required for language output will necessitate better estimates of the output itself. The difficulties interpreting silent or rest conditions have been well documented (e.g., Rosen et al., 2000) [5].

In each of the three questions discussed above, progress in functional neural imaging has been made but there are limits on future gains without better task specification. In the case of speaking aloud, this will require at least two things. First, some type of speech production monitoring must be included in the experimental protocols. Second, we must have a better idea about what processes are involved in the variety of different ways of speaking. Both of these requirements imply that future functional neural imaging will be tied to advances in future structural imaging and our understanding of acoustic-articulatory relations in speech. In order to understand the neural representation of the behavior, we must better understand the behavior itself.

Acknowledgements

This work was funded by NIH grant DC-00594 from the National Institute of Deafness and other Communications Disorders and NSERC. The author wishes to thank Jeff Jones for helpful comments on an earlier draft of this manuscript. Brad Story (Fig. 2) and Mark Tiede (Fig. 3) generously shared their vocal tract images.

References

Abbs, J. H., & Gracco, V. L. (1984). Control of complex motor gestures: orofacial muscle responses to load perturbations of the lip during speech. *Journal of Neurophysiology*, *51*, 705–723.

Ackermann, H., & Hertrich, I. (2000). The contribution of the cerebellum to speech processing. *Journal of Neurolinguistics*, *13*, 95–116.

Baer, T., Gore, J. C., Gracco, L. C., & Nye, P. W. (1991). Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. *Journal of the Acoustical Society of America*, *90*, 799–828.

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.

Baum, S., & Pell, M. (1999). The neural bases of prosody: insights from lesion studies and neuroimaging. *Aphasiology*, *13*, 581–608.

Bavelas, J., & Chovil, N., (1997). Faces in dialogue. In J. A. Russell, & J. M. Fernández-Dols (Eds.), *The psychology of facial expression* (pp. 334–346). Cambridge: Cambridge University Press.

Birn, R., Bandettini, P., Cox, R. W., Jesmanowicz, A., & Shaker, R. (1998). Magnetic field changes in the human brain due to swallowing or speaking. *Magnetic Resonance in Medicine*, *40*, 55–60.

---

[5] It should be noted that the lack of significant linguistic processing in many speech experiments is also a serious problem. In most physiological studies of articulation, subjects repeat nonsense utterances over and over. The absence of "language" in these studies may raise some of the same concerns as the use of "silent" tasks.

Birn, R., Bandettini, P. A., Cox, R. W., & Shaker, R. (1999). Event-related fMRI of tasks involving brief motion. *Human Brain Mapping*, *7*, 106–114.

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *Journal of the Acoustical Society of America*, *103*, 3153–3161.

Callan, D. E., Kent, R. D., Guenther, F. H., & Vorperian, H. K. (2000). An auditory-feedback based neural network model of speech production that is robust to developmental changes in the size and shape of articulatory systems. *Journal of Speech, Language & Hearing Research*, *43*, 721–736.

Caplan, D. (1992). *Language*. Cambridge, MA: The MIT Press.

Corfield, D. R., Murphy, K., Josephs, O., Fink, G., Frackowiak, R., Guz, A., Adams, L., & Turner, R. (1999). Cortical and subcortical control of tongue movement in humans: a functional neuroimaging study using fMRI. *Journal of Applied Physiology*, *86*(5), 1468–1477.

Cowie, R., & Douglas-Cowie, E. (1992). *Trends in linguistics, studies and monographs. Postlingually acquired deafness* (Vol. 62). New York: Mouton de Gruyter.

Creutzfeldt, O., Ojemann, G., & Lettich, E. (1989). Neuronal activity in the human lateral temporal lobe. II. Responses to the subjects own voice. *Experimental Brain Research*, *77*, 476–489.

Damasio, A. R. & Damasio, H. (1988). Advances in the neuroanatomical correlates of aphasia and the understanding of the neural substrates of language. In L. M. Hyman, & C. N. Li (Eds.), *Language, speech and mind: studies in honour of Victoria A. Fromkin* (pp. 103–117). London: Routledge.

Damasio, A. R., Damasio, H., Rizzo, M., Varney, N., & Gersh, F. (1982). Aphasia with lesions in the basal ganglia and internal capsule. *Archives in Neurology*, *39*, 15–20.

Dang, J., Honda, K., & Suzuki, H. (1994). Morphological and acoustical analysis of the nasal and the paranasal cavities. *Journal of the Acoustical Society of America*, *96*, 2088–2100.

