

# 13

## Computational primitives in phonology and their neural correlates

Philip J. Monahan, Ellen F. Lau, and William J. Idsardi

### 13.1 Introduction

Understanding the representational and algorithmic systems that underlie language has been the primary goal for linguists (e.g., Chomsky 1959a). As it stands, however, the neural bases of linguistic systems remain woefully underspecified. Moving toward an understanding of how such systems are encoded in the brain, however, demands a linking between the basic ontological primitives that underlie linguistic systems on the one hand with neurophysiology on the other (Poeppel and Embick 2005). The goal of the current chapter is to identify what we believe to be the core computational primitives that underlie phonological knowledge and present evidence from the domain of cognitive neuroscience that attempts to investigate the nature of the neural correlates of these primitives.

We take the goal of biolinguistics to be to understand the biological bases of human language with a strong emphasis on its evolutionary origins (Jenkins 2000). In practice, the focus has been placed on investigating the biological underpinnings of syntactic knowledge, e.g., the evolutionary origins of the syntactic operation MERGE (Hauser, Chomsky and Fitch 2002). Although this work has been enlightening, it is perhaps surprising that relatively less effort has been paid to the biolinguistic foundations of phonological systems (Samuels 2011) and their interface with speech perception and production. This is surprising because we believe that there are a number of reasons that biolinguistic inquiry into this domain should be more tractable. First, fewer levels of abstraction separate the fundamental representations of phonology from the basic sensory input representations. This means that knowledge about how basic auditory information is represented and processed in both humans and animals is more likely to provide important insights into how phonological information could be represented and processed. For this type of

investigation, we can use animal models that allow a fuller range of invasive measures such as recording from individual neurons. As we will show, there is also evidence that representations that are closer to the sensory input are more likely to be neurally implemented with a spatial-mapping coding scheme that is relatively easy to detect with current neuroimaging measures (see below). Second, there already exists an extensive literature to build on from cognitive psychology that has investigated the extent to which “speech is special” (Liberman 1996). The Motor Theory of speech perception is essentially a biolinguistic perspective: whether or not humans have specific biological adaptations that support the processing of auditory linguistic input or whether speech is processed and represented with purely domain-general machinery (Holt and Lotto 2008). Third, on most linguistic theories, phonological representations are the basic unit that connects sensory input and motor output. Therefore, by investigating the biological basis of phonological knowledge, we can benefit from existing evidence from other cognitive domains on the biological basis for sensory-motor translation, such as is needed for visually guided reaching.

Technological advances in the last 30 years have led to exceptionally powerful recording devices, allowing us to not only ask questions about when or where in the brain particular cognitive processes occur but, in theory, to provide answers with millisecond temporal resolution and/or millimeter spatial resolution (see Hämäläinen 1992; Luck 2005; Cabeza and Kingstone 2006 for reviews of some current techniques). Previously, much of what we knew about the brain with respect to language was inferred from neurological impairments, e.g., aphasias. The link between brain region and linguistic process required behavioral observation/testing and an often much later postmortem assessment of the lesions (Broca 1861; Wernicke 1874; see Geschwind 1970 for a review). The advent of hemodynamic and electrophysiological measures allows us to observe cortical and sub-cortical neural activity as it occurs in healthy adults and children. Yet, despite the popularity of these technologies in psycholinguistics and the cognitive neurosciences, our understanding of how the brain encodes speech sounds, words and higher-order linguistic representations has so far remained limited.

In this chapter, our aim is not to outline a complete program for a biolinguistic investigation of phonology, but rather to provide a summary of the existing “tools” for such investigation: the representational primitives from which we believe will be the easiest to begin, a representative selection of previous cognitive neuroimaging findings that may serve as a guide to promising entry points for further research, and in an appendix, the neuroimaging measures currently available. Although the primary concern of biolinguists may be to determine how phonological systems and grammars are encoded in cortical structures, we believe that phonological representations are likely to depend on how they are

deployed during speech perception and production, and thus that a better understanding of these processes is also required. Issues that must be addressed as we move forward in this line of research therefore include understanding what mechanisms the brain uses to parse acoustic input into words made up of abstract phonological pieces (e.g., Hickok and Poeppel 2007; Poeppel, Idsardi, and van Wassenhove 2008) and on the output side, what mechanisms are used by the brain to arrive at a set of articulation instructions governed by phonological rules of the language from a conceptualized message (see Levelt 1989; Levelt, Roelofs, and Meyer 1999 for psycholinguistic models). Our own research is primarily concerned with perception, and consequently, we will have nothing substantial to report with respect to the question of the neural bases of speech-motor planning.

## 13.2 Background

Phonological grammars provide a framework for how speech sounds are represented and the nature of the various combinatorial operations they undergo in mapping between *lexical representations* and their *surface-forms* (Chomsky and Halle 1968; Prince and Smolensky 2004; see Kenstowicz, 1994 for textbook discussion). These grammars have explicitly concerned themselves with the *competence* of an ideal speaker/hearer (Chomsky 1965; cf. Bybee 2001) and are generally unconcerned with psychological or neurobiological implications or what has been termed a speaker's *performance*. Consequently, phonologists are interested in developing theories of the *knowledge* of phonological primitives and rules that operate over them within the language. However, we may want to go beyond a theory of the knowledge that people possess to understand how that knowledge is actually instantiated in the brain. The question may be as simple as: in what brain region are phonological representations stored? Or, more complex, how do networks of neurons represent phonological categories? Because phonology is intimately tied to lexical representation, we might ask how *words* are represented neurally. And from here, we may go on to ask questions of evolutionary biology, e.g., why might the brain have gravitated to this particular way of representing phonology and the lexicon? Unfortunately, formal approaches to phonological grammars and developments in the cognitive neuroscience of language and audition tend to proceed in parallel, and only rarely do advances in one discipline frame discussion in the other. Moreover, the lack of linking hypotheses between representational and procedural primitives in phonology and those in the neurosciences potentially impedes progress in understanding how phonological grammars are neurally encoded or the range of plausible phonological computations that can be carried out in the nervous system (Poeppel and Embick 2005).

We can infer the nature of the knowledge of phonology that people possess from many sources of evidence: acceptability judgments, historical sources, production data, and reaction times on comprehension or production tasks. These measures are also, of course, indirect measurements of neural representations. One avenue of cognitive neuroscience research on phonology is devoted to identifying correlated evidence from slightly more direct measurements of brain activity, such as ERP, MEG, or fMRI (discussed in more detail in the appendix). The data gained from these more sophisticated measures is of a similar ilk as behavioral measurements in that they are aggregated measures whose response properties cannot be directly tied to biological events. Brain measures can have advantages over behavioral measures. For example, it is often not necessary to require an overt response on the part of the participants, and as such, the notion of the “task,” in principle, can be foregone with these measures. We provide a brief review of the tools that are currently used and some of their strengths and weaknesses in the appendix.

