

Compensation for visually specified coarticulation in liquid–stop contexts

Navin Viswanathan^{1,2} · Joseph D. W. Stephens³

Published online: 11 August 2016
© The Psychonomic Society, Inc. 2016

Abstract The question of whether speech perceivers use visual coarticulatory information in speech perception remains unanswered, despite numerous past studies. Across different coarticulatory contexts, studies have both detected (e.g., Mitterer in *Perception & Psychophysics*, 68, 1227–1240, 2006) and failed to detect (e.g., Vroomen & de Gelder in *Language and Cognitive Processes*, 16, 661–672. doi:10.1080/01690960143000092, 2001) visual effects. In this study, we focused on a liquid–stop coarticulatory context and attempted to resolve the contradictory findings of Fowler, Brown, and Mann (*Journal of Experimental Psychology: Human Perception and Performance*, 26, 877–888. doi:10.1037/0096-1523.26.3.877, 2000) and Holt, Stephens, and Lotto (*Perception & Psychophysics*, 67, 1102–1112. doi:10.3758/BF03193635, 2005). We used the original stimuli of Fowler et al. with modifications to the experimental paradigm to examine whether visual compensation can occur when acoustic coarticulatory information is absent (rather than merely ambiguous). We found that perceivers’ categorizations of the target changed when coarticulatory information was presented visually using a silent precursor, suggesting that visually presented coarticulatory information can induce compensation. However, we failed to detect this effect when

the same visual information was accompanied by an ambiguous auditory precursor, suggesting that these effects are weaker and less robust than auditory compensation. We discussed why this might be the case and examined implications for accounts of coarticulatory compensation.

Keywords Coarticulation · Speech perception · Audiovisual speech

After decades of research, the question of how listeners demonstrate stable perception despite variability in the acoustic signal remains unanswered. One source of variability is coarticulation: the temporal overlap of the gestures of segments that leads to context dependence in the production of a given segment. How listeners compensate for coarticulatory variability and what properties of neighboring segments they use are important questions for accounts of speech perception. A well-studied case of compensation for coarticulation was first reported by Mann (1980). She demonstrated that listeners classifying a [da]–[ga] continuum reported more “g” responses after the liquid [al] than after [aj], indicating that listeners’ perception of the target stop changed with its coarticulatory context.

The question of what properties of the preceding segment induce this shift has been a subject of considerable debate (Fowler 2006; Lotto & Holt, 2006). Two competing explanations, reflecting differing theoretical perspectives, have been offered. From an ecological gestural perspective, listeners perceive speech by perceiving the gestures of speech production: the causal source of the proximal acoustic signal (Fowler, 1986). In the present example, the coarticulatory variability in the acoustic signal arises from the overlap of the constriction gestures for the stops with those for the liquids. The compensatory perceptual shifts result from listeners perceiving the same intermediate continuum members differently, depending

✉ Navin Viswanathan
navin@ku.edu

✉ Joseph D. W. Stephens
jdstephe@ncat.edu

¹ Department of Speech-Language-Hearing, University of Kansas, Lawrence, KS, USA

² Haskins Laboratories, New Haven, CT, USA

³ Department of Psychology, North Carolina A & T State University, Greensboro, NC, USA

on the gestural properties of the context. That is, after [aɪ], listeners hear the intermediate acoustics as being caused by a constriction gesture for [ga] that has been pulled forward, and after [aʊ], as a constriction for [da] that has been pulled back.

