JSLHR

Research Article

The Role of Somatosensory Information in Speech Perception: Imitation Improves Recognition of Disordered Speech

Stephanie A. Borrie[a] and Martina C. M. Schäfer[b]

Purpose: Perceptual learning paradigms involving written feedback appear to be a viable clinical tool to reduce the intelligibility burden of dysarthria. The underlying theoretical assumption is that pairing the degraded acoustics with the intended lexical targets facilitates a remapping of existing mental representations in the lexicon. This study investigated whether ties to mental representations can be strengthened by way of a somatosensory motor trace.
Method: Following an intelligibility pretest, 100 participants were assigned to 1 of 5 experimental groups. The control group received no training, but the other 4 groups received training with dysarthric speech under conditions involving a unique combination of auditory targets, written feedback, and/or a vocal imitation task. All participants then completed an intelligibility posttest.
Results: Training improved intelligibility of dysarthric speech, with the largest improvements observed when the auditory targets were accompanied by both written feedback and an imitation task. Further, a significant relationship between intelligibility improvement and imitation accuracy was identified.
Conclusions: This study suggests that somatosensory information can strengthen the activation of speech sound maps of dysarthric speech. The findings, therefore, implicate a bidirectional relationship between speech perception and speech production as well as advance our understanding of the mechanisms that underlie perceptual learning of degraded speech.

Perceptual learning of speech, from a theoretical perspective, refers to "experience-evoked adjustments to the cognitive-perceptual processes required for recognizing spoken language" (Borrie, McAuliffe, & Liss, 2012, p. 291). Clinically, the term is used to describe the notion that listeners can *learn* to better recognize a speech signal that is initially difficult to understand. A series of recent studies by Borrie and colleagues (Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2012; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2013) provide empirical evidence that listeners can, indeed, learn to better understand neurologically degraded speech: intelligibility of dysarthric speech was significantly greater for listeners familiarized with the disordered input. It is assumed that exposure to the degraded acoustics facilitates a remapping of existing mental representations, including prelexical and lexical knowledge (e.g., Eisner & McQueen, 2005; Francis, Nusbaum, & Fenn, 2007) and that this morphed perceptual map enables the listener to better recognize dysarthric speech in subsequent encounters (Borrie, McAuliffe, & Liss, 2012). Borrie and colleagues (Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2012) further demonstrated that intelligibility gains following familiarization with dysarthric speech were superior in both magnitude and longevity over time when the degraded acoustics were paired with written transcripts of the spoken targets during the familiarization experience. This would suggest that provision of lexical information, which disambiguates the dysarthric signal, may serve to strengthen mental ties between the degraded acoustics and existing linguistic knowledge.

Taken together, there is mounting evidence to suggest that listener-targeted perceptual learning paradigms may be a viable clinical tool for reducing the intelligibility burden of dysarthria. Moreover, existing research has demonstrated that the listener's ability to decipher neurologically degraded speech can be significantly improved by integrating different modes of information processing into the familiarization paradigm. Although the use of external information to disambiguate the degraded acoustics (i.e., written

[a]Utah State University, Logan
[b]New Zealand Institute of Language, Brain and Behaviour, University of Canterbury, Christchurch, New Zealand
Correspondence to Stephanie A. Borrie: stephanie.borrie@usu.edu
Editor: Jody Kreiman
Associate Editor: Amy Neel
Received May 1, 2015
Revision received August 23, 2015
Accepted September 14, 2015
DOI: 10.1044/2015_JSLHR-S-15-0163

Disclosure: The authors have declared that no competing interests existed at the time of publication.

feedback) has been a predominant focus in theory and practice of perceptual learning paradigms aiming to improve intelligibility of dysarthric speech, this finding calls for a new perspective. In particular, it invites the investigation of other modes of information processing, including internally generated sources, to support mapping the degraded speech signal to linguistic knowledge.

One computational model that has extensively tested the relationship between externally and internally generated information and has comprehensively tested their interconnections for speech sound acquisition is the Directions Into Velocities of Articulators (DIVA) model of speech production (Guenther, Hampson, & Johnson, 1998). In the DIVA model, the identification of a teaching signal is key to the learning of corresponding acoustic and somatosensory information. Moreover, the model proposes that perceptual learning and motor learning are not separate, but instead, interconnected speech processes that occur in tandem. That is, both auditory and somatosensory information are required to guide the learning of consistent mappings between a speech sound map and its corresponding reference frames (Golfinopoulos, Tourville, & Guenther, 2010; Perkell, 2012). These include the auditory-perceptual target, tactile and proprioceptive information, muscle length, and articulatory movement to produce the target sound (Guenther et al., 1998). When applying these theoretical constructs of the DIVA model to perceptual learning experiments, the following two assumptions can be made: First, the neurologically degraded speech serves as a teaching signal. Thus, listening to dysarthric speech excites the neural network in recognition of a potential teaching signal. Second, the fine tuning of perceptual representations of the target sound requires the production thereof. In other words, imitation of the teaching signal facilitates the learning of both speech perception and speech production of the target sound by linking the acoustic target with the somatosensory information required to produce that sound. The corresponding speech sound map, a by-product of this learning process, enables accurate imitation as well as identification and recognition of the target sound in a listening environment. Although the DIVA model considers auditory and somatosensory feedback equally important for the learning of speech sounds, somatosensory feedback has not yet been explored in the context of perceptual learning of neurologically degraded speech. However, there is some evidence to suggest that somatosensory information may constitute an important learning mechanism to improve perception of ambiguous speech.

