Neurophysiological techniques in Laboratory Phonology

1. Neurophysiological and neuroimaging techniques: general advantages and disadvantages

The appeal of using the – now ubiquitous – noninvasive recording techniques lies in the remarkable scope of new (types of) data that can be gathered in the study of speech and language processing. Until the late 1980s, the principal evidence regarding neuronal implementation of speech derived from patient studies (neuropsychological deficit-lesion correlations). These studies are theoretically well-motivated and experimentally nuanced, but the neuronal information is necessarily restricted to spatial information (lesion location), and even those data remain rather coarse (on the order of centimeters), with no meaningful temporal information. The aggressive development of non-invasive recording using both electrophysiological (EEG, MEG) and hemodynamic (fMRI, PET, NIRS) recording techniques has made available a wealth of new data, at spatial and temporal resolutions that are permitting researchers to develop increasingly sophisticated hypotheses about the neuronal basis of perception and cognition. Notwithstanding the excitement surrounding the widespread availability of these new empirical approaches, robust skepticism is indicated. It is, indeed, not always obvious precisely in which way such data can enrich our understanding of speech and language processing. There is great promise, but the hurdles are significant.

Phillips (2001) provides an excellent review of the many issues in applying neuro-physiological techniques to the investigation of speech perception. As outlined there, researchers face an enormous number of issues in unpacking the mapping from the input (a continuous acoustic waveform) to the ultimate language percepts (features, sounds, words, sentences, etc.). For each available experimental technology, we face significant challenges at every level of analysis. One major advantage of neurophysiological techniques is that they offer measures that are related to, but separately collected from, the usual behavioral measures such as response accuracy and response reaction times, and thus provide quasi-independent verification of those measures. In addition, for some of the measures (EEG, MEG) the observable brain responses are hundreds of milliseconds earlier than the behavioral responses (which typically occur later than 500ms). The advantage gained by the ability to observe earlier responses is that we can begin to potentially dissect the complex computations involved in tasks such as word recognition. A response at 100ms is unlikely to be directly modulated by long-term memory representations (for example the lexicon) and thus provides evidence for the nature of phonetic and phonological processing that occurs in advance of lexical access.

The first portion of this chapter briefly reviews some of the available technologies. The second part reviews some studies linking brain measures with quantities of interest to LabPhon researchers. We focus here nearly exclusively on the use of these techniques for studying speech perception, although it is also possible to study some neural aspects of speech production with some of these
techniques as well. We conclude with some pragmatic recommendations based on the currently available technologies.

1.1 Available Technologies

The available non-invasive recording techniques trade off spatial resolution, temporal resolution, directness of the measure of brain activity (i.e. recording neuronal activity directly versus quantifying a proxy for neuronal activity), invasiveness, convenience and cost. Here we consider only gross characteristics of passive recording techniques; those based on blood flow (fMRI, PET, NIRS) are inherently slower than those based on the electromagnetic fields generated by neural activity (EEG, MEG). Simultaneous detailed resolution in time and space is possible only with intra-cranial recording, included here for comparison only as this technique is possible only with surgical populations (see Besle et al 2008 and Boatman 2004 for examples of its use in investigating speech perception). The trade-off is then fairly direct—time versus space. The electromagnetic measures (EEG, MEG) have millisecond-level temporal resolution, but resolve brain areas relatively poorly (though well enough to distinguish visual cortex from auditory cortex, say, or to detect hemispheric differences). The blood-flow techniques can resolve well spatially within specific brain regions, but at the expense of temporal resolution. Given the rapidity of many phonetic changes, resolution at the level of 1-10 seconds will not impress LabPhon researchers accustomed to 44.1kHz sampling rates for audio.

There are also active techniques such as surgical or pharmacological interventions which involve momentarily or permanently changing brain function; the only one considered here is transcranial magnetic stimulation (TMS, section 1.7). As well, we also disregard lesion studies and post-surgical populations. Pictures of the various machines and representative data plots can be found in the entries on Wikipedia (e.g. http://en.wikipedia.org/wiki/Electroencephalography).