A substantially different explanation has been proposed from a general auditory perspective (Diehl, Lotto, & Holt, 2004). Per this account, speech perception is based on the recognition of acoustic patterns, and these patterns may be altered by interactions of the acoustic signals within the auditory system. In the present example, the segments [aɪ] and [da] both have relatively high third formant (F3) frequencies, as compared to [aʊ] and [ga], respectively. When these liquid–stop disyllables are processed by the auditory system, they induce spectral contrast such that an intermediate F3 is interpreted to be relatively lower (i.e., more “ga-like”) when it follows the high F3 in [aɪ] than when it follows the low F3 in [aʊ]. Importantly, from this perspective, listeners’ compensatory shifts arise solely from the acoustic relationships of the precursor and the target. The demonstration that nonspeech tone analogues, with frequencies matched to F3 offsets of precursors, induce these boundary shifts despite lacking gestural information is cited as support for this perspective (Lotto & Kluender, 1998). For a detailed account of this debate, we refer the reader to the relevant studies (Holt, Stephens, & Lotto, 2005; Lotto & Holt, 2006; Stephens & Holt, 2003; Viswanathan, Fowler, & Magnuson, 2009; Viswanathan, Magnuson, & Fowler, 2010, 2013, 2014).

In this article, we focus on a question of interest to both perspectives: Do listeners compensate for visually provided information regarding coarticulation in a liquid–stop context? The gestural account predicts that coarticulatory information, irrespective of its modality, would induce perceptual compensation. A general auditory account requires a mechanism other than spectral contrast to explain the effects of visual coarticulatory information. Different studies have yielded inconsistent findings. Fowler, Brown, and Mann (2000, Exp. 3b) presented participants with hyperarticulated video segments of [alda] or [aʊda]. For each trial, these were both dubbed with an acoustic precursor syllable that was judged to be ambiguous between [aɪ] and [aʊ], and one member of the target [da]–[ga] continuum. Fowler et al. found that participants reliably used visual information about the precursor and demonstrated similar shifts (to those induced by the acoustic precursors [aɪ] and [aʊ]) in the identification of the target continuum. Holt, Stephens, and Lotto (2005) were able to replicate Fowler et al.’s finding with the original stimuli (Exp. 2), but they failed to detect visually induced perceptual boundary shifts with their own stimuli (Exp. 1). On the basis of two additional experiments that manipulated Fowler et al.’s stimuli, they argued that the Fowler et al. results were caused by differences in the visual characteristics of [da] between the [alda] and [aʊda] videos. That is, because differing visual features for each precursor accompanied the target syllables, it was impossible to assume that these features were perceptually attributed to the precursor syllable rather than to the target syllable. If visual features concurrent

with the target syllable were perceptually attributed to the target syllable, then the effect of visual information would be analogous to a McGurk effect (McGurk & MacDonald, 1976). In response, Fowler (2006) acknowledged that the videos of the target syllables were different, but she argued that the visual difference (specifically, lip rounding after [aʊ] but not after [aɪ]) could not possibly be interpreted as providing information about the target syllable, since neither [d] nor [g] is characterized by lip rounding. Instead, perceivers were attuned to the persistent effects of the lip rounding of the preceding [ɪ], even in the case that all information prior to the onset of the target syllable was deleted (as in Holt et al., 2005, Exp. 4). As such, the visual influence on perception of the target syllable could still represent compensation for coarticulation, because it was based on information specifying the preceding gesture. Lotto and Holt (2006) remained unconvinced that listeners compensate for subtle visual differences in this case (Exp. 4), but not in their original replication attempt, in which the precursor was actually present (Exp. 1).

These studies investigated whether visual information from the preceding liquid can influence the perception of the target stop, but they failed to provide a conclusive answer. Turning to other coarticulatory contexts does not clarify the picture. For instance, Vroomen and de Gelder (2001) examined this question in a fricative ([s]–[ʃ])–stop ([t]–[k]) context. They found that even though listeners demonstrated a reliable influence of visual fricative information on fricative categorization itself (a McGurk-like effect), they failed to detect shifts in the following stop. In contrast, Mitterer (2006, Exp. 3) examined native Dutch listeners’ categorization of fricative–vowel sequences. He found that listeners’ categorizations of a [s]–[ʃ] fricative continuum were altered by visual information about lip rounding of the following [i] or [y] vowel. Finally, Kang et al. (2016) reported that both English and French listeners demonstrated visual compensation for coarticulation for a [s]–[ʃ] fricative continuum when the fricatives were followed by vowels native to both groups ([i] and [u]). However, only the French listeners compensated for the coarticulatory influence of the French vowel [y], suggesting a language-specific locus of visual effects.

