Original Research

Cross-Modal Somatosensory Repetition Priming and Speech Processing

Aravind K Namasivayam[1,2,*], Tina Yan[1], Rohan Bali[1], Deborah Hayden[3], Pascal van Lieshout[1,4]

[1]Department of Speech-Language Pathology, University of Toronto, Toronto, ON M5G1V7, Canada
[2]Speech Research Centre Inc., Toronto, ON L7A2T1, Canada
[3]The PROMPT Institute, Santa Fe, NM 87505, USA
[4]Rehabilitation Sciences Institute, University of Toronto, Toronto, ON M5G1V7, Canada
*Correspondence: a.namasivayam@utoronto.ca (Aravind K Namasivayam)
Academic Editor: Gernot Riedel
Submitted: 18 April 2022   Revised: 19 June 2022   Accepted: 22 June 2022   Published: 19 August 2022

Abstract

Background: Motor speech treatment approaches have been applied in both adults with aphasia and apraxia of speech and children with speech-sound disorders. Identifying links between motor speech intervention techniques and the modes of action (MoA) targeted would improve our understanding of how and why motor speech interventions achieve their effects, along with identifying its effective components. The current study focuses on identifying potential MoAs for a specific motor speech intervention technique. **Objectives**: We aim to demonstrate that somatosensory inputs can influence lexical processing, thus providing further evidence that linguistic information stored in the brain and accessed as part of speech perception processes encodes information related to speech production. **Methods**: In a cross-modal repetition priming paradigm, we examined whether the processing of external somatosensory priming cues was modulated by both word-level (lexical frequency, low- or high-frequency) and speech sound articulatory features. The study participants were divided into two groups. The first group consisted of twenty-three native English speakers who received somatosensory priming stimulation to their oro-facial structures (either to labial corners or under the jaw). The second group consisted of ten native English speakers who participated in a control study where somatosensory priming stimulation was applied to their right or left forehead as a control condition. **Results**: The results showed significant somatosensory priming effects for the low-frequency words, where the congruent somatosensory condition yielded significantly shorter reaction times and numerically higher phoneme accuracy scores when compared to the incongruent somatosensory condition. Data from the control study did not reveal any systematic priming effects from forehead stimulation (non-speech related site), other than a general (and expected) tendency for longer reaction times with low-frequency words. **Conclusions**: These findings provide further support for the notion that speech production information is represented in the mental lexicon and can be accessed through exogenous Speech-Language Pathologist driven somatosensory inputs related to place of articulation.

Keywords: somatosensory; multi-sensory interaction; speech perception; speech production; lexical access; speech reaction time; motor speech intervention; PROMPT intervention

1. Introduction

Motor speech treatment approaches have been recommended and applied in both adults with aphasia and apraxia of speech and children with speech-sound disorders [1,2]. There has been a steady increase in the number of studies testing the efficacy of these interventions in these populations and the findings indicate that the effects of these interventions continue to be variable (e.g., [3,4]) and, in some cases, not maintained long-term (e.g., [5]). Continued progress in the development of more effective motor speech interventions could be achieved by building a better understanding of the "mode of action" (MoA), by which these interventions affect change [6,7]. A mode of action (MoA) refers to a functional (or anatomical) change in an organism resulting from exposure to a substance or event. This is distinguished from the mechanism of action (MOA), which describes these changes at a molecular level [8]. Identifying links between motor speech intervention techniques and the MoAs they target would improve our understanding of how and why motor speech interventions achieve their effects and what components (or "active ingredients") are more likely to be effective. The current study focuses on identifying potential MoAs for a specific motor speech intervention technique.

Specific techniques have been utilized within motor speech intervention approaches to aid sensory-motor learning of speech-motor targets, such as the Prompts for Restructuring Oral Muscular Phonetic Targets (PROMPT; e.g., [1,6]) and Dynamic Temporal and Tactile Cueing (DTTC; e.g., [2,7]). These techniques aim to enhance somatosensory inputs to facilitate the formation of sensory-motor pathways required for the acquisition and accurate production of speech. In the current study, the broad term "somatosensory information" was chosen because in a clinical motor speech intervention context, there is no empirical data to indicate which specific type of somatosensory information (touch, pressure, pain, temperature, position, movement, and vibration) is activated or specifically contributes to clinical improvements. This is in line with the contemporary use of this term in (non-intervention related) speech

motor control research [9,10]. For a review of somatosensory innervation in humans see Moayedi *et al.* [11]. Typically, in speech motor interventions somatosensory inputs involve multisensory (tactile, kinesthetic, proprioceptive, auditory and visual) cueing (or exogenous inputs) to emphasize place, voice and manner of speech production and slowing down speech rate to possibly enhance proprioceptive feedback and to highlight movement transitions [1,7].

There is evidence to suggest that endogenous somatosensory feedback is involved in the acquisition and maintenance of orosensory-to-auditory mapping (DIVA model; [12]) and for stability in speech-motor production [13]. However, there is sparse literature on the importance and role of such inputs delivered via an external agent (e.g., Speech-Language Pathologist [S-LP]) during motor speech interventions. In a study by Dale *et al.* [14], PROMPT treatment with and without somatosensory inputs was examined in a group of children with Childhood Apraxia of Speech (CAS). Their results demonstrated that PROMPT treatment with tactile input enhanced the production of untreated probe words relative to treatment without tactile input [14]. This finding suggests that somatosensory input can be considered an "active ingredient" (i.e., something that must be present for the intervention to be effective). Nonetheless, very little is known about the underlying mechanisms by which such inputs improve speech production.

PROMPT is a sensory-motor treatment approach often recommended and applied in children with speech sound disorders [15], autism [16], and cerebral palsy [17]. During PROMPT intervention, motor speech goals or treatment priorities are based on the notion of a hierarchical development of speech subsystems known as the Motor Speech Hierarchy [MSH; [1]]. Within this developmental MSH framework, speech production is thought to be the result of interactive development of seven key motor speech subsystems (i.e., Stage I: tone, Stage II: phonatory control, Stage III: mandibular control, Stage IV: labial-facial control, Stage V: lingual control, Stage VI: sequenced movements, and Stage VII: prosody). The PROMPT intervention enables the hierarchical establishment, refinement, and integration of normalized synergistic movement patterns within these speech subsystems through the application of somatosensory inputs or prompts to the client's face. As PROMPT treatment systematically progresses from the lower- to higher-levels of the hierarchy, speech production variables like place, manner, movement transitions, and timing of speech movements are expected to improve.

In addition to its use in children, PROMPT has also been applied to a limited extent on adults with neurological impairments. Four case studies in the literature suggest that PROMPT may improve the accuracy of words and phrases in acquired apraxia of speech [18–21]. For example, in both Square *et al.* [21] and Freed *et al.* [19] studies, participants with acquired apraxia of speech and aphasia were first presented with an auditory model of a word or functional phrase, and participants were required to repeat the target according to the auditory model. If participants could produce the utterance correctly, the next trial would be initiated. However, if participants were unable to repeat the model accurately, then the participants were first instructed not to respond while the clinician "mapped-in" (or primed) the correct sequence of motor patterns using somatosensory inputs for the target utterance. Following this, the clinician would direct the participants to produce aloud the target phoneme, word, or phrase again, during which the clinician simultaneously used somatosensory inputs to facilitate the appropriate movements. In general, their results indicated improved speech sequencing and precision of speech movements for both trained and untrained target words and sentences.

