

Supplementary Information
for Cummings, Karboga, Yang, & Jaeger
Compensation in audiovisual speech perception

T. Florian Jaeger and Shawn Cummings

January 8, 2025

Contents

1	Open science statement	1
2	Required software	1
2.1	Interested in using R markdown do create APA formatted documents that integrate your code with your writing?	2
3	Chronology of Experiments	2
4	Participant demographics	2
5	Experiments 1a-c	2
5.1	Materials: Selection of acoustic continuum steps	2
5.2	Exclusions	3
5.3	Exit survey	3
5.3.1	Experiment 1a	3
5.3.2	Experiment 1b	4
5.3.3	Experiment 1C	6
5.3.4	Survey results	7
5.4	Analyses	10
5.4.1	Experiment 1a	10
5.4.2	Experiment 1b	12
5.4.3	Experiment 1c	13
5.4.4	Liu and Jaeger (2018, Experiment 1a-b)	14
5.5	Subset analyses removing participants who commented on effects of the pen	15
5.5.1	Experiment 1a	15
5.5.2	Experiment 1b	16
5.5.3	Experiment 1c	17
6	Experiments 2 and 2b	18
6.1	Procedure	18
6.2	Exclusions	18
6.3	Analyses - Experiment 2	19
6.4	Comparison of Experiment 2 against Experiment 1c	21
6.5	Analysis - Experiment 2b	22
6.5.1	SPACE bar presses	22
6.5.2	Categorization responses	23
	References	25

1 Open science statement

This document contains the data preparation, visualization, and analyses reported in the main text. As of 2022, the Human Language Processing Lab at the University of Rochester is committed to producing articles and supplementary information in the form of R Markdown documents whenever possible. This particular project grew out of a yearlong undergraduate research class (BCS 206/7 “Undergraduate Research in Cognitive Science”). To accommodate differences in familiarity with R programming, and to facilitate project workflow, the authors agreed to write the main text in a common word processing software, and to provide this R Markdown document as supplementary information.

The R Markdown document that this PDF is generated from is available on the OSF repo for the paper (<https://osf.io/2asgw/>), along all experimental materials. This includes the original video and audio recordings as well as all audiovisual test stimuli for all experiments along with their phonetic annotations—lists, and trial-level data. The same holds for the JavaScript code for the experiments.

2 Required software

This document was compiled using `knitr` in RStudio with R:

```
platform      -  
platform      aarch64-apple-darwin20  
arch          aarch64  
os            darwin20  
system        aarch64, darwin20  
status  
major         4  
minor         4.2  
year          2024  
month         10  
day           31  
svn rev       87279  
language      R  
version.string R version 4.4.2 (2024-10-31)  
nickname      Pile of Leaves
```

We used the following R packages to create this document: R (Version 4.4.2; R Core Team 2024) and the R-packages *brms* (Version 2.22.0; Bürkner 2017, 2018, 2021), *cmdstanr* (Version 0.8.0; Gabry et al. 2024), *dplyr* (Version 1.1.4; Wickham et al. 2023), *forcats* (Version 1.0.0; Wickham 2023a), *ggplot2* (Version 3.5.1; Wickham 2016), *ggtext* (Version 0.1.2; Wilke and Wiernik 2022), *kableExtra* (Version 1.4.0; Zhu 2024), *knitr* (Version 1.49; Xie 2024), *lazerhawk* (Version 0.3.0; Clark 2024), *linguisticsdown* (Version 1.2.0; Liao 2019), *lubridate* (Version 1.9.4; Grolemund and Wickham 2011), *magrittr* (Version 2.0.3; Bache and Wickham 2022), *papaja* (Version 0.1.3; Aust and Barth 2024), *purrr* (Version 1.0.2; Wickham and Henry 2023), *Rcpp* (Version 1.0.13.1; Eddelbuettel et al. 2024), *readr* (Version 2.1.5; Wickham, Hester, and Bryan 2024), *rstan* (Version 2.32.6; Stan Development Team 2024), *StanHeaders* (Version 2.32.10; Stan Development Team 2020), *stringr* (Version 1.5.1; Wickham 2023b), *tibble* (Version 3.2.1; Müller and Wickham 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, and Girlich 2024), *tidyverse* (Version 2.0.0; Wickham et al. 2019) and *tinylabels* (Version 0.2.4; Barth 2023). If opened in RStudio, the top of the R markdown document should alert you to any libraries you will need to download, if you have not already installed them. The full session information is provided at the end of this document.

2.1 Interested in using R markdown do create APA formatted documents that integrate your code with your writing?

A project template, including R markdown files that result in APA-formatted PDFs, is available at <https://github.com/hlplab/template-R-project>. Feedback welcome. We aim to help others avoid the mistakes and detours we made when first deciding to embrace literal coding to increase transparency in our projects.

3 Chronology of Experiments

- **Experiment 1a:** test-only experiment ($N = 64$) conducted over MTurk between 01/20-22/2021.
- **Experiment 1b:** test-only experiment ($N = 64$) conducted over MTurk between 03/03-04/2021.
- **Experiment 1c:** test-only experiment ($N = 64$) conducted over MTurk between 06/03-05/2021.
- **Experiment 2:** test-only experiment ($N = 64$) conducted over MTurk between 07/14-15/2021.
- **Experiment 2b:** test-only experiment ($N = 64$) conducted over MTurk between 07/23-24/2021.

4 Participant demographics

Female	Male	unknown			
47.0	52.0	0.9			
	Hispanic	Non-Hispanic	unknown		
	8.5	89.7	1.9		
Asian	Black	multiple	other	unknown	White
7.8	12.5	3.1	0.6	3.1	72.7

Mean age = 36.9 (SD = 12.1). Interquartile range = 28-44. Declined to report: 1.6%

5 Experiments 1a-c

5.1 Materials: Selection of acoustic continuum steps

As described in the main text, we aimed to maximize the power to detect effects—like that of the pen in the mouth—along the *asi-ashi* continuum. Specifically, we aimed to select one step that, across all other manipulations, would yield approximated 25% *ashi* responses, four steps that would yield close to 50% *ashi* responses, and one step that would yield 75% *ashi* responses.