Such equivocal findings warrant further investigation. One possibility that Mitterer (2006) himself raised is that visual effects depend on the nature of the coarticulatory context being examined. This would entail multiple accounts of compensation (for different coarticulatory contexts) that would have to be mutually consistent within a broad theoretical framework. Alternatively, it is possible that a more parsimonious explanation could be arrived at with further investigation. In the present study, one researcher from each theoretical perspective collaborated to investigate these effects. We focused on the liquid–stop context described previously and examined whether visual coarticulatory effects could be detected even in the absence of the target visual information that researchers had previously argued was confounded with

precursor visual information. More generally, we also examined why visual coarticulatory effects are less robust than their auditory counterparts. We used the original stimuli of Fowler et al. (2000) but a design similar to that of Holt et al. (2005), with no concurrent visual information during target presentation. A crucial difference from both prior studies is that we presented the visual precursors with no accompanying acoustic signal (Exp. 1a), as well as with an ambiguous auditory precursor (Exp. 1b). An additional important difference is that we prevented access to the target gestural information by occluding the speakers’ oral gestures while retaining event continuity (see Fig. 1).

Method

Participants

Forty-three undergraduate students (19 in Exp. 1a, 24 in Exp. 1b) at North Carolina A&T State University, who were native English speakers and reported normal hearing, participated in the study. Because there was greater attrition in Experiment 1b (see the Discussion section), we ensured that both conditions had similar numbers of usable data points for comparison.

Materials

Auditory stimuli Both experiments used a ten-member target continuum ranging from [ga] to [da] with varying F3 onset

frequencies. Additionally, in Experiment 1b each target syllable was preceded by an intermediate precursor between [al] and [aɪ]. The auditory stimuli consisted of resynthesized speech created using the Praat software (Boersma & Weenink, 2006). Glottal sources for the [ga]–[da] series were generated by editing natural tokens of [da] from a male speaker (J.D.W.S.) to a length of 250 ms and inverse filtering by LPC coefficients. The intermediate precursor was similarly generated by editing natural tokens of [a] from the same speaker to a length of 450 ms and inverse filtering. These sources were then refiltered using formant parameters that had been used for the stimuli in two prior studies (Fowler et al., 2000; Holt et al., 2005): 750, 1200, 2450, 2850, and 3750 Hz. For [ga], the first three formants transitioned to the vowel parameters over the first 80 ms, from their initial frequencies of 300, 1650, and 1800 Hz (F4 and F5 remained constant). The other nine members of the consonant–vowel (CV) series were created by raising the initial F3 frequency in 100-Hz steps, such that the [da] endpoint had an initial F3 frequency of 2700 Hz. For the precursor syllable, the first three formants transitioned from the vowel parameters over the final 250 ms to their final frequencies of 556, 1300, and 2150 Hz (F4 and F5 remained constant). All tokens were matched in root-mean-squared intensity and resampled at 44.1 kHz. For Experiment 1b, the precursor was concatenated with each target syllable, with 50 ms of silence between syllables. Silence was also added to the beginning of each audio file so that the onset of the CV