The findings from these studies suggest that the somatosensory inputs applied by a trained clinician may increase oro-facial sensory awareness for movement trajectories and sequences, thereby improving spatial-temporal aspects of speech movements [18–21]. It is unknown whether these somatosensory inputs are also relevant for the cognitive-linguistic system (e.g., during lexical processing). There is some evidence from the speech motor control literature to indicate that somatosensory information may be used by the speech production/perception system beyond the level of simply increasing oro-facial awareness. In fact, somatosensory information may affect the lexicality of phoneme perception [9,22,23]. In these studies, modulation of speech perception has been observed because of stretching the facial skin around the mouth area [9] or stimulating mechanoreceptors with air puffs on the back of the hand, the centre of the neck, or on the subject's ankles [23].

More recently, Ogane *et al.* [10] demonstrated that somatosensory inputs influence processes at the level of word segmentation and lexical decision. They utilized an auditory identification test using a set of French phrases, which may be segmented differently depending on the placement of accents (i.e., based on acoustic prosodic cues like local pitch increases) within the phrase. They reasoned that since creating an acoustic-prosodic accent may be achieved by hyper-articulation [24], the cues for word segmentation could be obtained from not only acoustic but also from articulatory sensory information. They applied somatosensory stimulation (using a robotic device to stretch the skin at the sides of the participant's mouth) at different points in the utterance, which had been recorded with neutral accents (i.e., all syllables had a similar emphasis). They found that lexical decisions relating to word segmentation were systematically and significantly biased depending on the timing of the somatosensory stimulation. This effect was not found when somatosensory stimulation was applied to the skin of the participant's forearm. These findings suggest that neural speech processing broadly integrates event-related information across multiple modalities (i.e., auditory, tactile, vi-

sual), and perceptual-sensory-motor relationships may extend to the processing of lexical information in the human brain [10].

Ito *et al.* [9] have pointed out that such modulation of speech-perception may arise due to connections between facial somatosensory inputs and facial motor and premotor areas [25,26], or due to bidirectional linkages between the somatosensory and auditory cortices [27–30], and possibly as a result of auditory-somatosensory interaction at the subcortical superior colliculus level [31]. Other studies have found that multisensory input involving visual, auditory, and somatosensory stimuli show connections into the ventrolateral prefrontal cortex (VLPFC) and specifically converges in areas 47/12, as well as area 45 [32]. Area 45 (and 44) of the inferior frontal gyrus (commonly referred to as Broca's areas) is thought to be involved in myriad of motor and cognitive functions including lexical processing [33] and has also been associated with word frequency effects [33,34].

Word frequency effects have been demonstrated to be important in several areas of speech processing. Studies have shown that during the process of lexical (word) selection and encoding of sound structure, high lexical frequency facilitates processing, possibly due to their higher activation levels and strength of neural connections (e.g., [35]). This is reflected in decreased reaction times and higher accuracy (i.e., fewer semantic, phonological, articulatory errors and/or omissions) for high-frequency words relative to low frequency words [36]. In contrast, low-frequency words result in slower response times relative to high-frequency words, which indicates longer processing times for low-frequency words ([37]; for a review in adult aphasia literature, see Kittredge *et al.* [36]). Other evidence using event-related potentials (ERPs) have also shown larger N400 waves when processing low-frequency words relative to high-frequency words, further suggesting that low-frequency words require more extensive processing [38,39].

Apart from processing times, studies have also reported word frequency effects on speech articulation (see Bell *et al.* [40] for a review). For example, low-frequency words are produced with longer durations and more extreme articulatory and acoustic properties [40,41]. Such durational effects have been explained by models of speech and language processing with proposed interactions between lexical, phonological, and phonetic structures (e.g., [42]). Within these models, its argued that the pace of articulation/speech production may be moderated when processing (e.g., phonological encoding) is slowed as in the case of low-frequency words. It is argued that this facilitates the maintenance of temporal coordination between the lexical and articulatory aspects of speech [40]. Furthermore, under such accounts, word form representations are said to include a specification of the range of phonetic variation associated with the target forms. It is suggested

that speakers of a language may learn to associate high-frequency lexical items with a greater degree of phonetic variation relative to low-frequency items, a phenomenon referred to as lexically-conditioned phonetic variation [40]. This results in an activation of a representation specifying a narrow range of phonetic variation during the encoding of the sound structure of low-frequency words. Thus, low-frequency words are said to influence phonetic processing more strongly than high-frequency words [42]. Due to this, it is hypothesized that the processing of low-frequency words may benefit with not only more but also narrowly specified somatosensory input relative to high-frequency words. Such a finding would suggest that part of the effect of these somatosensory cues may occur at a level of cognitive-linguistic processing that involves interactions between lexical access and phonetic processing [40,42]. Clinically, it is important to understand the level at which exogenous clinician induced somatosensory information is integrated into the speech-production/perception system, as this may have implications for the speech training in children and in the rehabilitation of adult brain-injury patients (e.g., those with word-finding/word-retrieval issues in aphasia). It is likely that somatosensory input may provide a dual benefit wherein it may help improve speech motor patterns and, at the same time, facilitate the retrieval of lexical information.

Purpose of Study

In the current study, we aim to demonstrate that somatosensory inputs can influence lexical processing, thus providing further evidence that linguistic information stored in the brain and accessed as part of speech perception processes encodes information related to speech production. To this end, in a cross-modal repetition priming paradigm, we examined whether the processing of external somatosensory priming cues was modulated by both word-level (lexical frequency) and speech sound articulatory features, suggesting a broader role for speech production information in speech perception.

Participants were assigned to either a main experimental study or a control study. In both studies, participants listened to and repeated words (auditory targets) embedded in white noise following the presentation of congruent and incongruent somatosensory inputs (somatosensory primes) by a trained clinician. In the main experimental study, the participants received the somatosensory primes on their oro-facial structures (either to labial corners or under the jaw) while participants in the control study received the stimulation on their right or left forehead (control condition). In the experimental study, we expect that only congruent inputs (i.e., when somatosensory prime matches auditory targets) would yield the shortest speech-reaction time (SRT) and the highest phoneme recognition scores. For incongruent inputs, we expect an attenuation effect with longer reaction times and lower phoneme recognition

scores. Importantly, as the processing of low-frequency words may benefit with not only more but also narrowly specified somatosensory input relative to high-frequency words [42] we expect a stronger priming response in low-frequency words. In the control study, we do not expect any systematic priming effects of forehead stimulation (non-speech related site) on SRTs or phoneme recognition scores.

2. Materials and Methods

2.1 Participants

A total of thirty-three individuals participated in this study. These individuals were divided into two groups. The first group participated in the main experimental study and consisted of twenty-three native English speakers (M = 21.19 years, SD = 2.07; 18 participants self-identified as Female and 5 as Male) who received somatosensory priming stimulation to their oro-facial structures (either to labial [lip] corners or under the jaw; see Independent Variable section). The second group participated in a control study. This group consisted of ten native English speakers (M = 20.8 years, SD = 2.28; 8 participants self-identified as Female and 2 as Male) and somatosensory priming stimulation was applied to their right or left forehead as a control site.