Adank, Hagoort, and Bekkering (2010) showed that participants engaged in a training task involving imitation of an unfamiliar accent were significantly better at comprehending the accented speech in subsequent encounters. Although imitation of the speaker's accent was not measured acoustically, the use of a series of control conditions suggests that the perceptual benefit was consequential to the act of imitating vocally the precise pronunciation of the sentence (training instruction), rather than simply repeating the spoken stimuli in one's own accent or listening to auditory targets of the spoken stimuli. Moreover, a recent functional magnetic resonance imaging study in the same research laboratory using a similar experimental design found that perception and production areas of the brain are active after imitation but not after simple repetition of spoken stimuli in one's own accent (Adank, Rueschemeyer, & Bekkering, 2013). This suggests that vocal imitation improves not only intelligibility but also neuropathway activation. Although DIVA is a speech production model, and as such, does not make any predictions about or claims to explain speech perception, this finding is not inconsistent with DIVA's underlying neurophysiology. Indeed, Perkell (2012) mentions the phenomenon of *mirror neurons* (Watkins & Paus, 2004) as a functional part of the speech sound map and highlights its role in speech acquisition. Framing the finding of Adank et al. (2013) into the creation and activation of a mirror-neuron system allows for an interesting assumption, namely, that a mirror-neuron response only occurs during a listening experience aimed at deciphering an acoustic string of speech sounds once a speech sound map of that accent has been established (i.e., imitation vs. repetition). If true, this further supports the notion of a significant role of somatosensory information in the creation of mental representations to aid intelligibility of ambiguous speech. Thus, whether perceptual benefits can be accrued from imitation of disordered speech, as well as an exploration of the underlying mechanism, warrants scientific inquiry.

The purpose of the current study was to investigate whether ties to existing mental representations in the speech processing system can be strengthened by way of a somatosensory motor trace. Vocal imitation was used as a vehicle to examine this, specifically addressing the following research questions: (a) Does overt vocal imitation facilitate improved recognition of dysarthric speech? (b) Does the degree of overt imitation accuracy correlate with magnitude of intelligibility gains? Based on the DIVA model of speech production learning, it was hypothesized that vocal imitation would improve intelligibility of dysarthric speech, evidencing a clear role for somatosensory information in speech perception and perceptual learning. Further, in support of this mechanism, it was hypothesized that the greater the level of imitation accuracy, the greater the magnitude of perceptual benefit. Moderate spastic dysarthric speech, associated with bilateral upper motor neuron damage and underlying muscle hypertonia, was used as the entry point into investigations regarding the role of somatosensory information in perceptual learning of disordered speech.

Methods

Participants

One hundred young, healthy adults (64 women and 36 men) aged 19 to 37 years old (M = 25.48; SD = 4.19) participated in the experiment. All participants were native speakers of American English, with no history of speech, language, or cognitive disorders and had no known hearing problems. None of the participants had significant prior contact with persons having motor speech disorders. Participants were recruited from undergraduate classes at Utah State University as well as friends and family from the local Cache Valley community. Institutional review board consent was obtained from all participants.

Speech stimuli

Speech stimuli used in the current study were part of a larger set described in detail in an earlier article (Borrie, 2015). Relevant to the current study, 40 semantically plausible and 80 semantically anomalous phrases were elicited from a 26-year-old male native speaker of American English with dysarthric speech secondary to traumatic brain injury. The semantically plausible phrases (e.g., *the bread is stale* and *the nasty weather caused severe flooding*) ranged in length from four to eight words, containing between four and 12 syllables per phrase. This phrase set enabled the collection of data on selected acoustic variables of the speaker and also served as the training speech set in the experimental protocol. In contrast, the 80 semantically anomalous phrases (e.g., *amend estate approach* and *had eaten junk and train*), which served as the pretest and posttest speech sets, were more tightly controlled. All phrases in this set were syntactically plausible but semantically anomalous to reduce the influence of semantic and linguistic cues on word recognition (Liss, Spitzer, Caviness, Adler, & Edwards, 1998; Liss, Spitzer, Caviness, Adler, & Edwards, 2000). In addition, all phrases contained exactly six syllables, ranged in length from three to five words, and were balanced for phrasal stress patterns.

The speaker who provided the speech stimuli presented with a moderate spastic dysarthria, as diagnosed by three independent speech-language pathologists with expertise in assessment and diagnosis of motor speech disorders. Speech was characterized perceptually by monopitch, slow speaking rate, imprecise articulation, and a strained–strangled vocal quality—all of which are considered cardinal features of spastic dysarthria according to the Mayo Classification System (Darley, Aronson, & Brown, 1969a, 1969b; Duffy, 2005). For objective confirmation of some of the prominent perceptual features, the 40 semantically plausible phrases were also elicited from a 26-year-old male native speaker of American English, with no neurological history or presence of speech, language, or cognitive disturbances (control phrases).