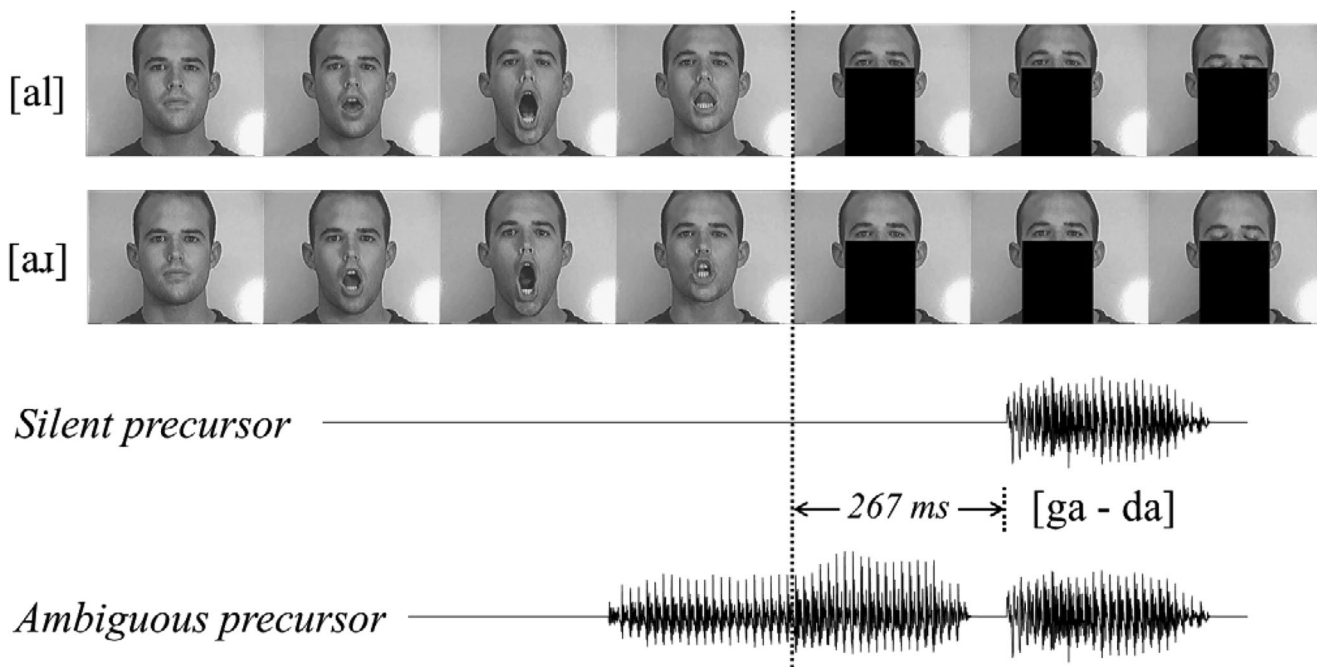


Fig. 1 Construction of the auditory and visual stimuli. By occluding only the lower half of the speaker’s face, the articulatory information for the consonant–vowel syllable was removed, while retaining some visual continuity throughout the utterance

test syllable would be properly synchronized to the appropriate video frame when the audio and video files were played simultaneously.

Visual stimuli The visual stimuli in both experiments consisted of videos originally created by Fowler et al. (2000), which feature a speaker hyperarticulating the syllables [alda] and [aɪda]. Holt et al. (2005) removed the visual precursor information in the Fowler et al. videos by presenting a blank screen starting 200 ms before the onset of the target syllable. For the present study, we aimed to retain more visual continuity between the precursors and targets, while still removing all visual articulatory information from the target portion of the videos. We did so by placing a black box over the lower half of the speaker's face, beginning 267 ms prior to the onset of the CV test syllable. This timing was intended to provide as much temporal separation as possible between the visual precursor cues and onset of the auditory cues for the stop consonant, given that McGurk effects have been observed with visual stimuli leading auditory stimuli by as much as 180 ms (e.g., Munhall, Gribble, Sacco, & Ward, 1996). Despite this early onset of the black box, enough of the visual precursor articulation was visible to easily distinguish the two videos; Fig. 1 illustrates the visual stimuli and their temporal alignment with the auditory stimuli.

Individual video frames were edited and then reassembled into .WMV files at the original rate of 30 frames per second. Each of the auditory stimuli was added as a soundtrack to the videos, resulting in 20 video files for the target trials in each experiment (with the Exp. 1a videos containing silence during the precursor and the Exp. 1b videos containing the ambiguous precursor). An additional catch-trial version of each video was also created, in which a small, red circle with a line through it was superimposed over the mouth for five frames prior to the appearance of the black box.