In summary, for LabPhon researchers, the space-time trade-off is the most relevant one, and the following table provides a simple (if crude) comparison of the space and time characteristics of the major technologies.

---INSERT TABLE 1 HERE---

1.2 Electroencephalography (EEG)

Electroencephalography (EEG) directly measures the electrical mass activity, using from one to hundreds of sensors attached to the scalp, generated by the coordinated activity of large groups of neurons in the brain ($10^5$-$10^6$ neurons). The source of the signal is hypothesized to be the post-synaptic current flow in the apical dendrites of pyramidal cells in cortex (not action potentials). EEG has a number of significant advantages. Having been in use for more than a century (Schwarz 1998), the equipment is now tried and true, and in this group of technologies, relatively inexpensive. Portable EEG systems are available, and this technique does not require extensive magnetic shielding to prevent interference from nearby electrical
equipment. The sensors must be placed on the scalp; to better increase the conductance to the sensors, they are enhanced with special gel or saline solutions (which does make the process somewhat messy). Although there is considerable debate on the ideal number and types of sensors, in practice researchers often report single channel data (when focusing on temporal aspects of the recorded signal), or data averaged over groups of sensors, rendering moot many of the theoretical issues (especially those involving localization). It is possible to combine EEG with fMRI to improve the spatial localization (Bagshaw et al 2006) either with simultaneous recording (obviously requiring much more expensive equipment) or by using separately collected fMRI scans to constrain the EEG localization solutions.

There are a number of ways to use EEG to get reliable brain measures. One of the most common is to measure the brain response arising from particular events controlled by the experimenter—event related potentials (ERPs, Luck 2005), a particularly common technique in many psycholinguistic experiments. Because of the large amount of uncorrelated electrical noise in individual trials, the responses to a series of replicated trials are averaged together to increase the signal to noise ratio (which may seem unusual to LabPhon researchers used to analyzing individual phonetic utterances). Other measures include steady state responses such as the auditory steady state response (aSSR, Burkard 2009) which will entrain to certain amplitude or frequency-modulated frequencies in auditory stimuli and other changes in the endogenous brain rhythms (Buzsaki 2006). A particularly useful technique is mismatch negativity (MMN, Näätänen et al 2007, Pulvermüller & Shtyrov 2006), an automatic response when an “oddball” is detected within a series of “standards”. The flexibility of MMN designs makes it a very useful technique, although the necessarily high ratio of standards to deviants (7:1 or more) makes this technique less efficient than ERP-style evoked responses.

1.3 Magnetoencephalography (MEG)

Magnetoencephalography (MEG, Hämäläinen et al 1993) measures the magnetic field generated by the electrical activity of the brain from just outside the scalp, also using from one to hundreds of sensors. The equipment is much more expensive than EEG, both in the initial cost and in the ongoing supply costs (which include liquid helium to cool the super-conducting sensors). The measures obtained in MEG are by and large analogous to those obtained with EEG: they can be event-related, steady-state, related to endogenous rhythms or mismatch fields. At the detailed level of the generation of electrical signals in the brain there are differences between the techniques, but for many high-level cognition experiments these differences are not readily apparent in the measured signals. The main advantage of MEG is the relative simplicity of the source localization algorithms (due to the relative magnetic transparency of the intervening materials—brain, dura, skull, scalp, hair and air), which allows all of the temporal resolution of EEG plus reasonable spatial resolution (~ 1 cm²), especially when combined with structural magnetic resonance scans. However, because the magnetic fields are so tiny (about 1 billionth the strength of the Earth’s magnetic field, less than 500 femtoTesla) the equipment requires extensive magnetic shielding and, as a practical matter for experiment design, a
larger number of replicated trials to achieve reasonable signal to noise ratios. How MEG is used in studies of audition and speech perception is reviewed in detail by Lütkenhöner & Poeppel (2010).