The intention for Experiment 1a was to employ the exact same acoustic steps as in Liu and Jaeger (2018): 12, 14, 15, 16, 17, and 19, where higher numbers indicate acoustically more *ashi*-like steps. However, a labeling mistake resulted in the steps being internally named with higher numbers indicating proportion of *asi*, rather than *ashi*. The actual steps that were used in Experiment 1a were 13, 15, 16, 17, 18, 20 unintentionally shifting the test continuum 1 step towards the *ashi* end, compared to Liu and Jaeger (2018). As we report below, Experiment 1a yielded overall more *asi* than *ashi* responses. Experiment 1b thus attempted to expand the test stimuli towards the *ashi* end of the continuum. However, as we still had not discovered the labeling mistake, we ended up expanding the continuum even further towards the *asi* end instead, with steps 10, 13, 14, 15, 16, 20. At this point, we identified the labeling mistake. Experiment 1c employed steps 13, 17, 18, 19, 20, and 24. We emphasize that the location of the test steps is not critical for Experiments 1a-c.

5.2 Exclusions

Table 1: Summary of exclusions.

Experiment	Exclude Participant.Reason	n	Percent
CISP-1a	Catch trials	1	1.6%
CISP-1a	none	60	93.8%
CISP-1a	Reaction time	3	4.7%
CISP-1b	Catch trials	2	3.1%
CISP-1b	No headphones	1	1.6%
CISP-1b	none	58	90.6%
CISP-1b	Repeat participant	2	3.1%
CISP-1b	Swapped response keys	1	1.6%
CISP-1c	Catch trials	1	1.6%
CISP-1c	No headphones	1	1.6%
CISP-1c	none	59	93.7%
CISP-1c	Swapped response keys	1	1.6%
CISP-1c	Technical difficulty	1	1.6%

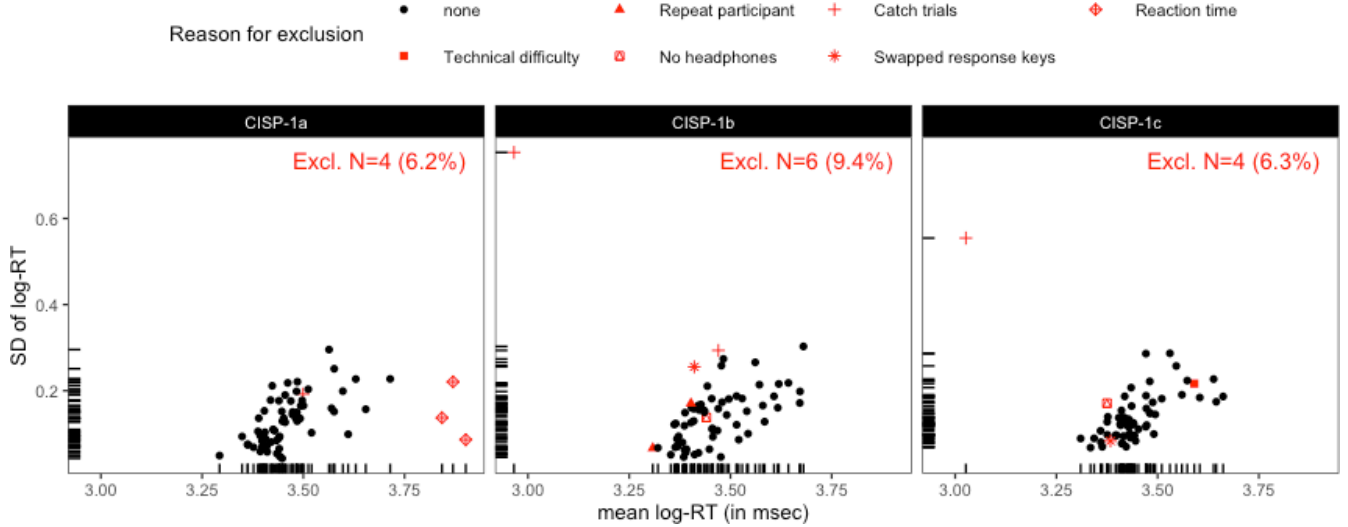


Figure 1: Mean and SD of log-transformed reaction times for participants in Experiments 1a-c.

5.3 Exit survey

5.3.1 Experiment 1a

The exit survey for Experiment 1a was identical to that of Liu and Jaeger (2018). Questions assessed the quality of the audio equipment, whether participants experienced stalling of audio or video (to help us catch problems with our code). Of particular note, the survey also contained a catch question, asking about the gender of the talker shown during the test phase. Following this survey, a final exit survey collected demographic information using the gender, age, race, and ethnicity categories required for NIH reporting. All responses in the demographic survey were indicated as optional.

- 1) Did any of the audio clips stall or lag during the experiment?
 - Yes, many
 - Yes, just a few
 - No, none
- 2) What kind of audio equipment did you use for the experiment?

- In-ear headphones
 - Over-the-ear headphones
 - External Speakers
 - Laptop Speakers
- 3) How does your audio equipment sound when you have Skype conversations, watch (high quality) music videos on YouTube, watch movies on Netflix, or engage in other similar activities involving audio?
- Poor (most words cannot be understood)
 - Okay (many sounds are distorted)
 - Good (occasionally some minor distorted sounds)
 - Excellent (crystal clear with no distorted sounds)
 - Professional quality
 - I don't know or don't do any of the above.
- 4) Did you notice anything odd about the speaker in the video?
- *text-box for subjects to answer*
- 5) Check any boxes that you think describe the speaker in the video.
- Slower than normal speaking rate
 - Normal speaking rate
 - Faster than normal speaking rate
 - Higher than normal voice pitch
 - Lower than normal voice pitch
 - Standard accented English
 - Accented or unusual pronunciations
 - Relaxed
 - Serious
- 6) Was the speaker in the video a man or a woman?
- Man
 - Woman
- 7) Did you notice anything in particular about how the speaker in the video said words that contained an “S” or “SH” sound?
- Yes
 - No
- 8) Which of the following best describes how the speaker in the video produced the “S” and “SH” sounds?
- Both S and SH sounded normal.
 - S sounded like SH.
 - SH sounded like S.
 - S sounded like SH and SH sounded like S.
- 9) You can enter any other comments you have below.
- *text-box for subjects to answer*

5.3.2 Experiment 1b

For Experiment 1b, we changed the post-experiment survey in three ways. First, we removed one question about audio quality which asked participants to rate the quality of their audio equipment as “good, professional, bad etc.”. The subjective nature of this question made it hard to interpret. Second, in order to encourage the participants to answer survey questions truthfully, we emphasised that they would get compensated independent of their answers, and that truthful answers would greatly help us with the interpretation of results. Third and finally, we added three questions that assessed the smoking habits of the participant, the exposure to smoking while growing up, and the exposure to smoking in the everyday life of the participant. These questions were included because some of the experimenters had independent interests in whether extended exposure to audiovisual speech inputs by talkers with pen-like objects in their mouth would affect how listeners perceived “asi” and “ashi” productions by a talker

with a pen in the mouth. However, given the potentially sensitive nature of the question, we decided to delete the data without analyzing it.

