1. Introduction

The processes underlying human speech perception have been widely examined and debated over the last six decades or so. By some accounts, speech is perceived as a function of its acoustic constituents and their impact on the auditory system (Klatt, 1979; Kuhl & Miller, 1975; Massaro & Cohen, 1990; Ohala, 1996; Stevens, 1981; Sussman, 1989). Additionally, these acoustic theories generally hold that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes. Criticisms of these perspectives were propagated by studies showing that speech perception and production are distinct processes.
create a biological link between senders and receivers of gestural goals, which may have served in the evolution of communication (Rizzolatti & Arbib, 1998). These notions were further bolstered by the discovery of a subclass of mirror neurons that fired not only in response to seeing an action, but also to hearing sounds associated with a specific action such as paper ripping (Kohler et al., 2002) or peanut breaking (Keysers et al., 2003). Hence, the authors suggested that mirror neurons code the intended goal of an action in an abstract amodal manner rather than a specific action itself, eliciting strong parallels to the nature of speech ‘gestures’ as described by motor theorists.

A growing body of research suggests that in humans, the mirror system may be involved in action recognition, imitation, empathy and theory of mind. Its role in speech perception also has been scrutinized using various measures. In the auditory modality, transcranial magnetic stimulation (TMS) has been used to show that listening to lingual speech sounds could evoke stronger motor evoked potentials (MEPs) in the tongue relative to non-speech sounds (Fadiga, Craighero, Buccino, & Rizzolatti, 2002) and stronger MEPs in lip muscles when listening to speech while watching white noise (Watkins & Paus, 2004; Watkins, Strafella, & Paus, 2003).

In addition, using functional magnetic resonance imaging (fMRI), listening to meaningless speech has been found to bilaterally activate portions of the ventral premotor cortex (though not Broca’s area), portions of the motor cortex, and the supplementary motor area (Wilson, Saygin, Sereno, & Iacoboni, 2004) and motor cortical regions in a somatotopic manner (Pfurtscheller et al., 2006) relative to non-speech stimuli. In the visual modality, Nishitani and Hari (2002) used magnetoencephalography (MEG) to reveal that observing still pictures of lips could activate Broca’s area and the motor cortex. Similar bilateral motor activation patterns have also been found to silent speech lip movements using fMRI (Campbell et al., 2001) and stillled speech (Calvert & Campbell, 2003; Buccino et al., 2004) also used fMRI to discover significantly higher activations in portions of the left inferior frontal gyrus in response to viewing speech reading and lip smacking, but not to viewing a dog barking. They suggested that activation of one’s own motor system via action observation occurs when the action in question is part of one’s own motor repertoire, again suggesting a biological underpinning for the mirror system.

Speech can be perceived unimodally via either audition or vision in isolation, or bimodally (audio-visual). Though the studies above provide evidence that speech perception through either audition or vision can activate the human mirror system, they did not examine the relative strength of each modality for inducing mirror activation or their relative strengths compared to audio-visual speech perception. Because the mirror system is thought to have close connections to the somatosensory (SI) system, Møttønen, Jarvelainen, Sams, and Hari (2004) used MEG to examine how viewing and hearing speech modulated activity in the left SI mouth cortex. Whereas viewing speech induced significant SI modulation, hearing speech did not. Similarly Sundara, Namasiavayam, and Chen (2001) found that visual and audio-visual presentations of the syllable/ba/both yielded significant increases in MEP amplitudes, whereas the MEP increase produced from auditory perception alone did not reach significance. Activation levels of motor areas during speech perception have also been examined using fMRI. Skipper, Nusbaum, and Small (2005) found that audio-visual speech activated the inferior frontal gyrus and premotor cortex to a greater extent than audio or visual speech alone. They also found that the activation level of the premotor cortex was modulated by a number of phonemes that participants could visually identify.