### 13.3 Representational primitives

Traditionally, generative phonology has asked two questions: (1) what representations subserve phonology (e.g., features, segments, syllables) and (2) what procedures map between surface forms and memory representations. The representational units that constitute part of the *knowledge* or *competence* of sound systems have been thought to be discrete and invariant in their most abstract form. The exact form of these representations, however, has been long debated within the field – though, we believe – there is a consensus that sub-segmental units, such as distinctive features, appear to be the appropriate level of description for the purpose of biolinguistic investigation. The specific features themselves continue to be revised, and while these revisions will bring important implications for how we investigate the biological instantiation of phonological knowledge and representation, the general point remains that we believe that electrophysiological measures will reveal a featural organization for phonology. This does not preclude investigating other aspects of phonological representations, e.g., syllables.

A hallmark proposal of phonological inquiry has been that the relevant representational primitives are not single segments, e.g., phonemes, but rather smaller units of which segments are composed (Jakobson, Fant, and Halle 1952). Initially, segments were thought to be bundles of divalent distinctive features, and in the mapping from lexical representation to surface forms, phonological rules would adjust their valence (Chomsky and Halle 1968). Evidence in support of distinctive features arises from extensive cross-linguistic observations that speech sounds pattern phonologically in natural classes. That is, phonological rules

rarely target individual speech sounds, but rather entire natural classes that share some common characteristic or feature, e.g., high-vowels ([+vocalic], [+high]) or voiced consonants ([+consonantal], [+voiced]). For example, the feature specifications [+syllabic, -consonant] designated all vowels in a language and all vowels shared this particular valuation (Halle, 1972) without an internal organization to the features. These features are arranged into a hierarchical organization (Clements and Hume 1995; McCarthy 1988)

For the purposes of perception, distinctive features are divided into two broad classes: articulator-free (e.g., [vowel], [consonant], [continuant]) and articulator-bound (e.g., [spread glottis], [high], [lateral]). Articulator-free features provide a “first-pass” segmentation of incoming speech by identifying “landmarks” in the signal for the perceptual system (Stevens 2002). In Stevens’s “Landmarks” model of speech perception and spoken word recognition, acoustic analysis around landmarks permits hypothesis generation for what the articulators were doing during production, i.e., recovery of the articulator-bound features. This latter class of articulator-bound features provides instructions to the articulators during production. For example, the specification of the feature [ $\pm$  nasal] provides instructions as to whether or not the velum, the soft tissue located at the roof of the mouth, should be elevated to prevent air passage through the nasal cavity, while [ $\pm$  labial] provides instructions to the lips as to whether or not a constriction should be formed. During natural articulation, these features must be orchestrated in time in a gestural score to produce a meaningful speech signal (Browman and Goldstein 1989).

Despite major differences in phonological theories over the past 50 years, phonologists continue to express generalizations in terms of distinctive features. In generative phonology, phonemes are nothing more than constellations of simultaneous distinctive features. Other theories (Mielke 2008), however, give segments a more primitive status, while still retaining features within the system. Much of the electrophysiological work, thus far, fails to distinguish between segmental and featural interpretations of the experiments. Well-designed experiments to tease apart the interpretations should be a priority as we move forward.

### 13.4 Where to look for phonological primitives

Now with over a century of work with aphasia and more recently, neuro-imaging, we have gained significant insights into the most basic neural implementation question - where the neurons that code the stored phonetic and phonological primitives are located. Today, there is fairly broad consensus that these neurons are located in the superior temporal gyrus (STG) bilaterally, a region neighboring primary auditory cortex (Binder 2000; see Hickok and Poeppel 2004 for a review).

Early neuroimaging studies using PET and fMRI universally showed more activity in STG (bilaterally) for listening to speech compared to rest (e.g. Petersen *et al.* 1988; Mazoyer *et al.* 1993; see Binder, 2000 for review). However, STG is a very large region that is likely to be involved in a huge number of different computations. The use of the basic speech-rest contrast did not allow researchers to distinguish between areas involved in non-specific auditory processing from areas involved in speech processing, or between areas involved in phonetic/phonological processing from areas involved in semantic and syntactic analysis. Subsequent imaging studies have contrasted speech with a variety of sophisticated conditions designed to isolate different aspects of auditory processing. For example, reversed speech is usually largely uninterpretable, but it maintains the same level of spectrotemporal complexity as speech, in that all the transitions are present but reversed. By comparing transformed conditions like reversed speech to auditory stimuli with less complexity (strings of tones) or less structure (noise), researchers have been able to show that bilateral regions of STG and the superior temporal sulcus (STS) bordering it are preferentially active for spectrotemporally complex stimuli like speech (Scott *et al.* 2000; Binder *et al.* 2000). Because these regions show increased activity even for speech-like stimuli that is semantically uninterpretable, it seems that they must be involved in a pre-lexical stage of processing.

The neuroimaging results fit in with the results of a long line of aphasia research showing that bilateral damage to STG results in a syndrome known as phonological word deafness (Barrett 1910; Henschen 1918). In this syndrome, early auditory processing is preserved – for example, patients can discriminate between tones of different frequencies and can recognize familiar non-speech sounds and music – but speech perception is severely damaged. Patients report that speech sounds like “meaningless noise” (Poeppel 2001) and frequently display phonemic paraphasias (errors in production) as well (Buchman *et al.* 1986). This pattern of deficits follows if STG supports both processing of speech and storage of the underlying representations.

So far this evidence is consistent with phonetic/phonological primitives being coded in STG/STS regions, but an alternative possibility is that these regions are involved simply in non-linguistic analysis of complex auditory signals. What evidence do we have that linguistic information is coded here? First, there is an argument from null results; phonetic/phonological representations presumably must be accessed in processing speech and speech-like stimuli, yet no other region consistently shows increased activity for speech and speech-like vs. non-speech stimuli. However, there are many reasons that neuroimaging studies might fail to show real brain activity – lack of power in neuroimaging studies is a constant concern – so this is a weak argument at best. Two lines of evidence are more compelling: work showing that STG is differentially active to

phonological contrasts as opposed to auditory contrasts matched for complexity, and work showing that this region is required for speech production as well.

First, many studies show that regions of STG are differentially sensitive to acoustic contrasts that are phonologically meaningful in the listener's native language. For example, a seminal fMRI study by Jacquemot and colleagues (2003) tested French speakers and Japanese speakers on sets of stimuli that differed acoustically along a dimension that was linguistically available in Japanese only (vowel length), or in French only (simple versus complex consonant strings). Behavioral evidence showed that French speakers had difficulty distinguishing stimuli differing on vowel length (e.g. *tokei* versus *tookei*) and Japanese speakers had difficulty distinguishing stimuli such as *ebza* versus *ebuza*, presumably because they perceive an epenthetic vowel in the former due to constraints on consonant form in Japanese. The fMRI data showed that activity in STG was associated with phonologically meaningful distinctions rather than acoustic differences; French speakers showed a significant increase in STG when the syllable form changed, and Japanese speakers showed a significant increase in STG when the vowel length changed. Similarly, the mismatch negativity in MEG, a response shown to be sensitive to phonological variables discussed in more detail below, consistently localizes to STG (Hickok and Poeppel 2004).