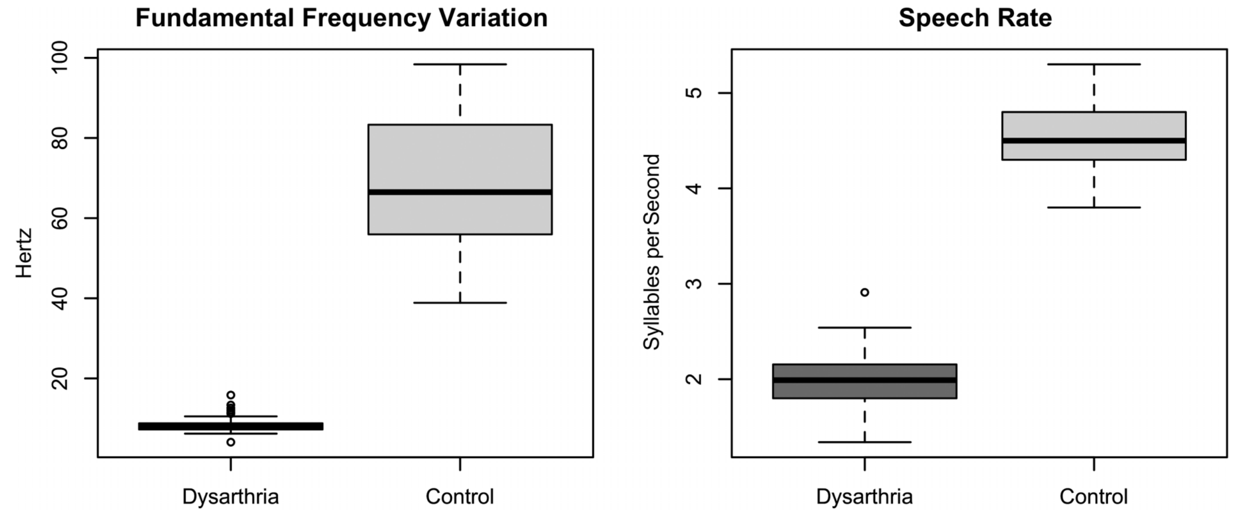
By using the acoustic analysis software, Praat (Boersma & Weenink, 2013), we were able to analyze the semantically plausible phrases produced by the speaker with dysarthria and the healthy control for two phrase-level measures: (a) fundamental frequency (F0) variation (standard deviation [in Hertz]) and (b) speech rate (in syllables per second). To perform this analysis, we selected the beginning and end point of each phrase by placing cursors before and after acoustic manifestations of speech on the spectrographic display. In addition, frequency traces were visually inspected to identify apparent anomalies, which were removed before analysis. Figure 1 shows box-and-whisker plots of F0 variation and speech rate of the dysarthric and control phrases. Paired t tests confirmed significant differences between the two acoustic measures of F0 variation, $t(39) = 22.92$, $p < .001$, $d = 2.91$, and speech rate, $t(39) = 37.90$, $p < .001$, $d = 8.35$. These findings confirmed that relative to the control phrases, the dysarthric phrases were characterized by reduced F0 variation (manifested perceptually as reduced pitch variation or monotone) and slow speaking rate. For the experimental protocol described in the following, only phrases produced by the speaker with dysarthria were used.

Procedure

The experiment was conducted in three phases: (a) pretest, (b) training, and (c) posttest. All participants were exposed to an identical pretest and posttest phase, but the training phase was different, depending on random assignment to one of the following five experimental groups ($n = 20$): (1) control (C), (2) auditory (A), (3) auditory-written (AW), (4) auditory-imitation (AI), or (5) auditory-written-imitation (AWI). Each participant attended a single experimental session held in a quiet laboratory, the Human Interaction Lab, at Utah State University. Upon obtaining informed consent, participants were seated in front of a computer preloaded with the experimental procedure and informed that task-specific instructions would be delivered via the computer program. This process was used to ensure identical stimulus presentation instructions across the 100 participants. All participants were fitted with sound-attenuating headphones (Sennheiser HD 650 PRO, Old Lyme, CT). In addition, participants assigned to the imitation groups (AI and AWI) were fitted with a lavalier microphone connected to a portable digital recorder (Zoom H4N, Ronkonkoma, NY). Speech stimuli across all phases were presented binaurally through the headphones at a comfortable listening level of 65 dB, and standard recording settings (48 kHz; 16-bit sampling rate) were used. Participants were instructed to begin the experiment by pressing the return key on the keyboard when they were ready to start.

The experiment began with the pretest phase for all participants, regardless of experimental group assignment. During the pretest phase, participants were informed that they would be presented with short phrases produced by someone with a speech disorder. They were told that the phrases all contained real English words but would not make sense (e.g., *darker painter baskets*). Phrases of the pretest speech set were presented one at a time, and following each presentation, participants were instructed to use the keyboard to type out exactly what they thought was being said. Participants were encouraged to make a guess at any words they did not recognize and use an *X* to represent any part of a provided phrase where a guess could not be made. Once they had finished typing their response, participants were prompted to press the return key to move on to the next phrase. The presentation order of the pretest speech set was randomized across all participants.