It has been suggested that the visually perceived gestures may play the stronger role in activating the motor system during speech perception (Skipper, von Wassenhove, Nusbaum, & Small, 2007), especially in conditions in which auditory speech is absent or degraded (Callan et al., 2003). In addition, it is not clear how brain regions involved in speech production are differentially activated during various perception tasks. For example, it has been suggested that motor recruitment might be influenced by the degree of linguistic processing necessary to a task (Callan, Jones, Callan, & Akahane-Yamada, 2004; Ojanen et al., 2005; Wilson & Iacoboni, 2006). Thus, the numerous measures of motor recruitment in speech perception and the variety of stimuli employed, combined with the diversity of findings, make it difficult to reconcile the discrepancies within the current body of research and explain the extent to which motor recruitment may be necessary in speech perception. As such, further investigation in this area is warranted.

Electroencephalography (EEG) has been suggested as a promising, cost-efficient and non-invasive means of indirectly examining the mirror neuron activity in humans. In particular, measurements of oscillation amplitudes in the mu frequencies (8–13 Hz) measured across the sensorimotor cortices acquired via surface level electrodes are thought to provide a valid index of mirror activity (Altschuler, Vankov, Wang, Ramachandran, & Pineda, 1997). Mu ‘rhythms’ are influenced by both motor activity and attention (see Pineda, 2005 for full review). When a person is at rest, amplitudes of waves in this band are highest because sensorimotor neurons responsible for generating these waves fire synchronously. Conversely, when a person performs an action, the pattern of firing is asynchronized, resulting in suppression of the mu wave and smaller amplitudes. However, a number of studies also have found that mu waves are suppressed when normal adults observe human hand movements (Muthukumaraswamy & Johnson, 2004; Muthukumaraswamy, Johnson, & McNair, 2004; Oberman et al., 2005; Virji-Babul et al., 2008) and implied point-light biological animation (Saygin, Wilson, Hagler, Bates, & Sereno, 2004; Ulloa & Pineda, 2007), and even when participants imagine biological motion (Pineda, Allison, & Vankov, 2000).

Because mu suppression can occur in these passive observation/imagery conditions in the absence of motor activity, the level of suppression is thought to provide an index of mirror neuron activity. When employing these paradigms, recordings from Cz, C3 and C4 are thought to be indirect measures of cortical activity in the supplementary motor areas and left and right sensorimotor cortices (S1-M1: Babiloni et al., 1999), respectively. Hence, EEG recordings from these electrodes are considered to be measuring “downstream modulation of sensorimotor areas by mirror neurons” (Oberman et al., 2005, p. 191). As the recordings are made from the scalp, it is difficult to map the sources of suppression to cortical landmarks. However, Nishitani and Hari (2000), in a study using MEG, found that the sources of mirror activity may be further ‘upstream’ in the primary motor cortex and in the inferior frontal cortex (e.g., BA 44). As these regions are often activated in speech perception tasks, it seems plausible that EEG recordings of mu rhythms at electrode sites Cz, C3 and C4 might also be suppressed when speech is perceived.

Though Muthukumaraswamy, Johnson, Gaetz, and Cheyne (2004) examined mu suppression to oro-facial movements (i.e., teeth-baring, blowing), to our knowledge, EEG has not yet been used to examine differential levels of mirror neuron activity to the perception of speech and non-speech stimuli. Previous studies that have identified selective mirror neuron functioning in response to observed dynamic biological stimuli have employed visual noise and ‘non-biological’ conditions as bases for comparison. Oberman et al. (2005) used a bouncing ball, whereas Ulloa and Pineda (2007) used scrambled versions of their point-light biological animations. As speech can be conveyed through both auditory and visual channels, in order to differentiate the effects of speech from noise and non-biological stimuli, it seems logical in the current study to employ all three types of stimuli (noise, speech, and non-biological) conditions for both input modalities. Based on these conditions, the results of this study will determine the extent to which speech and perception may be differentially influenced by mirror activity.
upon previous findings of motor activation as a function of speech perception, it is hypothesized that conditions in which speech is presented visually will result in the highest levels of mirror neuron activation and therefore, the most robust mu suppression relative to when noise is presented through both modalities. When non-speech stimulus conditions are presented, mu suppression is not expected. These data will also provide some preliminary findings of how auditory speech may suppress mu waves and contribute to the body of research investigating how auditory and visual speech signals interact in motor recruitment.

