# EFFECTS OF MOUTHING AND INTERLOCUTOR PRESENCE ON MOVEMENTS OF VISIBLE VS. NON-VISIBLE ARTICULATORS

**Katie Bicevskis[*1], Jonathan de Vries[1,2], Laurie Green[1], Johannes Heim[1], Jurij Božič[1], Joe D'Aquisto[1], Michael Fry[1], Emily Sadlier-Brown[1], Oksana Tkachman[1], Noriko Yamane[1], Bryan Gick[1,3]**

[1] Department of Linguistics, University of British Columbia, Vancouver, BC, Canada

[2] Interdisciplinary Studies Graduate Program, University of British Columbia, Vancouver, BC, Canada

[3] Haskins Laboratories, New Haven, CT, USA

## Résumé

Les locuteurs prennent en compte l'information qu'un partenaire de conversation nécessite pour mieux comprendre une expression. Malgré l'évidence grandissante que les mouvements d'articulateurs visibles (comme les lèvres) sont augmentés dans l'articulation silencieuse par rapport à l'articulation vocalisée, peux d'études ont comparé cet effet dans les articulateurs visibles contre les articulateurs non visibles. De plus, aucune étude n'a examiné si l'engagement de l'interlocuteur changera ces résultats. En élaborant un conception d'expérience présent/non présent, nous avons testé si la présence d'information audible et/ou d'un interlocuteur affecte les mouvements des lèvres et de la langue. Les participants ont parlé trois syllabes, avec et sans production audible, dans chacune des conditions interlocuteur-présent et interlocuteur-non présent. Les mouvements des lèvres et de la langue étaient enregistrés avec la vidéo et l'échographie. Nos résultats montrent que la protubérance des lèvres était plus grande dans les conditions non audibles par rapport à ceux audibles et que les mouvements de la langue étaient atténués (/wa/) ou non affectés (/ri/, /ra/) par ces mêmes conditions, indiquant les effets différents pour les articulateurs visibles et non-visibles dans l'absence d'un signal auditif. Une interaction significative entre les conditions d'engagement sociale et d'audibilité de vocalisation avec référence à la fermeture orale a montré que les participants ont produit des fermetures plus étroites dans les conditions de vocalisation audible, interlocuteur-non présent (par rapport à la condition interlocuteur-présent). Cependant, les mesures de protubérance des lèvres n'étaient pas affectées par condition d'engagement sociale. Nous concluons que les locuteurs utilisent à la fois les modalités auditives et visuelles dans la présence d'un interlocuteur, et lorsque l'information acoustique n'est pas disponible, les augmentations compensatoires sont réalisés dans le domain visuel. Nos résultats soulignent encore le caractère multimodal de discours, et posent des nouvelles questions au sujet des adaptations différentielles faites par les articulateurs visibles et non visibles dans les différentes conditions de parole.

**Mots clefs:** production de la parole, effets interlocuteur, parole silencieuse, feedback auditif et visuel, échographie

## Abstract

Speakers take into account what information a conversation partner requires in a given context in order to best understand an utterance. Despite growing evidence showing that movements of visible articulators such as the lips are augmented in mouthed speech relative to vocalized speech, little to date has been done comparing this effect in visible vs. non-visible articulators. In addition, no studies have examined whether interlocutor engagement differentially impacts these. Building on a basic present/not-present design, we investigated whether presence of audible speech information and/or an interlocutor affect the movements of the lips and the tongue. Participants were asked to a) speak or b) mouth three target syllables in interlocutor-present and interlocutor-not-present conditions, while lip and tongue movements were recorded using video and ultrasound imaging. Results show that lip protrusion was greater in mouthed conditions compared to vocalized ones and tongue movements were either attenuated (/wa/) or unaffected (/ri/, /ra/) by these same conditions, indicating differential effects for the visible and non-visible articulators in the absence of an auditory signal. A significant interaction between the social engagement and vocalizing conditions in reference to lip aperture showed that participants produced smaller lip apertures when vocalizing alone, as compared to when in the presence of an interlocutor. However, measures of lip protrusion failed to find an effect of social engagement. We conclude that speakers make use of both auditory and visual modalities in the presence of an interlocutor, and that when acoustic information is unavailable, compensatory increases are made in the visual domain. Our findings shed new light on the multimodal nature of speech, and pose new questions about differential adaptations made by visible and non-visible articulators in different speech conditions.

**Keywords:** speech production, interlocutor effects, mouthed speech, auditory and visual feedback, ultrasound imaging

[*] k.bicevskis@alumni.ubc.ca

## 1 Introduction

This study examines how the motion of visible articulators (e.g. the lips) and non-visible articulators (e.g. the tongue) are affected by two factors: (1) the presence of proprioceptive auditory feedback and (2) the presence or absence of an interlocutor. A large body of literature now points to the importance of the visual modality in speech perception [3, 8, 14, 15, 16, 21, 22, 24]. Perceptual accuracy generally increases when the perceiver can both hear and see a speaker. In light of such results, we ask whether an articulator's visibility (i.e. visible or less visible) will affect its magnitude of movement when information from the visual modality becomes more important.