1.4 Functional Magnetic Resonance Imaging (fMRI)

Magnetic resonance imaging (MRI) is akin to a very fancy X-ray machine, which detects differences in the resonance given off in response to extremely strong magnetic fields (> 1 Tesla, about 100,000 times the strength of the Earth’s magnetic field) applied to body tissues. Since the magnetic resonance properties of the tissue within the brain are inhomogeneous because the grey and white matter contain different amounts of water, it is possible to produce very detailed brain scans showing its structural details (structural MRI) and to diagnose tumors. However, we are more interested in the brain in action, and the magnetic resonance properties also differ between oxygenated (arterial) and de-oxygenated (venal) blood.

Oversimplifying wildly, when a brain area is working hard (or about to work hard, Sirotin and Das 2009) it requires more oxygenated blood, and consequently we can measure the blood oxygenation level dependent (BOLD) resonance properties as they change over time, say during an experiment. Unfortunately, changes in blood flow are relatively slow (the hemodynamic response, and its lag, peaking roughly 4–10 seconds after an ‘event’ of interest), and are only indirectly related to the electrical activity of the neurons. Thus, as described in the table, fMRI offers excellent spatial resolution on the order of 1 mm (producing compelling color pictures based on statistical work-up) but poor temporal resolution (~ 1s). Clever experimental designs can overcome some of the temporal limitations of the technology. The equipment is expensive, and requires specialized training to operate safely. However, there are many more fMRI installations than MEG centers due to the extensive clinical applications of magnetic resonance technologies. New analysis techniques, such as diffusion tensor imaging (DTI) allow for anatomical pathways of functional neural connections to be inferred from MRI data. There are as yet few such studies applied to language, but one notable one (Saur et al 2008) offers support for separate dorsal and ventral pathways in language perception (Hickok & Poeppel 2004, 2007).

However, LabPhon researchers are not only (or even mainly) interested in measuring the brain. MRI also offers significant potential as an imaging technique for the vocal tract. If the scanner is held in a mid-sagittal orientation, then low-resolution images can be collected at between 10 and 20 frames per second (i.e. 50 – 100 ms per frame). This speed is sufficient to study some aspects of speech production, such as velo-pharyngeal movements, tongue position, lip opening, and larynx height at spatial resolutions equal to or better than ultrasound techniques (Davidson, Chapter 20). As yet, MRI remains an unusual technology for studying speech production, and it is not yet clear what the overall advantages and disadvantages are of MRI for studying speech production. MRI offers two obvious advantages over the classic X-ray studies: (1) it is safer, and (2) it is more available than X-ray microbeam facilities. However MRI machines have a major disadvantage in that during the data collection phase (where the polarity of the magnet is rapidly

1.5 Positron Emission Tomography (PET)

Another spatial resolution brain imaging technology is positron emission tomography (PET). Subjects are given a dose of a radioactive compound (typically radio-labeled \(^{15}\text{O}\)) which migrates to the brain tissue. As the radioactive substance decays, it releases positrons (the positively charged counterpart to electrons). When the positrons encounter electrons they annihilate each other, and in the process release two photons which travel in opposite directions. The photons are detected outside the brain, and the position of the collision is inferred. For the present purposes, there is no significant difference between the information that can be obtained using PET and fMRI (Feng et al 2004) though there are substantial differences when studying neuro-chemistry and neuro-pharmacology. As emphasized by a reviewer, PET does continue to offer certain advantages over fMRI for speech studies. First, the PET scanning process is quiet, and thus offers a clear environment for speech production and perception by the subjects. Secondly, PET offers better imaging of certain brain areas (for example the anterior temporal lobe), with less distortion and less incidence of motion artifacts. However, as fMRI technology is improving rapidly, most of the imaging advantages are disappearing rapidly, leaving only the quiet environment as the major advantage of PET scanning. Therefore, for general speech and language studies there is no advantage to PET over fMRI and a significant disadvantage: the exposure of subjects to a radioactive tracer. For older but still relevant reviews of speech studies using PET see Poeppel 1996 and, in reply, Démonet et al 2002.