2. Methods

2.1. Participants

Fourteen right-handed female adults aged 20–51 (mean age = 28.03 years; SD = 8.24 years) who were native English speakers and had no diagnosed history of communicative, cognitive or attentional disorder. A female cohort was chosen because of the recent finding of stronger mu suppression in response to biological stimuli in females than in males (Cheng et al., 2008). As, to our knowledge this is the first study examining mu suppression to speech, it seemed appropriate to first examine the gender in which the responses might be strongest. In addition, in order to negate any potential ‘processing’ differences that second language acquisition may incur, native English speakers were chosen because an English speaker produced the speech stimuli. Prior to the experiment, informed consent (approved by The University of Tennessee Institutional Review Board) was obtained for all participants.

2.2. Stimuli and experimental conditions

The stimuli consisted of audio-visual presentations of (1) meaningless speech, (2) noise and (3) a second set of non-biological stimuli. Audio, video and audio-visual speech stimuli were constructed from recordings (using a Sony DCRH30 video camera) of the mouth (Fig. 1) of a native English speaking adult male producing a continuous stream of co-articulated meaningless CV syllables (e.g., /da/, /ba/, and /ga/). A male voice was chosen for the stimuli as males have lower fundamental frequencies and their voices are richer in harmonic content. The audio-visual clips were edited using Apple iMovie (version 5.0.2), which allowed for the separation of the audio and video tracks where necessary. The noise stimuli consisted of auditory and visual white noise. In the same vein that Ullsma and Pinèda (2007) used scrambled point-light stimuli as analogs of their point-light stimuli, the visual non-biological analog of speech movement was constructed by taking a still-frame of the mouth recordings and converting it into a kaleidoscope with five symmetrical portions of the mouth centered at 72◦ to each other using software from the website (http://www.krazydad.com/kaleido). Video recordings were made (Screen Movie Recorder 2.6) as undulating kaleidoscopic motion was applied in a random symmetrical and concentric manner (Fig. 2). The auditory non-biological analog was a 1000Hz pure tone. To help ensure uniform intensity levels, all audio stimuli were normalized using Adobe Audition (version 1.5) software. Hence, 90s audio-visual stimuli clips were constructed for nine different audio-visual conditions: (1) NN: noise (auditory)–noise (visual), (2) NS: noise–speech, (3) NK: noise–kaleidoscope, (4) SN: speech–noise, (5) SS: speech–speech, (6) SK: speech–kaleidoscope, (7) TN: tone–noise, (8) TS: tone–speech, and (9) TK: tone–kaleidoscope. The speech stimuli contained 61–69 syllables. Each clip was recorded onto a DVD using Apple iDVD (version 5.0.1).

2.3. Procedure

The experiment was conducted in an electronically and magnetically shielded, double-walled, sound-treated booth. Participants were seated in a comfortable reclining armchair with their heads and necks well supported. After the electrodes were placed, they were instructed to sit quietly with their eyes open and attend to the audio and visual stimuli. The stimuli were played from the DVD with a Phillips HDMI DVD player (model DVP5892). The visual output was displayed on a 32-in. Phillips LCD monitor (model 32PFL5332D). The audio output was routed through a Mackie DFX-6 mixer and was delivered binaurally to participants’ ears at their preferred intensity level using Ear Tone model ER-1-14A insert earphones. In order to help ensure that participants were paying attention, they were asked to silently count the number of times they heard or saw a syllable initiated by the /b/ phoneme. All experimental conditions were presented to participants twice in separate random sequences (i.e., two blocks). Experimental conditions were separated by 30 s breaks, during which participants were asked to report the number of /b/ initiated syllables they were able to recognize from the previous condition. All participants reported the hearing or seeing /b/ initiated syllables in the speech conditions and no participant reported hearing or seeing speech in any of the non-speech conditions.