Because of the non-trivial contribution of vision to speech perception, it is perhaps not surprising that speakers tend to increase facial movements in environments where the auditory signal is degraded [5, 6, 10]. Hazan & Kim [10] found that speakers visually enhanced their articulation of /æ/, /i/ and /ɛ/ (indicated by an increase in inter-lip area) when they were required to carry out a communicative task in noise. Increases in visible articulator movement could be interpreted as a mechanical side-effect of the increased effort required to speak louder in noisy settings. This increase in speech effort, usually referred to as Lombard Speech, was first noted by Lombard [13], who found an immediate and involuntary vocal increase as a response to noise. Interestingly, Herff, Janke, Wand & Schultz [11] found increased facial movement in noisy conditions in silent as well as vocalized articulation. These findings suggest that visible articulator movements increase in order to compensate for a degraded or absent auditory signal, even in the case of the relatively unnatural condition of silent speech. Furthermore, Ménard, Leclerc, Brisebois, Aubin & Brasseur's [17] study comparing blind and sighted speech found that in the production of French vowels, blind speakers demonstrated less difference in upper lip protrusion than sighted speakers and Cvejic, Kim & Davis [4] found that speakers made auditory cues (e.g. to prosody) more salient when it was known that visual cue information was unavailable to their conversation partner. Together, such findings imply that speakers take into account what type of information an interlocutor will require to best understand a given utterance in a given context. In the present study, rather than using noise to effect signal degradation, we include mouthed and vocalized utterances in order to examine how the *absence* or *presence* of an auditory signal affects the visible and non-visible articulators, respectively. Similar to previous work, we hypothesized that the movement of visible articulators would increase while mouthing, that is, when the auditory signal is absent.

While previous work has illustrated that the movement of visible articulators tends to increase when the visual modality is more important, such as when auditory information is degraded or absent, very little attention has been paid to the role of non-visible articulators (tongue).Though some work has been done examining the impact of visibility on articulator movement, samples have been small (i.e. a single participant in [7]). It has been suggested based on this data that tongue movements that are less visible do not increase in magnitude in noise, and that lip movements are not more enhanced in noise when interlocutors can see each other. However, these results should be seen as suggestive rather than conclusion due to the study's small sample size, a problem we attempt to rectify. A relatively clear prediction for the movement of articulators carrying less visual information may be formulated, namely that the movements of less visible articulators such as the tongue should be significantly less affected by changes in the environment which require increased attention to visual information. An alternative hypothesis would maintain that, as speech in noise is augmented in a variety of ways not exclusively visual [23], the augmentation should not be sufficiently sensitive to the modality-specific needs of an interlocutor, and should extend equally to both visible and non-visible articulators. To test our hypotheses, we employ simultaneous ultrasound and video imaging to capture the behaviour of the lips and tongue.

Considering visible and non-visible articulators also mandates consideration of social context, as previous studies indicate that visible articulator movements increase in saliency in the presence of an interlocutor [6, 10]. For example, Hazan & Kim's [10] study found that the size of lip gestures increased in magnitude when participants could see each other. The effect can also be found in hand gestures, which are larger when an interlocutor is present [1, 18]. The present study includes a social engagement condition where either an interlocutor is present and engaging with the participant, or the participant is alone. We hypothesized that while the movements of visible articulators would increase (interpreted as greater lip protrusion and smaller lip aperture) in the presence of an interlocutor, the movements of non-visible articulators should not be so affected.

Our experimental design involves simultaneous ultrasound imaging of the tongue and video imaging of the lips, capturing their behaviour in the presence of auditory information (vocalized condition), absence of auditory information (mouthed condition) and in the presence and absence of an interlocutor. We predict that: 1) tongue movements will be unaffected by speech condition (mouthed/vocalized) and the presence/absence of an interlocutor; 2) lip movements will increase in magnitude in mouthed conditions; 3) lip movements will increase in magnitude with the presence of an interlocutor.

## 2. Methods

### 2.1. Participants

22 students at the University of British Columbia participated in the study. All were native speakers of a North American variety of English. All participants self-reported normal or corrected-to-normal vision and hearing. All participants were paid for their services at a rate of $10 per hour.

Data from male participants with beards were excluded due to the effects of hair growth on ultrasound image quality. Since these exclusions significantly reduced the number of male participants compared to female participants, all males were ultimately excluded. Ultrasound image quality was also the major factor for excluding data obtained from a number of other participants: despite our efforts to keep subjects in a stable position, some subjects still moved away from the ultrasound probe, which resulted in poor image quality. Ultimately, 12 of 22 participants had to be excluded on these grounds. The final analysis was performed on the data obtained from 10 female participants (age range 18-24; *M* = 20; *SD* = 1.70).

