

The effect of manual action on vowel production in nouns and verbs

Nicole A. van Rootselaar, Marisa Lelekach, Fangfang Li, Claudia L.R. Gonzalez

The Brain in Action Lab, University of Lethbridge, The University of Lethbridge.
nicole.vanrootselaar@uleth.ca, fangfang.li@uleth.ca, claudia.gonzalez@uleth.ca

ABSTRACT

Studies show that word pronunciation alters grasping based either on facilitation between mouth and hand shape, or facilitation between speech meaning and hand movement. Less understood is the opposite of this relation, i.e., how *speech* is altered when meaningful words are pronounced during manual actions. To assess this, participants pronounced nouns and verbs that are homophones while performing different actions, including observing a block (look, served as baseline), placing fingers around a block (grasp), and reaching and grasping a block (reach-to-grasp). We found that manual actions influenced vowel production, significantly lowering the first formant during both the grasp and reach-to-grasp conditions. However, vowel production was not affected by words of different part of speech and no difference was observed between the verb/noun pairs. Our findings have implications for theoretical modelling of speech production processes to incorporate a linkage to the general kinaesthetic system.

Keywords: vowel production, formants, grasping, semantics

1. INTRODUCTION

Speech is a type of action. Surprisingly, most past and/or current speech production models do not take this into consideration and ignore the influence of other types of actions such as manual movement on speech production [9,21]. For example, the well-known Levelt's model [21] proposes three strata; the conception stratum where meanings are packed into expressible words and sentences, the morpho-syntactic stratum to encode words into grammatically correct forms, and a motor stratum to put articulators into actions. Whether such as a system is modular (i.e., automatic and independent of other processes) or can interact with other more general neurological and/or motor systems is not specified. This is an important question, given the growing research that documents how speech production affects manual movements when the two types of actions are executed simultaneously. Past research suggests a facilitative effect of simple syllable production on grasping when finger shape reflects position of the tongue [28]. Past research has also found that (pronunciation and observation of) nouns alter reaching trajectories and finger shaping

to reflect the word being pronounced/read rather than the object being grasped [16]. For example, when saying the word "apple" and grasping a grape, the fingers will open wider than necessary, reflecting the larger apple. There is mounting evidence of how manual movement is altered during both speech production and processing [for more, see works: 4, 5, 6, 10, 11, 12, 13, 15, 26, 27, 28], but the same cannot be said for the opposite interaction; how speech is altered during the execution of manual action. Past research is inconsistent in finding whether speech was altered during manual movement. In a study, participants produced syllables while grasping a joystick, and no changes in formants were observed, despite changes in hand shape and trajectory [26]. To the contrary, another experiment indicated that producing a syllable while grasping and moving a large or small fruit towards the mouth altered vowel production in accordance to the size of the fruit [13]. A larger item resulted in a larger mouth opening, and vice versa. There were accompanying changes in formants, which indicate the position of the tongue, and mouth shape during vowel production [17, 19, 22]. Furthermore, these authors found that speech is even sensitive to the observation of action. In that study, the aperture of the lips reflected the aperture of the hand grip when participants watched a volunteer grasp a large or small object and move it towards their (the volunteer's) mouth [14] or simply reach and pick it up [15]. These significant findings imply an interaction between *planning and executing* manual movement and speech which results in coordination between the movement of both the fingers and lips (and other articulators). Both right hand motor control and speech processing and production are known as lateralized functions, and in most of the population they are specialized in the left hemisphere of the brain [18]. It is probable that both manual and oral movements share similar neurological etiology that relates to their neurological organization (lateralization) [18].

Given this known relationship, it unexpected that both previously described studies did not note a change in formants during different manual actions. One notable commonality between these two studies is that participants pronounced consonant-vowel syllables rather than words (meaningful speech). There is scant evidence, at least to our knowledge, of

studies investigating changes to the pronunciation of words when produced simultaneously with an action. The lack of evidence may be related to the additional challenges, specifically, the task of accounting for the associated meaning. In studies measuring manual movement during word production, researchers found that pronouncing action words (verbs) alters hand movement to mirror speech meaning (i.e. saying “fast” results in the hand moving faster [4]). Such studies imply that not only is there a connection between manual movement and speech, but that the neural areas responsible for semantic processing also affect this system, enabling the meaning of spoken language to further interact with movement [23]. This idea fits well with what we know about the organization of grasping and language in the brain. In addition to both right hand motor control and speech motor control being lateralized, various imaging studies demonstrate that words belonging to different parts of speech (i.e. nouns and verbs) can be distinctly processed, but still lateralized to the left hemisphere [23, 29].