2.4. EEG data acquisition and analysis

Thirteen electrodes were used to acquire EEG data based on the international 10–20 method of electrode placement (Jasper, 1958) using a 32-channel, unlinked, sintered NeuroScan Quik Cap. Non-inverting electrodes included Cz, C3, C4, Pz, P3, P4, Fz, F3, F4, O1, O2, M1, and M2. The inverting electrode was placed on the nasion and the ground electrode was at Fpz. The electro-oculogram (EOG) was recorded by electrodes placed on the left superior orbit and the left inferior orbit. The impedances of all electrodes were measured at 100Hz before, during, and after testing and were never greater than 10kΩ. EEG was collected and analyzed using Compumedics NeuroScan Scan 4.3.3 software and Synamps 2 system. Data were obtained for approximately 180 s per condition (90 s for each of two runs), filtered (0.15–100Hz), and digitized via a 24-bit analog-to-digital converter at a sampling rate of 500 Hz. It has been reported that EEG in the 8–13 Hz frequency band recorded from electrodes in the region of the occipital cortex are confined by states of expectancy and awareness (Klimosch, Doppelmayr, Russegger, Pachinger, & Schweiger, 1998). In addition, there are shared frequencies between the mu rhythm and posterior alpha bands and the activity in the posterior alpha bands may be stronger than the mu rhythms. For this reason, researchers like Oberman et al. (2005) have suggested that EEG obtained from Cz, C1, and C4 might be confounded by posterior activity. However, this factor is present.
across conditions as they all they involve visual stimuli presentation to participants with their eyes open and hence does not affect any condition to a greater extent. To further combat this, the first and last 10 s of each of the two blocks of 90 s data corresponding to one experimental condition was removed. The remaining 70 s of data were used in the analysis. As data for each condition was collected in two blocks, a total 140 s of data were collected per condition. Eye blink and eye movements were identified in the EOG channel and EEG at all other channels during these intervals was removed. Then, the integrated power in the 8–13 Hz range was calculated using a Fast Fourier Transformation (1024 points) on the data, which were divided into 2 s epochs. A cosine window was used to minimize artifacts due to data splicing.

Mu suppression was calculated as the ratio of the power during the eight experimental conditions (NS, NT, KS, SS, SN, ST, KN, and KT) relative to the power in the NN condition, which was considered the baseline. Data are presented as ratios to minimize individual variability in absolute mu power that may result from differences in scalp thickness or electrode impedance, as opposed to mirror neuron activity (e.g., Oberman et al., 2005). A log transformation was applied to each ratio because of the non-normal distribution that ratio data yield. Thus, negative and positive log ratios are indicative of mu suppression and enhancement, respectively.

3. Results

Suppression of mu rhythms relative to baseline were measured from Cz, C3 and C4 as these are the sites coinciding with the mu wave source and most robust measurements (Pineda & Hecht, 2009). Because alpha bands were not suppressed in any other electrode site, we can safely rule out the possibility that recordings from these sites were influenced by posterior alpha activity. Mu suppression across three electrode sites (Cz, C3, and C4) and eight conditions were analyzed using a two-factor repeated-measures ANOVA. Greenhouse-Geisser corrections were applied where sphericity assumptions were violated and Bonferroni corrections were applied to multiple post hoc comparisons.

A significant main effect was found for condition [F(7, 91) = 6.367, p < 0.001]. However, no significant main effect was found for the electrode site [F(2, 226) = 0.105, p = 0.901] or the interaction between electrode site and condition, [F(14, 182) = 0.992, p = 0.463]. Hence, mu suppression data were collapsed across the three electrodes and are displayed as function of experimental condition in Fig. 3. The collapsed data were used in post hoc t-tests that examined how each condition suppressed the mu wave relative to the baseline (Oberman et al., 2005; Oberman, Ramachandran, & Pineda, 2008). The conditions in which visual speech were presented were the only ones to induce significant and robust levels of mu suppression [SS: t(13) = −4.223, p = .008; NS: t(13) = −3.904, p = .016; TS: t(13) = −4.760, p < .001].