## 2.2. Procedure

Participants were tested individually in a sound-attenuated booth. Seated in a dentist's chair, participants positioned their heads on a headrest to minimize head movement. An Aloka SSD-5000 Doppler Ultrasound Equipment with a UST-9118 endo-vaginal 180 degree electronic curved array probe on a microphone arm was positioned under a participant's chin. The ultrasound machine was connected to an iMac computer via a firewire port which displayed and recorded the video within the iMovie program. A small table with a computer screen was placed approximately 0.5m in front of the participant. A JVC GZ-E300AU camcorder was set up approximately 1.25m in front of the participant and adjusted to capture the entire face and head area. A 5mm x 5mm sticker was positioned on the zygomatic bone immediately anterior of the left ear in order to serve as a stable starting point from which to measure lip protrusion. An 18 x 21cm mirror was positioned at a 45 degree angle to the participant's face so that a side view of her lips was visible in the viewer of the camcorder. A Blue® Yeti USB Microphone (Model 1950) was placed inside the sound booth in omnidirectional mode. This was connected to a speaker outside the sound booth so that the experimenter could hear the participant's speech and the sound cue that signalled the end of a block. Participants were seated facing the door of the sound booth. This guaranteed the participants' awareness of the experimenter's presence inside the booth.

The experiment elicited both mouthed and vocalized utterances across a 4-stage continuum of interlocutor engagement (Social Engagement). In the first stage, there was no interlocutor present (Not Present); in the second, the interlocutor (a role performed by the experimenter, who was male) was present in the sound booth but did not engage with the participant (Not Engaged); in the third, the interlocutor was present in the sound booth and asked the participant some questions regarding the comfort of the equipment (Present and Engaged); in the fourth, the interlocutor was present and responded to each utterance with a matching hand gesture (Present and Gesturing). Each of the four stages constituted a Social Engagement condition. There were two conditions for speech production (Speech Production): vocalized and mouthed. This yielded a total of 8 conditions. A pilot study with 7 participants was run to test our experimental setup and conditions. An informal evaluation of that pilot data failed to yield promising results for Not Engaged and Present and Engaged. This was confirmed based on preliminary analysis of the first two experimental participants. In the resulting design, these two intermediate points were retained as fillers, and only the two endpoints of the interlocutor enhancement continuum (Not Present and Present and Gesturing) were included in the final analysis, yielding only 4 conditions.

We focused on three Target Syllables: /wa/, /ɹa/ and /ɹi/. The consonants /w/ and /ɹ/ were chosen as they are known to vary in their degree of lip and tongue constriction depending on their position in the syllable, exhibiting the greatest degree of constriction in onset position [2, 9]. /w/ was selected to induce lip aperture constriction (rounding) and tongue-dorsal movement while /ɹa/ and /ɹi/ were selected to induce lip protrusion and tongue-blade (for /ɹa/) and tongue-dorsal (for /ɹi/) movements. The reason for two /ɹ/ initial syllables was to avoid coarticulatory effects between the consonant and following vowel. In /ɹi/ the tongue anterior gesture of /ɹ/ is largely blended with that of the following high front vowel, while in /ɹa/ a similar blending occurs with the tongue root [20]. Analysis of each syllable was therefore focused on the position of the tongue less affected by vowel coarticulation.

Prior to the beginning of the experiment, participants were instructed to "read the item aloud in your normal speaking voice" for the vocalized conditions and "mouth the items without making a sound" for the mouthed conditions. Each block was initiated by the experimenter offering the participant a sip of water. Test items were presented using Psychtoolbox (version 3.0.11) (http://psychtoolbox.org) for MATLAB with a 1 second minimum presentation of each item. The order of the tokens with each block was pseudo-randomized. Participants controlled the transition between items with the space bar on a keyboard. Each run was comprised of 24 test blocks (3 for each of the 8 conditions) with 5 utterances per token per trial. This resulted in 15 tokens per utterance per condition. After the recording portion of the experiment, subjects completed a questionnaire on the experience of participating in the study. Participants rated the friendliness of the experimenter (*M* = 6.80, *SD* = 0.42), as well as the naturalness of their speech production for both mouthing (*M* = 4.3, *SD* = 1.06) and vocalizing conditions (*M* = 5.00, *SD* = 1.05), on a 7-point Likert scale.

## 2.3. Analysis

### Analysis of the lips

Using Final Cut Pro 10.1.1 (http://www.apple.com/final-cut-pro), one frame per token was extracted from the video at the most constricted closure point of /w/ for /wa/ tokens (as determined visually using the front view of the participant) and the most protruded point of /ɹ/ for /ɹi/ and /ɹa/ tokens (as determined visually using the side-view of the participant). Analysis proceeded in ImageJ 1.48 (http://imagej.nih.gov/ij/index.html). For each frame, the red channel was filtered out and "Default" or “Percentile” threshold settings applied to produce a bi-tone black and white image.

