Markedness-based account of motor cortex involvement and muscle excitability during speech perception

Experiment Proposal

[redacted]

LING 4230

Abstract

Since its proposal in 1985, the motor theory of speech perception (MTSP) has attempted to provide an alternative analysis of speech perception which is not based purely in the auditory domain but is instead articulatory grounded in the motor system; the ability to discriminate incoming speech sounds is resultant of the ability to map speech sounds to articulators and their correlative gestures. Contemporary studies have validated these claims, typically involving the use of transcranial magnetic stimulation (TMS) or functional magnetic resonance imaging (fMRI) to capture activation in the motor cortex during perception of speech, but the specific strength of the claim is still heavily debated; some conclude that motor activation is an essential part of perception while others label it a supplementary mechanic which is primarily concerned with aiding phonemic discrimination. The study proposed herein builds upon previous work which suggests that motor activation is actually a compensatory mechanic (Craighero et al. 2008; Du et al. 2014) and aims to investigate whether the degree of somatotopic activation in the motor cortex during speech perception is variably elicited based on phonological markedness within the native sound inventory.

1 Motor activation during perception

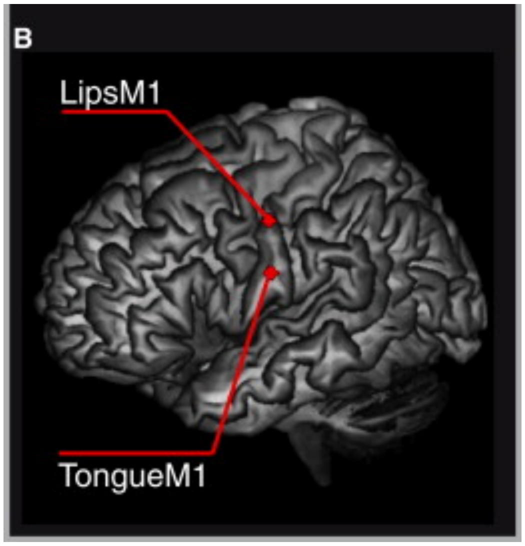
1.1 Mirror Neurons & Motor Resonance

Recent studies concerning the MTSP were inspired by findings of ‘mirror’ neurons in primates and the subsequent discovery of their existence in the human brain; neuroimaging studies have shown that viewing the motor actions of others elicits activity neural circuits in the brain of the viewer which are involved in the performance of the observed actions (Fadiga 1995) Thus, our common physiologies and correspondent neural mappings are utilized as a template by which these motor actions of others can be ‘understood’ or translated to our own particular physicality (Rizzolatti et al., 1996b). The MTSP proposes a homologous system specific to the perception of speech which thus involves activation of the neural circuits controlling articulators in the vocal tract.

Subsequent experiments by Fadiga et al. later confirmed this notion using TMS to map cortical representations of articulators and measure muscular excitability of tongue muscles during speech listening. Electrodes contacting the teeth and tongue measured motor-evoked potentials as Italian-speakers listened to Italian words and pseudowords of congruent shape (CVCCV) in which word-medial position was occupied by either 1) the strongly lingual [rr] cluster or 2) a non-lingual, labiodental fricative [ff]. Words were balanced according to their frequency in the Italian language in order to reduce variabilistic complications. Findings clearly show increased MEPs in response to [rr] words and pseudowords under all experimental conditions, while [ff] words showed less activation and [ff] pseudowords the least. Such findings indicate that passive listening triggers an automatic, “echolalic” resonance of the motor system which is best explained by the MTSP (Fadiga et al. 2002). Not only that, but it is demonstrated that such resonance is highly specific to those articulators involved in production of the segments being perceived.

1.2 Assistive Stimulation

Other studies have affirmed this positive relation through preemptive electrical stimulation of the motor cortex prior to and during speech perception and found a beneficial effect on phonemic discrimination; D’Ausilio et al. used TMS to prime either the lip or tongue representations in the primary motor cortex (M1) for a discrimination task testing categorical speech perception.