Demolin, D., Metens, T., & Soquet, A. (2000). Real time MRI and articulatory coordinations in vowels. In *Proceedings of the fifth seminar on speech production: Models and data* (pp. 93–96). Munich: Universität München.

Demonet, J., Fiez, J., Paulesu, E., Petersen, S., & Zatorre, R. (1996). PET studies of phonological processing: a critical reply to Poeppel. *Brain and Language*, *55*, 352–379.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Wilson, J., Moss, H., Matthews, P., & Tyler, L. K. (2000). Susceptibility-induced loss of signal: comparing PET and IMRI on a semantic task. *NeuroImage*, *11*, 589–600.

Dronkers, N. F. (1996). A new brain region for co-ordinating speech articulation. *Nature*, *384*, 159–161.

Dronkers, N. F., Redfern, B. B., & Knight, R. T. (2000). The neural architecture of language disorders. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 949–958). Cambridge, MA: MIT Press.

Fant, G. (1950). *Acoustic theory of speech production*. The Hague: Mouton.

Fiez, J. A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human Brain Mapping*, *5*, 79–83.

Fiez, J. A., & Raichle, M. E. (1997). Linguistic processing. *International Review of Neurobiology*, *41*, 233–254.

Folkins, J. W., & Abbs, J. H. (1975). Lip and jaw motor control during speech: responses to resistive loading of the jaw. *Journal of Speech & Hearing Research*, *18*, 207–220.

Fox, P. T., Ingham, R. J., Ingham, J. C., Zamarripa, F., Xiong, J., & Lancaster, J. (2000). Brain correlates of stuttering and syllable production. *Brain*, *123*, 1985–2004.

Fromkin, V. (1973). *Speech errors as linguistic evidence*. The Hague: Mouton.

Gagne, J. P., Masterson, V., Munhall, K. G., Bilida, N., & Querengesser, C. (1995). Across talker variability in speech intelligibility for conversational and clear speech. *Journal of the Academy of Rehabilitative Audiology*, *27*, 135–158.

Grabowski, T., & Damasio, A. (2000). Investigating language with functional imaging. In A. W. Toga, & J. C. Mazziotta (Eds.), *Brain mapping: the systems* (pp. 425–461). San Diego: Academic Press.

Gracco, V. L., & Abbs, J. H. (1985). Dynamic control of the perioral system during speech: kinematic analyses of autogenic and nonautogenic sensorimotor processes. *Journal of Neurophysiology*, *54*, 418–432.

Gracco, V. L., & Abbs, J. H. (1988). Central patterning of speech movements. *Experimental Brain Research*, *71*, 515–526.

Gracco, V. L., & Abbs, J. H. (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research*, *75*, 586–598.

Guenther, F. (1994). A neural network model of speech acquisition and motor equivalent production. *Biological Cybernetics*, *72*, 43–53.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konish, J. (1996). Cortical speech processing mechanisms while vocalizing visually presented languages. *NeuroReport*, *8*, 363–367.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konish, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport*, *8*, 2379–2382.

Hirayama, M., Vatikiotis-Bateson, E., & Kawato, M. (1994). Inverse dynamics of speech motor control. In S. J. Hanson, J. D. Cowan, & C. L. Giles (Eds.), *Advances in neural information processing systems* (Vol. 6, pp. 1043–1050). San Mateo, CA: Morgan Kaufmann Publishers.

Hoole, P., Nguyen-Trong, N., & Hardcastle, W. (1993). A comparative investigation of coarticulation in fricatives: electropalatographic, elctromagnetic and acoustic data. *Language and Speech*, *36*, 235–260.

Houde, J. F., Nagarajan, S. & Merzenich, M. (2000). Modulation of auditory cortex during speech production: an MEG study. In *Proceedings of the fifth seminar on speech production: models and data* (pp. 249–252). Munich: Universität München.

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*, 1213–1216.

Imamizu, H., Miyauchi, S., Tamada, T., Sasaki, Y., Takino, R., Pütz, B., Yoshioka, T., & Kawato, M. (2000). Human cerebellar activity reflecting an acquired internal model of a new tool. *Nature*, *403*, 192–195.

Indefrey, P., & Levelt, W. J. M. (2000). The neural correlates of language production. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 845–865). Cambridge, MA: MIT Press.

Jennings, J. M., McIntosh, A., Kapur, S., Tulving, E., & Houle, S. (1997). Cognitive subtractions may not add up: the interaction between semantic processing and response mode. *NeuroImage*, *5*, 229–239.