4. Discussion

In accord with our first hypothesis, significant suppression of the mu wave was found across electrodes Cz, C3 and C4 when participants visually perceived streams of meaningless co-articulated speech gestures (i.e., viewed continuous lip movements). All three visual speech conditions tested (SS, NS, and TS) produced similar levels of suppression, suggesting that visual speech perception recruited the motor system irrespective of the auditory stimulus with which it was paired. These findings are consistent both with studies showing similar patterns of mu suppression to observed human movements (Muthukumaraswamy, Johnson, & McNair, 2004; Oberman et al., 2005; Virji-Babul et al., 2008) and with studies using MEG or fMRI showing bilateral motor activity to visual speech perception (Calvert & Campbell, 2003; Campbell et al., 2001; Nishitani & Hari, 2002; Skipper et al., 2005). Furthermore, in accord with our second hypothesis, the non-speech conditions (TN, TK, and NK) failed to produce any significant levels of mu suppression, most likely, because in these conditions, participants did not perceive stimuli containing arrays of movements existing within their own motor repertoire (Buccino et al., 2001).

These findings also appear to be consistent with those of Möttönen et al. (2004) and Sundara et al. (2001) who found evidence of motor activation in response to visual but not auditory speech perception. As such, they support notions that motor recruitment for speech is often mediated primarily by visual systems (Skipper et al., 2007) and that when visual speech information is presented, perception of the gesture is supported by an internal simulation of the observed articulatory movements (Callan et al., 2003). This motoric supplementation which appears to be afforded by the presence of visual speech may contribute to previous findings showing that the addition of congruent visual cues to auditory speech is associated with increases in speech intelligibility (Sumby & Pollack, 1954), improved speech detection thresholds (Grant & Seitz, 2000), and accelerated neural processing (Van Wassenhove, Grant, & Poeppel, 2005).

Current findings may contrast with those that have found motor involvement during auditory-only speech perception tasks (e.g., Fadiga et al., 2002; Watkins et al., 2003; Watkins & Paus, 2004; Wilson et al., 2004). Skipper et al. (2005) suggest that any brief motor activation that occurs during auditory speech perception may not be observed in methods that employ temporal averaging. Perhaps however, a closer examination of acoustic speech signals may help foster a better understanding of their inconsistency in motor activation patterns. Acoustic speech signals are richly encoded and contain cues indicative of the gesture (e.g., formant structures, glottal pulsing, onset spectra) from every level of the vocal tract. In fact, the redundancy of information pointing to the intended gesture allows that the percept be recovered even when considerable portions of the acoustic signal are removed. An extreme case of this can be found when humans perceive linguistic information from sine wave analogs of speech that are devoid of traditional acoustic cues (Remez, Rubin, Pisoni, & Carrell, 1981). One perspective derived from cognitive psychology that might contribute to explaining this substantial human capacity is the view of speech perception as an ‘embodied’ process (Aziz-Zadeh & Ivy, 2009; Galantucci, Fowler, & Turvey, 2006; Skipper et al., 2005; Wilson, 2002). From this standpoint, the manner in which speech
is perceived is experientially shaped through multiple interactions. Connections between the acoustics of speech and the motoric gestural configurations responsible for producing those acoustics (i.e., parity) become strengthened over time. It seems logical that these sensorimotor connections are strengthened most during the early periods of language acquisition and development (Kalinowski & Saltuklaroglu, 2003). Therefore, it may be plausible that once a sufficient phonemic repertoire is acquired, motor involvement during basic speech perception becomes less essential; yet the motor system may remain an available referent during more demanding phonemic or linguistic processing tasks.