Second, there is significant evidence that STG is also accessed during production (see Indefrey and Levelt 2004 for a review). Conduction aphasia, a disorder characterized by phonemic paraphasias and naming difficulty, is associated with damage to left posterior STG (e.g., Damasio and Damasio 1980); and as alluded to above, subtle production difficulties are also often observed in pure word deafness. Neuroimaging studies have shown that regions of left STG are consistently activated in speech production tasks; this is the case even when production is covert (e.g. Hickok *et al.* 2000). The fact that part of this region is involved in both perception and production argues against an interpretation in which STG is only involved in processing complex sounds.

Therefore, there is sufficient evidence at this point to conclude that regions of STS/STG encode phonetic/phonological primitives, in addition to implementing numerous other computations required for processing complex spatiotemporal auditory signals. Although a larger discussion of those computations is beyond the scope of this chapter, it is interesting to note that left and right STG are sometimes thought to be specialized in speech perception at different levels of temporal analysis: left STG for phoneme-sized time-windows (20–40 ms) and right STG for syllable-sized time-windows (200–300 ms) (Poeppel 2001, 2003).

Finally, we note that so far there is not good evidence that any particular sub-region of STG is “speech-specific” in the sense that it only processes speech and not other similar input. In fact, it is not clear that we should expect to find such a sub-region, at least based on linguistic theory.

The only language-specific circuits motivated by classic phonetic and phonological theory are those needed to represent language-specific phonetic and phonological primitives in long-term memory. But these circuits may be very simple and non-centralized, and thus may encompass only a fraction of the machinery needed to process speech in real time. In the next section, we turn to existing evidence about the sub-organization of these long-term representations. While hemodynamic measures provide us with information about where to look, their temporal response properties are poor, especially given the quickly changing temporal dynamics of the speech signal. Electrophysiological data offer much better temporal resolution, and are known to robustly respond to activity in STS/STG.

### 13.5 How to look for phonological primitives

Electrophysiology (EEG, MEG) has proven to be an exceptionally useful tool for understanding the nature of auditory and speech representations. The early electrophysiological evoked components commonly associated with auditory and speech perception (N1/N1m/M100, N1-P2, MMN/MMNm/MMF) are pre-attentive and do not require a task, providing researchers with a task-independent probe into the early stages of processing. The automaticity of these components, combined with excellent, millisecond temporal resolution, makes the use of such methods extremely powerful in understanding the nature of linguistic and auditory representations and processes employed and entertained prior to contact with a lexical level of representation. An enormous amount of work has been done on not only understanding the response properties and the neurophysiological and cognitive sources of the MMN (Näätänen 1992; Näätänen, Jacobsen, and Winkler 2005; Näätänen *et al.* 2007), but also the nature of auditory and linguistic representations as indexed by the MMN (Näätänen 2001). In particular, MMN studies of speech perception have provided neurophysiological evidence for the existence of representations at the level of abstract phonology (e.g., phonemes, distinctive features) and shown sensitivity of native language phonological inventories (Näätänen *et al.* 1997; Winkler *et al.* 1999) and syllabic constraints (Dehaene-Lambertz, Dupoux, and Gout 2000).

The Mismatch Negativity (MMN; Mismatch Magnetic Field (MMF/MMNm) in MEG) is an electrophysiological component observed when there is a *discriminable physical change* within a series of auditorily presented standards that can be grouped based on some physical or psychological basis into a single category. In a typical MMN/MMF paradigm, participants are presented with a series of *standard* tokens interrupted by a *deviant*, which differs from the standard along some physical (or linguistic) dimension. The probability of hearing a deviant within a given experimental block is usually approximately 15%. If the deviant is perceived as being



perceptually distinct from the standards (the requisite physical difference between the standard and deviant is typically commensurate with behavioral discrimination thresholds for particular stimulus attributes (Näätänen *et al.* 2007)), then a large, negative-going waveform (in EEG; a larger magnetic field strength in MEG) in comparison with the electrophysiological response to the standard is observed approximately 150–300 ms post-onset of the deviant stimulus. The magnitude of the MMN elicited by the deviant is determined by subtracting the grand average waveform of the electrophysiological response to the standard from the grand average waveform of the electrophysiological response to the deviant (Näätänen 1992, 2001; Näätänen *et al.* 2007). Modulations of spectral properties of an auditory stimulus, such as the frequency, intensity, and duration have all been reported to reliably elicit an MMN/MMF (Gomes, Ritter, and Vaughan 1995; Sams *et al.* 1985; Winkler *et al.* 1990). Based on intracranial electrode recordings from monkey (Javitt *et al.* 1992; Javitt *et al.* 1994) and cat (Csépe, Karmos, and Molnár 1987), and MEG and EEG source modeling in humans (Hari *et al.* 1984; Alain, Cortese, and Picton 1998; Scherg, Vajsaar, and Picton 1989; see Näätänen and Alho 1995 for a review), the neural generators of the MMN/MMF component are located in the superior temporal plane in primary (or immediately adjacent to primary) auditory cortex, roughly 3–10 mm more anterior than the source of the N1m (M100; Näätänen *et al.* 2007). Moreover, its elicitation does not require attention on the part of the participant or active control modules. Instead, it can be elicited, for example, during sleep (M100; Näätänen *et al.*, 2007), and the magnitude and presence of the response are generally unaffected by attentional or task demands required of participants during the experimental procedure (Alho *et al.* 1998; Näätänen 1992; Ritter *et al.* 1992; Ritter and Ruchkin 1992). The Mismatch Negativity electrophysiological component has also been exploited extensively to probe the nature of speech representations and the impact of native language phonology on the perception of speech (see Näätänen 2001; Phillips 2001 for reviews). Research using the MMN in speech perception has demonstrated that this component reflects much higher levels of representation and processes than what can be inferred from the physical/acoustic attributes of the signal alone.