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: evidence from feedback perturbation. *Journal of the Acoustical Society of America*, *108*, 1246–1251.

Jordan, M. I. (1990). Motor learning and the degrees of freedom problem. In M. Jeannerod (Ed.), *Attention and performance* (Vol. 13, pp. 796–836). Hillsdale, NJ: Lawrence Erlbaum Associates.

Jordan, M. I. (1996). Computational aspects of motor control and motor learning. In H. Heuer, & S. Keele, (Eds.), *Handbook of perception and action* (Vol. 2, pp. 71–120). New York: Academic Press.

Kawahara, H. (1995). Transformed auditory feedback: the collection of data from 1993.1 to 1994.12 by a new set of analysis procedures. *ATR Technical Report, TR-H-120*.

Kawato, M. (1997). Bidirectional theory approach to consciousness. In M. Ito, Y. Miyashita, & E. T. Rolls (Eds.), *Cognition, computation, and consciousness* (pp. 223–248). Oxford: Oxford University Press.

Kawato, M. (1999). Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, *9*, 718–727.

Kawato, M., Furukawa, K., & Suzuki, R. (1987). A hierachical neural-network model for control and learning of voluntary movement. *Biological Cybernetics*, *57*, 169–185.

Kelso, J. A. S., Tuller, B., Vatikiotis-Bateson, E., & Fowler, C. (1984). Functionally specific articulatory cooperation following jaw perturbations during speech: evidence for coordinative structures. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 812–832.

Kent, R., & Moll, K. (1975). Articulatory timing in selected consonant sequences. *Brain and Language*, *2*, 304–323.

Kollia, H. B., Gracco, V. L., & Harris, K. S. (1992). Functional organization of velar movements following jaw perturbation. *Journal of the Acoustical Society of America*, *91*, 2474.

Kosslyn, S. M. (1999). If neuroimaging is the answer, what is the question? *Philosophical Transactions of the Royal Society of London*, *354*, 1283–1294.

Kroll, R., & de Nil, L. (1998). Positron emission tomography studies of stuttering: their relationship to our theoretical and clinical understanding of the disorder. *Journal of Speech, Language Pathology, and Audiology*, *22*, 261–270.

Kuratate, T., Munhall, K. G., Rubin, P. E., Vatikiotis-Bateson, E., & Yehia, H. (1999). Audio-visual synthesis of talking faces from speech production correlates. In *Proceedings of EuroSpeech'99* (pp. 1279–1282). ESCA.

Kuriki, S., Mori, T., & Hirata, Y. (1999). Motor planning center for speech articulation in the normal human brain. *NeuroReport*, *10*, 765–769.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech & Hearing Research*, *14*, 677–709.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation*. Cambridge, MA: MIT Press.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Science*, *4*, 187–196.

Lindblom, B., & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw & larynx movement. *Journal of the Acoustical Society of America*, *50*, 1166–1179.

Löfqvist, A., & Gracco, V. L. (1991). Discrete and continuous modes in speech motor control. *PERILUS*, *XIV*, 27–34.

Martin, R. E., Goodyear, B., Gati, J. & Menon, R. (in press). Cerebral cortical representation of automatic and volitional swallowing in humans. *Journal of Neurophysiology*.

Mathiak, K., Klose, U., Ackermann, H., Hertrich, I., Kincses, W. -E., & Grodd, W. (2000). Stroboscopic articulography using fast magnetic resonance imaging. In *Proceedings of the fifth seminar on speech production: models and data* (pp. 97–100). Munich: Universität München.

Mayer, M., Dogil, G., Ackermann, H., Erb, M., Riecker, A., Wildgruber, D., & Grodd, W. (2000). Prosody in speech production: a paradigm for functional imaging and first result. In *Proceedings of the fifth seminar on speech production: models and data* (pp. 281–184). Munich: Universität München.

McCarthy, G. Blamire, A. M., Rothman, D. L., Gruetter, R., & Shulman, R. G. (1993). Echo-planar magnetic resonance imaging studies of frontal cortex activation during word generation in humans. *Proceedings of the National Academy of Sciences in the USA, 90, 4952–4956.*

McGuire, P. K., Silbersweig, D. A., & Frith, C. D. (1996). Functional neuroanatomy of verbal self-monitoring. *Brain*, *119*, 907–917.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*, 1265–1279.

Miller, J., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *Journal of the Acoustical Society of America*, *73*, 1751–1755.

Mohr, J. P. (1976). Broca's area and Broca's aphasia. In H. Whitaker, & H. Whitaker (Eds.), *Studies in aphasia* (Vol. 1, pp. 201–233). New York: Academic Press.