All participants included in the study met the following criteria: (a) No reports of speech-language, or vision issues, (b) No hearing issues (all participants passed standardized hearing screening test and had thresholds less than 20 dB HL), (c) No vocabulary limitations (all participants met the criteria vocabulary competency on the Mill-Hill Vocabulary Test [[43]; a cut-off score of 9 of 21 items]), and (d) No evidence of auditory processing disorders (passed the screening criteria on the SCAN-3 test [44]).

All participants were recruited from the University of Toronto using flyers posted on campus. The study was approved by the University of Toronto's Health Sciences Research Ethics Board (Protocol # 31524), and all participants provided informed consent prior to participation.

2.2 Experimental Design

The study utilized a cross-modal repetition priming paradigm, similar to that reported by Buchwald *et al.* [45]. Priming paradigms are typically used in psycholinguistics studies to understand, if and when certain representations are active during speech and language processing [45]. Priming studies investigate changes in response performance (e.g., response time or response accuracy) in the processing of a target event (2nd stimulus) when the target is preceded by a prime event (1st stimulus) [46]. For example, faster response time and higher response accuracy are noted when the prime stimulus is semantically (e.g., presentation of the word "milk") or phonologically (e.g., presentation of the sound/phoneme /k/) congruent or associated with the subsequent target (e.g., "Cat"). In general, priming or facilitation occurs when there is a common processing component for both the target and prime representations,

and when this common processing is relevant to produce a response to the target [46]. In the current cross-modal repetition priming paradigm study, participants listened to, and repeated words (auditory target stimulus) embedded in white noise following the presentation of exogenous congruent and incongruent somatosensory inputs (somatosensory primes) by a trained S-LP therapist. We measured stimulus response time (SRT) and response accuracy as dependent variables.

2.3 Procedure

Participants were individually tested in a quiet room. Each participant listened to the auditory targets via AKG (K44) studio headphones at a comfortable listening level. Prior to the start of the experimental manipulations, participants were instructed to phonate the vowel /a/ as fast as possible when they hear a "go" tone. This task was repeated three times to get an estimate of phonation reaction time and to check audio recording procedures. Following this task, the auditory targets were then presented using an SRT paradigm controlled by DirectRT software [47] via a laptop (Lenovo-Yoga 13). An S-LP, specialized in using the PROMPT approach, provided the somatosensory inputs (see Fig. 1). The following control procedures were included in the study to decrease bias: (1) The study participants did not know in advance if the words were starting with /w/ or /d/ sounds (see 2.4 Stimuli section) (2) the presentation of congruent and incongruent somatosensory inputs with auditory targets was randomized and (3) the clinician providing the somatosensory priming was blinded to the auditory targets presented to the participants to avoid any bias.
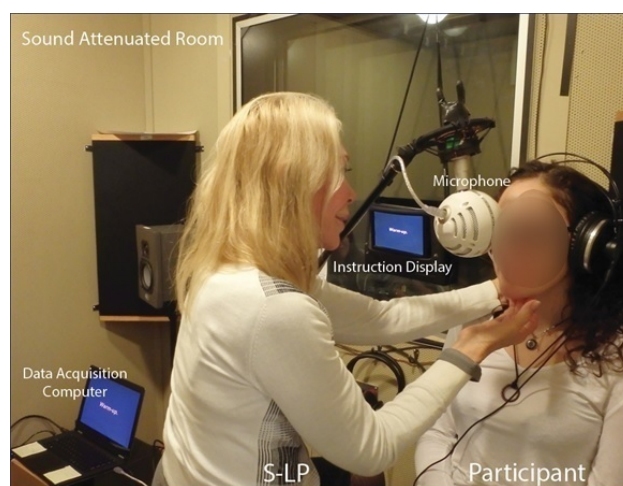
Fig. 1. Experimental Room Set-up. Experimental room set-up depicting a trained S-LP specialized in the PROMPT approach providing somatosensory inputs to participant. Experiment conducted in sound attenuated room with participants wearing sound attenuating circumaural headphones.

For each trial, there were two sets of instructions displayed for 1 second each. The first instruction prompted the S-LP and the participant to "Get Ready". The second instruction directed the S-LP on the somatosensory input for each trial, but the participants did not see this instruction. Following this, an alerting tone was presented (500 hertz, for 500 msec), after which the S-LP provided a specific somatosensory input three times within a 1.5–2 sec variable window. This random delay was inserted to avoid anticipatory responses from participants. There were no other speech/non-speech sounds associated or produced when somatosensory inputs were provided. Following this, participants heard an auditory target mixed in white noise over the headphones and responded verbally with the target word as fast and accurately as possible into a microphone (Blue Snowball iCE USB Microphone) connected to the DirectRT software (see Fig. 2 for a schematic of an experimental trial). Trials were terminated if response times were longer than 1 second. Participants pressed the space bar to progress to the next trial. Errors in the trials (i.e., false starts, voice key errors, and hesitations) were recorded by the experimenter in real-time. In addition, verbal responses were digitally recorded using a portable digital recorder (H4 Zoom) for offline accuracy analysis.
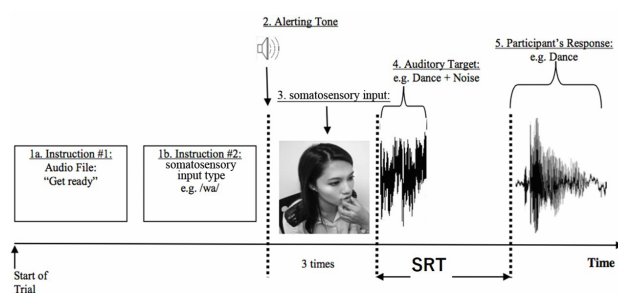
Fig. 2. Schematic of Trial. Process flow diagram (or Schematic) of a single incongruent trial. Instruction #1 "get ready" was heard by both S-LP and participant, while instruction #2 was presented on the instruction display screen and visible only to the S-LP. Prior to onset of the auditory target words the S-LP was instructed (via the display screen) to stop somatosensory stimulation. The participants then heard an auditory target mixed in white noise over the headphones and responded verbally word as fast and accurately as possible. Speech reaction time (SRT) was measured as the time delay from the start of the auditory target word to the onset of the participant's verbal response.

2.4 Stimuli

Sixteen monosyllabic words were categorized into two groups of 8 words (Tables 1,2). The words were grouped according to the place of articulation of the onset consonant in the first syllable, either alveolar (e.g., dance, dine; words starting with consonant /d/) or bilabial (e.g.,

wipe, wed; words starting with consonant /w/). The present study assessed speech reaction times (SRT) and accuracy responses for low-frequency (0–10 per million) and high-frequency (greater than 10 per million) words.