One additional shared feature of previous research examining the link between speech and grasping is the presence of non-functional grasping actions. Authors had participants place fingers around a cylinder [27], observe a grasp [5, 14, 15] or close fingers around object without first reaching for the object [14, 15]. These various studies beg the question; does the nature of the manual action differentially affect speech? Previous imaging studies suggest that the brain processes natural, ecologically relevant grasping actions differently than awkward, inefficient grasps [2]. This altered processing may translate into unique interactions with the semantic and speech planning systems.

Through this study, we attempted to understand if the interaction between meaningful speech production and natural manual movement reflects this theory, and to develop further support for altering speech production models (such as Levelt’s Model [21]) by including other factors relevant to speech production. We investigated if pronunciation of meaningful words is altered when performed with different, simultaneous manual actions. Specifically, regarding action, we predicted that performing a less functional grasping action would alter vowel production more than a functional action or observation of a block. Secondly, we predicted that pronouncing nouns or verbs would lead to distinct changes in speech production, because each part of speech is associated with different embodied meanings between verbs and nouns.

2. METHODS

2.1. Participants and materials

Thirty (17 females) right-handed university students participated in the present study, and each provided written informed consent.

Each participant sat at a table with a large computer monitor (51 cm display) located 80 cm away. We placed a small black mat between them on the table (22 x 14.5 cm) within comfortable reaching distance. A piece of tape on the edge of the table in front of the participant marked the centre of the table and the rest position for participant’s hands. A Shure SM87A microphone was suspended approximately 15 cm below the participant’s mouth to avoid interfering with reaching and grasping actions. The microphone was attached to a Marantz flashcard PMD661 recorder. We used two items for grasping targets; a small block (2x2x2 cm) and a large block (5x5x2 cm). To guide participants through which word form to pronounce (verb or noun), we used Super Lab (version 4.5) to display the target word and picture on the computer monitor.

2.2. Procedure

Before beginning the experimental trials, the participant viewed the different image slides to become familiar with each of the four words and their associated meanings (verb versus noun). On a computer monitor, a slide show was presented which featured targets ([fɔl] and [bɔl]) and distractor words paired with an image to give context to the meaning of the homophone pairs (i.e. “Fall” paired with a person slipping, or “Fall” shown with a tree with orange and red leaves during the season of fall)). After viewing these images, participants also received instructions on how to perform the different grasping actions.

There were four different grasping protocols. The first was a *control* condition, where no targets were present, and participants did not move their hands. The second was the *look* condition, where a block was placed on the target mat, but participants did not perform any movements towards it. For the third condition, known as *grasp*, a block was placed on the edge of the table between the participant’s fingers in the rest position. Participants placed their right-hand fingers around a block without moving their arm. During the fourth condition (*reach-to-grasp*), we placed a block on the target mat, and participants reached forward with their right hand to grasp and pick up the block. The small and large blocks were pseudo-randomly presented during the experiment. The participant would complete a block composed of eight trials, performing the same grasping condition throughout.

Before beginning the trials, participants went through several practice trials, combining the different grasping actions and words. The practice and experimental trials were executed identically. Before starting the block, the participant was told which grasping action to execute for the following trials. A trial began with the participant viewing a picture paired with the word on the computer screen. After reading and memorizing the word, they closed their eyes. Next, the experimenter played two beeps. At the first beep, the participant opened their eyes and observed the target grasping area (either on the mat (*control*, *look*, *reach-to-grasp*), or between their fingers (*grasp*). 1000 ms later, the second beep played, and the participant executed the appropriate manual action while pronouncing the target word. Each participant pronounced the target words (fall/fall and bawl/ball) thirty-two times, or four times per each grasping condition.