Lip protrusion and lip aperture were measured with the straight line tool. Lip protrusion was measured by drawing a line (on the side mirror image) from the sticker on the side of the participant’s face to the most protruded point (taken to be the most rightward pixel) on a participant’s upper lip. When the most protruded point spanned more than one pixel, the most protruded pixel closest to the mouth opening was selected. Lip aperture was measured by drawing a 90 degree line in approximately the centre of the lip opening as seen in the front view image. As ImageJ measures in pixels, the measurements were then converted to centimetres. A scale was possible by comparing the width (in pixels) of the ultrasound probe tip in the image to its known physical width of two centimetres.

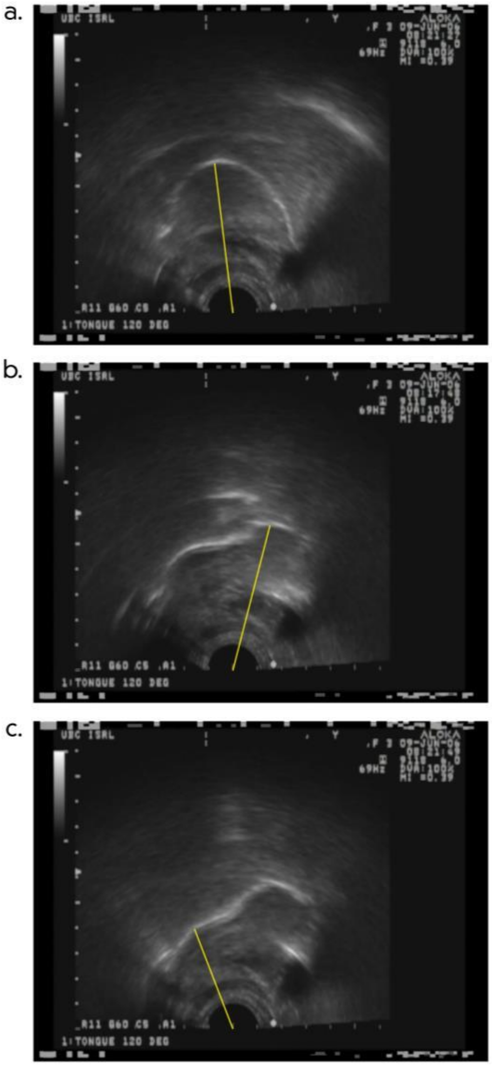

**Figure 1: Tongue measurement points:** (a) tongue dorsum for /wa/; (b) tongue tip/blade for /ɹa/; (c) tongue root for /ɹi/.

**Analysis of the tongue**

Tongue frames were extracted from the ultrasound video using Final Cut Pro. For each token, the extracted frame represented the point of most extreme constriction within the consonant prior to the transition into the vowel. For /wa/, this was the frame in which the tongue dorsum was highest relative to the middle of the transducer arc; for /ɹa/, this was the frame where the tongue blade was highest relative to the transducer arc; and for /ɹi/, where the visible portion of the tongue root was in its most posterior position relative to the same point on the transducer arc (see section 2.2). Analysis proceeded in ImageJ. Using the straight line tool, the distance from the transducer arc to the relevant point in each token was measured (see Figure 1). As with measurements for the lips, values were then scaled to centimetres.

## 3. Results

In order to investigate the validity of our hypotheses regarding the effects of Speech Production type, degree of Social Engagement and Target Syllable, three separate 2x2x3 repeated measures ANOVAs were conducted with normalized values (Student’s t-statistic for each participant) for tongue height, lip protrusion and aperture as the dependent variables respectively. The statistical analyses were primarily conducted utilizing the GLM syntax in SPSS (http://www-01.ibm.com/software/analytics/spss/), with minor further investigations employing the affix package in R (http://www.r-project.org). Maulchy’s test for Sphericity was employed and where sphericity was violated the Greenhouse-Geisser method was utilized to correct degrees of freedom. Additionally, simple main-effects analysis with a Bonferroni correction (significance at $p < 0.05$), was employed to further investigate any significant effects found in the repeated measures ANOVAs.