Procedure

Experiment 1a Participants completed the experiment while seated at a computer in a quiet room. The auditory stimuli

were presented through a Behringer HA400 amplifier via Sony MDR-7506 headphones. The visual stimuli were presented on a 19-in. LCD monitor with a resolution of 640 × 480 pixels, viewed from a distance of approximately 26 in. The videos themselves had a resolution of 320 × 240 pixels.

Participants were instructed to watch the screen, listen to the sounds, and indicate whether each test syllable was “DA” or “GA” by using keys on the computer keyboard labeled “DA” and “GA.” Additionally, the participants were instructed to withhold their response on catch trials in which a red “no” symbol appeared on the face. A practice block was presented containing four standard trials and four catch trials, in random order, and feedback was provided to indicate whether the participant had been correct in either making or withholding a response. Following this, the main experiment consisted of ten blocks. Each block consisted of 20 standard trials (each precursor followed by one member of the continuum) and two catch trials, presented in random order. No feedback was provided. Finally, participants categorized the silent visual precursors [al] and [aɪ], which were each presented ten times in random order.

Experiment 1b Experiment 1b was identical to Experiment 1a, except that the stimuli presented during the audiovisual task included the intermediate auditory precursor. Participants were instructed that they would hear two syllables, and that they should identify the second syllable as either “DA” or “GA.” Finally, as in Experiment 1a, participants categorized the intermediate precursors as [al] or [aɪ].

Results

The data from 17 of the participants in Experiment 1a and 18 in Experiment 1b, who exhibited greater than 80 % accuracy on the catch trials, were included in the analyses. Figure 2 depicts the performance of both groups of listeners in the experimental block. Visual inspection of the results indicated that the listeners from Experiment 1a (silent precursor), but not those from Experiment 1b (intermediate acoustic precursor), appeared to

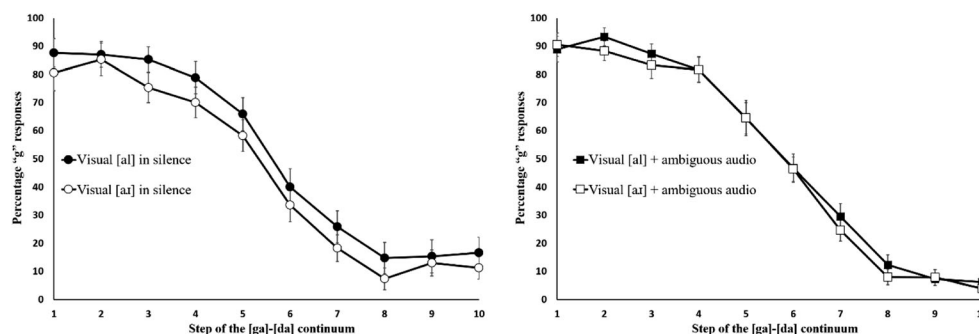


Fig. 2 The figure demonstrates that the categorization of the target continuum shifted after silent visual precursor (*left*) but not after visual precursor with ambiguous audio (*right*)

shift their perceptual categorization on the basis of the identity of the visual precursor. The proportions of “g” responses from each experiment were logit-transformed before being submitted to two separate 2 (visual precursor) × 10 (step) repeated measures analyses of variance (ANOVAs). Proportions of 0 and 1 were replaced by .01 and .99, respectively, to avoid singularities in the transformed data (see Viswanathan et al., 2014).