Participants were presented with CV syllables whose onset was either labial [p b] or alveolar [t d] and were tasked with discerning which onset was heard by pushing one of four buttons; when the M1 representation of the congruent articulator was stimulated via TMS just before presentation of stimuli, participants performed better compared to those stimuli produced with an incongruent articulator (D'Ausilio 2009). Such benefits speak to the existence of a mechanic which actively cross references incoming speech segments with the gestural inventory of particular articulators in search of a match, the finding of which may thus provide circumstantial evidence to validate the audio-based perception. Again, this is reminiscent of Rizzolatti et al.'s assertions that the motor cortex is working to 'translate' perceived mechanical gestures into the personal mechanical inventory of the perceiver (Rizzolatti et al., 1996b).

1.3 Inhibition via Disruption

An alternative method to demonstrate the inherent connection between speech perception and recruitment of motor neurons is converse to D'Ausilio et al.'s study in that both broad and narrow disruption of M1 representations actively impairs categorical perception. Application of repetitive TMS (rTMS) at frequency of 1Hz has been shown to be effective at disrupting neural activity in a number of cortical areas but is specifically documented as affecting cortical excitability of the primary motor cortex (Gerschlager et al. 2001; Mottaghy et al. 2003). Meister et al. used fMRI to first localize premotor cortices activated in both speech production and speech perception before applying rTMS prior to a syllable identification task. The non-trivial impairment of participants discrimination abilities led the authors to describe the role of the M1 as “essential” (Meister et al. 2004). Disruption of specific articulator representations showed the same effect restricted so as to only affect perception of segments produced with said articulator, the direct negative of the relation demonstrated by D’Ausilio et al. (Möttönen et al. 2009).

1.4 Cross-modal Interfacing

That the auditory pathway converges with other primary neural systems as part of an active, cross-referencing mechanic to assist in the categorical perception of speech is certainly uncontroversial (Derrick et al. 2018). The most notable interfacing occurs with the visual system vis a vis the McGurk Effect; presentation of incongruent audio and visual stimuli causes a ‘fusion’ effect by which the low level acoustic information is hybridized with visual information (MacDonald & McGurk 1976). The resultant percept is thus distinct from the audio and video stimuli (most often /ba/ and /ga/ respectively) and is perceived as /da/. The illusion has been replicated extensively but research has shown it is a variable phenomenon which not everyone experiences and is further sensitive to temporal modulation among other factors (Wassenhove et al 2007; Mallick et al., 2015). The visual system is even less directly related to speech than the motor system and yet it demonstrably affects low level perception. There is even research which suggests that motor involvement is directly related to this audio-visual integration system - Miyakoshi et al. found that “processing auditory-only words via this motor mechanism is ineffective,” while noting its beneficiary contribution during perception of audiovisual words (Miyakoshi et al. 2021). Such patterns would be expected if this cross-modal interfacing was not a sequence of bimodal comparisons (audio-visual, audio-motor, motor-visual, etc.) but resultant of a multisensory integration process interfacing all three modalities; there is evidence that the posterior superior temporal sulcus/gyrus (pSTS/G) or the superior parietal lobule may be localized centres for such processes (Beauchamp et al. 2004; Erickson 2014).

1.5 A Compensatory Nature

While there is still debate regarding the necessity of motor cortex involvement during speech perception, there are several studies attesting to the fact that such involvement is in fact compensatory in nature and therefore supplementary to the primary processing pathway that is the auditory system (Du et al. 2014; Schmitz et al. 2019) Some of the most telling research comes from Du et al. who investigated this perception-based motor activation in the presence of noise. Using fMRI, researchers captured blood-oxygen-level-dependent (BOLD) signals during a phoneme discrimination task. In typical fashion, monosyllables with contrastive onset consonants were presented to participants who were tasked with identification; in this instance, however, syllables were presented with varying degrees of noise in order to selectively degrade the acoustic signal. Data shows strong evidence of a, “greater robustness of the SMS than auditory regions for categorical speech perception in noise,” and note a “stronger BOLD activation in inferior frontal and premotor regions as well as weaker activation in temporal regions when phonemes were presented with increasing noise” (Du et al. 2014). This is perhaps the most transparent evidence in a decade that shows the motor system is recruited in a specifically assistive manner, a notion which has been found to hold on micro (phonetic) and macro (lexical) levels.