Figure 1. Comparison of acoustic features by speech type: From left to right, the panels reflect significant differences between fundamental frequency variation and speech rate of semantically plausible phrases produced by a speaker with dysarthria and an age- and gender-matched healthy control.



Immediately following the pretest phase, participants in all but the C group actively engaged in a training phase. Although participants in the A, AW, AI, and AWI groups were trained on the same stimuli (i.e., the training speech set), groups were differentiated by the nature of the training task. Participants in all four groups were told that they would hear a series of phrases produced by a person with a speech disorder. They were told that the phrases would make sense (e.g., *the bakery opened early*) and that they would need to listen closely to each phrase and try to understand what was being said. Participants in the A group did not receive any further instructions; their task was to simply listen to the auditory productions. Participants in the AW group were told that written subtitles would be provided on the screen and that they should use this information to help them understand what was being said. Participants in the AI group were told to repeat each phrase back, as best as possible, in the same manner (i.e., speaking rate, intonation, stress, and pronunciation) it was produced. Similar to the AW group, the AWI group was told to use the written subtitles provided on the screen to help them understand what was being said. In addition, however, and like the AI group, they were asked to repeat each phrase back, as best as possible, in the same manner it was produced. Imitation attempts, termed *training response productions*, were recorded for participants in the AI and AWI groups. All participants were prompted to press the return key when they had finished their task and were ready to move on to the next phrase.

The experiment ended with a posttest phase for all participants, regardless of experimental group assignment. This phase was identical to the pretest phase, with the only difference being the use of novel speech stimuli, the posttest speech set. Upon completion of the posttest phase, participants were thanked and debriefed before exiting the lab.

File analysis

The total data set consisted of (a) 100 participant transcripts of the pretest speech set, (b) 100 participant transcripts of the posttest speech set, and (c) 40 participant training response productions (.wav files) from participants assigned to the AI and AWI experimental groups. Participant transcripts were analyzed for a standard measure of speech intelligibility, percent words correct (PWC). As per previous studies that have examined perception of dysarthric speech (e.g., Borrie, McAuliffe, Liss, Kirk, et al., 2012; Liss, Spitzer, Caviness, & Adler, 2002), words were counted correct if they accurately matched the intended target, or differed only by tense (*-ed*) or plurality (*-s*). Substitutions between *a* and *the* were also counted as correct. Two PWC scores were tabulated for each participant based on the transcripts of pretest and posttest stimuli. These scores reflect a measure of intelligibility for each participant before and after the training phase. An intelligibility improvement score was also calculated for each participant by subtracting pretest PWC scores from posttest PWC scores.

Training response productions from participants engaged in imitation were analyzed acoustically for a measure of F0 variation and speech rate, with identical procedures and definitions to those applied in the acoustic analysis of the speech stimuli used in the experiment. F0 variation and speech rate of each training production were then compared with the same metrics from the corresponding dysarthric stimuli production. That is, the difference (or distance) between the two phrase productions was calculated by subtracting the acoustic measure obtained from the dysarthric production from the same acoustic measure obtained from the participant's training response production. Thus, the closer a difference score was to 0, the more accurately the participant imitated the dysarthric stimuli production. As there were 40 training stimuli, this resulted in 40 difference scores for F0 variation and 40 difference scores for speech rate per participant. These difference scores were then transformed to absolute values and averaged for each participant to form a single difference score, labeled the *imitation accuracy score,* and used for statistical analysis.

Reliability analysis

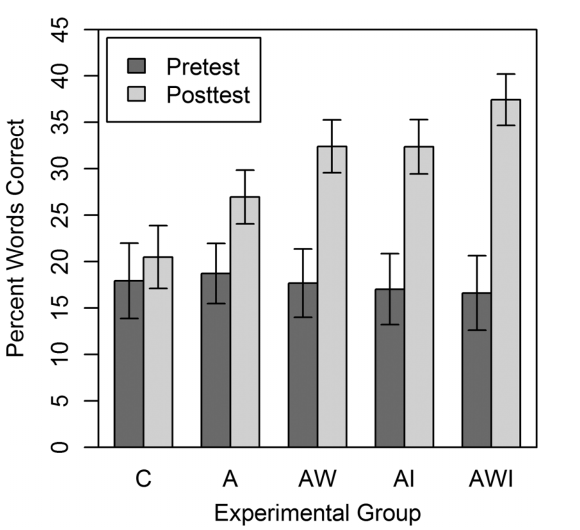
Twenty-five percent of the total data set (50 transcripts and 10 audio recording sets) were randomly selected according to computer-generated random number lists and reanalyzed by the original judge (intrajudge) and a second trained judge (interjudge) to obtain reliability estimates for the dependent variables of PWC (transcripts), and acoustic measures of F0 variation and speech rate (audio recording sets). Reliability analysis confirmed that the agreement rate between the reanalyzed data and the original data was high (all correlations $r > .97$).

Results

Intelligibility

Intelligibility data, expressed by the mean PWC scores of the pretest and posttest speech sets, were collected for each participant across the five experimental groups. To test the effect of test type (pretest and posttest) and experimental group (C, A, AW, AI, and AWI) on intelligibility of dysarthric speech, PWC data was analyzed by using a two-factor mixed design analysis of variance. The within-participant factor was test type and the between-participant factor was experimental group. Figure 2 provides a summary of these results.