Moll, K. (1960). Cinefluorographic techniques in speech research. *Journal of Speech & Hearing Research*, *3*, 227–241.

Moore, C. A. (1992). The correspondence of vocal tract resonance with volumes obtained from magnetic resonance images. *Journal of Speech & Hearing Research*, *35*, 1009–1023.

Muller-Preuss, P., & Ploog, D. (1981). Inhibition of auditory cortical neurous during phonation. *Brain Research*, *215*, 61–76.

Munhall, K. G., Kawato, M., & Vatikiotis-Bateson, E. (2000). Coarticulation and physical model of speech production. In M. B. Broe ,& J. B. Pierrehumbert (Eds.), *Laboratory phonology V: acquisition and the lexicon* (pp. 9–28). Cambridge: Cambridge University Press.

Munhall, K., Löfqvist, A., & Kelso, J. A. S. (1994). Lip-larynx coordination in speech: effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America*, *96*, 3605–3616.

Munhall, K. G., Vatikiotis-Bateson, E., & Tohkura, Y. (1994). *X-ray film database for speech research* (Vol. 1). Canada: Queen's University.

Munhall, K. G., Vatikiotis-Bateson, E., & Tohkura, Y. (1995). X-ray film database for speech research. *Journal of the Acoustical Society of America*, *98*, 1222–1224.

Nadeau, S., & Crosson, B. (1997). Subcortical aphasias. *Brain and Language*, *58*, 355–402.

Narayanan, S. S., Alwan, A. A., & Haker, K. (1995). An articulatory study of fricative consonants using magnetic resonance imaging. *Journal of the Acoustical Society of America*, *98*, 1325–1347.

Numminen, J., & Curio, G. (1999). Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. *Neuroscience Letters*, *272*(1), 29–32.

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters*, *265*(2), 119–122.

Ojemann, G. A. (1991). Cortical organization of language. *Journal of Neuroscience*, *11*(8), 2281–2287.

Ojemann, G. A., & Mateer, C. (1979). Human language cortex: localization of memory, syntax, and sequential motor-phoneme identification systems. *Science*, *205*, 1401–1403.

Op de Beeck, H., Wagemans, J., & Vogels, R. (2001). Can neuroimaging really tell us what the brain is doing? The relevance of indirect measures of population activity. *Acta Psychologica*, *107*, 323–351.

Parsons, L. (2001). Integrating cognitive psychology, neurology, and neuroimaging. *Acta Psychologica*, *107*, 155–181.

Paus, T., Perry, D. W., Zatorre, R. J., Worsley, K. J., & Evans, A. C. (1996). Modulation of cerebral blood flow in the human auditory cortex during speech: role of motor-to-sensory discharges. *European Journal of Neuroscience*, *8*, 2236–2246.

Penfield, W., & Roberts, L. (1959). *Speech and brain-mechanisms*. Princeton, NJ: Princeton University Press.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., & Guoid, P. (1997). Speech motor control: acoustic goals,saturation effects, auditory feedback and internal models. *Speech Communication*, *22*, 227–250.

Perry, D. W., Zatorre, R. J., Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1999). Localization of cerebral activity during simple singing. *NeuroReport*, *10*, 3979–3984.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintum, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical analtomy of single word processing. *Nature*, *311*, 585–589.

Petersen, S. E., van Mier, H., Fiez, J. A., & Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. *Proceeding of the National Academy of Sciences in the USA*, *95*, 853–860.

Poeppel, D. A. (1996). Critical review of PET studies of phonological processing. *Brain and Language*, *55*, 317–351.

Poldrack, R. A. (2000). Imaging brain plasticity: conceptual and methodological issues – a theoretical review. *NeuroImage*, *12*, 1–13.

Price, C. J., Moore, C. J., & Frackowiak, R. S. (1996a). The effect of varying stimulus rate and duration on brain activity during reading. *NeuroImage*, *3*, 40–52.

Price, C. J., Wise, R. J. S., Warburton, E. A., Moore, C. J., Howard, D., Patterson, K., Frackowiak, R. S. J., & Friston, K. J. (1996b). Hearing and saying: the functional neuro-anatomy of auditory word processing. *Brain*, *119*, 919–931.

Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz, J. A., Cox, R. W., & Binder, J. R. (1997). Distributed neural systems underlying the timing of movements. *Journal of Neuroscience*, *17*, 5528–5535.

Raymond, J., Lisberger, S., & Mauk, M. (1996). The cerebellum: a neuronal learning machine. *Science*, *272*, 1126–1131.