QUESTIONS “Please answer the questions on this survey truthfully. Your answers will not affect your compensation.”

- 1) Did any of the videos stall or lag during the experiment?
 - Yes, many
 - Yes, just a few
 - No, none
- 2) Did any of the audio clips stall or lag during the experiment?
 - Yes, many
 - Yes, just a few
 - No, none
- 3) What kind of audio equipment did you use for the experiment?
 - In-ear headphones
 - Over-the-ear headphones
 - External Speakers
 - Laptop Speakers
- 4) Did you notice anything odd about the person in the video?
 - *text-box for subjects to answer*
- 5) Check any boxes that you think describe the speaker in the video.
 - Slower than normal speaking rate
 - Normal speaking rate
 - Faster than normal speaking rate
 - Higher than normal voice pitch
 - Lower than normal voice pitch
 - Standard accented English
 - Accented or unusual pronunciations
 - Relaxed
 - Serious
- 6) Was the speaker in the video a man or a woman?
 - Man
 - Woman
- 7) Did you notice anything in particular about how the speaker in the video said words that contained an “S” or “SH” sound?
 - Yes
 - No
- 8) Which of the following best describes how the speaker in the video produced the “S” and “SH” sounds?
 - Both S and SH sounded normal.
 - S sounded like SH.
 - SH sounded like S.
 - S sounded like SH and SH sounded like S.
- 9) Which of these best defines your personal smoking habits?
 - I don’t smoke or I smoke less frequently than one cigarette per day.
 - I smoke more than one cigarette per day.
 - I smoke more than 10 cigarettes per day.
- 10) In your current daily life, how often do you interact with people while they are smoking?

- Never or less frequently than once per day.
 - More than once per day.
 - More than 10 times per day.
- 11) Which of these best defines the environment in which you were raised?
- No one close to me was a smoker.
 - One parent/guardian/person I saw on a daily basis was a smoker.
 - Multiple people I saw on a daily basis were smokers.
- 12) You can enter any other comments you have below.
- *text-box for subjects to answer*

5.3.3 Experiment 1C

For Experiment 1C, we added an additional question to the post-experiment survey. This question asked the participants to select their current timezone to let us know in what part of the US they were located and at what time of the day they took the HIT. We also made minor changes to the wording of some questions. Specifically, we changed the two questions that asked about video or sound issues during the experiment. We clarified that these questions were asking solely about technical issues, rather than any oddities between the alignment of the video and audio (as separate questions already assessed that aspect). The revised questions were introduced by the following statement: “The videos and sounds in this experiment were manipulated by aligning the same video with different sound sources. As a consequence, you might have noticed some ‘jumps’ or slightly odd-looking moments in the video. Here we are interested in *potential technical issues with the internet connection* beyond any oddities you might have noticed about how the video and sound aligned.” (emphasis in original) Then two questions assessed whether the videos or sounds failed to load or stopped and restarted playing during exposure. As these revised questions explicitly mentioned that the videos and sounds had been manipulated, they were moved towards the end of the survey, so that they followed any questions that inquired about the pronunciation of the talker in the video (recall that participants could not go back to previously answered questions).

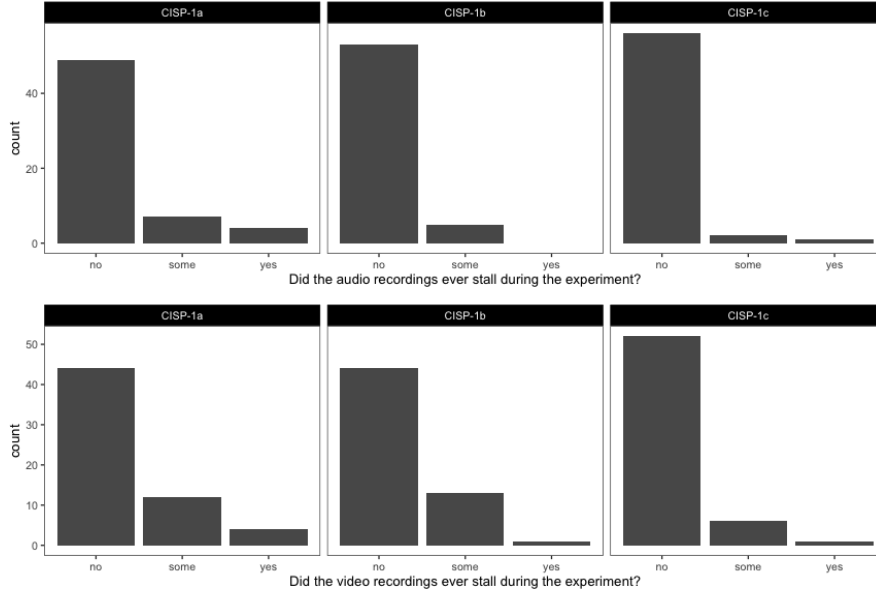
- 1) What kind of audio equipment did you use for the experiment?
 - In-ear headphones
 - Over-the-ear headphones
 - External Speakers
 - Laptop Speakers
- 2) Did you notice anything odd about the person in the video? (Please note that we will ask about audio/video lag in subsequent questions. Here we are interested in understanding anything you noticed about the person in the video.)
 - *text-box for subjects to answer*
- 3) Check any boxes that you think describe the speaker in the video.
 - Slower than normal speaking rate
 - Normal speaking rate
 - Faster than normal speaking rate
 - Higher than normal voice pitch
 - Lower than normal voice pitch
 - Standard accented English
 - Accented or unusual pronunciations
 - Relaxed
 - Serious
- 4) Was the speaker in the video a man or a woman?
 - Man
 - Woman
- 5) Did you notice anything in particular about how the speaker in the video said words that contained an “S” or “SH” sound?