Figure 2. Mean percent words correct for listeners by experimental group (n = 20) at pretest and posttest. Error bars delineate +1 *SD* of the mean. Note that C, A, AW, AI, and AWI refer to training type defined as control, auditory, auditory-written, auditory-imitation, auditory-written-imitation, respectively.



The analysis of variance showed a statistically significant interaction between test type and experimental group on intelligibility of dysarthric speech, $F(4,95) = 123.19$, $p < .001$, partial $\eta^2 = .84$. Therefore, the simple main effects of group were used to establish differences in PWC scores between the five experimental groups at each level of the within-participant factor test type. As there are two levels of the within-participant factor, two separate tests were carried out. For the pretest data, there was no statistically significant difference in PWC scores between experimental groups, $F(4, 95) = 1.14$, $p = .34$, partial $\eta^2 = .05$. Thus, participants across all experimental groups performed similarly on the intelligibility test prior to training. For the posttest data, however, there were statistically significant differences in PWC scores between experimental groups, $F(4, 95) = 35.69$, $p < .001$, partial $\eta^2 = .61$. Thus, participants performed differently on the intelligibility test post-training, with differences mediated by experimental group.

Follow-up post hoc independent t tests with Bonferroni correction demonstrated that the posttest PWC scores of the AWI group were significantly higher than those of the AI group, $t(38) = 3.50$, $p = .001$, $d = 1.11$, the AW group, $t(38) = 3.26$, $p = .002$, $d = 1.03$, the A group, $t(38) = 6.21$, $p < .001$, $d = 1.96$, and the C group, $t(38) = 11.50$, $p < .001$, $d = 3.63$. Thus, participants in the AWI group realized the greatest perceptual benefit from training. Posttest PWC scores between the AI group and the AW group were not significantly different from one another, but t tests revealed that the posttest PWC scores of the AI group were significantly higher than those of the A group, $t(38) = 3.38$, $p = .002$, $d = 1.07$, and the C group, $t(38) = 8.66$, $p < .001$, $d = 2.74$, and similarly, the posttest PWC scores of the AW group were significantly higher than those of the A group, $t(38) = 3.30$, $p = .002$, $d = 1.04$, and the C group, $t(38) = 8.25$, $p < .001$, $d = 2.61$. Thus, the addition of an imitation task or the presence of written information during training afforded a perceptual benefit over and above that achieved by the auditory signal alone; however, neither of these two information sources was more beneficial than the other in isolation. As a final point, the posttest PWC scores of the A group were significantly higher than those of the C group, $t(38) = 4.01$, $p < .001$, $d = 1.23$, validating a clear effect of perceptual training with dysarthric speech.

Simple main effects were also conducted to test for differences in intelligibility of dysarthric speech between pretest and posttest for each level of the between-participant factor, experimental group. As there were five experimental groups, five separate paired-samples t tests with Bonferroni correction were carried out. For the C group, there was no significant difference between pretest and posttest PWC scores. However, significant differences were found in the other four experimental groups, A, $t(19) = 31.69$, $p < .001$; AW, $t(19) = 22.47$, $p < .001$, $d = 3.52$; AI, $t(19) = 16.46$, $p < .001$, $d = 3.86$; and AWI, $t(19) = 27.87$, $p < .001$, $d = 5.12$. Thus, intelligibility of dysarthric speech significantly improved for listeners in the four experimental groups that received training with the disordered speech; however, this improvement was not evident for the control group who received no training.

Imitation accuracy

For the participants assigned to the two experimental groups that included an imitation component (AI and AWI), Pearson correlation coefficients were used to examine whether imitation accuracy[1] was related to intelligibility improvement. The results revealed that imitation accuracy of both F0 variation, $r(38) = -.77$, $p < .001$, and speech rate, $r(38) = -.69$, $p < .001$, was significantly correlated with intelligibility improvement. These relationships are illustrated in Figure 3. Thus, these results demonstrate that participants who were most accurate in emulating the F0 variation and speech rate of the disordered speech stimuli achieved larger intelligibility gains than participants who were less accurate at emulating these degraded acoustic-prosodic properties.

Discussion

Training with dysarthric speech improved subsequent recognition of dysarthric speech; however, the perceptual benefits afforded by training were greatly enhanced when the degraded auditory signal was accompanied by additional sensory input. Consistent with previous research by Borrie and colleagues (Borrie, McAuliffe, Liss, Kirk, et al., 2012; Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2012), recognition of dysarthric speech was superior when the training phase included written transcripts of the spoken targets. Novel, however, is that the most significant intelligibility gains were realized when training included overt vocal imitation of the degraded speech. Moreover, a robust relationship between intelligibility improvement and imitation accuracy was identified. Taken together, these results support a relationship between somatosensory information and the mechanisms that underlie perceptual learning of dysarthric speech.

Using the DIVA model of speech production (Guenther et al., 1998) as a theoretical framework, the results of the current study confirmed the following assumptions regarding the proposed link between speech perception and speech production. First, participants were able to identify dysarthric speech as a teaching signal. This is consistent with DIVA's perspective on speech sound acquisition; once the cognitive system is developed enough to identify a speech sound consistently as part of the language repertoire, an infant uses the external signal as a teaching signal (Guenther, 1995). In other words, the infant learns to establish a link between somatosensory feedback and its auditory perceptual consequences; particularly, he or she uses trial-and-error feedback from imitation attempts to make necessary articulatory adjustments to better match the target sound. In the present study, participants would have used similar feedback mechanisms during the imitation task in their attempt to more closely approximate the degraded speech signal.

