Interspeech 2018
2-6 September 2018, Hyderabad

Sensorimotor response to tongue displacement imagery by talkers with Parkinson's disease

William F. Katz[1], *Patrick Reidy*[2], *Divya Prabhakaran*[3]

[1,2]University of Texas at Dallas
[3]Vanderbilt University, Tennessee

wkatz@utdallas.edu, reidy@utdallas.edu, divya.prabhakaran@vanderbilt.com

Abstract

In a previous study, we asked healthy adult speakers to produce the word *head* under noise-masked (visual only) conditions and while watching videos of a 3D tongue avatar that gradually morphed from producing *head* to *had*. Results indicated that during the visual mismatch phases all participants entrained to the visually presented word, *head*, without being aware that their vowel quality had changed. Here, we explore whether similar effects occur for individuals with presumed sensorineural processing disorders, patients with Parkinson's disease (PD). We also examine the effects of PD treatment on this entrainment behavior. Participants were 14 individuals with PD, with eight in ongoing speech/language therapy, and six reporting no recent therapy. Participants heard pink noise over headphones and produced the word *head* under four viewing conditions: First, while viewing repetitions of *head* (baseline); next, during "morphed" videos shifting gradually from *head* to *had* (ramp); then videos of *had* (maximum hold); and finally videos of *head* (after effects). Analysis with a linear mixed-effects model indicated a significant F1 difference between baseline and maximum hold phases for the productions of the treated PD group, but not for the untreated group. Implications for the causes and treatment of PD speech disorders are discussed.

Index Terms: speech production and perception, visual feedback, electromagnetic articulography, sensorimotor adaptation, Parkinson's disease, dysarthria

1. Introduction

Individuals with Parkinson's disease (PD) present with gait and balance difficulties, tremor, rigidity, slowed movement, and speech and swallowing problems [1]. The speech of most (70-75%) individuals with PD involves hypokinetic dysarthria, including reduced amplitude and irregular timing [2]. The basis for this dysarthria appears to be sensorimotor, involving both perceptual and motoric processing deficits. For instance, neural imaging (fMRI) data reveal that PD patients show evidence of reduced monitoring of auditory feedback and suggest dysarthria may result from imprecise shaping of motor representations by this improperly processed feedback [3]. Speech perturbation experiments have provided important evidence concerning the linkage between sensory monitoring and changes in motor output. These studies provide data on speech behavior by altering sensory information so that underlying control processes and short-term learning may be observed. Perturbation delivered in an unexpected and random fashion is assumed to tap moment-to-moment control processes (compensation), while perturbation applied in a more predictable and constant manner is thought to assess a form of short-term learning (adaptation).

Several acoustic feedback studies have investigated vowel production in healthy individuals by having subjects hear their voice (mixed with noise) over headphones while a rapid, online acoustic perturbation that changes the status of one or more speech parameters is introduced. Sensorimotor compensation experiments have generally found that healthy subjects rapidly adjust in the opposite direction of the perturbation. This has been noted for shifts in formant frequencies [4] and F0 [5]. Sensorimotor adaptation experiments have demonstrated more gradual changes in procedural learning, also occurring in the opposite direction to the feedback shift [6,7].

Taken together, perturbation studies suggest that, for healthy individuals, both immediate control processes as well as short-term learning act together to maintain vowel phonetic quality during speech. Individuals with PD appear to respond in an abnormal fashion, showing for F0 larger compensatory responses than individuals without PD [8,9] and for formant frequency changes reduced sensorimotor adaptation compared to healthy individuals [10].

Notably, these studies have been restricted to the effects of auditory feedback; that is, on-line shifting of either the F0, the formant frequencies or the amplitudes of speech signals delivered acoustically to subjects during speech. However, it is well known that speech involves both the auditory and visual channels [11]. Recent evidence also suggests that individual's speech perception and production may be subtly affected by viewing the tongue, an internal articulator that is not often visible without instrumental means [12].

To address this issue, we conducted a visual speech perturbation experiment in which healthy participants repeated the word *head* while a visual tongue avatar gradually morphed from *head* to *had* [13]. Results indicated that all participants altered their vowel quality to match the visually-presented word, *head*, suggesting entrainment.