### 13.5.1 Vowels

Näätänen *et al.* (1997) assessed the extent to which one's native language vowel inventory affects elicitation of the MMN, and more generally, the early stages of perceptual processing of speech. They tested native speakers of Finnish and Estonian, two closely related languages with nearly identical vowel inventories. The primary difference between the two vowel inventories is that Estonian contains the vowel /õ/, while Finnish does not. Näätänen *et al.* (1997) synthesized vowel tokens corresponding to /e/, /õ/, /ō/, and /o/, whose fundamental acoustic difference is the frequency of their second

formant (F2). The semi-synthetic tokens were matched on their fundamental frequency (F0) and first (F1), third (F3), and fourth (F4) formants. In the MMN paradigm, there were two primary comparisons between the groups: (1) the response to the deviant synthetic vowel tokens (i.e., /*ö*/, /*õ*/, /*o*/) from the prototype standard /*e*/ synthetic vowel token and (2) the response to deviant sinusoids of the same frequencies of the prototype F2 values for the vowels /*ö*/, /*õ*/, and /*o*/ from the F2 value of the standard /*e*. Given the findings from Tiitinen *et al.* (1994), who found monotonic relationships between properties of the MMN and the distance of difference between the standard and deviant tokens, Näätänen *et al.* (1997) expected a monotonic rise in the magnitude of the MMN the further away the F2 value of the deviant was from the F2 value of the standard. Therefore, /*o*/ should elicit the largest MMN (because its F2 value is furthest from the standard's F2 value), while /*ö*/ should elicit the smallest MMN compared with the electrophysiological response to the standard /*e*. The critical comparison is the response to the deviant /*õ*. An MMN magnitude mid-way between /*ö*/ and /*o*/ should be elicited in the Estonians and not in the Finnish participants if the MMN is reflecting phonetic/phonemic processing. They found a monotonic increase across the three vowel tokens in the magnitude of the MMN for the Estonian participants but not for the Finnish, while both groups showed a monotonic rise to pure sinusoids whose frequencies matched the F2 formant frequencies of the vowels, suggesting that this difference is not solely attributable to the physical properties of the stimuli, but must also be a function of the phonemic vowel inventories (and therefore, presumably, the distinctive features) of the two groups of participants. Winkler and colleagues (1999) followed up on these results comparing Finnish and Hungarian participants with contrasts that were within-category in one language and across-category in the other. The range of the Finnish vowel /*e*/ occupies portions of the four-dimensional (F1-F4) vowel space occupied by /*é*/ and /*ɛ*/ in Hungarian. Meanwhile, the Finnish vowels /*e*/ and /*æ*/ are located in the region of vowel space occupied by /*ɛ*/ in Hungarian. Consequently, they synthesized a pair of vowel tokens that would be perceived as /*é*/ and /*ɛ*/ by Hungarian participants, but only as /*e*/ by Finnish participants. They also synthesized a pair of vowel tokens that would be perceived as /*e*/ and /*æ*/ by Finnish participants, but only as /*ɛ*/ by Hungarian participants. Winkler and colleagues (1999) report an MMN in all comparisons across groups. The MMNs in the across-category conditions, however, were significantly larger than the within-category condition in both groups of participants. While these findings are consistent with distinctive feature theories, they do not preclude purely segmental explanations.

### 13.5.2 Consonants

Sharma and Dorman (1999) used consonant-vowel (CV; /*ta*/-/*da*/) sequences to better understand the influence of phonetic categories on the MMN.

They found a minimal MMN in the within-category condition, but a significantly larger MMN in the across-category condition, despite the fact that the tokens were equally acoustically distant (20 ms VOT difference between each token in both the within- and across-category conditions). Subsequently, Sharma and Dorman (2000) compared the MMN responses of Hindi and American English speakers on a VOT contrast native to Hindi but absent in English. They manipulated the amount of pre-voicing duration to create a /pa/-/ba/ continuum between 0 and -90 ms VOT. In the MMN experiment, the standard was the -10 ms VOT stimulus and the deviant was the -50 ms VOT stimulus. As predicted, a large and reliable MMN beginning roughly 175 ms post-onset of the deviant stimulus was found in the Hindi participants, but absent in the English participants. Consistent with the findings from Näätänen *et al.* (1997) and Winkler *et al.* (1999), these results suggest that the MMN is sensitive to phonetic category distributions of the native language of the participants. The interpretation of these results are complicated by the fact that the stimuli were words in Hindi, allowing the Hindi participants to map the sounds onto lexical entries, but not for the English speakers. While it is clear that these studies report electrophysiological sensitivity to properties of the stimulus not reflected in their physical attributes (e.g., native language inventory, differential effects contingent upon category boundaries, etc.), it remains to be seen whether or not listeners are constructing phonological representations of the standards in these cases, or whether they are reflecting phonetic category-level distributions.

The experiments conducted up to that point could not distinguish between a phonological account from a purely acoustic or phonetic one. Using MEG in an oddball paradigm, Phillips and colleagues (2000) employed a many-to-one design (e.g., Aulanko *et al.* 1993; Gomes, Ritter, and Vaughan 1995; Winkler *et al.* 1990) in the discrimination of /dæ/ and /tæ/, which differ in the duration of voice onset time (VOT; the duration between the release of the stop closure on /t/ and /d/ and the onset of voicing in the vowel). This time instead of varying pitch (cf. Aulanko *et al.* 1993), which is not the primary acoustic/phonetic contrast between /b/ and /g/, Phillips *et al.* modulated the duration of voice onset time, which is a primary cue in the distinction between /d/ and /t/ (Liberman, Delattre, and Cooper 1958). The VOT duration for /d/ is of the range 0 ms to 25 ms and /t/ is of the range 30 ms to 105 ms (Lisker and Abramson 1964). In a many-to-one oddball design, there is a many-to-one relationship at the phonological but not acoustic level of representation (cf. Sharma and Dorman 1999). Phillips *et al.* synthesized a series of stimuli along the /dæ/-/tæ/ continuum which varied in the duration of the VOT in 8 ms increments. In the first half of the experiment, 87.5% of the acoustic tokens were randomly sampled from one side of the category boundary continuum (12.5% from the continuum of the other side of the category boundary), and in the second half of the experiment, the majority of acoustic tokens (87.5% again)

were randomly sampled from the other side of the category boundary. Consequently, at the level of acoustic representation, there was no standard, as each successive stimulus presentation was acoustically distinct from the token that either preceded or followed. Phillips and colleagues reported a reliable MMN to the deviant, suggesting that listeners were able to construct a category representation at the phonological level for the standard despite the acoustic variation in the individual tokens. That is, listeners seemed to be able to perceptually group these acoustically distinct tokens together to form a category, and when they perceived a token from the other side of the category boundary, they were able to detect the change (as indexed by the MMN).

To demonstrate that listeners were constructing representations consistent with their linguistic representations and not simply grouping the tokens based on “long” or “short” VOTs, Phillips and colleagues (2000) conducted a follow-up whereby 20 ms of VOT were added to all the VOT values, such that now all the tokens had “long” VOTs, and there was no longer a many-to-one relationship at the phonological category level. They found no MMN, suggesting that in the earlier experiment, listeners were in fact grouping the standards together in a manner consistent with their linguistic categories. These results do not, however, necessarily point to a phonological explanation over a phonetic category explanation. In exemplar models (e.g. Pierrehumbert 2002), phonetic representations reflect Gaussian distributions along a number of acoustic phonetic parameters. Consequently, perception of a token that is sampled from a distinct phonetic category could give rise to the MMN reported in Phillips *et al.* (2000). An additional, alternative explanation is one based entirely on neurophysiology. The categorical boundary in English stop consonants is roughly 30 ms VOT. The idea is that this VOT boundary is a consequence of auditory neuron response properties to independent acoustic events occurring in quick succession (Sinex and McDonald 1988, 1989; Steinschneider *et al.* 1995). Certain groups of auditory neurons respond to both the noise burst of the stop consonant and the onset of the voicing of the vowel. The refractory period for some of these neurons is roughly 30 ms, which lines up well with the typically cross-linguistically observed VOT durations.