1.6 Repetitive Transcranial Magnetic Stimulation (rTMS)

In addition to recording electrical, magnetic or hemodynamic correlates of brain activity, it is also possible to disrupt normal brain activity with electrical stimulation (as in electro-convulsive therapy), with moderate cooling of brain tissue (Malhotra and Lomber 2007) and, more usefully, with transcranial magnetic stimulation (TMS, Pascual-Leone et al 2002). By applying a fluctuating magnetic field across the skull and into the brain a weak electrical current can be induced. The motor areas of the cortex are particularly susceptible to such modulation with TMS, easily inducing involuntary movements of the limbs in subjects. TMS was recently approved by the FDA as a treatment for depression, so it considered safe enough for clinical use. The
possibility of long-term dangers of TMS are presently unknown, making this a relatively unlikely technique for LabPhon researchers. In addition, some subject report moderate amounts of pain due to the inadvertent stimulation of facial nerves when TMS is used to stimulate certain brain regions associated with speech. However, the susceptibility of the motor areas allows for the possibility of testing motor theories of speech perception (Liberman and Mattingly 1985) and mirror neuron conjectures (Arbib 2006), and a few studies supporting motor involvement in speech perception have been published (Iacoboni 2008, Roy et al 2008), but Lotto et al (2009) detail the limitations of such studies in addressing the exact nature of the connection between perceptual and motor areas.

1.7 Summary

For most LabPhon speech perception questions, better time resolution is more important than better spatial resolution. MEG, especially when combined with structural MR scans, offers excellent temporal resolution while maintaining a good compromise for spatial resolution. However, EEG is a viable alternative in many circumstances, and has much lower setup and maintenance costs. Naturally, the choice is ultimately dictated by the question at hand, so if the hypotheses require an answer in terms of anatomic information, fMRI is the most available and most appropriate technique; if the research centers on a processing models or any issue requiring a temporal answer, MEG and EEG are optimal. It is worth noting that for acquisition studies, even with infants, Near Infra-Red Spectroscopy (NIRS, or optical tomography; the little cousin of fMRI) and EEG can be used effectively. These techniques are less susceptible to movement artifacts, are silent, and generate the types of data that permit evaluation of hypotheses regarding the processing of speech information in learners. For a recent example testing newborns, see Telkemeyer et al (2009).

2. A brief survey of EEG and MEG findings relevant for LabPhon

We now turn to some particular experimental paradigms using EEG and MEG. We concentrate on these two techniques because of their relative availability (especially EEG) and suitability for testing questions involving the time course of speech perception and phonological processing. The majority of studies examine either the first prominent evoked responses (N100/M100) or mismatch (oddball detection) responses (MMN/MMF). Phillips (2001) reviews the same material and provides much useful phonological and psycholinguistic context for the general nature of these studies; consequently this section primarily updates that review with more recent publications.

2.1 N1/N100/N1m/M100

Any auditory stimulus with a well-defined onset will elicit a characteristic pattern of brain responses. Among these responses – including the P50, N1, P2, and others, all occurring in a characteristic cascade - is a relatively prominent and clear response
peaking around 100 ms after the stimulus onset, located in auditory cortex in the superior temporal lobe. This response has various designations in the literature depending on the technique employed (EEG or MEG). In EEG, the deflection has negative polarity, and is named with an N-prefix, either N1 or N100. In MEG common names include N1m (for N1 magnetic) and M100. Obviously, very little of the auditory signal can be analyzed in time to produce a clear cortical signal 100 ms after its onset; estimates (Gage and Roberts 2000) are that approximately the first 20-50 ms of the signal conditions the brain response. This makes this response very useful for assessing the information available and used at the beginning of a signal, and also aligns well with useful acoustic correlates of phonetic properties that fall within the first 50 ms (e.g. vowel formants, burst spectra, VOT). Non-linguistically, there are M100 latency differences in (sinusoidal) tone perception such that the shortest latency for the M100 is found near 1000 Hz (Roberts & Poeppel 1996); more recently Monahan et al (2008) show the same pattern of responses for pitch inferred from higher harmonics. Here we will conflate the EEG and MEG results and the various names for this early response.