2.3. Analysis

The recordings were uploaded and analysed using Praat (version 5.3.51 [3]). Each individual word was identified, and the vowel was manually segmented from the surrounding consonants. We extracted the median values of the first three formants (F1, F2, F3). Additionally, the duration of the vowel was determined by subtracting the time when vowel production began from the time when the vowel was marked as finished. To determine the centroid of the vowel, we identified the target word and manually scanned for the initiation of the sinusoidal pattern which corresponded with the appearance of formants (dark horizontal bands representing higher intensity). Next, we marked the conclusion of the vowel by marking the location of breakdown of the sinusoidal pattern and diminishing of the formants and their intensity.

Statistical analysis was completed using R Studio [25], and statistical package lme4 [1]. We applied a generalized linear mixed effects model to investigate the effects of part of speech and manual action on vowels in meaningful speech. We included part of speech (nouns versus verbs), manual action (*control*, *look*, *grasp*, and *reach-to-grasp*), block size (small, large), condition order (indicates which words were spoken earlier [more novel] or later in trial [more repetitive], and sex (male, female) as fixed effects. Participant and words (“ball/bawl” versus “fall/fall”) were included as random effects (for example: 1|participant). Through visual inspection of residual plots, we determined that residuals were homoscedastic and normally distributed. We determined p-values by using likelihood ratio tests from comparisons of the full model with each relevant effect against a model constructed without the effect. Follow up tests were

conducted to compare significant interactions. Where appropriate, we applied Bonferroni corrections.

3. RESULTS

3.1. First Formant

The model revealed main effects of action ($p = 0.002$), and sex ($p < 0.001$). For action, follow-up tests indicated that in the *grasp* condition, the value of the first formant was approaching significance control condition ($p = 0.076$). *Reach-to-grasp* followed this pattern as well ($p = 0.020$). *Look* was not significantly different for either actions or the control ($p > 0.05$).

For the remaining variables, we did not find any relevant results. With the second formant, the significant main effect of sex is present, but no other main effects or interactions of interest were found. In the third formant, there was a main effect of sex and order of pronunciation. No relevant results were found for vowel duration.

3. DISCUSSION

According to previous work examining the interaction between manual action and speech production, we developed two hypotheses. First, we predicted that less functional grasping (the *grasp* condition) would alter simultaneous vowel production more than a natural grasp (*reach-to-grasp*) or observation of an object (*look*). Second, we hypothesized that the meaning of the pronounced words would distinctly alter speech during grasping. Specifically, pronouncing verbs while performing manual action would significantly change formants compared to noun production. Our first hypothesis was partially confirmed—there was an effect of action on the first formant. F1 is inversely proportional to jaw opening [19,22], so our findings suggest that regardless of the word or associated meaning, performing a grasping action during speech results in a more closed jaw (and a higher tongue placement in the mouth). This is different from previous studies which measured vowel production, and only found changes to the second and third formants [14,15]. However, there is no significant difference between the *look* and *control* conditions. Thus, the observation of a block is not enough to elicit a change in pronunciation. Gentilucci et al. [14,15] demonstrated changes in formants when participants observed someone pick up or pantomime picking up an object. The combination of our null findings in the *look* condition and supporting research [5,14,15] suggests that it is observation of body movement, and not the object itself which evokes changes in speech.

Between the two grasping conditions—*grasp* (an awkward position), and *reach-to-grasp* (a natural movement), there was no differences in F1. It is

possible that we did not observe differences between the grasping conditions because both used a similar hand shape. In future studies, using a more unnatural hand position (i.e. grasp with the thumb and pinkie or ring finger) may prove drastic enough to distinctly alter vowel production.

A notable difference between the current and previous studies is that the changes in formant were related to the size of the stimuli grasped, rather than completely different grasp positions [14,15]. Size of the block was not a significant factor in the present study. This lack of consistency with previous research may stem from several differences between ours and the other studies: For example, the use of consonant-vowel syllables instead of meaningful words (consonant-vowel-consonant), the use of different vowels, and unlike our study, a non-grasping control condition (producing speech without performing or observing a movement) is usually not included in other studies. All these methodological differences may contribute to the discrepancy between our findings and previously published results.

One way to determine whether (for example) the stimuli type is a factor that affects the results is to select open syllable words such as “fly” (noun: a bug, or verb: to move through air). Using pairs like this one will minimize formant changes that could be attributed to surrounding consonants but not word meaning or manual action. And we would expect to observe vowel changes during executed actions if the mouth motor systems are susceptible to both manual movement and semantic messages of words.