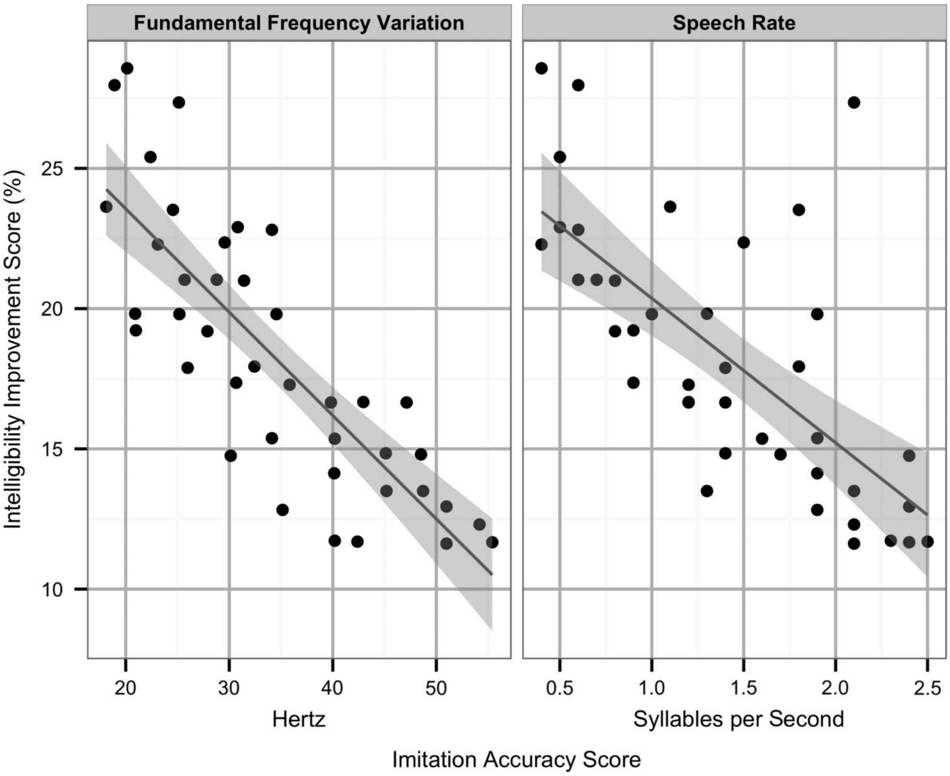
Second, imitation of dysarthric speech strengthened the mental representation of dysarthric speech through linking auditory and somatosensory reference frames to a speech sound map. Note that mental representations of dysarthric speech are not assumed to replace or adapt existing speech sound maps, but instead, represent a separate entity within the neural network. In DIVA, the speech sound map refers to a set of cells, each one representing a different phoneme, high-frequency multiphonemic syllable, or word (Guenther, Gosh, & Tourville, 2006). This speech sound map typically only becomes active after the speech recognition system of the model is able to identify a signal as an auditory-perceptual target (see first assumption). Once the auditory target region of a speech sound has been tuned through repeated activation of the speech sound map upon recognition of a teaching signal, reliable mappings between reference frames for the accurate production of the target sound can be established. These consistent mappings are then used to initiate feedforward commands to articularly realize the auditory-perceptual target (Golfinopoulos et al., 2010; Guenther, 2006). Although the DIVA model has not been computationally applied to the learning of neurologically degraded speech, the results of the present study suggest that speech sound maps of dysarthric speech were established, as witnessed in both improved intelligibility and closer acoustic approximations.

In addition, the results of the current study are in support of the proposed hypothesis that imitation of a degraded speech signal improves the perception thereof. This finding cannot be embedded in the DIVA model in its current form as it only accounts for speech production and, therefore, proposes a one-directional effect of auditory and somatosensory feedback mechanisms on the speech sound map (Golfinopoulos et al., 2010; Guenther, 2006). The concept and neurophysiological location of a mirror-neuron system is, however, reflected in the speech sound map and its activation pattern during perception and production of a speech sound. Because DIVA is a speech production model, this bidirectional activation is limited to speech sound acquisition. In contrast, results of the present study and those of Adank and colleagues (2010, 2013) suggest a bidirectional effect of imitation beyond speech sound acquisition, namely, that speech production also informs speech perception. In particular, the information obtained from imitation improves decoding of the neurologically disordered speech signal in a subsequent transcription task, presumably by activation of perceptual and motor areas of the brain that reflect the auditory and somatosensory information of the established speech sound map.

The idea that sensorimotor experiences create and refine a mirror-neuron response is not a new one. Catmur (2013) and others have noted that observation of an action in another person activates the same motor program in the observer, and this motor program is formed into a mirror-neuron system when contingent or predictive relationships

[1]Given that imitation accuracy is the discrepancy between the training response production (imitation attempt) and the dysarthric stimuli production, numbers closer to 0 reflect greater imitation accuracy.