### 2.1.1 Vowels

Given the MEG findings for tones below 1000 Hz – low-frequency tones of 100-300 Hz are associated with M100 latencies up to 30 ms longer than higher-frequency tones of 500-3000 Hz -- we have a reasonable expectation of tracking properties of the first formant (F1). We should expect longer M100 latencies for high vowels (with F1 distant from 1000 Hz) and the shortest latencies for low vowels (with F1 closest to 1000 Hz). And indeed this is exactly what was found for English listeners hearing synthesized tokens of /u/ and /a/ (Poeppel et al 1997). They failed, however, to find any consistent influence of the vowel pitch on the M100 response, perhaps due to the relative complexity of the vowel stimuli as compared to the simpler tone stimuli. Roberts et al (2004) more closely examined a continuum of synthesized back vowel tokens ranging from English /u/ to /a/. Rather than finding a smooth 1/f curve (as in the case of matched sinusoidal stimuli) they instead found a staircase effect, which they interpret as evidence of categorical perception of the vowel stimuli. Monahan & Idsardi (2009) demonstrate that the M100 response is not a simple response to only the first formant, but integrates information from F3 (but not from F2). By manipulating F3 while holding F1 and F2 constant they were able to modulate the M100 latency in accord with predictions based on an F1/F3 ratio derived from the previously reported results. This suggests a brain response by 100 ms to a derived vowel measure, one at least partially normalized for speaker.

The responses to simple tones have been investigated across a wide range of frequencies, and we see increasing latencies in these responses to tones above 1000 Hz (Roberts & Poeppel 1996). Consequently, we should also expect to be able to find some aspect of the M100 response which tracks the second formant (F2). However, thus far no such finding has been reported. One possible explanation is that because the amplitude of F1 is substantially greater than that for F2 in most cases, the effect of F2 on the M100 is concomitantly weaker and more difficult to detect given the limitations of the recording techniques. Nevertheless, the lack of a response latency...
correlated with F2 dramatically limits the present usefulness of the M100 latency in mapping vowel space perception.

Animal studies (using various methods) have often revealed tonotopic organization within auditory cortex (grey squirrel, Merzenich et al 1976, ferret Kelly et al 1986), and this has been extended to human auditory cortex as well (Romani et al 1982, Pantev et al 1989). Although the resolution of MEG is not sufficient to resolve place-coding of the granularity revealed by the single-unit studies in animals, it is possible that populations of neurons will have different “centers” for different formant frequencies. Obleser et al (2004), using MEG, and building on earlier related work by Diesch et al., calculated the Equivalent Current Dipole (ECD) of the source for seven distinct German vowels. They found that front vowels tend to map onto a more anterior portion of auditory cortex while back vowels map onto a more posterior region of auditory cortex. Results for vowel height were not as clear, though the Euclidean distance between the dipole locations for high and low vowels was greater than that for high and mid vowels. The allure of a cortical vowel map is plainly powerful (Poeppel 2008) but as yet premature.

2.1.4 Consonants

Turning to consonants, we might hope to find spectral differences signaling place of articulation differences in consonants reflected in various aspects of the M100 response. Obleser et al (2003) report differences for dorsal and coronal stops parallel to those reviewed above for back and front vowels (also, in their terms, dorsal and coronal). The dorsal consonants were localized more posteriorly in auditory cortex. Gage et al (2002) report M100 latency differences for /ba/ (longest), /da/ and /ga/ (shortest); however this was significant only in the right hemisphere.