Another potential avenue is to choose a homophone pair which involves the hand. Previous studies demonstrate that words related to hand movement evoke a greater change in hand kinematics [6, 10,11]. Perhaps our lack of significant results related to word meaning is due to not using hand specific verbs (“bawl” and “fall”). Producing a verb such as “grasp” should alter hand movement more [10], and in turn may produce greater changes in speech. We plan to consider this in the future to better ascertain the factors that impact speech production during manual actions.

A limitation of this study was the varying frequency of the words used. Out of the top 60,000 ranked words used on a variety of websites, (according to [7]), the verb “bawl” ranked 32245, while “ball”, the noun, ranked 962. “Fall” (verb) ranked 638, and the noun “fall” was 1262. It is possible that these low frequency words (particularly bawl) may be less salient, evoking weaker action associations, resulting in reaction in the manual movement. In follow-up studies, we will also try to select word stimuli where both noun/verb

components have higher, but similar levels of frequency in everyday use.

Current speech production models tend to focus only on components that are linguistically relevant. As mentioned, Levelt’s model [21] focuses on three levels/stratum: the conception, the morpho-syntactic, and a motor. In the model, motor refers to speech articulators and does not consider other motor actions which are commonly executed simultaneously. It is now time to revise this model by incorporating the most recent studies including our own that demonstrate clear interaction between speech articulation (or possible word meaning) and other motor actions.

In short, despite the inconclusive results we found from the current study (e.g. alterations only to F1 and no other formants as in previous studies), the effect of manual action on speech production has been confirmed. We propose to revise the current speech production models to incorporate a linkage between speech articulation and manual action. Our research will shed light on the possible gaps in building such models and therefore has implications to guide future research in order to refine these models.

Table 1. Sample data from a female participant. Data presented according to grasping condition, part of speech, and dependant variables (formants and vowel duration). Values are averaged across size.

			Control	Look	Grasp	Grasp-to-place
Bawl	F1	Noun	585.02	582.51	560.28	562.07
		Verb	565.91	572.99	567.03	569.40
	F2	Noun	918.59	901.77	877.85	862.03
		Verb	884.78	910.72	884.96	904.81
	F3	Noun	2711.52	2790.07	2816.88	2872.10
		Verb	2817.43	2771.36	2806.13	2787.06
Duration	Noun	0.1741	0.1888	0.2267	0.1918	
	Verb	0.1948	0.1670	0.2124	0.1903	
Fall	F1	Noun	592.93	575.69	568.07	579.36
		Verb	572.51	574.16	579.01	584.30
	F2	Noun	878.10	877.58	873.71	872.68
		Verb	878.99	886.56	860.55	869.76
	F3	Noun	2850.75	2822.64	2814.58	2856.29
		Verb	2808.21	2838.20	2833.21	2847.58
	Duration	Noun	0.1459	0.1792	0.1692	0.1769
		Verb	0.1709	0.1716	0.2095	0.1950