For Experiment 1a, the effect of the silent visual precursor was significant [$F(1, 14) = 12.35, p = .003, \eta_p^2 = .47$], indicating that participants reported more “ga”s after the silent [aI] than after silent [aI]. The expected effect of step was significant [$F(9, 126) = 98.39, p < .0001, \eta_p^2 = .84$], whereas the Visual Precursor × Step interaction did not approach significance [$F(9, 126) = 1.25, p = .27, \eta_p^2 = .08$]. For Experiment 1b, neither the effect of the visual precursor (accompanied by an identical intermediate auditory precursor) [$F(1, 17) = 1.38, p = .26, \eta_p^2 = .08$], nor the Visual Precursor × Step interaction [$F(9, 153) = 1.04, p = .41, \eta_p^2 = .06$], approached significance. Only the expected effect of step was significant [$F(9, 153) = 104.06, p < .0001, \eta_p^2 = .86$]. Analysis of identification responses of the precursors presented in isolation indicated that participants were able to use visual information similarly to categorize the precursor in both conditions (the average accuracies in the silent condition [81.67 %] and the ambiguous auditory-precursor condition [84.23 %] were not reliably different [$p = .66$]). As in Viswanathan et al. (2009), the data from both conditions were submitted to an omnibus mixed ANOVA with Condition as a between-subjects factor and Precursor and Step as within-subjects factors. The boundary shifts across precursor conditions, indicated by the Precursor × Condition interaction, were reliably different ($p = .023$) in the middle of the continuum (Steps 3–8), but not across the whole

continuum (Steps 1–10; $p = .19$). Given that typical compensation effects are strongest in the ambiguous region of the continuum (e.g., Viswanathan et al., 2009, Fig. 3) and that the interaction term was between a relatively weak within-subjects factor (Precursor) and a between-subjects factor (Condition), we suggest that the present design lacked the power to detect the weak interaction. Nevertheless, the difference in the effects across conditions must be interpreted cautiously.

Discussion

The question of whether listeners’ perceptions of speech are altered by coarticulatory visual information has remained unresolved. From either of the main theoretical perspectives on speech perception, there is a need to reach some empirical consensus on the extent to which such effects occur. In this study, we focused on the well-studied liquid–stop context and asked two specific questions: First, we closely examined Holt et al. (2005)’s suggestion that the visual coarticulatory effects found by Fowler et al. (2000) were due solely to McGurk-like effects from concurrently presented visual information. In Experiment 1a, we used the original stimuli of Fowler et al. (2000), but made two crucial changes. First, we occluded the speaker’s mouth during the production of the target, while leaving part of the head visible to preserve event continuity. Second, we presented no auditory information during the precursor, ensuring that there was no discord between the visual and auditory information during the precursor. The results of Experiment 1a demonstrated a reliable effect of the silent precursor on target categorization. This suggests that the visual information specifying a preceding liquid

Table 1 Comparison of experiments investigating audiovisual compensation effects in the liquid–stop context

Article	Exp.	Auditory Precursor Before [ga]–[da] Series	Visual Stimuli	N	% “g” Responses Following [aI]–Following [aI]
Fowler et al. (2000)	3a	[aI]–[aI] precursor	None (Auditory Control condition)	13	12.8***
Fowler et al. (2000)	3b	Intermediate [aI]–[aI] precursor	Hyperarticulated [alda] and [aıda] videos throughout precursor and test syllables	13	11.1***
Holt et al. (2005)	1	Intermediate [aI]–[aI] precursor	Normally articulated precursor [aI] and [aI], blank screen during test syllable	9	–3.86
Holt et al. (2005)	2	Intermediate [aI]–[aI] precursor	Hyperarticulated [alda] and [aıda] videos throughout precursor and test syllables (Fowler et al. materials)	11	8.27*
Holt et al. (2005)	3	Intermediate [aI]–[aI] precursor	Hyperarticulated [aI] and [aI] during precursor (from Fowler), blank screen during test syllable	12	2.67
Holt et al. (2005)	4	None	No precursor; da] portions of Fowler [alda] and [aıda] videos during test syllable	12	6.83**
Present study	1a	None	Hyperarticulated [alda] and [aıda] videos with a black box occluding speaker’s articulation during the test syllable	17	6.45**
Present study	1b	Intermediate [aI]–[aI] precursor	Hyperarticulated [alda] and [aıda] videos with a black box occluding speaker’s articulation during the test syllable	18	1.84

* $p < .05$, ** $p < .01$, *** $p < .001$

can indeed produce compensation for coarticulation in the identification of a subsequent stop consonant. Second, we examined why visual coarticulatory effects are relatively hard to detect. In Experiment 1b, we retained the oral occlusion from Experiment 1a but presented an intermediate auditory precursor accompanying the visual precursor. In this case, we failed to detect an effect of the precursor on target categorization (but see the omnibus analyses). This result suggests that visual effects, in this coarticulatory context, are indeed less robust than their auditory counterparts, as was argued by Lotto and Holt (2006). In Experiment 1b, the effect of the intermediate auditory precursor appeared to neutralize or weaken the effect of the visual precursor.