One of the core definitional properties of being phonological in nature is the direct relationship to meaning (Halle 2002), however. The phoneme, a unit of representation undeniably phonological, is traditionally defined as the smallest unit of linguistic representation that can serve to distinguish lexical meaning. Consequently, it is important to compare contrasts that do and do not serve to distinguish lexical meanings in different languages. Kazanina and colleagues (2006) compared speakers from two languages in a mismatch experiment nearly identical to Phillips *et al.* (2000). The novelty of this experiment is that Kazanina *et al.* tested both Russian and Korean speakers on the /d/-/t/ continuum, speakers of languages which differ in

the phonemic status assigned to /t/ and /d/. In Russian, both /t/ and /d/ have phonemic status: [tom] “volume” and [dom] “house” are two distinct lexical entries. In Korean, however, /t/ and /d/ share an allophonic relationship and appear in complementary distribution. The voiced allophone /d/ occurs intervocally (e.g., /paTa/ → [pada] “sea”), whereas the voiceless unaspirated counterpart occurs word-initially (e.g., /Tarimi/ → [tarimi] “iron”). Thus, both languages share a bimodal distribution of /d/ and /t/ at the phonetic level. The expectation, then, is that in a many-to-one oddball design, if participants are constructing a purely phonological representation of the standard, then we expect to find an MMN in the Russian participants but not in the Korean participants, because these sounds are represented independently at the level of phonology in Russian but not in Korean. They reported a reliable MMN for the Russian participants, and no reliable difference in the RMS of the MEG temporal waveform for the Korean participants in any time window between 20 ms and 340 ms. Unlike the findings from Phillips *et al.* (2000), the results in Kazanina *et al.* (2006) are considerably more difficult to explain by appealing to phonetic category distributions alone. Given that both [t] and [d] occur phonetically in Korean and Russian, if participants were simply constructing a phonetic representation of the standard, then both the Korean and Russian participants should show an MMF to deviant stimuli. One caveat, however, is that given that Korean listeners rarely, if ever, hear [d] word-initially, their failure to elicit an MMF to the deviant tokens in the experiment could be explained by the fact that [d] never occurs word-initially in Korean, and therefore, word-initial [d] does not exist in the phonetic distribution in the language (see Silverman 2006 for the idea that allophonic variants are stored together with their contextual licensing environments). The fact that allophones occur in complementary distributions makes it difficult to design a completely balanced experiment.

Gomes and colleagues (1995) suggest that listeners can perceptually group standards in an oddball MMN design along one particular physical dimension that all the standards share. If this interpretation is correct, then the MMN is an excellent tool for investigating the representative nature of distinctive features. In the only experiment of which we are aware, Yeung and Phillips (2004) asked if participants would be able to perceptually group relatively disparate standards sharing one distinctive feature (i.e., [+voice]). In 37.5% of the trials, they heard /bæ/, and in another 37.5% of the trials they heard /gæ/. In 12.5% of the trials they heard the pseudo-deviant /dæ/, and in the final 12.5% of the trials they heard the deviant /tæ/. The consonants /b/, /d/ and /g/ all form a natural class: voiced stop consonants. While the consonant /t/ is also a stop, it is produced without vibration of the vocal folds in the glottis. Therefore, in an oddball paradigm, if listeners can construct representations of the standard at the level of the distinctive feature, then they predict to find an MMF to the /tæ/ syllable and not the /dæ/ syllable even though their

likelihood of occurrence is identical. However, this design is heavily dependent on the exact details of the feature theory. For example, /b/ and /g/ can form a natural class to the exclusion of /d/ if assuming a feature such as [peripheral] or [grave] (Jakobson, Fant and Halle 1952). If the standard is then [+voice]; [+grave], then both /d/ and /t/ would be appropriate deviants. Ultimately, Yeung and Phillips found significant effects in the mismatch region between standards and deviants and also failed to find a difference between the standard and pseudo-deviants in any region. Curiously, however, they also found an effect in the mismatch time window for the acoustic condition (cf. Phillips *et al.* 2000). This result makes these findings difficult to interpret, but highlight quite nicely the power of the MMN/MMF paradigm in trying to assess the representational nature of speech sounds. More studies using this design are encouraged.

### 13.5.3 Underspecification

Investigating a perhaps more nuanced linguistically motivated hypothesis, Eulitz and Lahiri (2004) used the MMN to test whether phonemic representations in the lexicon are underspecified for non-contrastive distinctive feature values in the language. They used the German vowels /e/, /ø/ and /o/ in an oddball paradigm with German participants. It has been assumed that the feature [coronal] is not specified in the phonological lexicon (Archangeli 1988; Lahiri and Reetz 2002). Under this hypothesis, then the vowel /e/ is underspecified for its place of articulation in the phonological lexicon, while /o/ is specified for both [dorsal] and [labial], since it is both round (i.e., [labial]) and back (i.e., [dorsal]). Given that /ø/ is both front and round, it is specified for [labial] but underspecified for [coronal]. The comparison of interest lies in the /o/-/ø/ pair. When /o/ is the standard and /ø/ is the deviant, a conflict at the level of phonological representation occurs. This is because the [coronal] feature extracted from the auditory signal of /ø/ mismatches with the stored representation of [dorsal] for the standard /o/. A contrast should not occur in the opposite direction because [coronal] is omitted from the lexical representation. If /ø/ is underspecified for its place of articulation, then the constructed representation of the standard does not contain a specification for place, and therefore, the specified [dorsal] feature on /o/ would not conflict. For the /e/-/ø/ pair, since neither is specified for place of articulation, no conflict should exist at the level of phonological representation. Therefore, they predict a larger MMN when /o/ is the standard and /ø/ is the deviant compared to when /ø/ is the standard and /o/ is the deviant. Moreover, they predicted no difference in the /e/-/ø/ pair. They found a clear MMN component in the grand average waveform for all conditions. There was no difference in the latency or amplitude of the MMN for the /e/-/ø/ pair. That is, an approximately equivalent MMN was elicited irrespective of

which phoneme was the standard and which phoneme was the deviant. They did, however, find a differential MMN in the /ø/-/o/ pair: a larger and earlier MMN when /o/ was the standard and /ø/ was the deviant than in the opposite configuration. That is, despite the fact that the acoustic difference is identical, a larger and earlier MMN is elicited in one standard/deviant configuration than the other, suggesting that the MMN is indexing more than just the physical properties of the stimulus. Eulitz and Lahiri (2004) suggest that these findings support the predictions of a featurally underspecified lexicon model (Lahiri and Reetz 2002, 2010), whereby some features, those that do not play a contrastive role in lexical representation, are not phonologically specified.