Participants were presented with low-frequency targets intermixed with high-frequency words as fillers to avoid any perceptual learning effects of targets. Target word characteristics such as neighbourhood density (Freq_KF), part of speech (nouns, auxiliary verbs, verbs, adjectives, adverbs, pronouns), and syllabic structure (CV, CVC, CV:C, CVCC) are presented in Tables 1,2. Target words did not include proper names. All word characteristics of the target words were obtained via the English Lexicon Project [48,49].

2.5 Stimuli List Recording and Playback

Audio recordings of the stimulus words were made by a male native Canadian English speaker (age 24 years). The male speaker was instructed to produce each word at a comfortable loudness and timed to a visual metronome [50] set at 30 beats per minute. The items were first read aloud before recording to ensure proper pronunciation of the test items. The recording took place in a sound attenuated booth with a directional microphone placed approximately 20 cm away from the speaker's mouth. Two calibration tones were first played before recording for later loudness adjustments. The audio recordings were then segmented into individual word recordings. A 100 msec space was inserted before and after each recording. All stimulus words were loudness normalized to 70 dB SPL. As individuals may differ in their ability to perceive speech in noise [51], individually customized levels of random Gaussian white noise were mixed with target words (i.e., individually adjusted signal-to-noise (SNR) ratios ranging from 0 to –10; e.g., [51]) using PRAAT scripts [52]. SNR was adjusted until we obtained a production accuracy of approximately 50–65% for each speaker during practice trials with an unrelated set of words. It is necessary to reduce auditory-only open-set word-recognition scores to below-ceiling levels to detect cross-modal priming effects [45].

2.6 Independent Variable

For the main study, somatosensory priming was provided in one of two ways: as a congruent condition (e.g., alveolar somatosensory input for a word with initial alveolar sound embedded in noise) or as an incongruent condition (e.g., bilabial somatosensory input for a word with initial alveolar sound embedded in noise). Somatosensory inputs were provided for the word-initial phoneme before the word-recognition-production task. There were two different somatosensory inputs: /d/ or /w/. For the lingual alveolar consonant /d/, somatosensory input was provided under the jaw into the skin covering the mylohyoid muscle applying upward pressure using a slightly curved middle finger. For the bilabial /w/, the index and thumb were placed at the

Table 1. Characteristics of high-frequency target words.

	Bilabials				Alveolars		
Target Word	Part of Speech	Syllable Structure	Neighborhood Density (Freq_KF)	Target Word	Part of Speech	Syllable Structure	Neighborhood Density (Freq_KF)
Was	Verb	CVC	9816	Dead	Noun	CVC	174
Wait	Verb	CV:C	94	Dance	Noun/Verb	CVCC	90
Wife	Noun	CV:C	228	Deal	Noun/Verb	CV:C	142
White	Adjective/Noun	CV:C	365	Date	Noun/Verb	CV:C	103
		Mean	2625.75			Mean	127.25
		Range	94–9816			Range	90–174

Table 2. Characteristics of low-frequency target words.

	Bilabials				Alveolars		
Target Word	Part of Speech	Syllable Structure	Neighborhood Density (Freq_KF)	Target Word	Part of Speech	Syllable Structure	Neighborhood Density (Freq_KF)
Web	Noun	CVC	6	Dam	Noun	CVC	5
Wed	Verb	CVC	2	Dale	Noun	CV:C	5
Wipe	Verb	CV:C	10	Dine	Verb	CV:C	2
Wade	Verb	CV:C	2	Dank	Adjective	CVCC	1
		Mean	5			Mean	3.25
		Range	2–10			Range	1–5

labial (lip) corners (at the intersection between zygomatic major and orbicularis oris muscles) to facilitate lip rounding manually (see Fig. 3). The presentation of congruent and incongruent somatosensory inputs and auditory targets were randomized and presented twice (16 words × 2 conditions [congruent and incongruent] × 2 repetitions = 64 trials per participant), while the S-LP therapist providing the somatosensory inputs was blinded to the auditory targets.

Fig. 3. Delivery of somatosensory inputs. Demonstration of delivery of somatosensory inputs for bilabial /w/ and lingual /d/ consonant by a speech-language pathologist on co-author (TY). For the bilabial /w/, the index and thumb were placed at the labial (lip) corners (at the intersection between zygomatic major and orbicularis oris muscles) to facilitate lip rounding manually. For the lingual alveolar consonant /d/, somatosensory input was provided under the jaw into the skin covering the mylohyoid muscle applying upward pressure using a slightly curved middle finger.

Since the current study used only two groups of words and two types of priming stimulation, it can be relatively easy to map (or perceptually match) one priming stimulation on one consonant group. To examine whether the perceptual modulation was due to simple matching of the events between sound stimuli and priming stimuli we conducted a separate control study with 10 participants. In this control study, a clinician applied the same somatosensory stimulation to a different body part (e.g., [10], namely right and left forehead of the participant). The clinician trained the participant to identify and associate /d/ sounds with a firm pressure on the right forehead and /w/ sounds on the left forehead. Participants continued this training till 100% correct association was established across 3 consecutive trials. A lack of a systematic priming effect in the forehead somatosensory priming (control) study would strengthen our claim regarding the association between specific-orofacial somatosensory priming stimulation and processing of stimulus sounds.

2.7 Dependent Variables
2.7.1 Phoneme Accuracy

Phoneme accuracy was calculated as a percentage of word-initial phonemes produced correctly from the audio recordings. To obtain inter-rater reliability, twenty-six percent of data corresponding to participants' verbal responses were transcribed by two independent phonetically trained graduate research assistants, and point-by-point inter-observer agreement was 87.31%.

Table 3. Mean and Standard Deviations (in parenthesis) for Speech Reaction Time and Phoneme Accuracy for Low- and High-Frequency Words.

	High-Frequency Words		Low-Frequency Words	
Oral Somatosensory condition	Speech Reaction Time (ms)	Phoneme Accuracy (%)	Speech Reaction Time (ms)	Phoneme Accuracy (%)
Congruent Input	753.5 (168.9)	93.4 (9.8)	802.4 (164.6)	92.3 (14.9)
Incongruent Input	749.5 (210.8)	92.9 (12.9)	847.4 (206.1)	88.0 (14.3)
Forehead control condition				
Congruent Input	603.3 (92.6)	93.7 (10.6)	692.4 (101.6)	91.4 (10.2)
Incongruent Input	627.3 (70.1)	92.5 (16.8)	686.1 (110.9)	90.2 (9.8)

2.7.2 Speech Reaction Time (SRT)

SRT was automatically calculated by the DirectRT software (ver. 2012, Empirisoft Corporation, New York, NY, USA). SRT was measured as the time delay from the start of the auditory target word to the onset of the participant's verbal response. Additionally, we measured phonation (or vocal) reaction time to ensure accurate instrument functioning (i.e., to check microphone/audio trigger) and data validity. As speech reaction time requires higher-level (speech/lexical) processing we expect SRT to be significantly longer than phonation reaction times. The phonation reaction time data served as a lower boundary marker for speech reaction time to eliminate false starts during data analysis. Trials were terminated if response times were longer than 1 second (failure of audio trigger).