From a general auditory perspective, we interpreted these results as evidence that coarticulatory effects are driven primarily by bottom-up auditory cues. If spectral contrast were the primary mechanism for these effects, then no difference would be expected in the ambiguous-precursor condition, because the acoustic signal was the same. The effect of the silent precursor in Experiment 1a argues for the existence of mechanisms beyond spectral contrast, because in this case spectral relations were entirely absent (Lotto & Holt, 2006). For example, interactive processes could allow for visual cues to indirectly activate auditory representations via their mutual associations with phonetic categories (similar to lexical compensation effects; e.g., McClelland, Mirman, & Holt, 2006).

The results can also be argued to be consistent with the gestural perspective. First, the expected effect due to the coarticulatory influence of the precursor was obtained in the silent-precursor condition, providing clear support. The lack of effect in the ambiguous condition could be explained by the fact that this manipulation introduced dissonance between the auditory and visual precursor information that potentially weakened the effect of the precursor. Of course, Fowler et al. (2000) did detect visual effects under these conditions. They may have obtained a stronger effect because their video of coarticulation across the precursor and target syllables was completely unmodified and continuous. As we discussed earlier, we had to modify the design to rule out a target-bias explanation, and may have weakened the effect as a result. A final possibility, pertinent to both perspectives, is that the presence of an auditory precursor encouraged listeners to ignore the visual stimuli. The greater likelihood of participant attrition in our Experiment 1b, due to failure to withhold response in visual catch trials, supports this suggestion.

Beyond these specific stimuli, the present results help put in perspective prior studies of this issue (see Table 1 for a comparison of all studies of audiovisual compensation in this liquid–stop context) and explain why visual effects, especially in the liquid–stop context, have been difficult to detect. First, the constriction locations for [al] and [aʊ] are more difficult to visually discriminate than, for instance, the lip rounding gestures of the vowels used in Mitterer (2006). In fact, when Holt

et al. (2005) used stimuli that were not hyperarticulated, effects were not detected. Furthermore, even with hyperarticulation, our participants only demonstrated slightly higher than 80 % accuracy in a two-alternative forced choice task of categorizing the precursors on the basis of visual information. Consequently, we suggest that in typical communicative conditions, listeners are less likely to rely on visual gestural information for these liquids than in other contexts (e.g., lip rounding) in which the gestural information is more salient. Second, it is likely that visual effects are weaker than their auditory counterparts (e.g., around a 6.5 % shift in responses in Exp. 1a vs. around 12 % in the auditory control condition of Fowler et al., 2000). If so, given our inferential techniques, we should expect the weaker visual effects not to be replicated consistently. In fact, as was argued by Francis (2012), inconsistent replication of a weak effect should be seen as evidence for, rather than against, the presence of the effect.¹ Finally, our experimental paradigm introduced conditions that may minimize the use of visual information by listeners. As we highlighted, visual discontinuity between the target and the precursor, the dissonance between auditory and visual precursor information, and the lack of reliable precursor discriminability may all contribute to the decreased reliance on visual information in these contexts.

In summary, the present study provides new evidence both that visual coarticulatory effects are real and that they are less robust than their auditory counterparts. The relative weakness of these effects can be explained in different ways by the two theoretical perspectives considered here. Therefore, this phenomenon by itself does not adjudicate between the two accounts, unless a more precise difference in the predictions of the two perspectives can be established. This study provides a starting point for further research that can aim to dissociate these accounts on the basis of this phenomenon. Additionally, the present findings contribute to our empirical knowledge of how listeners utilize visual coarticulatory information, which is critical to an understanding of the architecture and the mechanisms that serve the perception of speech.