More recently, Hacquard, Walter, and Marantz (2007) exploited the MMF to investigate the role of vowel inventory and size on the perception of vowels. The size of a vowel inventory within a given language influences the acoustic consequences of articulation. In particular, it has been reported that languages with larger vowel inventories also tend to have a larger acoustic vowel space relative to languages with smaller vowel inventories (Bradlow 1995). To understand the influence of vowel inventory size and organization on perception, Hacquard *et al.* (2007) tested native speakers of Spanish and French in an oddball MMF paradigm. The vowel spaces of Spanish and French differ on both their size and organization: (1) Spanish is a five-vowel system, while French has 12 vowels (including the five vowels found in Spanish) and (2) French has a series of vowels intervening in F2/F1 space between the vowels of Spanish (e.g., /ɛ/ intervenes between /e/ and /a/; /ɔ/ intervenes between /o/ and /a/). They hypothesized that if inventory organization had an effect on perception, then the effect of the MMF should be roughly equivalent for the /o/-/u/ pair as the /o/-/a/ pair in Spanish. That is because these two pairs are equally distant in terms of intervening vowels. In French, however, since /ɔ/ intervenes between /o/ and /a/, they predicted the MMF to be larger in the /o/-/a/ pair than in the /o/-/u/ pair if inventory size played a role. Across languages, if inventory organization was the primary factor in perception, no difference would be predicted in the magnitude of the MMF for the /o/-/u/ pair between the French and Spanish pair, while a difference would be predicted between the /o/-/a/ pair, since French has an intervening vowel category and Spanish does not. If, on the other hand, inventory size was the primary factor driving the MMF response, Hacquard *et al.* expected the French participants to show a larger MMF across the board, since French has the larger vowel inventory. What they found was that a larger MMF was elicited in French participants compared to Spanish participants for all vowel comparisons except the /ɛ/-/e/ pair, suggesting that vowel inventory size (the number of vowel categories in the language) and not inventory organization affects the perception of vowels. Hacquard *et al.* took this particular pattern of results to support a model of expanding vowel space whereby point vowels (e.g., vowels on the edge of the vowel space) are

produced more distinctly from one another than vowels more centrally located in the perceptual space.

In sum, the MMN/MMF has proven to be an extremely powerful tool in assessing the types of auditory and linguistic representations supported by auditory cortex. While the ultimate focus of these studies was to investigate the available representations supported by auditory cortex, as well as properties of phonetic category distributions and native language inventories, they also serve to demonstrate that the MMN indexes abstract properties of the stimulus. It should also be noted that the MMN oddball paradigm has been used to investigate the role of phonological constraints on syllable structure in native and non-native speech perception (Dehaene-Lambertz, Dupoux, and Gout 2000), the nature of lexical access (Assadollahi and Pulvermüller, 2003; Shtyrov and Pulvermüller, 2002a; 2002b) and certain aspects of syntactic processing (Pulvermüller and Shtyrov, 2006; Shtyrov *et al.* 2003). Designing experiments geared toward directly testing the status of phonological features, *per se*, remains challenging.

### 13.6 Phonotopy

The most intuitive neural coding scheme in any domain of cognition is a topographical one, the most famous example being retinotopy in vision (Palmer 1999). The neurons in primary visual cortex are mapped in such a way that brain space corresponds to retinal space. For example, two neurons in neighboring areas of cortex each represent neighboring parts of the retina. This kind of coding is known as place coding, because the significance of a particular neuron firing can be straightforwardly related to the region of cortex in which it is located. Place coding is probably the easiest neural coding scheme to detect with non-invasive neuroimaging techniques, and we can look for place coding in any domain in which we believe the representations to be ordered along some kind of semi-continuous stimulus dimension. In the auditory domain, we know that receptors in the cochlea are organized “tonotopically,” *i.e.*, along a frequency gradient (Schnupp, Nelkin, and King, 2010). Tonotopic mapping is preserved in the projection to auditory cortex (Formisano *et al.* 2003; Talavage *et al.* 2004; Langers *et al.* 2007; Humphries *et al.* 2010) both in individual neurons and in aggregations of neurons. Using MEG it has been shown that the estimated location of the source of the M100 auditory response to a tone depends on the tone’s pitch along a medial-lateral dimension (Pantev *et al.* 1989; Huotilainen *et al.* 1995; Langner *et al.* 1997).

Since the neural representation of basic acoustic properties appears to use place coding, a natural question is whether this coding continues to be used as the representations to be encoded become more complex and speech-specific. Phonemic representations have often been defined by virtue of their position along a number of proposed featural dimensions,



most notably the vertical and horizontal position of articulators within the mouth. Might phonemic representations therefore be subject to place-coding as well? Many studies have examined this question in recent years, mainly by using MEG to localize the M100 response to vowel or syllable presentation. The most consistent finding across these studies is that early MEG responses localize along an anterior–posterior gradient in STG according to the place of articulation on the front–back dimension (e.g., Diesch *et al.* 1996; 2000; Obleser *et al.* 2003; 2004; Scharinger *et al.* 2010). For example, Obleser *et al.* (2004) examined the M100 response to German vowels [a], [e], [i], [o], [u], [ø], and [y]. They showed that the M100 response to front/coronal vowels such as [i], [y], [ø] and [e] localized to a more anterior position than the response to back/dorsal vowels such as [u] and [o]. Convergent results have been observed in fMRI (Obleser *et al.* 2006).

Although these kinds of results show that the location of neural responses tracks featural dimensions that have been proposed by phonological theory, a similar pattern of results could arise from a place code that is purely acoustic in nature. Speech sounds are spectrally complex, of course, and featural dimensions proposed on articulatory grounds have acoustic correlates in the form of frequency shifts across primary formants like F1 and F2. In this way, what looks like “phonotopy” may simply be multidimensional “tonotopy.” In order to show that the spatial selectivity observed reflects the encoding of *linguistic* representations, one must demonstrate that the topographical pattern of activity cannot be explained by acoustic parameters alone. Several of these studies have made progress towards this by showing that a model including phonological features achieves a better fit to the observed localizations than a purely acoustic model (Diesch *et al.* 1996; Obleser *et al.* 2004). For example, Scharinger *et al.* (2011) mapped the M100 response for the entire vowel space of Turkish, and observed spatial gradients that could capture the featural dimensions of height, place of articulation, and roundedness. However, the map for front vowels was oriented orthogonally to the map for back vowels. The authors show with statistical modeling that while the cortical maps reflect acoustic properties of the signal, the map is warped towards linguistically relevant categories. This suggests that place-based encoding of language-specific phonemic representations may develop on the basis of experience from “tonotopic” encodings that are originally purely acoustic. More practically, this study indicates that future experiments in other languages may need to more exhaustively probe their vowel spaces in order to correctly interpret how the topographical pattern maps onto the linguistically relevant feature dimensions.