- Yes
 - No
- 6) Which of the following best describes how the speaker in the video produced the “S” and “SH” sounds?
- Both S and SH sounded normal.
 - S sounded like SH.
 - SH sounded like S.
 - S sounded like SH and SH sounded like S.
- 7) The videos and sounds in this experiment were manipulated by aligning the same video with different sound sources. As a consequence, you might have noticed some ‘jumps’ or slightly odd looking moments in the video. Here we are interested in potential technical issues with the internet connection beyond any oddities you might have noticed about how the video and sound aligned. Did any of the videos fail to load or stop playing for a moment and then restart during the experiment? (We will ask about the sounds separately on the next page.)
- Yes, many
 - Yes, just a few
 - No, none
- 8) Did any of the audio clips fail to load or stop playing for a moment and then restart during the experiment?
- Yes, many
 - Yes, just a few
 - No, none
- 9) Which of these best defines your personal smoking habits?
- I don’t smoke or I smoke less frequently than one cigarette per day.
 - I smoke more than one cigarette per day.
 - I smoke more than 10 cigarettes per day.
- 10) In your current daily life, how often do you interact with people while they are smoking?
- Never or less frequently than once per day.
 - More than once per day.
 - More than 10 times per day.
- 11) Which of these best defines the environment in which you were raised?
- No one close to me was a smoker.
 - One parent/guardian/person I saw on a daily basis was a smoker.
 - Multiple people I saw on a daily basis were smokers.
- 12) What is your current timezone?
- Hawaii Standard Time (UTC -10)
 - Hawaii-Aleutian Daylight Time (UTC -9)
 - Alaska Daylight Time (UTC -8)
 - Pacific Daylight Time (UTC -7)
 - Mountain Standard Time (UTC -7)
 - Mountain Daylight Time (UTC -6)
 - Central Daylight Time (UTC -5)
 - Eastern Daylight Time (UTC -4)
 - other
- 13) You can enter any other comments you have below.
- *text-box for subjects to answer*

5.3.4 Survey results

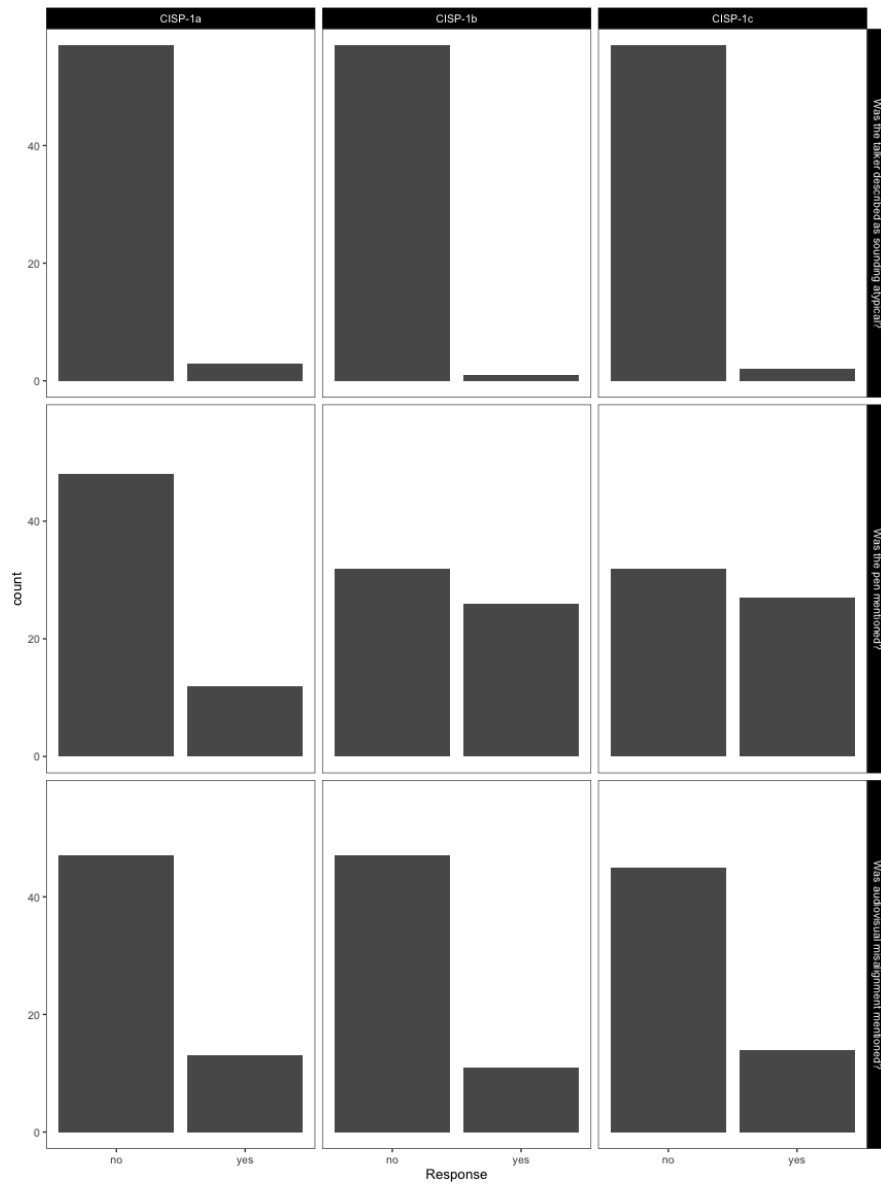
All survey responses are shared as part of the OSF repo for this project. The first questions assessed the issues with the audio or video playback. The clear majority of all participants did not report any issues with the audio or

video recordings.