Author note This research was supported by NIDCD Grant No. R15 DC011875-01 to N.V. and J.D.W.S. and NSF Grant No. BCS-1431105 to N.V. J.D.W.S. has also been supported by the Air Force Research Laboratory and OSD under Agreement No. FA8750-15-2-0116. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes, notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the National Institutes of Health, Air Force Research Laboratory and OSD, or the U.S. Government.

¹ Given the theoretical interest in both the presence and the absence of a visual effect, this phenomenon may have escaped the *file drawer* issue, which refers to the decreased likelihood of null effects being published, resulting in overestimation of the occurrence of reported effects (Rosenthal, 1979).

References

- Boersma, P., & Weenink, D. (2006). Praat: Doing phonetics by computer (Version 4.4.16) [Computer program]. Retrieved April 1, 2006, from www.praat.org
- Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179. doi:10.1146/annurev.psych.55.090902.142028
- Fowler, C. A. (1986). An event approach to a theory of speech perception from a direct-realist perspective. *Journal of Phonetics*, *14*, 3–28.
- Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics*, *68*, 161–177. doi:10.3758/BF03193666
- Fowler, C. A., Brown, J. M., & Mann, V. A. (2000). Contrast effects do not underlie effects of preceding liquids on stop-consonant identification by humans. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 877–888. doi:10.1037/0096-1523.26.3.877
- Francis, G. (2012). The psychology of replication and replication in psychology. *Perspectives on Psychological Science*, *7*, 585–594. doi:10.1177/1745691612459520
- Holt, L. L., Stephens, J. D. W., & Lotto, A. J. (2005). A critical evaluation of visually moderated phonetic context effects. *Perception & Psychophysics*, *67*, 1102–1112. doi:10.3758/BF03193635
- Kang, S., Johnson, K., & Finley, G. (2016). Effects of native language on compensation for coarticulation. *Speech Communication*, *77*, 84–100.
- Lotto, A. J., & Holt, L. L. (2006). Putting phonetic context effects into context: A commentary on Fowler (2006). *Perception & Psychophysics*, *68*, 178–183. doi:10.3758/BF03193667
- Lotto, A. J., & Kluender, K. R. (1998). General contrast effects of speech perception: Effect of preceding liquid on stop consonant identification. *Perception & Psychophysics*, *60*, 602–619. doi:10.3758/BF03206049
- Mann, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, *28*, 407–412.
- McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*, 363–369. doi:10.1016/j.tics.2006.06.007
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748. doi:10.1038/264746a0
- Mitterer, H. (2006). On the causes of compensation for coarticulation: Evidence for phonological mediation. *Perception & Psychophysics*, *68*, 1227–1240.
- Munhall, K. G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, *58*, 351–362. doi:10.3758/BF03206811
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*, 638–641. doi:10.1037/0033-2909.86.3.638
- Stephens, J. D. W., & Holt, L. L. (2003). Preceding phonetic context affects perception of nonspeech. *Journal of the Acoustical Society of America*, *114*, 3036–3039.
- Viswanathan, N., Fowler, C. A., & Magnuson, J. S. (2009). A critical examination of the spectral contrast account of compensation for coarticulation. *Psychonomic Bulletin & Review*, *16*, 74–79. doi:10.3758/PBR.16.1.74
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2010). Compensation for coarticulation: Disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *Journal of Experimental Psychology: Human Perception and Performance*, *36*, 1005–1015.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2013). Similar response patterns do not imply identical origins: An energetic masking account of nonspeech effects in compensation for coarticulation. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 1181–1192.
- Viswanathan, N., Magnuson, J. S., & Fowler, C. A. (2014). Information for coarticulation: Static signal properties or formant dynamics? *Journal of Experimental Psychology: Human Perception and Performance*, *40*, 1228–1236.
- Vroomen, J., & de Gelder, B. (2001). Lipreading and the compensation for coarticulation mechanism. *Language and Cognitive Processes*, *16*, 661–672. doi:10.1080/01690960143000092