Remez, R., Lipton, J., & Fellowes, J. (1996). Perceiving the difference between spontaneous and read speech: the role of physical duration. *Journal of the Acoustical Society of America*, *99*, 2493.

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.

Rokkaku, M., Hashimoto, K., Imaizumi, S., Niimi, S., & Kiritani, S. (1986). Measurement of the three-dimensional shape of the vocal tract based on the magnetic resonance imaging technique. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, *20*, 47–54.

Rosen, H. J., Ojemann, J. G., Ollinger, J. M., & Petersen, S. E. (2000). Comparison of brain activation during word retrieval done silently and aloud using fMRI. *Brain and Cognition*, *42*, 201–217.

Salmelin, R., Schnitzler, A., Schmitz, F., & Freud, H.-J. (2000). Single word reading in developmental stutterers and fluent speakers. *Brain*, *123*, 1184–1202.

Saltzman, E., Lofqvist, A., & Mitra, S. (2000). "Glue" and "clocks": Intergestural cohesion and global timing. In M. B. Broe, & J. B. Pierrehumbert (Eds.), *Laboratory phonology V: acquisition and the lexicon* (pp. 88–101). Cambridge: Cambridge University Press.

Saltzman, E., & Munhall, K. G. (1989). A dynamic approach to gestural patterning in speech production. *Ecological Psychology*, *1*, 333–382.

Saltzman, E., & Munhall, K. G. (1992). Towards a dynamics of skill acquisition and development. *Journal of Motor Behavior*, *24*, 49–57.

Savoy, R. (2001). History and future directions of human brain mapping and functional neuroimaging. *Acta Psychologica*, *107*, 9–42.

Schiff, N., Ribary, U., Plum, F., & Llinas, R. (1999). Words without mind. *Journal of Cognitive Neuroscience*, *11*, 650–656.

Shadle, C. H., Mohammad, M., Carter, J. N., & Jackson, P. J. B. (1999). Multi-planar dynamic magnetic resonance imaging: new tools for speech research. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 623–626). San Francisco, CA.

Shaiman, S. (1989). Kinematic and electromyographic responses to perturbation of the jaw. *Journal of the Acoustical Society of America*, *86*, 78–87.

Shiller, D. M., Ostry, D. J., & Gribble, P. L. (1999). Effects of gravitational load on jaw movements in speech. *Journal of Neuroscience*, *19*(20), 9073–9080.

Stone, M., Dick, D., Douglas, A. S., Davis, E. P. & Ozturk, C. (2000). Modelling the internal tongue using principal strains. In *Proceedings of the fifth seminar on speech production: models and data* (pp. 133–136). Munich: Universität München.

Stone, M., Morrish, K. A., Sonies, B. C., & Shawker, T. H. (1987). Tongue curvature: a model of shape during vowel production. *Folia Phoniatrica*, *39*, 302–315.

Story, B. H., Titze, I. R., & Hoffman, E. A. (1996). Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, *100*, 537–554.

Story, B. H., Titze, I. R., & Hoffman, E. A. (1998). Vocal tract area functions for an adult female speaker based on volumetric imaging. *Journal of the Acoustical Society of America*, *104*, 471–487.

Tiede, M. K. (1996). An MRI-based study of pharyngeal volume contrasts in Akan and English. *Journal of Phonetics*, *24*, 399–421.

Tiede, M. K., Masaki, S. & Vatikiotis-Bateson, E. (2000). Contrasts in speech articulation observed in sitting and supine conditions. In *Proceedings of the fifth seminar on speech production: models and data* (pp. 25–28). Munich: Universität München.

Veltman, D. J., Friston, K. J., Sanders, G., & Price, C. J. (2000). Regionally specific sensitivity differences in fMRI and PET: where do they come from? *NeuroImage*, *11*, 578–588.

Wise, R. J., Greene, J., Buchel, C., & Scott, S. K. (1999). Brain regions involved in articulation. *The Lancet*, *353*, 1057–1061.

Yang, C. -S., & Kasuya, H. (1994). Accurate measurement of vocal tract shapes from magnetic resonance images of child, female, and male subjects. In *Proceedings of the International Conference on Spoken Language Processing* (Vol. 2, pp. 623–626).

Zelkowicz, B. J., Herbster, A. N., Nebes, R. D., Mintun, M. A., & Becker, J. T. (1998). An examination of regional cerebral blood flow during object naming tasks. *Journal of the International Neuropsychological Society*, *4*, 160–166.