Fadiga et al.'s findings (as discussed in 1.1) that pseudowords elicited lesser MEPs in tongue muscles than extant words also speaks to some interaction with the semantic interface (Fadiga et al. 2002). This phenomenon was subsequently investigated by Craighero et al. who sought to isolate extant meaning from lexical frequency and thus showed the latter indeed influences the degree of premotor involvement during speech perception. Akin to the original study, TMS was used to show that the presentation of rare words evoked greater MEPs than frequent words, which in turn evoked greater MEPs than pseudowords (Craighero et al. 2008). Frequency is also a factor on the micro level, as vowels perceived to be non-native also elicit greater activation than those which are judged to be native, as found by Schmitz et al. in a 2019 study which specifically describes the role of the roles of the SMS as “active and compensatory [...] during listening to perceptually/articulatory unfamiliar phonemes,” (Schmitz et al. 2019).

Given the sensitivity of the SMS to factors such as lexical frequency, extant meaning, and perceived nativeness, there is every reason to believe that the intimate multisensory integration of auditory, visual, and sensorimotor information during passive listening also occurs during even the most habitual cases; the motor cortex is likely also recruited to help parse even native speech during passive and active listening. The study proposed herein will attempt to demonstrate the habitual sensitivity of the SMS to such factors by measuring premotor activation in response to phonological markedness during perception of a listener’s native language.

2 Markedness

2.1 Definitions

An oft debated yet widely accepted notion in linguistics is that of markedness — contemporary usage would define markedness as a general measure of “linguistic naturalness” and, although originally proposed in 1939 as a phonological notion, has since been used in numerous areas of generative linguistics including syntax (Bermúdez-Ortero & Börjars, 2004). There are a constellation of properties associated with marked sounds and marked structures, but for purposes of this experiment it can be considered an approximation of a segment’s frequency, distribution, and articulatory difficulty. Those segments which are marked are thus infrequent or have highly restricted distribution; they are rarely resultant of common processes of assimilation, epenthesis, deletion, reduction, etc. (Hume 2015).

2.2 In English

Based on these and other criteria it can be argued that the most marked single consonantal segment in English is the voiced postalveolar fricative [ʒ]; it has the most restricted distribution and is the least frequent consonant (Dewey 1923). It has no dedicated grapheme and frequently appears in borrowings such as *massage* [məsɑ**ʒ**], *espionage* [ɛspijənɑ**ʒ**], etc. or as the product of [j]-coalescence in words such as *pleasure* [plɛ**ʒ**ɚ] *measure* [mɛ**ʒ**ɚ] and *treasure* [tɹɛ**ʒ**ɚ]/[t͡ʃɹɛ**ʒ**ɚ]. Moreover, it is articulatorily similar to other, substantially less marked segments [s], [z], and [ʃ] which makes it an ideal target segment for the experiment proposed herein. These alveolar/postalveolar fricatives are extremely common across English dialects and appear as part of native words and affixes (both inflectional and derivational), most prominently as allophonic variations of the plural marker /-s/; realization is based on immediate phonological environment and the suffix will manifest as [-s], [-z] or [-əz] in words such as *cats* [kæt**s**] *dogs* [dɑg**z**] and *horses* [hɒɹsə**z**]. The segment [ʃ] is included to match specific place of articulation of [ʒ] but is still significantly more common and unmarked than the latter (Dewey 1923).

That these segments match or greatly approximate each other in terms of gestural complexity, relevant articulators, and sonorance is important for reducing variability in this experiment. There is rather clear evidence that both the inferior frontal gyrus (IFG) and premotor cortex record gestural and phonological complexity during speech production and, to a lesser extent, during perception (Moringlane 2019; Park 2019). It has been suggested that an increased BOLD response in the IFG is therefore reflective of the variable contribution of various articulatory mechanisms. The segments of this study share many primary articulatory features in that they are strident fricatives articulated with the tip/blade of the tongue against the alveolar ridge, differing only slightly in placement (with [ʃ ʒ] being produced slightly closer to the palate compared to [s z]). Voicing is also considered in order to account for differences in glottal engagement, which in turn directly increases gestural complexity; [z ʒ] share the feature [+voice]. Thus there are very minimal articulatory differences between these segments but a vast difference in markedness (frequency, distribution), making them ideal for use in this study.