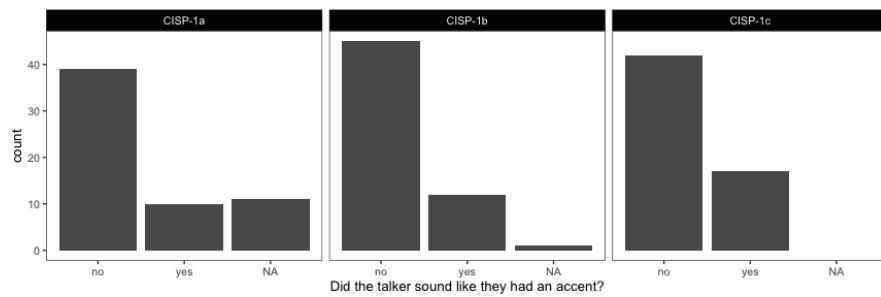


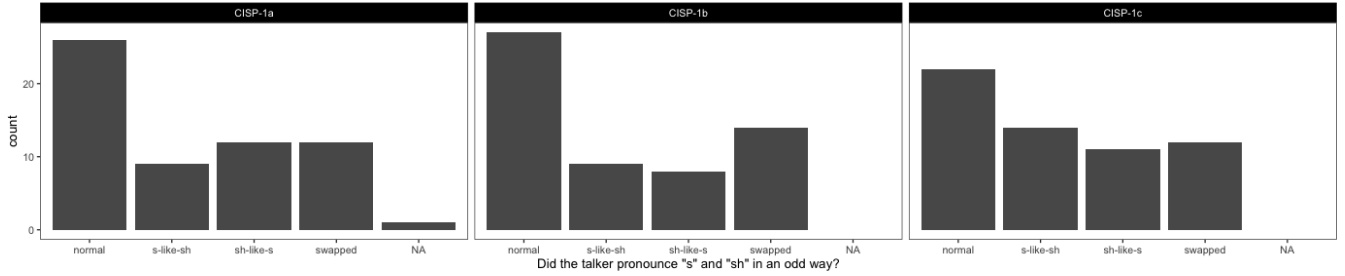
The next set of questions assessed how participants perceived the speech of the talker, including potential effects of the pen. The first question simply asked participants whether they noticed anything odd about the talker, followed by an open-ended text prompt. We annotated participants’ responses in terms of whether their statement mentioned that 1) the talker sounded atypical (e.g., having a lisp), 2) the talker had a pen in the mouth, and 3) the audio and video were misaligned. For statements that mentioned the pen, we also annotated what participants noted about the pen.

While only 6 (3.4%) participants across the three experiments considered the talkers’ speech atypical (e.g., “accented”, exhibiting a “lisp”), 65 (36.7%) participants commented on the fact that the talker often had a pen in the mouth. The latter is expected given that the pen was in the talker’s mouth on half of the trials, and thus rather salient. Of the responses that mentioned the pen, some claimed that the pen affected the pronunciation of the talker (in reality the audio recording was by design independent of the location of the pen). Most participants simply noticed the pen. Some participants, however, further noted that the pen affected their ability to tell whether the talker was saying *asi* or *ashi* (7 (4%)), or that the pen affected the talker’s pronunciation (6 (3.4%)), and some hypothesized that the experiment was about the pen (3 (1.7%)). While these responses suggests that relatively few participants had conscious hypotheses about the effects of the pen, Section 5.5 below presents subset analyses of Experiments 1a-c that exclude those participants. Finally, a substantial minority of participants—38 (21.5%)—noted misalignments between the sound and the video. This is an inevitable consequence of the approach we chose to investigate the effect of the pen (see main text for our rationale for this approach). Critically thought, it is unclear how awareness of audiovisual misalignment would bias the results we reliably observed across Experiments 1a-c.



Subsequent multiple choice questions further explored participants' perception of the talker. The clear majority of the participants considered the talker's pronunciations to be typical. At the same time, however, the majority of participants answered on a later question that the talker had "swapped" "s" and "sh". It is possible that some of participants' responses to this latter question reflect confusion. For instance, since "s" and "sh" were always presented in the context of the nonce-word *asi/ashi*, it is unclear how participants would come to the conclusion that the talker swapped "s" and "sh". Yet, about one sixth of all participants gave that answer.





5.4 Analyses

2 summarizes the results of Experiments 1a-c. The data suggests qualitatively similar effects of pen location, acoustic continuum, and the visual bias across all three experiments—as also confirmed by the analyses presented below. In particular, participants in all three experiments were less likely to respond “ashi” when the pen was in the mouth during the production of the nonce-word.

For reference, responses from a audio-only experiment using some of the same stimuli (Liu and Jaeger 2018) are shown in gray. This reference line also illustrates a well-known dependency of listeners’ categorization responses on the range of stimuli that participants are exposed to. Such dependencies might arise, for example, because participants assume that the range they hear is meant to express the range from typical *asi* to typical *ashi*, or because they think that both response options should be used about equally often. A bias of either type will result in the pattern we observe across Experiments 1a-c. For instances, whereas participants’ responses in Experiment 1a resemble those from Liu and Jaeger when the pen was in the hand, the opposite is true for Experiment 1b. This is the case because, for the same acoustic input, participants in Experiment 1b were more likely to respond “ashi” than participants in Experiment 1a. Notably, the acoustic continuum in Experiment 1b was overall shifted towards more *asi*-like steps, compared to Experiment 1a. Conversely, Experiment 1c shifted the acoustic continuum towards more *ashi*-like steps, compared to Experiment 1a. And, in line with the reasoning offered above, participants in Experiment 1c were overall less likely to respond “ashi” for the same acoustic input than participants in Experiment 1a.

This dependency on the range of inputs is important to keep in mind when interpreting the results. If the effect of the pen in the mouth is due to compensation, we would expect decreased rates of “ashi” responses relative to a baseline in which the talker’s articulation is not obstructed by anything. While the pen-in-the-hand condition arguably provides this baseline, a skeptic might argue that Experiments 1a-c leave open whether the pen in the mouth decreases the probability of *ashi*-responses, or whether—for unspecified reasons—the pen in the hand *increases* the probability of *ashi*-responses. In this context, it is encouraging to see that Experiment 1a—the experiment that most closely matches the acoustic continuum from Liu and Jaeger’s audio-only experiment—suggests that the pen in the mouth indeed decreases the probability of *ashi*-responses, as predicted by Fowler’s compensation account (Fowler 2006). We address this question more directly in Experiment 2, which removes other potential confounds that might result from comparing audio-only to audiovisual experiments.