2.8 Statistical Analysis

Means and standard deviations for the dependent variables (SRT and phoneme accuracy) across congruent and incongruent conditions for both the high- and low-frequency words were calculated from the DirectRT software and audio recordings. A two-way repeated measures analysis of variance (ANOVA) was conducted with somatosensory input condition as factor 1 (congruent or incongruent) and word-frequency as factor 2 (high- or low-frequency). Separate two-way ANOVAs were carried out for each of the dependent variables and for the control data. Bonferroni adjusted post-hoc t-tests were carried out as necessary. Phonation reaction time was compared to the SRT data using paired t-tests. Effect sizes are reported as partial eta squared (η^2p) for the ANOVAs and as Cohen's d statistic corrected for dependence between the means for the t-tests. For ANOVAs, effect sizes of 0.01, 0.06 and 0.14 were considered small, medium, and large effects, respectively. For t-tests, 0.2, 0.5 and 0.8 are typically referred to as small, medium, and large effects, respectively [53,54].

3. Results

Phonation reaction times (M = 371.1; SD = 67.1) as expected were significantly ($t(9) = 5.6, p = 0.0003$) shorter than the fastest speech reaction time data (M = 603.3; SD = 92.6) recorded. Means and standard deviations (in parenthesis) for SRT and word-initial phoneme accuracy across

congruent and incongruent conditions were calculated and reported in Table 3 for oro-facial somatosensory priming and for forehead control conditions.

3.1 Speech Reaction Time (SRT)

For the main study, SRT results revealed a significant main effect of word frequency ($F(1,22) = 22.45, p = 0.001, \eta^2p = 0.505$), but no main effect for somatosensory input condition ($F(1,22) = 1.77, p = 0.196, \eta^2p = 0.075$). Importantly, a significant condition × word frequency interaction ($F(1,22) = 7.39, p = 0.013, \eta^2p = 0.252$) was found. Post-hoc paired t-tests indicated that significant effects of somatosensory priming were found only for low-frequency words ($t(22) = -2.22, p = 0.037$, Cohen's $d = -0.72$) as opposed to the high-frequency words ($t(22) = 0.27, p = 0.789$, Cohen's $d = 0.07$).

In the control study, SRT results revealed a significant main effect of word frequency ($F(1,9) = 22.26, p = 0.001, \eta^2p = 0.712$), but no main effect for somatosensory input condition ($F(1,9) = 0.10, p = 0.749, \eta^2p = 0.012$), or a condition × word frequency interaction ($F(1,9) = 0.50, p = 0.496, \eta^2p = 0.053$).

3.2 Phoneme Accuracy

For the main study, phoneme accuracy results revealed a main effect of somatosensory input condition ($F(1,22) = 4.43, p = 0.047, \eta^2p = 0.168$), but no main effect for word frequency ($F(1,22) = 1.26, p = 0.273, \eta^2p = 0.054$) or an input condition × word frequency interaction ($F(1,22) = 2.04, p = 0.166, \eta^2p = 0.085$). However, Table 3 suggests the presence of a potential congruency effect for low-frequency words relative to high-frequency words. Given the exploratory nature of this study, with a relatively modest participant sample (see limitations section), post hoc comparisons are relevant despite lack of higher-level interaction effects [55]. A further exploration of this effect using paired t-tests indeed indicated an advantage for congruent somatosensory priming ($t(22) = 2.57, p = 0.017$, Cohen's $d = 0.74$) for the low-frequency words as opposed to the high-frequency words ($t(22) = 0.29, p = 0.770$, Cohen's $d = 0.09$).

In the control study, phoneme accuracy results revealed no main effects for word frequency ($F(1,9) = 0.32, p = 0.580, \eta^2p = 0.035$), somatosensory input condition

$(F(1,9) = 0.17, p > 0.686, \eta^2p = 0.019)$ or condition $\times$ word frequency interaction $(F(1,9) = 0.00, p = 0.990, \eta^2p = 0.000)$.

4. Discussion

The experiment reported in this paper used a cross-modal repetition priming paradigm to examine whether the processing of external S-LP driven somatosensory inputs was modulated by both word-level (lexical frequency) and speech sound articulatory features, suggesting a broader role for speech production information in speech perception. The main effects corroborate data from previous research indicating that the high-frequency words had significantly shorter reaction times than low-frequency words (e.g., [37,56]; for a review in adult aphasia literature, see Kittredge *et al.* [36]). Importantly, a statistically significant condition $\times$ word frequency interaction indicated that significant effects of somatosensory priming were found only for low-frequency words as opposed to the high-frequency words.

For phoneme accuracy, the results revealed a main effect of somatosensory input condition. A further exploration of this effect revealed that the percent phoneme accuracy scores were lower for incongruent conditions than congruent conditions for the low-frequency words, but the interaction failed to reach significance. Overall, the results showed significant somatosensory priming effects for the low-frequency words, where the congruent somatosensory condition yielded significantly shorter reaction times and numerically higher phoneme accuracy scores when compared to the incongruent somatosensory condition (see Table 3). In general, for the low-frequency words, incongruent somatosensory input increased SRTs by approximately 45 ms and decreased percentage phoneme accuracy scores by 4–5%. For high-frequency words, SRTs and phoneme accuracy did not differ as a function of somatosensory congruency.

Importantly, data from the control study did not reveal any systematic priming effects from forehead stimulation (non-speech related site), other than a general (and expected) tendency for longer reaction times with low-frequency words. This finding suggests that the perceptual modulation effects in the current study are less likely from simple pattern matching of events between sound stimuli and somatosensory (priming) inputs. Although, reaction time data for forehead stimulation was faster when compared to oral somatosensory condition, readers should note that forehead data was collected from a separate group of control participants at a later date, and thus some differences are expected.

The present findings not only align with contemporary theories of speech perception that support the role of the somatosensory system in speech perception [9,23] but also extends this to a broader and more integrative role of somatosensory information in lexical access as proposed

by Ogane *et al.* [10]. Specifically, the interaction effects observed between word frequency and type of somatosensory input related to place of articulation suggest that external somatosensory input (as provided by a therapist) may play a role at the level of lexical-phonetic interactions in speech processing [10]. Within interactive activation approaches to word production, such lexical-phonetic interactions may arise in a distributed language network where nodes are bidirectionally (i.e., feedforward and feedback manner) connected across semantic, lexical, and phonetic levels of representation [42]. In some spoken-word recognition models (as discussed in the Introduction section), word frequency is assumed to affect the activation levels during lexical access; thus, it is likely that articulatory place information as stimulated by these somatosensory inputs is one way of raising the activation levels of words which have naturally lower levels of activation (i.e., low-frequency words compared to high-frequency words; [40,57]).

In the context of restorative neurology, the provision of congruent somatosensory cues may be facilitative for word-finding or naming difficulties (e.g., in anomia, adult apraxia of speech, Broca's aphasia) in adult aphasia [21,58,59]. Typically, phonological cueing techniques have been utilized to improve naming speed and accuracy in individuals with aphasia [58,59]. In the phonological cueing technique, persons with aphasia are provided with a verbal cue of the first phoneme for an object or picture they are required to name [60]. For example, verbal cue /k/ for the word cat. In such cueing intervention approaches, the addition of somatosensory inputs by a speech therapist [19,21] or silent observation of visual articulatory gestures [61] has added benefits to improving naming performance in individuals with aphasia. The current findings on healthy control participants, when taken together with the data from multisensory cueing studies in aphasia research [19,21,61], suggest that exogenous therapist driven, tightly controlled or specific somatosensory inputs may facilitate rehabilitation of adult brain-injury patients such as those with word-finding/word-retrieval issues in aphasia [18–21].