### 13.7 Phonological processes

To this point, we have primarily focused on the nature of phonological representations, the primitives that have been investigated to date in

the cognitive neuroscience literature. The other half of phonology has concerned itself with the processes that operate over these representations. Far less work has been conducted on the cognitive neuroscience of phonological processes, and consequently, we have less to say about these issues. They do, however, remain ripe for further investigation. Electrophysiological techniques have been used to investigate the neural time course of violations of phonological rules/phonotactic restrictions (Flagg, Oram Cardy, and Roberts 2006; Mitterer and Blomert, 2003; Tavabi *et al.* 2009). For example, Flagg and colleagues (2006) exploited the fact that pre-nasal vowels are nasalized in English and measured the electrophysiological latencies of the response peaks in MEG to congruent (i.e., [aba], [ãma]) and incongruent (i.e., [ãba], [ama]) VCV sequences. Overall, latencies in the time-window of 50–100 ms post-onset of the consonant were shorter for the congruent as opposed to incongruent sound sequences.

In an oddball detection experiment, Mitterer and Blomert (2003) found in Dutch listeners that violations of expected nasal place assimilation patterns between words elicited a larger mismatch negativity response (MMN, approximately 100–200 ms after onset of the consonant) than did adherence to nasal place assimilation patterns. These results suggest that although cortical responses that reflect phonological processing are early, the types of responses measured (latency versus amplitude) and the time-window in which differences were found varied.

In recent MEG work (Monahan, Hwang and Idsardi 2008) based on previous behavioral findings (Hwang, Monahan, and Idsardi 2010), we found a reliable difference between consonant clusters that were congruent in their voicing specification (i.e., [utz]) with those that were incongruent (i.e., [uds]) as early as 150 ms post-onset of the violating segment, in this case, the fricative (i.e., [s]). We interpreted these differences between to suggest that listeners can exploit their knowledge of phonological processes and representation to constrain early perceptual parses of the sensory input, as well as using this detailed knowledge to serve as the basis for generating hypotheses and predictions about the nature of the upcoming speech signal (see Poeppel and Monahan 2011 for more discussion on how we believe these results fit into a larger, analysis-by-synthesis architecture).

### 13.8 Conclusions

Although a vast amount of work has examined the biological basis of speech perception and speech production, a critical missing piece has been our lack of knowledge about the biological basis of the phonological representations that map between lower-level sensory and motor representations to more abstract lexical representations. We believe that

linguists have an important role to play in the effort to better understand the neural implementation of phonological knowledge, and that phonology may in fact be a more fruitful starting place than syntax for the general biolinguistic enterprise. Here we have reviewed some recent work that makes some beginning steps forward in this direction and hope that in doing so we have encouraged others to continue to expand this effort.

## Appendix: Cognitive neuroscience methodologies

Acceptability judgments and various fieldwork methods have been the preferred methods for the field of phonology over the past several decades. In the domain of behavioral psycholinguistics, a variety of tasks and methodologies have been employed to understand the psychological processes that underlie spoken word recognition (see Grosjean and Frauenfelder 1996 for a collection of papers on the various techniques in the field), and by and large, the primary dependent measures in those tasks are reaction time data, accuracy, or normalized discriminability scores (e.g.,  $d'$ ; see Macmillan and Creelman, 1991 for review). While we advocate the position that no type of measure or technique is inherently better than any other and that an interpretable answer to any problem requires a combination of the right question with the appropriate technique, the techniques of cognitive neuroscience may remain less familiar than others to some readers of this volume. Therefore in the following we briefly introduce the primary techniques used in this field.

### A.1 Aphasias and patient-work

#### Overview

Neuropsychological studies of patients with language deficits are one important source of evidence. In these studies, investigators develop tasks to carefully probe those aspects of linguistic knowledge or processing that are relatively preserved and those that are not. If one observes that performance on two seemingly independent tasks are typically correlated in patients, one can conclude that the same brain area is likely responsible for task performance, even before determining which brain area is recruited; conversely, a dissociation in performance implicates a dissociation in regions responsible for performing those tasks. Much of the early work proceeded in this way, with investigators identifying deficits that tended to cluster together or dissociate. Given the current prevalence of MRI, most neuropsychological studies additionally include measures that attempt to determine exactly which brain area is responsible for the deficits. Structural MRIs of numerous patients presenting with various patterns of linguistic deficits can be conducted, lesions can be identified

in each, and the number of lesions in a particular area can be correlated with behavioral measures in a method known as voxel-based lesion mapping (e.g., Dronkers *et al.* 2004). Perhaps the key benefit of neuropsychological studies in understanding language in the brain is that they permit at least limited conclusions about causality. With measures of brain activity, the absence of observed activity in an area could always be due to the insensitivity of the technique; the presence of observed activity could reflect some process correlated with the process of interest. However, with patient studies, one can conclude that a particular brain area is in some way crucial for performing a task because when that area is damaged or not functioning, the patient simply cannot do the task.

### Practical issues

Neuropsychological studies require a very good theory of the processes and knowledge required for performing the tasks that compose the dependent measure (see Hickok and Poeppel 2004 for a discussion of how misunderstanding of the task may have led to incorrect conclusions about the brain areas involved in phonological processing). The small sample sizes and the heterogeneity of the brain areas damaged and the deficits observed across patients frequently make it difficult to extrapolate findings to the general population.

## A.2 Electroencephalography (EEG)

### Overview

Perhaps the most widely used measure of brain activity, EEG is typically conducted with an array of electrodes attached to a cap that is placed on the scalp. Salt water or water-soluble gel is used to improve the connection between the scalp and the electrodes, and then the voltage between each electrode and a reference electrode is recorded. Momentary neural events like action potentials (~1 ms) or extremely local changes in activity (across a few neurons) are unlikely to sum to a change in potential large enough to be observed at the scalp. Therefore, changes in the EEG are thought to mainly reflect modulatory activity across entire neuronal populations in rough synchrony (Nunez 1981; Luck 2005). In cognitive studies, one is often interested in the brain response to stimuli (events) presented at particular points in time, so one might present many events and average the EEG in the time-window following each event; the resulting waveform is referred to as an event-related potential (ERP), and studies that focus on changes in the average EEG response to a stimulus are therefore often referred to as ERP studies. As a measure for probing phonological processing, EEG has a number of benefits. In contrast to behavioral measures like reaction times, an ERP study does not require an explicit task, which can alleviate concerns of invoking unnatural processing strategies. Moreover, the time-sensitivity of the technique is excellent. With sufficient statistical

power, one can detect differences on the order of a few milliseconds, or more commonly, a few tens of milliseconds, and therefore, it is one of the best techniques available for investigating questions about the time course of processing.