5.4.1 Experiment 1a

Table 3: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1a	Pen location Mouth -> fewer <i>ashi</i> -responses	-0.68	0.26	-1.12	-0.26	165.7	0.994	*
1a	Pen effect increases for more <i>ashi</i> -like acoustic input	0.01	0.09	-0.13	0.15	0.8	0.442	
1a	Pen effect increases for visually <i>ashi</i> -biased input	-0.37	0.31	-0.88	0.14	7.5	0.883	
1a	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.13	0.10	-0.29	0.03	9.8	0.907	

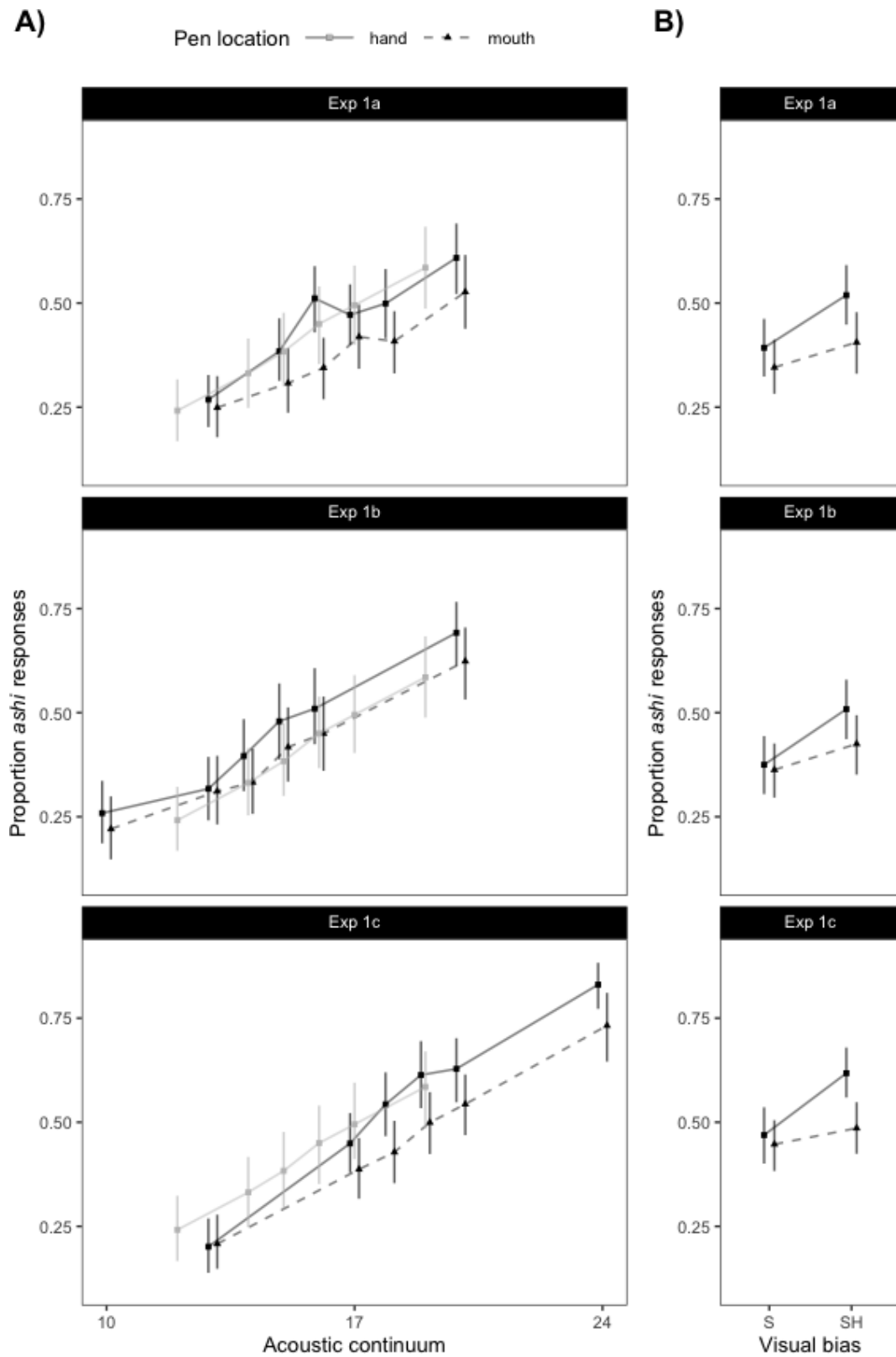


Figure 2: Summary of participants' responses in Experiments 1a-c, depending on pen location and acoustic continuum step (Panel A) or visual bias (Panel B). For comparison, responses from a audio-only experiment using some of the same stimuli (Liu and Jaeger 2018) are shown in gray.

Table 2: Results of Bayesian mixed-effect logistic regression for Experiment 1a.

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.85	0.30	-2.46	-1.28
Condition.Test.OriginalLabelSH	0.81	0.27	0.24	1.31
Condition.Test.PenM	-0.71	0.34	-1.38	-0.03
Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.05	0.35	-0.74	0.66
moBlock	0.02	0.13	-0.23	0.29
moCondition.Test.Audio	0.29	0.05	0.19	0.38
moBlock:Condition.Test.OriginalLabelSH	-0.11	0.09	-0.28	0.09
moBlock:Condition.Test.PenM	0.11	0.10	-0.09	0.29
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.12	0.08	-0.27	0.07
moCondition.Test.Audio:Condition.Test.PenM	0.01	0.09	-0.15	0.18
moBlock:moCondition.Test.Audio	0.00	0.06	-0.12	0.12
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.01	0.13	-0.26	0.27
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.03	0.10	-0.32	0.06
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.02	0.02	-0.03	0.06
moBlock:moCondition.Test.Audio:Condition.Test.PenM	0.03	0.02	-0.03	0.07
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.06	-0.12	0.13

Table 4: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1a	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.29	0.05	0.21	0.37	> 3999	1.000	*
1a	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	0.50	0.22	0.14	0.87	71.7	0.986	*
1a	Acoustic and visual bias effects are independent	-0.12	0.08	-0.27	0.07	8.4	0.894	