Figure 3. Imitation accuracy and intelligibility improvement: From left to right, the panels reflect significant correlations between listener intelligibility improvement and imitation accuracy scores of fundamental frequency variation and speech rate.



between the observed and performed action is warranted. The important contribution of Adank et al. (2013) was that this neuropathway activation during a subsequent intelligibility task was observed after imitation, but not repetition, of a novel accent. In accordance with neural pathway activation, participants engaged in the imitation task also demonstrated significantly greater intelligibility gains than those engaged in the repetition task. This suggests that existing speech sound maps are not sufficient to aid improved intelligibility of unfamiliar accented speech. Instead, specific sensorimotor experiences are required to ensure the intelligibility gain. This is consistent with recent perceptual theories that assume that separate generative models for different speakers, accents, and groups are formed to ensure faster and more robust language processing (Kleinschmidt & Jaeger, 2015).

The current study cannot make any inferences about neuropathway activation; however, it adds further support for the benefit of imitation over repetition for the perceptual learning of dysarthric speech. Although a specific repetition group was not included, objective evidence of the equivalent finding can be drawn from the acoustic analysis of training response productions. That is, participants who most accurately emulated the acoustic features of the degraded speech signal also showed greater intelligibility gains post-training. In contrast, smaller perceptual benefits were observed in participants who performed poorly at imitating the acoustic blueprint of dysarthric speech, which more or less equates to them repeating, rather than imitating, the spoken stimuli. These differences in the robustness of mental representations suggest that the more somatosensory information of the degraded signal was realized and mapped onto a speech sound map, the more successful subsequent decoding of neurologically degraded speech became. Hence, the most important contribution of the current investigation is the confirmation that perceptual learning can be strengthened by way of a somatosensory motor trace.

Important questions that arise in this context are the following: (a) Why did some people imitate the dysarthric productions more accurately than others? (b) Can those individuals who did not imitate as well be trained to better emulate degraded acoustics? Perkell, Guenther, et al. (2004) and Perkell, Matthies, et al. (2004) established a link between speech perception and speech production by which better auditory acuity was related to greater articulatory contrast. The authors found that the articulatory target region of two neighboring consonants was smaller and more distinct in individuals who showed better discrimination of the two sounds. In contrast, speech sound productions of individuals with poor discrimination skills were less distinct and more variable as they had larger target regions. Because these studies used the DIVA model of speech production as the underlying theoretical framework, the interpretation of these results was also one-directional. That is, more acute speech perception results in more precise speech production. For the present study, this may imply that participants who more closely matched the target sounds were able to identify, and thus replicate, more distinct differences in the acoustic signal. This may warrant further exploration of perceptual training modules for those with smaller intelligibility gains, such as the discrimination of different speech production parameters to improve imitation accuracy. As indicated earlier, however, improved intelligibility following imitation is indicative of a bidirectional learning mechanism. Thus, perceptual learning modules to improve imitation accuracy would ultimately serve the utilization of somatosensory information to strengthen the activation of speech sound maps of dysarthric speech.

Note that the greatest intelligibility gains were observed when training involved all sensory feedback channels; that is, auditory feedback from the speech signal, somatosensory information from the imitation, and lexical information from the written transcript. Although the DIVA model does not include written feedback, learning within this theoretical framework is conditional upon linking auditory and somatosensory reference frames to a speech sound map. Given that dysarthric speech of moderate severity is inherently ambiguous, the identification of a speech sound as a teaching signal is likely to be difficult. Hence, written feedback can serve to disambiguate the degraded acoustics, allowing for the sensory linking mechanism to a speech sound map to be strengthened. This indicates that disambiguating the degraded acoustics is key to the relative ease and accuracy with which the speech sounds are mapped in the mental lexicon.

This finding also implies that speech signals with less severe acoustic degradation (i.e., more intelligible at baseline) may not require the same level or combination of external (written) and internal (auditory and somatosensory) feedback to establish robust mental representations of disordered speech. Future studies involving speech disorders of different types and severities will be required to further examine this speculation. In addition, future studies are needed to establish whether novel mental representations can continue to be accessed at later stages (i.e., 12 hr, 1 week, or 3 months post-training), as shown in similar intelligibility gains, and if so, what combination of sensory stimuli afford the most robust results. Likewise, investigating the value of follow-up sessions to ensure long-lasting adjustments in the speech perception system would further enhance our knowledge about how best to maintain perceptual learning gains over time. We know that when perceptual training involves just written feedback, intelligibility gains remain significantly greater than baseline data (albeit some decay) at 1-week follow-up testing (Borrie, McAuliffe, Liss, Kirk, et al., 2012). Examining the longevity of learning following imitation-based training paradigms would advance our mechanistic account of somatosensory-based perceptual learning as well as the clinical application of this work.