### Practical issues

Setting up an EEG lab is relatively inexpensive (~ \$30,000), and the only continuing costs are minor supplies for set-up and replacing the caps (\$500–\$1000) that hold the electrodes and are worn by participants after several hundred uses. The analysis methods, while requiring some training, are easier to grasp and are less computationally intensive than those for fMRI, and are more standardized (and therefore easier to build on previous results and to get published) than those for MEG. Therefore, EEG can be a good choice for new investigators. A well-known weakness of EEG is that it does not provide reliable information about the location of changes in neural activity. This is because the conductive layers of skull and tissue between the brain and the electrodes tend to smear electrical activity. Some degree of source localization can be achieved if one records from a greater number of electrodes, but these methods are still under development and replication of results has been challenging.

## A.3 Magnetoencephalography (MEG)

### Overview

MEG measures fluctuations in magnetic field strength recorded externally from the scalp. Like all electrical signals, the changing electrical current measured by EEG is accompanied by magnetic fields. Thus, MEG captures the same kind of synchronous modulatory neural activity, and as a consequence, it has the same excellent temporal resolution; however, the relationship of the relative orientation of the source of the electromagnetic potentials/fields to the scalp with the optimal orientation for the sensitivity of the techniques is slightly different (Hansen, Kringelbach, and Salmelin 2010). While EEG records electrical activity oriented in any direction, MEG is blind to activity in certain orientations, in particular magnetic dipoles that are radial to the surface of the scalp. For the same reason, MEG is less sensitive to activity deeper in the brain, in subcortical areas. Although this seems to suggest that EEG “sees more” than MEG, the fact that more activity is visible to EEG also means that EEG is more susceptible to cancellation between sources (activity in two areas simultaneously with opposite orientation will sum to zero at the scalp). The most well-known benefit of MEG compared to EEG is that it provides better information about the location of neural activity because, unlike electrical fields, magnetic fields are not distorted by intervening tissue between the electric activity and sensors/electrodes. For example, even without

sophisticated source localization methods, in MEG it is often possible to reasonably conclude from observing greater activity in left hemisphere sensors that there is greater neural activity in the left hemisphere.

### Practical issues

MEG setup is typically much faster and more comfortable for participants, and requires less training for the experimenters. In MEG the magnetic sensors are arranged in a fixed helmet, because they must be constantly cooled by liquid helium. Unlike EEG, they do not need to be physically attached to the participant, and there is no need to create a conductive bridge between the sensor and the scalp. A major downside of MEG is its cost; start-up can be as much as \$4,000,000, in large part due to the costs associated with the magnetically shielded room, and the liquid helium required to maintain the machine is approximately \$50,000 annually. Analysis is much less standardized than in EEG, and therefore often requires a much greater time investment for the beginning investigator who may need to develop individualized analysis procedures. Source localization, while more accurate than for EEG, is still a computationally ill-formed problem with an infinite number of solutions, and it is still largely unknown how well current methods do in cases in which multiple sources are active simultaneously. Collecting a separate structural MRI is also necessary for accurate source localization in EEG or MEG, which further increases cost and complexity. However, it is important to remember that although MEG is often touted on the basis of its improved localization, MEG measurements can still be very informative about the time course of processing even without precise source localization analyses.

## A.4 Functional magnetic resonance imaging (fMRI)

### Overview

fMRI is the most widely available tool for localizing neural activity to specific brain regions. Different types of biological tissue and substances differ in magnetic susceptibility, and MRI can use these differences to create maps that highlight the location of the tissue or substance of interest. Functional MRI typically measures the blood-oxygen-level dependent (BOLD) signal, the change in magnetic susceptibility that arises when the blood oxygen level changes. It is known that local neural activity results in systematic changes in blood oxygenation, and therefore changes in BOLD in a given brain region are often taken as a proxy for changes in neural activity in that region. However, one limitation of fMRI is that the exact nature of the relationship between BOLD and neural firing is still not completely understood (see Huettel *et al.* 2004 for an introduction to this issue), and many factors such as the proximity of large blood vessels and excitatory versus inhibitory firing may complicate the interpretation of

BOLD changes. Because blood flow is relatively slow, the peak of a stimulus-evoked increase in BOLD occurs about 6 seconds after the item is presented, and temporal precision is poor for the same reason, so that fMRI is generally not a good method for estimating the time-course of processing. However, no other non-invasive technique can currently match fMRI in spatial precision. Unlike MEG source localization, which is only as accurate as the model, fMRI delivers an unambiguous measurement of magnetic susceptibility for each unit of space in the measurement. This means that fMRI is often still the best choice for addressing “where” questions.

### Practical issues

In the past, one particular challenge for fMRI work on phonology has been that the shifting magnetic gradients that are required for imaging are very loud, thus inducing noise and possible confounds in studies using auditory stimuli. Studies of phonological production are similarly challenging because changes in articulator position can result in time-locked artifacts. However, clever imaging sequences are now available in which acquisition is delayed until several seconds after stimulus presentation or production, thus mitigating some of these problems. Because of both cost and the safety issues involved in working in an environment with a powerful magnet, most MRI machines are maintained by large institutions rather than individual laboratories. At the same time, there are now so many MRI machines for medical use that researchers in developed countries can usually find one nearby that they can arrange to use during off-hours. Although the cost per hour for MRI use is often high (~\$500), many institutions will allow researchers with a well-developed proposal to collect pilot data at little or no cost, which can be used to motivate grant proposals that would allow completion of the project. fMRI data analysis is computationally intensive and less intuitive than ERP analysis, but it is currently more standardized and better documented than MEG analysis.

### A.5 Alternative methods

Several other techniques have also been used to probe brain activity associated with phonological processing. Positron Emission Tomography (PET) can measure brain metabolism of different compounds through the use of radioactive tracers; for example one could measure the rate of glucose metabolism in different brain areas while participants perform cognitive tasks in order to determine which areas become more active for which tasks. Some benefits of PET for phonological studies are that it provides fairly good spatial precision but is not noisy to acquire like fMRI and is not disrupted by articulator movement. Some drawbacks are that only block designs can be used with PET (as it takes a number of minutes for the tracer to leave the system), and high cost, technical difficulty in

working with short-lived radioactive tracers, and radiation exposure to participants. Near Infrared Spectroscopy (NIRS) uses infrared light to measure changes in blood oxygenation associated with neural activity. It is less widely available than fMRI but is beginning to be used more frequently for studies of processing in infants, because of safety concerns in exposing infants to strong magnetic fields, the reduced acoustic noise compared to fMRI, and less susceptibility to movement artifact. Event-related optical signaling (EROS) is an even newer method that also uses infrared light but measures changes in the optical properties of active neurons themselves, thus increasing temporal resolution.