The question remains whether specific exogenous S-LP driven somatosensory inputs in clinical rehabilitation settings constitute an "active ingredient" underlying MoAs in motor speech interventions such as the PROMPT approach. As described in Hayden *et al.* [1] recent research studies from independent laboratories have demonstrated changes in cortical spatial-temporal neural activity and brain structure following PROMPT intervention [62–64]. A magnetic resonance imaging study by Kadis *et al.* [63] revealed that PROMPT intervention induced thinning of the left posterior superior temporal gyrus (Wernicke's area) in eight of nine children with CAS. Similarly, in Yu *et al.* [64] study, participants demonstrated significant differences in cortical neural activity during speech tasks as a function of PROMPT intervention on magnetoencephalog-

raphy recordings. In both these studies, significant improvements in speech motor control and speech articulation were also observed following the intervention. More recently, Fiori *et al.* [62] studied PROMPT intervention-related brain structural connectivity changes in white matter fiber tracts using diffusion-weighted imaging (diffusion tractography) in a group of 10 children with CAS. Five CAS children received the PROMPT intervention, while the other five received Language and Non-speech Oral Motor (LNSOM) treatment. Their results revealed significant and positive changes in the standardized assessment of speech motor control in the group that received PROMPT intervention relative to the LNSOM-treated group. Although in both PROMPT- and LNSOM-treated groups, the speech improvements paralleled changes in the ventral corticobulbar tracts (innervates tongue/larynx), only the PROMPT intervention group showed additional neuroplastic changes such as an increase in connectivity in the left dorsal corticobulbar tracts (i.e., those associated with lip/facial muscles). The authors argue that this may be an intervention-specific response as jaw control and labial-facial control were priority speech motor goals for children receiving PROMPT therapy.

Importantly, such discoveries of experience-dependent neuroplasticity in children receiving somatosensory based motor speech intervention (e.g., PROMPT) for speech sound disorders demonstrate a mode of action and improves our understanding of how motor speech intervention produces its effects. For federal agencies such as the U.S. Food and Drug Administration (FDA), identifying MoA's is a key component of discovering and developing therapeutic interventions. The discovery of MoAs of motor speech intervention will hopefully permit the development of future research directed at modifying existing techniques to replicate such neural effects, adjust the dosage to elicit systematic effects in targeted neural pathways and potentially discover populations that are responsive to motor speech interventions [1]. Thus, converging evidence from behavioural data in children [14], adults [18–21], perceptual data from cross-modal research [9,10], and neuroimaging studies [62–64] seem to reinforce the notion that somatosensory inputs may be a potential active ingredient in motor speech interventions.

Limitations

The results of the study must be interpreted with some caution. Several important caveats are warranted. First, studies using priming paradigms generally use the computer-controlled presentation of primes and target stimuli (e.g., [46]). Although, there was a potential option to provide the somatosensory inputs via mechanical skin stretch by an automated robotic device such as that used by Ito and colleagues (e.g., [9]) or pneumatic tactile stimulator devices (e.g., [65]), the provision of somatosensory inputs (primes) by a trained S-LP offers greater ecological va-

lidity to the study methods. S-LP provided somatosensory input improves the ability to generalize study findings to real-world clinical settings in the field of speech-language pathology.

Second, we only primed at the single phoneme level (either /d/ or /w/) while the target (initial phoneme) was embedded in a larger word-level unit. Generally, priming effects are stronger as the degree of similarity between prime-target pairs increases. For example, Meyer [66] demonstrated that phonological priming effects increased with the amount of overlap between prime and target words. Word-form overlap only in the onset consonant (e.g., "kever - kilo") resulted in a facilitatory effect of about 30 ms, relative to an unrelated condition (e.g., "hamer - kilo"). However, when the entire first syllable overlapped (e.g., "kilo - kiwi") the facilitatory effect was greater at about 50 ms [66,67]. Furthermore, clinically in the PROMPT approach, syllabic, word level, and phrase-level cues (or prompts) are preferred over cueing at a single phoneme level [1]. Thus, there is a potential that the current study may have underestimated the priming effects in the experiment and the real world with the use of smaller phoneme level somatosensory primes.

A third limitation is the small sample size in the current study. Typical priming studies encompass between 40 and 150 participants to detect subtle priming effects (e.g., [45,46]). Eliciting significant statistical interaction effects with small sample sizes is difficult and, therefore, may lead to a higher probability of type II beta-errors (due to lower statistical power). Thus, the results of the present study should be considered preliminary.

Fourth, prior to start of the experiment we manipulated SNR to reduce auditory-only open-set word-recognition scores to below-ceiling levels [45]. We obtained a production accuracy of approximately 50–65% for each speaker during practice trials with an unrelated set of words. However, as the current study used only two groups of 8 words (16 in total) a relatively high accuracy rate of phoneme identification (near ceiling level ~90%), was inadvertently obtained for our sample of healthy young adult control speakers. As the current effect is an attenuation effect by an incongruent pair, not having much room above the obtained scores (ceiling values) does not critically affect the interpretation of results. Nevertheless, we strongly recommend follow-up studies to carefully manipulate SNR to limit production accuracy to below ceiling levels to examine enhancement effects. Further, we recommend future studies to have an additional control condition where participants are tested with the same SNR but without somatosensory stimulation.

Fifth, as stimuli contained only 64 trials per participant including repetitions, hearing the same stimulus multiple times may facilitate identification regardless of any priming or low-/high-frequency effects. However, this is less likely as we randomized presentation of stimuli and neither partic-

ipant nor the clinician (providing the somatosensory priming) knew in advance if the words were starting with /w/ or /d/ sounds. Further, our control study with 10 additional participants did not reveal any systematic priming effects from forehead stimulation (non-speech related site), other than a general (and expected) tendency for longer reaction times with low-frequency words. If the effects observed were based on simple pattern matching then the results obtained from oro-facial somatosensory priming would have been no different from the forehead stimulation, which was not the case in the current study.

Finally, the study was carried out on a healthy young adult cohort; the findings may not necessarily extend to children with speech sound disorders or older adults with neurological impairments. For example, Trudeau-Fisette *et al.* [68] demonstrated that somatosensory information's effects on sound categorization were smaller in children than in adults. They argue that the stronger auditory and somatosensory integration found in adults may result from the development of sensory-perceptual systems, sensory-perceptual experience, and/or attentional issues [68]. In the case of adults with neurological impairments (e.g., following cerebral stroke), sensorimotor integration is shown to be negatively impacted and correlated with the level of recovery [69]. The findings of such studies imply that the integration of sensory information may evolve continuously through the course of human development and recovery from injury. Thus, this issue needs clarification and further research with children and adults with and without speech and language disorders. Further, the study results may be restricted to the specific group of speakers (young university graduate students) and tasks (speech in noise) we utilized, thereby limiting the generality of our findings.