**Tongue Height**

Maulchy's test for sphericity regarding the 2x2x3 ANOVA for tongue height indicated a violation. A 2x2x3 repeated measures ANOVA yielded statistically significant differences between the means of Target Syllables, $F(1.12, 10.04) =55.41$, $p = 0.0001$, $\eta^2_G = 0.84$, as well as significant interaction between the Target Syllables and the Speech Production method, $F(1.43, 12.91) = 7.51$, $p = 0.01$, $\eta^2_G = 0.03$. As illustrated in Figure 2, simple main effects post-hoc tests (Bonferroni corrected) on the estimated marginal means revealed significant mean differences ($p < 0.05$) in both vocalized ($p < 0.001, < 0.001$, $M$=1.919, 2.034, $SE$ = 0.172, 0.164, 95% CIs [1.415, 2.423], [1.553, 2.516]) and mouthed ($p < 0.001, < 0.001$, $M$=1.757, 1.729, $SE$ = 0.174, 0.121, 95% CIs [1.247, 2.267], [1.375, 2.083]) syllables of /ɹa/ and /wa/ compared to /ɹi/ respectively. Additionally, post-hoc tests indicated that the mean difference between vocalized and mouthed conditions only proved statistically significant for /wa/ ($p = 0.004$, $M$=0.159, $SE$ = 0.042, 95% CI [0.064, 2.53]) as displayed in Figure 2 (error bars in the graphs correspond to the standard error of the mean in all figures). These results suggest that participants exhibited an attenuation in tongue height during mouthing versus vocalized conditions in /wa/ utterances. However, the current measurement of participant tongue height appeared to be statistically unaffected by the Social Engagement conditions, in line with the initial hypothesis. These results appear to indicate that tongue height attenuation during mouthed compared to vocalized speech is observed for certain syllables, and is unaffected for others.

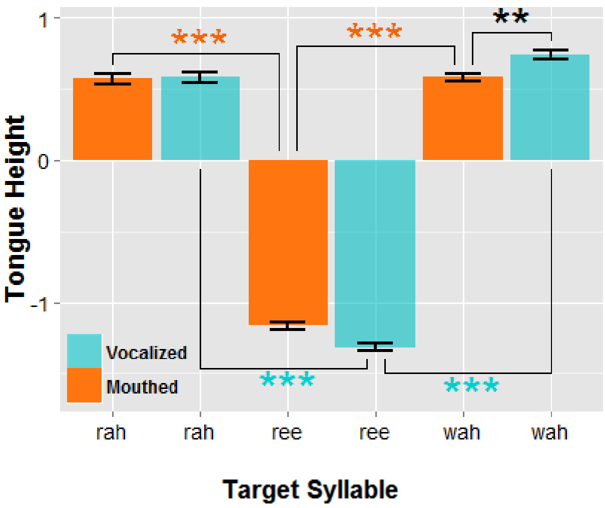

**Figure 2**: Post-hoc pairwise comparison regarding the interaction between Target Syllable and Speech Production method yielded significant mean differences in both vocalized ($p$ < 0.001, < 0.001) and mouthed ($p$ < 0.001, < 0.001) conditions. Note the significance values pertain to comparisons indicated by the brackets and are colour coded by Speech Production method. $p$ < 0.05*, 0.01**, 0.001***.

### Lip Protrusion

Similar to the results for Tongue Height, Maulchy's test for sphericity indicated that the Greenhouse-Geisser correction should be employed. As per the analysis of tongue height, a 2x2x3 repeated measures ANOVA regarding lip protrusion was conducted. Critically only the main effects of the method of Speech Production, $F(1, 9) = 10.85$, $p = 0.009$, $\eta^2_G = 0.10$, as well as Target Syllable, $F(1.27, 11.39) = 17.20$, $p = 0.0009$, $\eta^2_G = 0.19$, proved statistically significant. Bonferroni adjusted pair-wise post-hoc comparisons (see Figure 3) indicated an increase in lip protrusion for mouthed compared to vocalized utterances ($p$ = 0.009, $M$=0.308, $SE$ = 0.093, 95% CI [0.096, 0.519]), as well as for /wa/ compared against /ɹa/ and /ɹi/ ($p$ = 0.009, 0.001, $M$=0.5, 0.456, $SE$ = 0.125, 0.081, 95% CIs [0.133, 0.866], [0.219, 0.694]) respectively.

Participants appeared to exhibit more lip protrusion during mouthed compared to vocalized utterances. The differences in lip protrusion between the syllables appear to pattern in a related, but inverse manner to the tongue height data. Specifically, /wa/ exhibited an increased degree of lip protrusion comparative to /ɹa / and /ɹi/, as shown in Figure 4. However, lip protrusion measures in participants appear to be inert to the Social Engagement conditions, contra to our hypotheses.

While providing merits in isolation, measurements of lip protrusion only provide a single metric of assessing the external regions of the vocal tract, hence, the results of this data should be considered in correspondence with those of lip aperture.

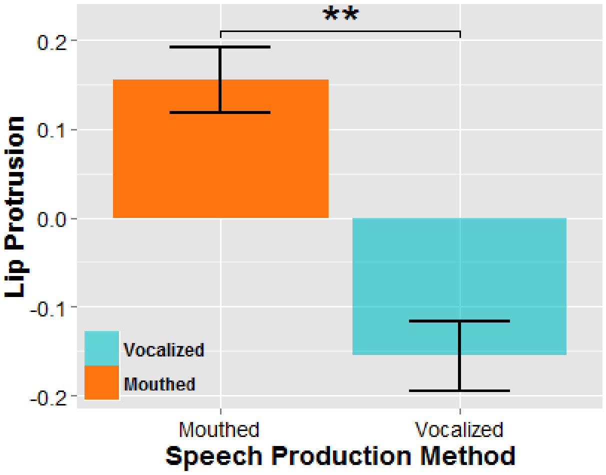

**Figure 3**: Post-hoc pairwise comparison regarding the main effect of Speech Production method. Lip protrusion increased significantly for mouthed compared to vocalized utterances ($p$ = 0.009). $p$ <0.05*, 0.01**, 0.001***