Differences in the M100 amplitude and latency have been reported for short-lag versus long-lag VOT differences (Phillips et al 1995, Simos et al 1998, Sharma and Dorman 1999). These studies were designed so that the materials crossed a phonemic boundary for the listeners, and clear VOT differences were visible only with “double on” responses. The “double on” responses exhibit in essence two separate M100 responses, one to the burst and one to the onset of the vowel, and occur only with reasonably long VOTs (> 40ms). This provided support for a “refractory period” explanation of category boundary near this VOT lag—two abrupt events less than 40 ms apart were treated as a single event and those further than 40ms apart would show direct tracking of the VOT lag in the second response. However, more recently, Frye et al (2007) report a decrease in amplitude and increase in latency as VOT increases, even for the “single on” responses, suggesting that there are auditory cortex mechanisms to track a wide range of VOT values. Thus, so far, the M100 seems to track relatively low-level aspects of the VOT. While it would be desirable to see if the response can be influenced by factors such as speech rate, the necessity of presenting simple materials time-locked to the onset of stimulus presentation makes speech-rate manipulations logistically difficult.
2.2 Mismatch responses (MMN/MMNm/MMF)

Mismatch designs rely on a kind of “surprise” (violation of expectation) response to an uncommon “oddball” or “deviant” in an ongoing series of common “standards”. This is a very general and useful experimental paradigm, being applicable to a wide range of stimulus types and conditions (Näätänen et al 2007). The mismatch response (abbreviated in various ways including MMN, MMNm and MMF depending on the underlying technology employed) is robust and clear in both EEG and MEG (which we will again conflate here) and remarkably unaffected by listeners’ states of attention or even wakefulness. Because of the flexibility of the contextual definitions of “standard” and “deviant” it is possible to construct sophisticated tests of classes of speech sounds with this technique. The method also has clear affinities to classic habituation techniques in infant speech perception (Eimas et al 1971, Maye Chapter 22). This allows for the exploration of categorical perception effects, the dissection of natural classes and the degree of abstraction in phonological representations. Thus, for LabPhon researchers, the mismatch paradigm is probably the most generally useful of the neurophysiological techniques.

2.2.1 Inventories

Since vowels can reasonably be presented as unitary speech events, it is easy to construct mismatch experiments for vowels. In a series of studies (Näätänen et al 1997, Winkler et al 1999), listeners have consistently shown larger mismatch responses to standard-deviant contrasts that map onto vowel prototypes in their native language. Peltola et al (2005) found no effect of second-language instruction on the mismatch responses—advanced Finnish learners of English behaved like monolingual Finns.

A more sophisticated design is employed in Phillips et al (2000). That study examined the status of English [tə] and [də] while introducing substantial acoustic variations in the tokens. If the tokens were encoded categorically, then the task was a standard mismatch design. If, however, each token was treated as a separate exemplar, then no particular standard dominated the sequence, and there should be no mismatch response. In fact, a clear mismatch response was observed, consistent with a categorical encoding of the stimuli. Furthermore, no mismatch response was observed when the VOTs were all shifted upwards so that all tokens were in the [tə] range, indicating the subjects’ inability to distinguish standards and deviants in that condition.

---INSERT FIGURE 1 HERE---

A criticism of these studies is that they confound phonetic and phonological inventories—the non-native contrasts are not allophonically present to any substantial extent. Kazanina et al (2006) addresses this deficiency by testing a contrast [də]/[tə] between Russian listeners (for whom it is phonemically contrastive) and Korean listeners (for whom it is a common allophonically
conditioned variation—[d] is the intervocalic allophone of /t/). Russian speakers showed a clear mismatch response whereas the Koreans did not, a result consistent with the failure of the Koreans to register a difference between “standards” and “deviants” (though they must control this distinction in production in order to produce the contextually appropriate sounds). This showed that the mismatch negativity response is sensitive to abstract organizations of speech sounds into phonemic categories.

Eulitz & Lahiri (2004) discovered an asymmetrical mismatch effect. They found that for German vowels coronal (=front) deviants presented amongst dorsal (=back) standards produced larger mismatch responses than did dorsal deviants presented amongst coronal standards. This was a notable finding as most mismatch responses are symmetrical when the condition is reversed (e.g. blue vs. red = red vs. blue). They attribute this difference to the different status of coronal in the FUL model of speech perception (Lahiri & Reetz 2002) in which [coronal] is represented in the auditory input, but not in memory representations, leading to a 3-valued matching procedure (match, mismatch, no mismatch). The explanation of the results is that coronal standards induce a “blank” standard, which dorsal deviants neither match nor mismatch, producing a reduced response relative to the full mismatch of coronal deviants against the (working) memory representation of the dorsal standard. Ikeda et al (2002) report a similar asymmetry with native language prototype standards producing a larger mismatch with non-prototypical deviants than non-prototypical standards with prototypical deviants.