5. Conclusions

These findings provide preliminary support for the notion that speech production information is represented in the mental lexicon and can be accessed through exogenous S-LP-driven somatosensory inputs related to place of articulation. Low-frequency words usually take longer to be recognized, and the additional information provided by somatosensory input can facilitate this process. Our findings suggest that the way somatosensory inputs influence lexical selection is biased by the usage sensitivity of how lexical information is stored in the mental lexicon [70,71]. The study results highlight a broader and more integrative role of somatosensory information in lexical access, as proposed by Ogane *et al.* [10].

References

[1] Hayden D, Namasivayam AK, Ward R, Eigen J, Clark A. The PROMPT approach: Theory, Evidence, Use and Application. In Williams L, McLeod S, McCauley R (eds.) Interventions for Speech Sound Disorders (pp. 477–504). 2nd edn. Brookes Publishing: Baltimore, US. 2020.

[2] Strand EA, Stoeckel R, Baas B. Treatment of severe childhood apraxia of speech: A treatment efficacy study. Journal of Medical Speech-Language Pathology. 2006; 14: 297–307.

[3] Preston JL, Maas E, Whittle J, Leece MC, McCabe P. Limited acquisition and generalisation of rhotics with ultrasound visual feedback in childhood apraxia. Clinical Linguistics & Phonetics. 2016; 30: 363–381.

[4] Hitchcock ER, Byun TM, Swartz M, Lazarus R. Efficacy of Electropalatography for Treating Misarticulation of /r/. American Journal of Speech-Language Pathology. 2017; 26: 1141–1158.

[5] Fawcett S, Bacsfalvi P, Bernhardt BM. Ultrasound as visual feedback in speech therapy for/r/with adults with Down syndrome. Down Syndrome Quarterly. 2008; 10: 4–12.

[6] Hayden DA, Square PA. Motor Speech Treatment Hierarchy: a systems approach. Clinics in Communication Disorders. 1994; 4: 162–174.

[7] Strand EA. Dynamic temporal and tactile cueing: A treatment strategy for childhood apraxia of speech. American Journal of Speech-Language Pathology. 2020; 29: 30–48.

[8] Grant RL, Combs AB, Acosta D. Experimental models for the investigation of toxicological mechanisms. In McQueen CA (ed.) Comprehensive Toxicology (pp. 203–224). 2nd edn. Elsevier Ltd: Oxford, UK. 2010.

[9] Ito T, Tiede M, Ostry DJ. Somatosensory function in speech perception. Proceedings of the National Academy of Sciences. 2009; 106: 1245–1248.

[10] Ogane R, Schwartz J, Ito T. Orofacial somatosensory inputs modulate word segmentation in lexical decision. Cognition. 2020; 197: 104163.

[11] Moayedi Y, Michlig S, Park M, Koch A, Lumpkin EA. Somatosensory innervation of healthy human oral tissues. Journal of Comparative Neurology. 2021; 529: 3046–3061.

[12] Guenther FH. A neural network model of speech acquisition and motor equivalent speech production. Biological Cybernetics. 1994; 72: 43–53.

[13] Namasivayam AK, van Lieshout P, McIlroy WE, De Nil L. Sensory feedback dependence hypothesis in persons who stutter. Human Movement Science. 2009; 28: 688–707.

[14] Dale PS, Hayden DA. Treating Speech Subsystems in Childhood Apraxia of Speech with Tactual Input: the PROMPT Approach. American Journal of Speech-Language Pathology. 2013; 22: 644–661.

[15] Square PA, Namasivayam AK, Bose A, Goshulak D, Hayden D. Multi-sensory treatment for children with developmental motor speech disorders. International Journal of Language & Communication Disorders. 2014; 49: 527–542.

[16] Rogers SJ, Hayden D, Hepburn S, Charlifue-Smith R, Hall T, Hayes A. Teaching Young Nonverbal Children with Autism Useful Speech: a Pilot Study of the Denver Model and PROMPT Interventions. Journal of Autism and Developmental Disorders. 2006; 36: 1007–1024.

[17] Ward R, Strauss G, Leitão S. Kinematic changes in jaw and lip control of children with cerebral palsy following participation in a motor-speech (PROMPT) intervention. International Journal of Speech-Language Pathology. 2013; 15: 136–155.

[18] Bose A, Square PA, Schlosser R, van Lieshout P. Effects of PROMPT therapy on speech motor function in a person with aphasia and apraxia of speech. Aphasiology. 2001; 15: 767–785.

[19] Freed DB, Marshall RC, Frazier KE. Long term effectiveness of PROMPT treatment in a severely apraxic-aphasic speaker. Aphasiology. 1997; 11: 365–342.

[20] Square PA, Chumpelik DA, Adams SG. Efficacy of the PROMPT system of therapy for the treatment of acquired apraxia of speech. In Brookshire RH (ed.) Clinical aphasiology: Conference proceedings (pp. 319–320). BRK Publishers: Minneapolis, US. 1985.

[21] Square PA, Chumpelik DA, Morningstar D, Adams SG. Efficacy of the PROMPT system of therapy for the treatment of apraxia of speech: A follow-up investigation. In Brookshire RH (ed.) Clinical aphasiology: Conference proceedings (pp. 221–226). BBK Publishers: Minneapolis, US. 1986.

[22] Derrick D, Gick B. Full body aero-tactile integration in speech perception. INTERSPEECH. 2010; 122–125.

[23] Gick B, Derrick D. Aero-tactile integration in speech perception. Nature. 2009; 462: 502–504.

[24] Fougeron C. Articulatory properties of initial segments in several prosodic constituents in French. Journal of Phonetics. 2001; 29: 109–135.

[25] Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. Psychological Review. 1967; 74: 431–461.

[26] Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition. 1985; 21: 1–36.

[27] Foxe JJ, Wylie GR, Martinez A, Schroeder CE, Javitt DC, Guilfoyle D, et al. Auditory-Somatosensory Multisensory Processing in Auditory Association Cortex: an fMRI Study. Journal of Neurophysiology. 2002; 88: 540–543.

[28] Jousmäki V, Hari R. Parchment-skin illusion: sound-biased touch. Current Biology. 1998; 8: R190–R191.

[29] Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, et al. Grabbing your Ear: Rapid Auditory–Somatosensory Multisensory Interactions in Low-level Sensory Cortices are not Constrained by Stimulus Alignment. Cerebral Cortex. 2005; 15: 963–974.

[30] Schürmann M, Caetano G, Jousmäki V, Hari R. Hands help hearing: Facilitatory audiotactile interaction at low sound-intensity levels. The Journal of the Acoustical Society of America. 2004; 115: 830–832.

[31] Stein BE, Meredith MA. The merging of the senses. The MIT Press: Cambridge, US. 1993.

[32] Romanski LM. Integration of faces and vocalizations in ventral prefrontal cortex: Implications for the evolution of audiovisual speech. Proceedings of the National Academy of Sciences. 2012; 109: 10717–10724.

[33] Heim S, Wehnelt A, Grande M, Huber W, Amunts K. Effects of lexicality and word frequency on brain activation in dyslexic readers. Brain and Language. 2013; 125: 194–202.