Table 5: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1a	Pen effect is stable over blocks	0.19	0.11	-0.04	0.39	18.3	0.948	
1a	Acoustic effect is stable over blocks	0.00	0.06	-0.12	0.12	36.7	0.973	
1a	Visual bias effect is stable over blocks	-0.07	0.11	-0.28	0.15	57.8	0.983	

5.4.2 Experiment 1b

Table 7: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1b	Pen location Mouth -> fewer <i>ashi</i> -responses	-0.40	0.25	-0.82	0.00	18.9	0.950	
1b	Pen effect increases for more <i>ashi</i> -like acoustic input	-0.04	0.05	-0.13	0.04	3.6	0.783	
1b	Pen effect increases for visually <i>ashi</i> -biased input	-0.30	0.28	-0.74	0.16	6.6	0.868	
1b	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	0.01	0.08	-0.12	0.15	0.8	0.432	

Table 6: Results of Bayesian mixed-effect logistic regression for Experiment 1b.

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.97	0.31	-2.60	-1.36
Condition.Test.OriginalLabelSH	1.07	0.33	0.44	1.73
Condition.Test.PenM	-0.30	0.31	-0.91	0.30
Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.34	0.34	-0.97	0.36
moBlock	0.04	0.20	-0.33	0.45
moCondition.Test.Audio	0.28	0.02	0.24	0.33
moBlock:Condition.Test.OriginalLabelSH	-0.13	0.10	-0.33	0.06
moBlock:Condition.Test.PenM	0.04	0.09	-0.14	0.22
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.01	0.06	-0.10	0.12
moCondition.Test.Audio:Condition.Test.PenM	-0.04	0.05	-0.15	0.06
moBlock:moCondition.Test.Audio	0.00	0.04	-0.07	0.07
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.02	0.20	-0.39	0.38
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.08	-0.15	0.17
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.01	0.02	-0.05	0.03
moBlock:moCondition.Test.Audio:Condition.Test.PenM	-0.01	0.02	-0.04	0.02
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.04	-0.07	0.07

Table 8: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1b	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.28	0.02	0.25	0.32	> 3999	1.00	*
1b	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	1.08	0.24	0.70	1.49	> 3999	1.00	*
1b	Acoustic and visual bias effects are independent	0.01	0.06	-0.10	0.12	48.4	0.98	

Table 9: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1b	Pen effect is stable over blocks	0.02	0.10	-0.18	0.22	74.2	0.987	
1b	Acoustic effect is stable over blocks	0.00	0.04	-0.07	0.07	81	0.988	
1b	Visual bias effect is stable over blocks	-0.16	0.11	-0.39	0.05	24.5	0.961	

5.4.3 Experiment 1c

Table 11: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1c	Pen location Mouth -> fewer <i>ashi</i> -responses	-0.49	0.29	-0.98	-0.02	21.1	0.955	*
1c	Pen effect increases for more <i>ashi</i> -like acoustic input	-0.09	0.05	-0.17	-0.01	26	0.963	*
1c	Pen effect increases for visually <i>ashi</i> -biased input	-0.77	0.31	-1.27	-0.25	132.3	0.992	*
1c	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.10	0.07	-0.22	0.02	11.2	0.918	

Table 10: Results of Bayesian mixed-effect logistic regression for Experiment 1c.

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-2.07	0.24	-2.56	-1.60
Condition.Test.OriginalLabelSH	0.38	0.31	-0.25	0.98
Condition.Test.PenM	-0.27	0.36	-0.97	0.43
Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.52	0.37	-1.25	0.22
moBlock	0.03	0.21	-0.40	0.42
moCondition.Test.Audio	0.34	0.02	0.31	0.38
moBlock:Condition.Test.OriginalLabelSH	0.04	0.09	-0.14	0.22
moBlock:Condition.Test.PenM	0.09	0.09	-0.08	0.28
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.01	0.06	-0.11	0.11
moCondition.Test.Audio:Condition.Test.PenM	-0.09	0.05	-0.19	0.01
moBlock:moCondition.Test.Audio	0.00	0.03	-0.06	0.06
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.05	0.21	-0.35	0.47
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.07	0.07	-0.24	0.05
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.02	0.02	-0.01	0.05
moBlock:moCondition.Test.Audio:Condition.Test.PenM	0.00	0.01	-0.03	0.03
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.00	-0.06	0.06

Table 12: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1c	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.34	0.02	0.31	0.37	> 3999	1.000	*
1c	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	0.41	0.24	0.01	0.80	20.3	0.953	*
1c	Acoustic and visual bias effects are independent	0.01	0.06	-0.11	0.11	48	0.980	

Table 13: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1c	Pen effect is stable over blocks	0.09	0.10	-0.09	0.29	52.1	0.981	
1c	Acoustic effect is stable over blocks	0.00	0.03	-0.06	0.06	116.1	0.991	
1c	Visual bias effect is stable over blocks	0.09	0.10	-0.12	0.29	51	0.981	

5.4.4 Liu and Jaeger (2018, Experiment 1a-b)

For comparison, we also analyzed the “baseline” experiment from Liu and Jaeger (2018, Experiment 1a-b) with the same analysis approach employed in the present study. Like all previous work in this line of research, this baseline experiment employed audio-only test stimuli. Unlike our Experiments 1a-c, Liu and Jaeger’s control experiment contained an exposure phase (with only typical fricative pronunciations). We further note that none of Experiments 1a-c employed the exact same continuum steps as Liu and Jaeger (2018). It is known that the selection of acoustic continuum steps that participants experience in an experiment can affect their categorization responses: even the exact same continuum step can be categorized differently depending on the other steps included in the experiment (Yamata & Tohkura, 1992). These factors limit the extent to which our Experiments 1a-c can be directly compared to Liu and Jaeger’s audio-only experiment. We thus focus only on the magnitude of the acoustic effects in Liu and Jaeger, compared to the present experiments.