5. REFERENCES

- [1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.
- [2] Begliomini, C., Caria, A., Grodd, W., Castiello, U. 2007. Comparing natural and constrained movements: new insights into the visuomotor control of grasping. *PloS one*, 2(10), e1108.
- [3] Boersma, P., Weenink, D. 2013. Praat: doing phonetics by computer [Computer program]. Version 5.3.51, retrieved January 2015 from <http://www.praat.org/>
- [4] Boulenger, V., Roy, A. C., Paulignan, Y., Deprez, V., Jeannerod, M., and Nazir, T. A. 2006. Cross-talk between Language Processes and Overt Motor Behavior in the First 200 msec of Processing. *Journal of Cognitive Neuroscience*, 18(10), 1607-1615.
- [5] Campione, G. C., Gentilucci, M. 2011. Is the observation of the human kinematics sufficient to activate automatic imitation of transitive actions? *Beh Brain Res*, 225(1), 201-208.
- [6] Dalla Volta, R., Gianelli, C., Campione, G. C., Gentilucci, M. 2009. Action word understanding and overt motor behavior. *Expel brain res*, 196(3), 403-412.
- [7] Davies, M. 2018. Ma. *iWeb Corpus*. Retrieved from <https://corpus.byu.edu/iweb>.
- [8] Decker, D. M. 1999. *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge University Press.
- [9] Dell, G.S. 1997. "Analysis of Sentence Production". *Psychology of Learning and Motivation*: 133-177.
- [10] Fargier, R., Menoret, M., Boulenger, V., Nazir, T. A., Paulignan, Y. 2012. Grasp It Loudly! Supporting Actions with Semantically Congruent Spoken Action Words. *PLoS ONE*, 7(1), e30663.
- [11] Frak, V., Nazir, T., Goyette, M., Cohen, H., Jeannerod, M. 2010. Grip Force Is Part of the Semantic Representation of Manual Action Verbs. *PLoS ONE*, 5(3), e9728.
- [12] García, A. M., Ibáñez, A. 2016. A touch with words: dynamic synergies between manual actions and language. *Neuroscience Biobehavioral Reviews*, 68, 59-95.
- [13] Gentilucci, M., Benuzzi, F., Gangitano, M., Grimaldi, S. 2001. Grasp with hand and mouth: a kinematic study on healthy subjects. *J of Neurophysiology*, 86(4), 1685-1699.
- [14] Gentilucci, M., Santunione, P., Roy, A. C., Stefanini, S. 2004. Execution and observation of bringing a fruit to the mouth affect syllable pronunciation. *Euro J of Neuro*, 19(1), 190-202.
- [15] Gentilucci, M., Stefanini, S., Roy, A.C., Santunione, P., 2004. Action observation and speech production: study on children and adults. *Neuropsychologia* 42, 1554- 1567
- [16] Glover, S., Rosenbaum, D. A., Graham, J., Dixon, P. 2004. Grasping the meaning of words. *Exper Brain Research*, 154(1), 103-108.
- [17] Hardcastle, W.J., Laver, J. and Gibbon, F.E. eds., 2012. *The handbook of phonetic sciences* (Vol. 116). John Wiley & Sons.
- [18] Hauk, O., Johnsrude, I., Pulvermüller, F. 2004. Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron*, 41(2), 301-307.
- [19] Hoemeke, K. A., Diehl, R. L. 1994. Perception of vowel height: The role of F1-F0 distance. *The Journal of the Acoustical Society of America*, 96(2), 661-674. doi:10.1121/1.410305
- [20] Kolb, B., & Whishaw, I. Q. (2009). *Fundamentals of human neuropsychology*. Macmillan.
- [21] Levelt, W. J.M. 1999. Models of Word Production. *Trends in Cog Sciences*. 3, 223-232.
- [22] Lindblom, B. E. F., Sundberg, J. E. F. 2005. Acoustical consequences of lip, tongue, jaw, and larynx movement. *J of Acous Soc of Am*, 50(4), 1166-1179.
- [23] Moseley, R. L., Pulvermüller, F. 2014. Nouns, verbs, objects, actions, and abstractions: local fMRI activity indexes semantics, not lexical categories. *Brain and language*, 132, 28-42.
- [24] Preissl, H., Pulvermüller, F., Lutzenberger, W., Birbaumer, N. 1995. Evoked potentials distinguish between nouns and verbs. *Neuroscience Letters*, 197(1), 81-83.
- [25] R Core Team 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- [26] Tiainen, M., Tiippana, K., Vainio, M., Komeilipoor, N., & Vainio, L. (2017). Interaction in planning vocalizations and grasping. *Quar J of Exper Psychology*, 70(8), 1590-1602.
- [27] Vainio, L., Rantala, A., Tiainen, M., Tiippana, K., Komeilipoor, N., Vainio, M. 2017. Systematic Influence of Perceived Grasp Shape on Speech Production. *PLoS ONE*, 12(1), e0170221.
- [28] Vainio, L., Schulman, M., Tiippana, K., Vainio, M. 2013. Effect of Syllable Articulation on Precision and Power Grip Performance. *PLoS ONE*, 8(1), 10.
- [29] Zhang, Z., Sun, Y., Wang, Z. 2018. Representation of action semantics in the motor cortex and Broca's area. *Br and lang*, 179, 33-41.