[34] Fiebach CJ, Friederici AD, Müller K, Cramon DYV. FMRI Evidence for Dual Routes to the Mental Lexicon in Visual Word Recognition. Journal of Cognitive Neuroscience. 2002; 14: 11–23.

[35] Griffin ZM, Bock K. Constraint, Word Frequency, and the Relationship between Lexical Processing Levels in Spoken Word Production. Journal of Memory and Language. 1998; 38: 313–338.

[36] Kittredge AK, Dell GS, Verkuilen J, Schwartz MF. Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. Cognitive Neuropsychology. 2008; 25: 463–492.

[37] Pagliuca G, Arduino LS, Barca L, Burani C. Fully transparent orthography, yet lexical reading aloud: the lexicality effect in Italian. Language and Cognitive Processes. 2008; 23: 422–433.

[38] Famoyegun A. Word Frequency Effects in L2 Speakers: An ERP Study [master's thesis]. Arizona (US): The University of Arizona University. 2012.

[39] Rugg M. Event-related brain potentials dissociated repetition ef-

fects of high-and low- frequency words. Memory & Cognition. 1990; 18: 367–379.

[40] Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. Journal of Memory and Language. 2009; 60: 92–111.

[41] Bose A, Van Lieshout P, Square P. Effects of word and bigram frequency on linguistic processing and speech motor performance in individuals with aphasia and normal speakers. Journal of Neurolinguistics. 2007; 20: 65–88.

[42] Goldrick M, Ross Baker H, Murphy A, Baese-Berk M. Interaction and representational integration: Evidence from speech errors. Cognition. 2011; 121: 58–72.

[43] Raven J, Raven JC, Court JH. Manual for Raven's progressive matrices and vocabulary scales. Section 5: The Mill Hill vocabulary scale. Harcourt Assessment: San Antonio, TX. 1998.

[44] Keith R. SCAN–3 for Adolescents and Adults: Tests forAuditory Processing Disorders. Pearson Education: San Antonio, TX. 2009.

[45] Buchwald AB, Winters SJ, Pisoni DB. Visual speech primes open-set recognition of spoken words. Language and Cognitive Processes. 2009; 24: 580–610.

[46] Kim J, Davis C, Krins P. A modal processing of visual speech as revealed by priming. Cognition. 2004; 93: B39–B47.

[47] Jarvis BG. DirectRT (Version 2012) [Computer Software]. Empirisoft Corporation: New York, NY. 2012.

[48] Balota DA, Yap MJ, Hutchison KA, Cortese MJ, Kessler B, Loftis B, et al. The English Lexicon Project. Behavior Research Methods. 2007; 39: 445–459.

[49] Loftis B. English lexicon project web site. Washington University in St. Louis. 2014. Available at: https://elexicon.wustl.edu/ (Accessed: 21 June 2014).

[50] Repp BH, Su Y. Sensorimotor synchronization: a review of recent research (2006–2012) Psychonomic Bulletin & Review. 2013; 20: 403–452.

[51] McLaughlin DJ, Baese-Berk MM, Bent T, Borrie SA, Van Engen KJ. Coping with adversity: Individual differences in the perception of noisy and accented speech. Attention, Perception, & Psychophysics. 2018; 80: 1559–1570.

[52] Boersma P, Weenink DJM. Praat: doing phonetics by computer. Version 5.3.79. Praat.org. 2014. Available at: http://www.praat.org/ (Accessed: 21 June 2014).

[53] Cohen J. Statistical power analysis for the behavioral sciences. 2nd edn. Lawrence Earlbaum Associates: USA. 1988.

[54] Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods. 2002; 7: 105–125.

[55] Wei J, Carroll RJ, Harden KK, Wu G. Comparisons of treatment means when factors do not interact in two-factorial studies. Amino Acids. 2012; 42: 2031–2035.

[56] Forster KI, Chambers SM. Lexical access and naming time. Journal of Verbal Learning and Verbal Behavior. 1973; 12: 627–635.

[57] Dahan D, Magnuson JS, Tanenhaus MK. Time Course of Frequency Effects in Spoken-Word Recognition: Evidence from Eye Movements. Cognitive Psychology. 2001; 42: 317–367.

[58] Meteyard L, Bose A. What does a Cue do? Comparing Phonological and Semantic Cues for Picture Naming in Aphasia. Journal of Speech, Language, and Hearing Research. 2018; 61: 658–674.

[59] Wisenburn B, Mahoney K. A meta-analysis of word-finding treatments for aphasia. Aphasiology. 2009; 23: 1338–1352.

[60] Best W, Herbert R, Hickin J, Osborne F, Howard D. Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. Aphasiology. 2002; 16: 151–168.

[61] Grechuta K, Rubio Ballester B, Espín Munné R, Usabiaga Bernal T, Molina Hervás B, Mohr B, et al. Multisensory cueing facilitates naming in aphasia. Journal of NeuroEngineering and Rehabilitation. 2020; 17: 122.

[62] Fiori S, Pannek K, Podda I, Cipriani P, Lorenzoni V, Franchi B, et al. Neural Changes Induced by a Speech Motor Treatment in Childhood Apraxia of Speech: a Case Series. Journal of Child Neurology. 2021; 36: 958–967.

[63] Kadis DS, Goshulak D, Namasivayam A, Pukonen M, Kroll R, De Nil LF, et al. Cortical Thickness in Children Receiving Intensive Therapy for Idiopathic Apraxia of Speech. Brain Topography. 2014; 27: 240–247.

[64] Yu VY, Kadis DS, Goshulak D, Namasivayam AK, Pukonen M, Kroll RM, et al. Impact of Motor Speech Intervention on Neural Activity in Children with Speech Sound Disorders: Use of Magnetoencephalography. Journal of Behavioral and Brain Science. 2018; 08: 415–429.

[65] Wang L, Liu Y, Liu Q, Zhu Y, Wang H, Xie Z, et al. A metalelectrode-free, fully integrated, soft triboelectric sensor array for self-powered tactile sensing. Microsystems & Nanoengineering. 2020; 6: 59.

[66] Meyer AS. The time course of phonological encoding in language production: Phonological encoding inside a syllable. Journal of Memory and Language. 1991; 30: 69–89.

[67] Shao Z, Meyer AS. Word priming and interference paradigms. In De Groot AMB, Hagoort P (eds.) Research methods in psycholinguistics and the neurobiology of language: A practical guide (pp. 111–129). Wiley Blackwell: New Jersey, US. 2018.

[68] Trudeau-Fisette P, Ito T, Ménard L. Auditory and somatosensory interaction in speech perception in children and adults. Frontiers in Human Neuroscience. 2019; 13: 344.

[69] Edwards LL, King EM, Buetefisch CM, Borich MR. Putting the "sensory" into sensorimotor control: the role of sensorimotor integration in goal-directed hand movements after stroke. Frontiers in Integrative Neuroscience. 2019; 13: 1–16.

[70] Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. Ear and Hearing. 1998; 19: 1–36.

[71] Marslen-Wilson WD. Functional parallelism in spoken word-recognition. Cognition. 1987; 25: 71–102.