The goal of the present study is to determine whether the sensory feedback monitoring problems thought to underlie the speech production difficulties of individuals with PD also extend to the visual modality. Specifically, we predict reduced tongue movement entrainment (or mirroring) effects in individuals with PD, compared with healthy participants. In addition, we test whether a type of widely used therapy affects this visual speech monitoring behavior. Based on studies suggesting that speech amplitude-based scaling training (e.g., Lee Silverman Voice Treatment [LSVT], "SPEAK OUT!") induces intra-systemic reorganization across speech production processes, we predict that, similar to healthy speakers, individuals undergoing this type of training will entrain and show AV shadowing of speech articulators [14].

10.21437/Interspeech.2018-2592

2. Methods

2.1 Participants

Fourteen speakers diagnosed with PD participated in the experiment. All were monolingual speakers of American English from the Dallas/Fort Worth community. None had any prior experience with the virtual tongue model. All participants were required to be off their Parkinsonism medications during testing (and for at least 12 hours prior) so that PD behaviors could be heightened. PD severity was assessed using part III (motor examination) of the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS UPDRS) [15], a commonly-used scale that includes measures of thought, behavior and mood, self-evaluation of daily life activities, motor exam, and assessment of motor complications. Cognitive abilities were screened using the Montreal Cognitive Assessment (MOCA), a rapid screening instrument for mild cognitive dysfunction (30 points possible, 26 or above = normal)[16]. Participant details are listed in Table 1. Eight participants (5 men, 3 women) were involved in speech/language therapy at the time of testing, indicated with a 't'. These patients had undergone at least three months of training in the SPEAK OUT! program at the Parkinson's Voice Center, Dallas Texas [17]. Six participants (5 men, 1 woman) were recruited from a local Parkinsonism support group (Dallas Area Parkinson's Society) and reported no current involvement in any speech/language therapy, nor any for the past year.

2.2 Visual Stimuli

The experiment used images from an animated 3D tongue avatar, with video data captured from actual tongue movements produced by a male native speaker of American English (WK) speaking the words *hid, head,* and *had*. Tongue images were created using an interactive articulatory feedback system, Opti-Speech [18], based on data input from the WAVE magnetometer system (Wave; NDI, Waterloo, Ontario, Canada). The Opti-Speech system represents speech movement as an avatar consisting of flesh-point markers and a modeled surface placed in a synchronously moving, transparent head. We used the words *head* and *had* because these lax vowels correspond with easily observed tongue movements and they have yielded robust shifts in previous perturbation experiments [13]. Video editing software (Camtasia 2, Techsmith, 2015) was used to record moving images of the tongue model while the /hVd/ words were produced by an adult male native speaker of American English. Animation software (Adobe After Effects, Adobe Systems, 2015) was subsequently used to morph video clips of the tongue avatar in a five-step continuum from *head* to *had*. In order to encourage simultaneous speech production while viewing the avatar tongue movements, each /hVd/ video clip was preceded by a "3,2,1" countdown and a green "get ready" signal (Figure 1, see also [13]).

The video materials were assembled in timed presentations for playout. During the experiment, stimuli were blocked, with a one-second inter-stimulus interval (ISI) between video clips, and a five-second inter-block interval (IBI). The entire speaking task took approximately 17 minutes.

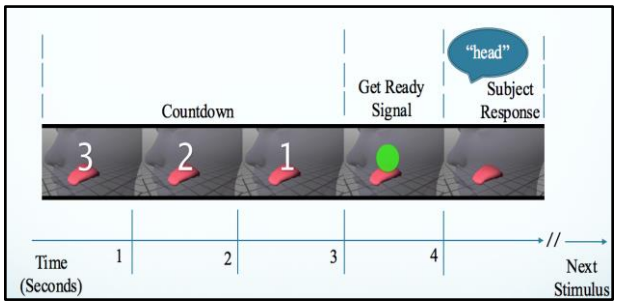

Figure 1. *Overview of the synchronized tongue viewing and speaking task. [13].*

Table 1: *Participant characteristics*

Subject code	Age	Years post diagnosis	MOCA	MDS UPDRS
t-M1	67	4	23	42
t-M2	76	3	25	33
t-M3	64	4	24	15
t-M4	69	7	--	40
t-M5	67	4	29	38
t-F1	57	10	24	33
t-F2	71	4	29	29
t-F3	75	8	--	20
M1	65	2	30	17
M2	71	3	23	28
M3	65	2	23	37
M4	73	2	25	18
M5	74	6	27	48
F1	45	3	29	38