Dehaene-Lambertz et al. (2000), based on a behavioral study by Dupoux et al. (1999), were able to use a variant of the mismatch design to test whether in Japanese speakers versus French speakers, syllable structure of the language conditions vowel perception. They could show that the language-typical vowel epenthesis for Japanese speakers has a distinct neural correlate in their mismatch design, providing further evidence that such metrics can be productively employed to probe inventories.

Overall, the mismatch paradigm provides a rich and flexible paradigm in which to investigate the structure of phonetic and phonological classes.

### 2.2.3 Sequences

Investigations of phonetic and phonological inventories reveal certain aspects of natural language sound systems, but they do not address the dynamic character of speech, nor the contextual realizations of speech sounds due to co-articulation effects and phonotactic restrictions. However, to date there have been relatively few studies which have examined constraints on speech sound sequences. Flagg et al (2006) examined the interaction in English for nasality between vowels and the following consonants. English vowels are typically nasalized before nasals, yielding the simple (if idealized) pattern of licit [ab, âm] and illicit *[âb, am] sequences (though pronunciations such as [kât] for “can’t” with effacement of the nasal consonant are common at fast speech rates). They report mismatch-like effects in the early auditory responses to the consonant—the responses are faster for licit than for illicit sequences (though the effect is asymmetric: [ab] is facilitated, but
*æb*, *æm* and *äm* are all statistically equivalent). Hwang et al (2008) extend such findings to English final voicing sequences, with voiced stops facilitating the processing of a subsequent voiced fricative, but interfering with the processing of a following voiceless one: *dz* < *ts*, *tz* < *ds*. Monahan et al (2009) also report similar results for velar stops and following vowels: unfronted velar stops facilitate the processing of following back vowels.

### 2.3 Summary

Evoked responses and mismatch designs in both EEG and MEG have proven useful for the exploration of the neural processing of speech sounds. M100 latency varies according to the spectral properties of the incoming signal revealing integration of information across different spectral regions (F1 and F3) and modulation by abstract category structure. The localization of the M100 response within auditory cortex may also reveal a cortical vowel map. Mismatch responses have been used to investigate phonetic and phonological inventories and phonotactic constraints on sound sequences. To date very little work has been done on under-studied languages; there is enormous opportunity in this area for the investigation of typologically unusual situations and generalizations.

### 3. Pragmatic Recommendations

In our opinion, MEG represents the best current compromise of spatial and temporal resolutions, non-invasiveness and data acquisition in quiet conditions for work with adult subjects. When combined with structural MRI scans MEG is capable of reasonable spatial resolution, certainly enough to distinguish areas associated with the motor control of different articulators (for example the larynx, tongue, and lips) or to confirm that responses are indeed in auditory cortex, and perhaps to begin to find cortical maps for speech sounds (if such maps exist). One practical consideration is the efficiency of data collection, given the necessity of averaging brain responses over a fairly large number of repetitions of stimulus presentations. For this purpose, evoked responses such as the M100 are more efficient, but are only known for a limited number of attributes (some spectral information and for VOT). Mismatch studies are, in contrast, very flexible, allowing various manipulations of category structures, but are open to some nagging questions of interpretation (what is a “standard” representation exactly?) and are relatively inefficient in terms of data-points obtained per stimulus presentation (about 0.1 data point per presentation). It still remains the case, however, that EEG equipment is less costly and more available than MEG machines, and we can expect to continue to see EEG and MEG research on evoked responses relevant to speech for the foreseeable future. In the next decade or so we should expect to see an increasing number of such studies, and we also expect to see a shift in focus to studies investigating the online processing of the unfolding sequence of sounds in speech. That is, we expect a proliferation of studies of phonotactic constraints in various languages.