While the results of this study expand our theoretical understanding of perceptual learning of disordered speech, they also hold significant clinical value. That is, they offer insight into training conditions that yield the greatest intelligibility gains for listeners of dysarthric speech—an important step to developing a perceptual learning approach to the management of dysarthria (Borrie, McAuliffe, & Liss, 2012). Indeed there is now theoretical support for the inclusion of both written information and an associated sensorimotor task in the development of a listener-targeted treatment protocol to reduce the intelligibility burden of dysarthria. Future studies in our lab will target the development (and testing) of a novel treatment tool that will focus on the listener and the mechanisms that underpin improved speech perception. Borrie, McAuliffe, and Liss (2012) have already highlighted that treatment focusing on the neurologically intact listener (e.g., family members, friends, and health care professionals) to improve intelligibility of disordered speech may prove key to optimizing management of dysarthria, given that physical, cognitive, and memory deficits frequently co-occur in this clinical population. In general, the current finding of the role of somatosensory information in speech perception is also applicable to the development of rehabilitation programs for adaptation to hearing devices and cochlear implants, as well as programs that promote second language learning and improved understanding of foreign-accented speech.

Conclusion

That overt vocal imitation improved listener recognition of disordered speech suggests that ties to the mental representation of the lexicon can be strengthened by way of a somatosensory motor trace. The findings, therefore, implicate a bidirectional relationship between speech perception and speech production as well as advance our understanding of the mechanisms that underlie perceptual learning of degraded speech. The findings contribute to a robust theoretical framework that supports the development of a novel clinical treatment tool, one that exploits perceptual learning as a means of reducing the intelligibility burden of dysarthria.

Acknowledgments

This research was supported by the Utah State University Research Catalyst Grant FY15 (awarded to the first author). Research assistants in the Human Interaction Lab (Utah State University) are gratefully acknowledged for their assistance with data collection and transcript analysis.

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science, 21,* 1903–1909.

Adank, P., Rueschemeyer, S. A., & Bekkering, H. (2013). The role of accent imitation in sensorimotor integration during processing of intelligible speech. *Frontiers in Human Neuroscience, 7,* 634. doi:10.3389/fnhum.2013.00634

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 2.0.08) [Computer software]. Retrieved from http://www.praat.org

Borrie, S. A. (2015). Visual information: A help or hindrance to perceptual processing of dysarthric speech. *The Journal of the Acoustical Society of America, 137,* 1473–1480.

Borrie, S. A., McAuliffe, M. J., & Liss, J. M. (2012). Perceptual learning of dysarthric speech: A review of experimental studies. *Journal of Speech, Language, and Hearing Research, 55,* 290–305.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., Kirk, C., O'Beirne, G. A., & Anderson, T. (2012). Familiarisation conditions and the mechanisms that underlie improved recognition of dysarthric speech. *Language and Cognitive Processes, 27,* 1039–1055.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. (2012). A follow-up investigation into the mechanisms that underlie improved recognition of dysarthric speech. *The Journal of the Acoustical Society of America, 132,* EL102–EL108.

Borrie, S. A., McAuliffe, M. J., Liss, J. M., O'Beirne, G. A., & Anderson, T. (2013). The role of linguistic and indexical information in improved recognition of dysarthric speech. *The Journal of the Acoustical Society of America, 133,* 474–482.

Catmur, C. (2013). Sensorimotor learning and the ontogeny of the mirror neuron system. *Neuroscience Letters, 540*(1), 21–27.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969a). Differential diagnosis patterns of dysarthria. *Journal of Speech and Hearing Research, 12,* 246–269.

Darley, F. L., Aronson, A. E., & Brown, J. R. (1969b). Clusters of deviant speech dimensions in the dysarthrias. *Journal of Speech and Hearing Research, 12,* 262–496.

Duffy, J. R. (2005). Motor speech disorders: Substrates, differential diagnosis, and management (2nd ed.). St. Louis, MO: Elsevier Mosby.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67,* 224–238.

Francis, A. L., Nusbaum, H. C., & Fenn, K. (2007). Effects of training on the acoustic phonetic representation of synthetic speech. *Journal of Speech, Language, and Hearing Research, 50,* 1445–1465.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large scale neural network modeling and functional brain imaging in speech motor control. *Neuroimage, 52,* 862–874.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review, 102,* 594–621.

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders, 39,* 350–365.

Guenther, F. H., Gosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96,* 280–301.

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105,* 611–633.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review, 122,* 148–203.

Liss, J. M., Spitzer, S. M., Caviness, J. N., & Adler, C. (2002). The effects of familiarization on intelligibility and lexical segmentation in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 112,* 3022–3030.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. (1998). Syllabic strength and lexical boundary decisions in the perception of hypokinetic dysarthric speech. *The Journal of the Acoustical Society of America, 104,* 2457–2466.

Liss, J. M., Spitzer, S. M., Caviness, J. N., Adler, C., & Edwards, B. (2000). Lexical boundary error analysis in hypokinetic and ataxic dysarthria. *The Journal of the Acoustical Society of America, 107,* 3415–3424.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics, 25,* 382–407.

Perkell, J. S., Guenther, F. H., Lane, H., Matthies, M., Stockmann, E., Tiede, M., & Zandipur, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts. *The Journal of the Acoustical Society of America, 116,* 2338–2344.

Perkell, J. S., Matthies, M. L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., . . . Guenther, F. H. (2004). The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect. *Journal of Speech, Language, and Hearing Research, 47,* 1259–1269.

Watkins, K. E., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of Broca's area. *Journal of Cognitive Neuroscience, 16,* 978–987.