3.0 Method

3.1 Participants

Based on previous similar experiments utilizing fMRI and TMS to map and measure the motor cortex and MEPs (described below), it is predicted that 10-20 participants would be appropriate for a preliminary study such as this (D'Ausilio 2009; Du 2014; Meister 2007; Schmitz 2019). All participants would be right-handed and have no demonstrable hearing deficiencies; vision will not be a substantial factor but ideally should be normal or corrected-to-normal for all participants, Lastly, an average age of 22.5 years and 50/50 distribution of males and females would be ideal as well.

3.2 MEP Procedure

Akin to Fadiga et al.'s study measuring corticobulbar excitability of tongue muscles during passive listening, this experiment will also utilize electrodes to measure and record MEPs of tongue muscles while listening to speech. Ag±AgCl electrodes will rest on the premolars in contact with the tongue while participants are seated in a well lit, quiet room (Fadiga 2002). Additionally,

3.3 Stimuli & Task

Participants will be presented with audio stimuli recorded from a native speaker of English producing the four target consonants with one of two vowel qualities, resulting in eight distinct CV syllables. These target stimuli will be intermixed with syllables consisting of matching vowels but non-lingual consonants /b p f v/. These labial & labiodental consonants do not require any action by the tongue during production and should therefore not elicit any excitability in tongue muscles. Consonant-vowel combinations are listed below.

	Labial		Labiodental		Alveolar		Postalveolar	
	p	**b**	**f**	**v**	**s**	**z**	**ʃ**	**ʒ**
Condition 1 V = ɑ	[pɑ]	[bɑ]	[fɑ]	[vɑ]	[sɑ]	[zɑ]	[ʃɑ]	[ʒɑ]
Condition 2 V = i	[pi]	[bi]	[fa]	[vi]	[si]	[zi]	[ʃi]	[ʒi]

Syllables will be randomly or pseudo-randomly presented over repeated trials while participants are occasionally tasked with identifying whether the most recent stimuli was spoken by a male or a female talker. Response will be indicated via one of two buttons. This will ensure participants pay continued attention to stimuli but are not actively thinking about or trying to identify specific articulators. Participants will also be aware that there is no obligation to produce or reproduce any speech sounds used in the study or otherwise. These considerations are such that there is no additional bias or cause for participants to engage the motor system during listening; any activity and/or differences in MEP should therefore be purely indicative of the degree of recruitment of the motor system as a compensatory mechanic to aid in perception.

4 Expectations, Further Study

Based on all of the aforementioned literature and previous experiments, it is expected that the presentation of the marked [ʒ] segment will elicit the greatest MEP in tongue muscles of participants immediately after presentation of stimuli. This would converge with previous findings that motor cortex involvement is 1) passively recruited during perception; 2) articulator-specific;

3) not essential but compensatory in nature. This study seeks to fill in gaps that have arisen from the work of others by assessing the specific sensitivity of this recruitment during perception of entirely native speech sounds; previous studies have always utilized non-native sounds, pseudowords, etc. which would obviously involve greater perceptual challenges & impose a greater cognitive load on the listener when tasked with identification or discrimination. Thus the segments used in this study are chosen to eliminate any extraneous work that arises from presentation of such outliers.

Numerous studies have attempted to define the validity and boundaries of the MTSP since its proposal in the 1970's and there is still contentious debate regarding both the exact nature of the SMS during perception and its necessity, with some claiming it as essential and others hesitating to do so. This proposed experiment accords with a weak form of the hypothesis which would thus characterize motor involvement as secondary, supplementary, and compensatory. However, it also hinges on the assumption that involvement is active, sensitive, and ongoing, contributing to perception even when listening conditions are not adverse and speech sounds are familiar. Further study could incorporate other marked segments for a more fine-grained phonological account of motor activity. Additionally, the amount of biomechanical energy required to produce specific speech sounds may be considered in order to elucidate motivation for recruiting mechanical systems when there is no need for tangible engagement of these language-adjacent faculties.

Acknowledgements

The author would like to thank the professor and classmates for a great semester and interesting discussions. There are no conflicts of interest to be declared. Unfortunately, no funding or grants were acquired to help produce this proposal.