Figure 2: *Individual with PD producing the word "head" in synchrony with Opti-Speech tongue avatar video clip. Participant's own speech is masked with pink noise.*

2.3 Procedure

Each participant was seated in a quiet room facing a computer monitor while wearing closed-cell headphones (Sennheiser HD 500) which transmitted masking (pink) noise at ~72 dB (Figure 2). A Tascam DR-05 recorder was used to record audio data. Participants were instructed to produce each word "in time with the moving tongue on the screen" following a three-second

countdown. No information was provided about the tongue avatar movement varying as a function of vowel type. After some warmup trials, the experimental trials were begun. For each trial the participant was textually cued to produce one of three target words (*hid*, *head*, or *had*) in synchrony with a visual model of the tongue that was producing one of those words, or was following a trajectory that interpolated two of those words. The experiment has four phases, completed in the following order: baseline, ramp, maximum hold, and after effect. During the baseline phase, participants were cued to produce five sets of the words *hid*, *head*, and *had*, in that order; on each trial, the tongue model was congruous with the target word. During the ramp phase, the participants were cued to produce 40 productions of *head*; however, the tongue model gradually traversed a five-point scale whose steps interpolated a canonical *head* and a canonical *had*. Beginning with the step that corresponded to a canonical *head*, each step in the scale was presented eight times in succession before moving to the next step.. During the maximum hold phase, the participant was cued to produce 100 productions of *head*, in synchrony with a tongue model that produced canonical *had*; these productions were grouped into five blocks of 20 productions. During the after effect phase, the participant was cued to produce 15 productions of *head*, in synchrony with a tongue model that produced canonical *head.*

A debriefing session was held immediately after the experiment finished. Participants were asked, "What did you notice about this experiment?" in order to obtain participants' impressions concerning the difficulty of the task and to detect whether participants were aware that the visual tongue positions had changed vowel quality. After recording participants' responses, we informed participants that the avatar had actually shifted from *head* to *had*, and participants were further queried whether they noticed such a change taking place.

2.4 Acoustical analyses

Each experimental session was recorded digitally as a single audio file. Individual productions of the target /hVd/ words were subsequently parsed into separate audio files and screened for background noise, resulting in 2322 target productions for analysis. The productions were randomized and anonymized so that neither the talker's identity nor the position of the production within the experimental sequence was known during the acoustical analysis. Linear predictive coding (LPC) in Praat was used to estimate the frequency of the first formant (F1) from the middle half of the vowel of each production. A custom Praat script allowed the user to view the waveform and spectrogram of each production, select the midpoint and endpoint of the target vowel, and interactively modify parameters of the LPC algorithm (i.e., frequency range, LPC order, window length) before estimating the F1 value in Hz. For each production, LPC parameters were varied until the resulting model matched the first three formants as represented in the spectrogram. Once the LPC parameters were set, the mean F1 value within the middle half of the vowel was recorded.

2.5 Statistical analysis

Prior to statistical analysis and plotting, each participant's F1 values were Lobanov z-scored relative to the mean and standard deviation of the F1 values measured from all their productions of *hid*, *head*, and *had*; each participant's z-scores were then centered according to their mean z-score for baseline productions of *head*. This reduced between-talker variability, including male/female differences due to dissimilarities in vocal tract length.

To test the effects of Treatment Group and Experimental Phase on F1 frequency, a linear mixed-effects model was fitted to the centered z-scored F1 values of productions of *head* from the baseline, maximum hold, and after effects phases. These three levels of the Experimental Phase variable were coded as treatment contrasts, with maximum hold as the reference level. The Treatment Group variable was likewise coded as a treatment contrast, with the SPEAK OUT! therapy group as the reference level. The model included uncorrelated random intercepts and random effects for Experimental Phase, grouped within each talker. The significance of model coefficients was determined with 95% confidence intervals.