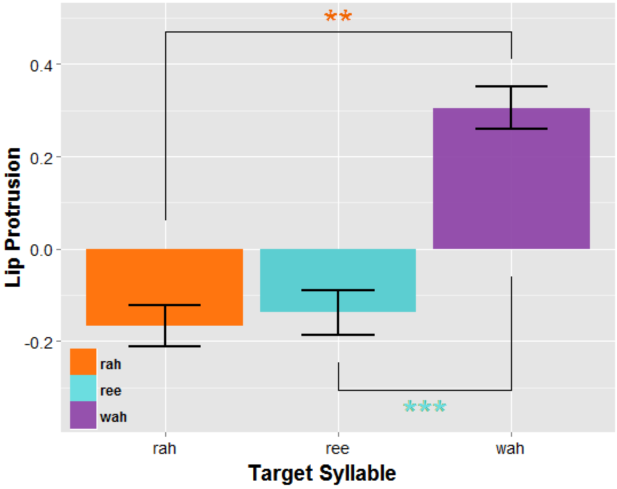

**Figure 4**: Post-hoc pairwise comparison regarding the main effect of Target Syllable. Lip protrusion was significantly different for /wa/ compared against /ɹa/ and /ɹi/ ($p$ = 0.009, 0.001) respectively. Note the significance values pertain to comparisons indicated by the brackets and are colour coded to indicate the comparative difference in means regarding /wa/. $p$ < 0.05*, 0.01**, 0.001***.

### Lip Aperture

Results from the 2x2x3 repeated measures ANOVA regarding standardized measurements of lip aperture indicated statistically significant results for the main effect of Target Syllable, $F(1.65, 14.85) = 5.02$, $p = 0.03$, $\eta^2_G$ = 0.21, as well as a significant interaction between whether participants were vocalizing or mouthing and the Social Engagement condition, $F(1, 9) = 6.62$, $p = 0.03$, $\eta^2_G = 0.01$. Bonferroni corrected post-hoc pairwise comparisons regarding the Target Syllables yielded non-significant results for all pairwise comparisons. Similar applications of the post-hoc procedure to the interaction yield a singular statistically significant mean difference between Social

Engagement conditions when participants were vocalizing. Specifically, participants exhibited smaller lip apertures during vocalization in the Not Present condition compared to when an interlocutor was Present and Gesturing ($p$ = 0.025, $M$=0.187, $SE$ = 0.069, 95% CI [0.030, 0.344]) as displayed in Figure 5. Interpretation of these results may benefit from disclosure that a visual inspection of this data indicated a greater degree of participant variability compared to the tongue height and lip protrusion metrics. For instance, the standard error regarding the difference between the means of participants mouthing in the Not Present condition and those obtained from participants when vocalizing in the Present and Gesturing condition are approximately three times greater than those of the statistically significant comparison despite visually similar disparities in magnitude (see Figure 5).

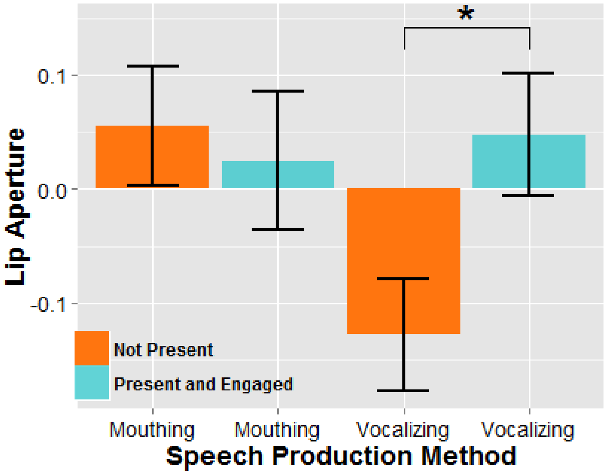

**Figure 5**: Post-hoc pairwise comparison regarding the interaction between Speech Production method and Social Engagement conditions. Participants exhibited smaller lip apertures during vocalization alone compared to when an interlocutor was present ($p$ = 0.025). $p$ < 0.05*, 0.01**, 0.001***.

## 4. Discussion

This study examined the effects of mouthing vs. vocalizing and interlocutor presence vs. absence on the movements of visible and non-visible articulators. Previous studies on speech in noise [6, 10] found increased movement in the visible articulators during speech in noisy environments. Ménard's [17] study on blind speech supports the notion that visible articulators are used by sighted speakers to convey speech information. In this context, our study was designed to shed more light on the possible differential uses of visible and non-visible articulators by sighted speakers in the absence of noise. We will discuss our findings in relation to our hypotheses provided in the introduction and will conclude with some elaborations that go beyond these hypotheses.