References

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–823. https://doi.org/10.1016/s0896-6273(04)00070-4

Bermúdez-Otero, R., & Börjars, K. (2006). Markedness in phonology and in syntax: The problem of grounding. *Lingua*, *116*(5), 710–756. https://doi.org/10.1016/j.lingua.2004.08.016

Craighero, L., Fabbri-Destro, M., & Fadiga, L. (2008). Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. *Journal of Physiology-Paris*, *102*(1–3), 101–105. https://doi.org/10.1016/j.jphysparis.2008.03.006

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The Motor Somatotopy of speech perception. *Current Biology*, *19*(5), 381–385. https://doi.org/10.1016/j.cub.2009.01.017

Dewey, G. (1923). Relativ frequency of English speech sounds. *Harvard Studies in Education*, *4*. https://doi.org/10.4159/harvard.9780674419193

Du, Y., Buchsbaum, B. R., Grady, C. L., & Alain, C. (2014). Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proceedings of the National Academy of Sciences*, *111*(19), 7126–7131. https://doi.org/10.1073/pnas.1318738111

Erickson, L. C., Zielinski, B. A., Zielinski, J. E., Liu, G., Turkeltaub, P. E., Leaver, A. M., & Rauschecker, J. P. (2014). Distinct cortical locations for integration of audiovisual speech and the McGurk effect. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00534

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience*, *15*(2), 399–402. https://doi.org/10.1046/j.0953-816x.2001.01874.x

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608

Gerschlager, W., Siebner, H., & Rothwell, J. (2001). Decreased corticospinal excitability after subthreshold 1 Hz RTMS over lateral premotor cortex. *NeuroImage*, *13*(6), 1170. https://doi.org/10.1016/s1053-8119(01)92490-5

Hume, E. (2015). Phonological Markedness and its Relation to the Uncertainty of Words. *Journal of the Phonological Society of Japan*. https://www.academia.edu/16383799/Phonological_Markedness_and_its_Relation_to_the_Uncertainty_of_Words

Keough, M., Gick, B., & Derrick, D. (2014). Cross-modal perception of Speech. *Encyclopedia of Language Development*. https://doi.org/10.4135/9781483346441.n40

Mallick, D., F. Magnotti, J., & S. Beauchamp, M. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review*, *22*(5), 1299–1307. https://doi.org/10.3758/s13423-015-0817-4

Mcgurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Current Biology*, *17*(19), 1692–1696. https://doi.org/10.1016/j.cub.2007.08.064

Miyakoshi, M., Norato, G., Medvedev, A. V., & Turkeltaub, P. E. (2021). Motor engagement relates to accurate perception of phonemes and audiovisual words, but not auditory words. *Communications Biology*, *4*(1). https://doi.org/10.1038/s42003-020-01634-5

Moringlane, J. R., Sikos, L., Reith, W., & Krick, C. M. (2019). FMRI and acoustic analyses reveal neural correlates of gestural complexity and articulatory effort within bilateral inferior frontal gyrus during speech production. *Neuropsychologia*, *132*, 107129. https://doi.org/10.1016/j.neuropsychologia.2019.107129

Mottaghy, F. M., Gangitano, M., Horkan, C., Chen, Y., Pascual-Leone, A., & Schlaug, G. (2003). Repetitive TMS temporarily alters brain diffusion. *Neurology*, *60*(9), 1539–1541. https://doi.org/10.1212/01.wnl.0000058903.15205.46

Park, H. (2019). Neural correlates of the phonological complexity of Korean vowels. *Korean Journal of Linguistics*, *44*(1), 79–100. https://doi.org/10.18855/lisoko.2019.44.1.004

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Schmitz, J., Bartoli, E., Maffongelli, L., Fadiga, L., Sebastian-Galles, N., & D'Ausilio, A. (2019). Motor cortex compensates for lack of sensory and motor experience during auditory speech perception. *Neuropsychologia*, *128*, 290–296. https://doi.org/10.1016/j.neuropsychologia.2018.01.006

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia*, *45*(3), 598–607. https://doi.org/10.1016/j.neuropsychologia.2006.01.001