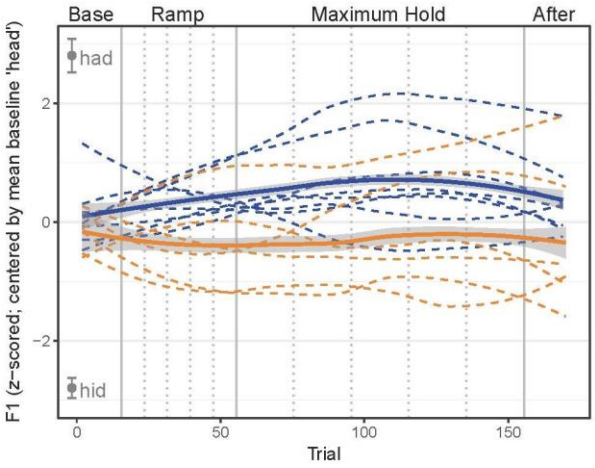

Figure 3: *Trajectories of normalized F1 frequencies for "head" productions by treated (n=8) and untreated (n=6) individuals with PD across the experimental phases. Treated individuals are shown with dark blue lines; untreated individual data with orange lines. Local nonparametric smooths are also shown.*

3. Results

Figure 3 plots local non-parametric smooths of the z-scored and centered F1 frequencies as a function of experiment trial number. The solid vertical lines divide the trials into baseline, ramp, maximum hold, and after effect phases consecutively. The vertical dotted lines indicate intra-phase divisions into blocks. The curves for the individual participants are shown as dashed lines, while the group-level curves are shown as solid lines with confidence intervals. Inspection of the group-level curves indicates that the participants in the SPEAK OUT! program demonstrated overall higher F1 values than the untreated speakers, perhaps reflecting lower jaw positions corresponding with overall louder productions noted for this treated group. Participants in the SPEAK OUT also showed a slight but significant increase in F1 frequency between the baseline and maximum hold phases, suggesting entrainment to the tongue-lowering manipulation of the tongue avatar. Furthermore, 7 of the 8 speakers in this group demonstrated such an increase in F1 frequency. Conversely, the group curve for the speakers who did not participate in a SPEAK OUT! treatment exhibits insignificant variation across the duration of the experiment, suggesting neither entrainment nor adaptation to the visual tongue-lowering manipulation. Moreover, only 2

of the 6 speakers in this group demonstrated an increase in F1 frequency between the baseline and maximum hold phases.

The fitted linear mixed-effects model indicated that for the treated SPEAK OUT! group, F1 frequency during the baseline phase was significantly less than during the maximum hold phase (b = -0.625, SE = 0.276, conf. int. = [-1.162, -0.089]). Comparing between groups in the maximum hold phase, the fitted model indicated that speakers in the untreated group produced vowels with significantly lower F1 than speakers in the treated SPEAK OUT! group (b = -0.922, SE = 0.423, conf. int. = [-1.745, -0.099]). Additionally, the interaction between Treatment Group and Experimental Phase for the baseline phase was comparable in magnitude, but opposite in sign from the simple effect of group (b =0.918, SE = 0.421, conf. int. = [0.099, 1.737]), indicating no difference between the baseline and maximum hold phases for the untreated group.

In terms of variation between speakers, the standard deviation of the effect of baseline between speakers was estimated in the fitted model to be 0.725 (conf. int. = [0.430, 1.064]). Therefore, while the fixed-effects coefficients in the fitted model indicated a significant increase in F1 between the baseline and maximum hold phases for the treated SPEAK OUT! group, this finding should be interpreted cautiously, in light of the estimated variance component for the random effect of baseline between speakers, whose magnitude is comparable to that of the corresponding fixed effect.

Upon debriefing, no participants indicated they were aware that the tongue avatar had switched from *head* to *had*. Some commented that the tongue position had changed, and one mentioned a change in "height." However, when the participants were told that the avatar had actually shifted to the word *had* and were also asked whether they were aware of having produced this word, all replied "no".

4. Discussion

In order to examine how visual information influences the vowel processing of individuals with PD during speech, a group of 14 talkers participated in an experiment in which /hVd/ words were elicited while viewing a synchronous, moving tongue image. Productions of *hid*, *head*, and *had* were first obtained during a baseline phase in which the moving tongue model was congruous with the target word. Next, during a ramp phase, the tongue model gradually morphed to producing *had* while the participants were instructed to produce *head*. In a following maximum hold phase, productions of *head* were elicited while the tongue model produced *had*. Lastly, in an after effect phase productions of *head* were elicited with congruous movement of the tongue model producing *head*.