Firstly, we predicted that tongue movement would be unaffected by Speech Production method (mouthed/vocalized) and Social Engagement condition (presence/absence of an interlocutor). Considering /wa/, this was not the case with regard to Speech Production method. Mouthed speech showed significantly less articulatory movement as compared to vocalized speech. This finding may be explained by the fact that in the absence of an acoustic signal, it is not necessary for the tongue to hit an articulatory target. For the remaining two syllables, however, our hypothesis was confirmed: tongue height was unaffected by the changes in speech condition. Further, none of the Target Syllables were significantly affected by the Social Engagement conditions. Hence, we interpret these results as a partial validation of our initial hypothesis. The major differences in tongue height between the individual syllables /ɹa/, /wa/ and /ɹi/ can probably be ascribed to articulation differences due to the following vowel. One reason why /wa/ stands out as the only syllable showing a significant effect might be that the lips are perceptually more prominent during the articulation of /w/ versus /ɹ/. We can therefore not rule out that the tongue height findings are associated with the differences in lip movement. The finding that tongue height is statistically unaffected by the Social Engagement conditions does not come as a surprise since non-visible articulators are not expected to be affected by the presence of an interlocutor.

Secondly, we predicted that lip movements would increase in magnitude in mouthed conditions. In line with this prediction, results indicated that participants increased lip protrusion during mouthed utterances compared to vocalized utterances. The individual differences for the Target Syllables resemble the pattern that emerged for the tongue height data. Specifically, /wa/ exhibited a significantly increased degree of lip protrusion compared to /ɹa/ and /ɹi/. This implies a trade-off between tongue position and lip protrusion in /wa/. A similar trade-off has been previously observed between the tongue body and lip rounding for the vowel /u/ [19]. The measurements in lip aperture, however, did not produce any valuable insight for the distinction between mouthing and vocalizing.

Thirdly, we predicted that lip movements would increase in magnitude with the presence of an interlocutor. We therefore expected participants to produce articulations with greater protrusion and smaller aperture when an interlocutor was present. The findings for lip protrusion were not affected by Social Engagement condition. However, lip aperture showed a significant effect of Social Engagement, albeit in the direction opposite to what we predicted. During vocalized speech, participants produced smaller lip apertures when they were vocalizing alone, compared with when an interlocutor was Present and Gesturing. This was a surprising finding considering our prediction, but the relatively smaller aperture in the Not Present condition may be related to the lack of a communicative partner. Under this condition, because there is no communicative reason to make visual cues salient, participants may produce less dynamic articulations in general, maintaining a relatively more closed mouth across the entire utterance. In contrast, the presence of an interlocutor introduces a situation under which visual cues are useful and participants therefore respond more dynamically.

Though participants behaved in a way that contradicted our third prediction, the data can still be interpreted as demonstrating the sensitivity of visible articulators to the Social Engagement conditions in a way that supports a multimodal view of speech. Specifically, the observed interaction for lip aperture may only arise as the visual domain becomes relevant for communicative purposes. Lesser lip protrusion in blind participants compared to sighted [17] as well as the increase of lip protrusion under the effects of noise [5, 6, 10, 12] would appear to support these observations. However, under this interpretation it is unclear why no effect is observed in the mouthing condition when an interlocutor is present.

It is worth noting the limitations of our methodology. The measurement techniques we employed measured the maximal point of constriction of the Target Syllables. However, this measurement is static rather than dynamic, we were therefore unable to capture the amount of overall movement in each articulation. A more dynamic method of measurement which is able to capture movement could potentially be beneficial in obtaining data which more accurately depicts levels of movement/activation in speech gestures under these different speech conditions. Regarding the third hypothesis, a suggestion from an editor of this paper was that the perceived friendliness of the interlocutor could have influenced participant tendencies to display positive affect using the visible articulators (i.e. via smiling), and that this impacted lip aperture values. While we did look at naturalness and friendliness to ensure reliability and validity of our experimental conditions, our study was not designed to examine naturalness or friendliness as statistical factors. However, these would be interesting directions for future study.

Our findings suggest that speakers make use of both the auditory and visual speech signals and are aware of the information available to their interlocutor. To aid in communication, compensations are made when information from one of these signals is unavailable. Potential future research should investigate how the various visible and non-visible articulators respond dynamically under social engagement conditions.

## Conflict of Interest Statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Author and Contributors

Authorship has been separated into two tiers, each ordered alphabetically (excluding the last author). Contributions within each tier were approximately equal.

## Acknowledgements

The authors gratefully acknowledge funding provided by an NSERC Discovery Grant awarded to Bryan Gick, and by NIH Grant DC-02717 to Haskins Laboratories. We also thank Dr. Lawrence Ward and Grace Truong for consultation relating to analysis, Nicole Anger, Leighanne Chan, Lewis Haas and Taylor Poulin for their involvement in video frame extraction and measurement and Sarah Tarnowsky for assistance with editing the figures.