The relationship between processing demands and motor activation may be further considered when examining the role of Broca’s area in speech perception. Though Broca’s area is classically considered a key speech production area, it is also considered by some to mediate functioning of the mirror system (Hari, Levinan, & Raji, 2000; Nishitani & Hari, 2002) and play a strong role in multisensory processing. During passive speech perception tasks, Broca’s area has been found to be more active in conditions that require increased amounts of processing such as when an auditory signal is degraded or absent (Callan et al., 2003), when auditory and visual signals are not matched (Ojanne et al., 2005), when listening to non-native phonemes (Wilson & Iacoboni, 2006) and when listening to sentences paired with incongruent gestures (Willems, Ozurek, & Hagoort, 2007). Hence if the suppression of the mu rhythm is influenced by activity in Broca’s area, it may be postulated that in the current study, the processing demands incurred by the auditory-only speech conditions were insufficient for necessitating recruitment of the motor system and therefore, did not result in any significant suppression of the mu rhythm. That is, the meaningless acoustic speech signal was produced by an English speaker and was presented clearly and audibly. Its combination with either visual noise or the kaleidoscopic stimuli most likely did not tax the speech processing demands. In addition, the participants were all female native English speakers, free from speech, language or hearing disorders, and recruited from a university setting, suggesting that they had little difficulty interpreting the phonemic information that was presented under these conditions.

Skipper et al. (2005) suggest that rather than being static in nature, the mirror system may be a dynamic entity with variable thresholds for engagement across stimuli, both within and between individuals and dependent on the nature and goals of a given task. Considering this dynamic nature and the possibility that speech mechanisms within individuals may be tied to their embodiment in world experiences, it is not surprising that the original tenets of motor theory have been recently re-examined (e.g., Galantucci et al., 2006; Lotto, Hickok, & Holt, 2009). The current paradigm may be appropriate for shedding further light on questions surrounding the role of motor involvement in speech perception. Future studies examining mu suppression to speech may consider increasing processing demands, perhaps by adding different types of noise to the speech signal (e.g., Callan et al., 2003), using non-native linguistic forms (e.g., Wilson & Iacoboni, 2006) or employing speech from disordered populations such as those who stutter or have dysarthric speech. Another application may be found in the different populations that may be tested, Much remains to be learned about the speech processing in disordered populations such those with autism, specific language disorders and aphasia. In addition, Nystrom (2008) recently used high density EEG to find similar mirroring mechanisms within children and comparing them to adults, especially if theories about embodied processing prove to be valid.

To our knowledge, this is the first study to examine mu suppression during auditory and visual speech perception. Though we believe the findings to be of importance, certain limitations should be recognized in interpreting the findings. First, though the mu rhythm is thought to emanate from sensorimotor areas, the limited spatial resolution of surface EEG recordings does not allow accurate site mapping of the neural activity. As such, future investigations continue to require that motor activity measured by mu suppression is indexed as a function of the task and relative to other conditions under investigation. Second, this study used an exclusively female cohort. Though females were chosen because of their higher sensitivity in recording mu suppression, these findings need to be replicated in males. Third, though the visual non-biologic analog was dynamic (i.e., the kaleidoscope), the non-biologic analog to auditory speech (i.e., the pure tone) was not. In future studies, to better understand the differential abilities of speech versus non-speech acoustic signals to induce mirror activity, using more dynamic non-biologic auditory stimuli as controls may be advantageous.

These preliminary findings showed mu suppression, considered evidence of cortical motor activation, to visually perceived streams of meaningless syllables, irrespective of acoustic pairing. Though suppression to acoustic only speech stimuli was not found herein, we are cautiously optimistic that it might be found in future studies using this technique. As such, indexing mu suppression across audio-visual tasks appears to be a non-invasive, cost-efficient technique that may continue to provide clues to the many unanswered questions about the nature of human speech perception.

Uncited reference


References


Please cite this article in press as: Crawcour, S., et al. Mu wave suppression during the perception of meaningless syllables: EEG evidence of motor recruitment. Neupropsychologia (2009), doi:10.1016/j.neupropsychologia.2009.05.001