Previous findings established that healthy individuals shift towards the visually presented stimuli. That is, during the ramp phase, talkers' *head* become more like *had*, remained similar to *had* values during maximum hold, and then returned to *head* F1 values during the after effect phase. For individuals with PD, we predicted that speech sensorimotor problems would mitigate this entrainment behavior, with untreated individuals showing more attenuated behavior than those receiving treatment for reduced vocal intensity. It was also predicted that SPEAK OUT! treatment would correspond with individuals behaving more like healthy adults and showing entrainment towards the visually presented word, *head*.

Although the current findings must be interpreted cautiously due to the relatively high between-speaker variance and the small number of participants tested, results are nevertheless consistent with our predictions. F1 values for the untreated group of individuals with PD showed little evidence of change as a function of experimental phase, suggesting minimal effects of viewing the tongue avatar on their vowel quality. In contrast, the F1 values of the individuals with PD who participated in the SPEAK OUT! increased significantly from the baseline phase to the maximum hold phase, indicating a shift from *head* to *had*. Together, these patterns support the notion that (untreated) individuals with PD have difficulty in processing audiovisual imagery relevant to speech articulation. In addition, the data provide support for the second prediction, that individuals with PD engaged in amplitude-based scaling therapy will show improved speech processing, including the processing of audiovisual information.

While a cross-study comparison must necessarily be considered indirect, the SPEAK OUT! treated group here showed a similar degree of entrainment as noted for healthy adult participants in our previous experiment using this same paradigm [13]. The current findings that the SPEAK OUT! group showed entrainment effects may relate to recent reports that individuals with PD are able to use metrical (temporal) auditory speech cues to entrain to speech [22]. Together, the data suggest possible strategies using entrainment in the auditory and visual channels as a possible means of addressing the speech difficulties of individuals with PD.

Additional research is needed to replicate the current findings and to better describe the audiovisual speech capabilities of individuals with PD. In addition to sensory feedback deficits, there are other potential reasons why individuals with PD might perform poorly at this task, including attentional impairment [19], visual deficits in motion perception [20], and difficulty with dual task performance [21]. Finally, the current experiments using video presentation of perturbed speech should be contrasted with actual, online kinematic perturbation (e.g,, via EMA or WAVE) in order to more fully examine healthy and PD speech behavior patterns.

5. Conclusions

The present findings suggest that talkers with PD have difficulty in sensorimotor processing associated with linking a synchronous visual image of the moving tongue to their own speech production. When tasked with saying *head* while repeat images of the tongue morph from *head* to *had*, they do not entrain in a manner similar to healthy adults. Rather, they maintain constant vowel quality. In contrast, individuals with PD treated with SPEAK OUT! therapy perform more like healthy adult talkers, showing entrainment toward the visually displayed sound. Future studies are needed to explore the timing of these effects and to determine whether they generalize across different stimuli and for individuals with varying levels of PD severity.

6. Acknowledgements

We thank Prof. Thomas Campbell and members of the University of Texas at Dallas C-Tech Center for providing resources for this research. We also thank Diane Walsh, the members of the Parkinson's Voice Project (DVP), the Dallas Area Parkinsonism Society (DAPS), and the Bedford Public Library for contributing their valuable time and resources to this research.

7. References

[1] S. Sveinbjornsdottir. "The clinical symptoms of Parkinson's disease," *Journal of neurochemistry*, vol. 139(S1), pp. 318-324, 2016.

[2] K. Tjaden. "Speech and swallowing in Parkinson's disease," *Topics in geriatric rehabilitation*, vol. 24(2), pg. 115, 2008.

[3] C. Arnold, J. Gehrig, S. Gispert, C. Seifried, and C.A. Kell. "Pathomechanisms and compensatory efforts related to Parkinsonian speech", *NeuroImage: Clinical*, vol. 4, pp. 82-97, 2014.

[4] S, Cai, S. S. Ghosh, F. H. Guenther, and J. S. Perkell, "Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing," *Journal of Neurosci*, vol. 31, no. 45, pp. 16483-16490, 2011.