## References

[1] J. B. Bavelas, J. Gerwing, C. Sutton, and D. Prevost. Gesturing on the telephone: Independent effects of dialogue and visibility. *J. Mem. Lang.* 58: 495-520, 2008.

[2] F. Campbell, B. Gick, I. Wilson, and E. Vatikiotis-Bateson. Spatial and temporal properties of gestures in North American English /r/. *Lang. and Speech* 53: 49-69, 2010.

[3] G. Calvert, C. Spence, B. E. Stein, and MIT CogNet. *The handbook of multisensory processes*, 2004. Cambridge: MIT Press.

[4] E. Cvejic, J. Kim, and C. Davis. Effects of seeing the interlocutor on the production of prosodic contrasts (L). *J. Acoust. Soc. Am.* 131(2): 1011-1014, 2012.

[5] C. Davis, J. Kim, K. Grauwinkel, and H. Mixdorff. Lombard speech: Auditory (A), Visual (V) and AV effects. *Speech Prosody*, Paper 252, 2006. Dresden, Germany.

[6] M. Fitzpatrick, J. Kim, and C. Davis. The effect of seeing the interlocutor on speech production in different noise types. *Interspeech* 12: 2829-2832, 2011. Florence, Italy.

[7] M. Garnier, L. Ménard, and G. Richard. Effect of being seen on the production of visible speech cues. A pilot study on Lombard speech. *InterSpeech* 13: 611-614, 2012.

[8] B. Gick. From Quantal Biomechanics to Whole Events: Toward a Multidimensional Model for Emergent Language. *Can. Acoust.* 40.3: 24-25, 2012.

[9] B. Gick. Articulatory correlates of ambisyllabicity in English glides and liquids. *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation*, eds. J. Local, R. Ogden and R. Temple, 222-236, 2003. Cambridge: Cambridge University Press.

[10] V. Hazan, and J. Kim. Acoustic and visual adaptations in speech produced to counter adverse listening conditions. *Proceedings of Auditory-Visual Speech Processing*, 93-98, 2013. Inria: Rocquencourt, France.

[11] C. Herff, M. Janke, M. Wand, and T. Schultz. Impact of different feedback mechanisms in EMG-based speech recognition. *Interspeech* 12: 2213-2216, 2011. Florence, Italy.

[12] J. Kim, C. Davis, G. Vignali, and H. Hill. A visual concomitant of the Lombard reflex. In *Audio-Visual Speech Processing* 17-22, 2005. British Columbia, Canada.

[13] E. Lombard. Le signe de l'élévation de la voix. *Annales des maladies de l'oreille et du larynx*, 37: 101–119, 1911.

[14] D. W. Massaro, and M. M. Cohen. Perception of synthesized audible and visible speech. *Psychol. Sci.* 1: 55-63, 1990.

[15] Mayer, C., Gick, B., Weigel, T., & Whalen, D. H. Perceptual effects of visual evidence of the airstream. Proceedings of Acoustics Week in Canada. *Can. Acoust.* 41: 23-27, 2013.

[16] H. McGurk, and J. MacDonald. Hearing lips and seeing voices. *Nature* 264: 746-748, 1976.

[17] L. Ménard, A. Leclerc, A. Brisebois, J. Aubin, and A. Brasseur. Production and perception of French vowels in blind speakers and sighted speakers. International Seminar

on Speech Production (ISSP 2008), 197-200, 2008. Strasbourg, France.

[18] L. Mol, E. Krahmer, A. Maes, and M. Swerts. Seeing and being seen: the effects on gesture production. *J .Comp. Med. Comm.* 17(1):77-100, 2011.

[19] J. S. Perkell, M. L. Matthies, M. A. Svirsky and M. I. Jordan. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *J. Acoust. Soc. Am.* 93(5):2948-2961, 1993.

[20] I. Stavness, B. Gick, D. Derrick, and S. S. Fels. Biomechanical modeling of English /r/ variants. *J. Acoust. Soc. Am. Express Letters* 131(5): 355-360, 2012.

[21] W. H. Sumby, and I. Pollack. Visual contribution to speech intelligibility in noise. *J. Acoust. Soc. Am.* 26(2): 212-215, 1954.

[22] A. Q. Summerfield. Use of visual information for phonetic processing. *Phonetica* 36: 314–331, 1979.

[23] W. V. Summers, D. B. Pisoni, R. H. Bernacki, R. I. Pedlow, and M. A. Stokes. Effects of noise on speech production: acoustic and perceptual analyses. *J. Acoust. Soc. Am.* 84(3): 917-28, 1988.

[24] L. A. Ross, D. Saint-Amour, V. M. Leavitt, D. C. Javitt, and J. J. Foxe. Do You See What I Am Saying? Exploring Visual Enhancement of Speech Comprehension in Noisy Environments. *Cereb. Cortex* 17: 1147-1153, 2007.