[5] H. Liu and C. R. Larson, "Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex," *Journal of the Acoustical Society of America*, vol. 122, pp. 3671, 2007.

[6] J. F. Houde and M. I. Jordan, "Sensorimotor adaptation of speech I: Compensation and adaptation," *Journal of Speech Language and Hearing Research*, vol. 45, pp. 295-310, 2002.

[7] M. Shum, D. M. Shiller, S. R. Baum, and V. L. Gracco, "Sensorimotor integration for speech motor learning involves the inferior parietal cortex," *European Journal of Neuroscience*, vol. 34, no. 11, 2011.

[8] X. Chen, X. Zhu, E.Q. Wang, L. Chen, W. Li, Z. Chen, and H. Liu. "Sensorimotor control of vocal pitch production in Parkinson's disease," *Brain Research*, 1527, 99-107, 2013.

[9] H. Liu, E.Q. Wang, L.V. Metman, and C.R. Larson."Vocal responses to perturbations in voice auditory feedback in individuals with Parkinson's disease," *PloS one*, 7(3), p.e33629, 2012.

[10] F. Mollaei, D. M. Shiller, and V. L. Gracco, "Sensorimotor adaptation of speech in Parkinson's disease," *Movement Disorders*, vol. 28, pp. 1668-1674, 2013.

[11] L. Menard, "Multimodal speech production," *The Handbook of Speech Production*, M. A. Redford. UK: J. Wiley, 200-221, 2015.

[12] W. Katz., "New horizons in clinical linguistics," *The Routledge Handbook of Phonetics*: Routledge, in press, 2018.

[13] W. Katz and D. Prabhakaran. "Sensorimotor response to visual imagery of tongue displacement. *In 17th Annual Conference of the International Speech Communication Association , Sept 8-16, San Francisco, California, Proceedings*, 2016,pp. 2090-2094, 2016.

[14] C.R. Watts. "A retrospective study of long-term treatment outcomes for reduced vocal intensity in hypokinetic dysarthria," *BMC Ear, Nose and Throat Disorders*, vol. 16(1), 2, 2016.

[15] Movement Disorder Society Task Force on Rating Scales for Parkinson's Disease. "The Unified Parkinson's Disease Rating Scale (UPDRS): status and recommendations," *Movement disorders: official journal of the Movement Disorder Society*, vol. 18(7), p.738. 2003.

[16] Z.S. Nasreddine, N.A. Phillips, V. Bédirian, S. Charbonneau, V. Whitehead, I. Collin, J.L. Cummings, and H. Chertkow. "The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment," *Journal of the American Geriatrics Society*, vol 53(4), pp.695-699, 2005.

[17] K. Wiley, S. Elandary. "SPEAK OUT!: A practical approach to treating Parkinson's," *San Antonio: Texas Speech and Hearing Association Annual Convention*, 2014.

[18] Katz, W., Campbell, T. F., Wang, J., Farrar, E., Eubanks, J. C., Balasubramanian, A., and Rennaker, R. "Opti-speech: A real-time, 3D visual feedback system for speech training" *In Fifteenth Annual Conference of the International Speech Communication Association, Sept 14-18, Singapore, Proceedings*, pp. 1174-1178, 2014.

[19] Z.R. Brown and C. Marsden. "Dual task performance and processing resources in normal subject and patients with Parkinson's disease," *Brain*, vol. 114, pp. 215-231 , 1991.

[20] R.S.Weil, A.E. Schrag, J.D. Warren, S.J. Crutch, A.J. Lees, and H.R. Morris, "Visual dysfunction in Parkinson's disease," *Brain*, vol. 139(11), pp.2827-2843. 2016.

[21] J. A Foley, R. Kaschel, and S.D. Sala. "Dual task performance in Parkinson's disease," *Behavioural Neurology*, vol. 27(2), 183-191, 2013.

[22] M. Späth, I. Aichert, A.O. Ceballos-Baumann, E. Wagner-Sonntag, N. Miller, and W. Ziegler, "Entraining with another person's speech rhythm: Evidence from healthy speakers and individuals with Parkinson's disease," *Clinical linguistics and phonetics*, vol. 30(1), pp.68-85. 2016.