Table 14: Effects of acoustic continuum.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
LJ18-NORM	Acoustic continuum more ASHI-like -> more ASHI-responses	0.30	0.07	0.19	0.41	> 3999	1.000	*
LJ18-NORM	Continuum effect is stable over blocks	0.05	0.01	0.03	0.08	0.5	0.353	*

5.5 Subset analyses removing participants who commented on effects of the pen

As noted in the description of the survey results in Section 5.3.4, some participants commented on potential effects of the pen. We thus conducted a subset analysis to address the potential that those participants might have responded in a way that reflects conscious deliberations about the effect of the pen, rather than the presumably automatic, subconscious processes entailed by compensation. For this analysis, we excluded all participants who commented that 1) the pen made it difficult to see what the talker said, 2) the pen affected pronunciation, or 3) the experiment was about the effects of the pen. This removed 39 additional participants, leaving only 138 participants for analysis. As reported in the following subsections, all results reported in the main analyses held for this much smaller subset of participants. In fact, the effect estimates ($\hat{\beta}$) for the subset analyses were as large or larger than in the main analyses.

5.5.1 Experiment 1a

Table 15: Results of Bayesian mixed-effect logistic regression for Experiment 1a after excluding participants who responded that the pen affected the talker’s pronunciation (in reality, this was not the case in our recordings).

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.61	0.37	-2.34	-0.89
Condition.Test.OriginalLabelSH	0.57	0.37	-0.19	1.25
Condition.Test.PenM	-0.79	0.36	-1.50	-0.08
Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.19	0.55	-0.89	1.26
moBlock	-0.06	0.04	-0.16	0.02
moCondition.Test.Audio	0.17	0.04	0.09	0.25
moBlock:Condition.Test.OriginalLabelSH	0.06	0.12	-0.16	0.28
moBlock:Condition.Test.PenM	0.08	0.09	-0.10	0.26
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.06	0.10	-0.26	0.12
moCondition.Test.Audio:Condition.Test.PenM	0.00	0.08	-0.15	0.16
moBlock:moCondition.Test.Audio	0.03	0.01	0.01	0.05
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.01	0.18	-0.35	0.36
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.22	0.15	-0.51	0.08
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.02	0.03	-0.08	0.05
moBlock:moCondition.Test.Audio:Condition.Test.PenM	0.01	0.02	-0.04	0.05
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.03	0.04	-0.06	0.12

Table 16: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1a	Pen location Mouth -> fewer <i>ashi</i> -responses	-0.79	0.30	-1.27	-0.30	209.5	0.995	*
1a	Pen effect increases for more <i>ashi</i> -like acoustic input	0.00	0.08	-0.13	0.13	1	0.506	
1a	Pen effect increases for visually <i>ashi</i> -biased input	-0.36	0.43	-1.05	0.35	4	0.801	
1a	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.22	0.15	-0.46	0.03	12.7	0.927	

Table 17: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1a	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.17	0.04	0.10	0.24	> 3999	1.000	*
1a	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	0.42	0.26	-0.01	0.84	17.4	0.946	
1a	Acoustic and visual bias effects are independent	-0.06	0.10	-0.26	0.12	26.3	0.963	

Table 18: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1a	Pen effect is stable over blocks	0.09	0.06	-0.02	0.20	38.3	0.975	
1a	Acoustic effect is stable over blocks	0.03	0.01	0.01	0.05	6.3	0.862	*
1a	Visual bias effect is stable over blocks	0.03	0.06	-0.09	0.15	118.9	0.992	

5.5.2 Experiment 1b

Table 19: Results of Bayesian mixed-effect logistic regression for Experiment 1b after excluding participants who responded that the pen affected the talker’s pronunciation (in reality, this was not the case in our recordings).

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.72	0.42	-2.54	-0.89
Condition.Test.OriginalLabelSH	1.35	0.38	0.59	2.09
Condition.Test.PenM	-0.34	0.35	-1.07	0.35
Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.21	0.66	-1.49	1.12
moBlock	0.03	0.06	-0.09	0.14
moCondition.Test.Audio	0.26	0.04	0.18	0.34
moBlock:Condition.Test.OriginalLabelSH	-0.28	0.11	-0.49	-0.04
moBlock:Condition.Test.PenM	0.01	0.11	-0.19	0.23
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.06	0.06	-0.17	0.06
moCondition.Test.Audio:Condition.Test.PenM	-0.03	0.06	-0.16	0.08
moBlock:moCondition.Test.Audio	0.00	0.01	-0.02	0.02
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.04	0.23	-0.39	0.50
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.02	0.13	-0.27	0.21
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.02	0.02	-0.02	0.05
moBlock:moCondition.Test.Audio:Condition.Test.PenM	0.00	0.02	-0.03	0.04
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.01	0.04	-0.09	0.07

Table 20: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1b	Pen location Mouth -> fewer <i>ashi</i> -responses	-0.42	0.28	-0.88	0.03	15.1	0.938	
1b	Pen effect increases for more <i>ashi</i> -like acoustic input	-0.03	0.06	-0.14	0.06	2.4	0.708	
1b	Pen effect increases for visually <i>ashi</i> -biased input	-0.26	0.47	-1.01	0.48	2.5	0.712	
1b	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.02	0.13	-0.23	0.18	1.2	0.551	

Table 21: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1b	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.26	0.04	0.19	0.33	> 3999	1.000	*
1b	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	1.19	0.29	0.71	1.68	> 3999	1.000	*
1b	Acoustic and visual bias effects are independent	-0.06	0.06	-0.17	0.06	23.5	0.959	

Table 22: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1b	Pen effect is stable over blocks	0.01	0.07	-0.13	0.16	109.4	0.991	
1b	Acoustic effect is stable over blocks	0.00	0.01	-0.02	0.02	231.4	0.996	
1b	Visual bias effect is stable over blocks	-0.23	0.08	-0.38	-0.07	1.5	0.594	*

5.5.3 Experiment 1c

Table 23: Results of Bayesian mixed-effect logistic regression for Experiment 1c after excluding participants who responded that the pen affected the talker’s pronunciation (in reality, this was not the case in our recordings).

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.63	0.37	-2.32	-0.88
Condition.Test.OriginalLabelSH	0.50	0.37	-0.22	1.25
Condition.Test.PenM	-0.25	0.41	-1.04	0.54
Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.82	0.67	-2.17	0.44
moBlock	-0.12	0.07	-0.27	0.00
moCondition.Test.Audio	0.29	0.05	0.19	0.39
moBlock:Condition.Test.OriginalLabelSH	-0.11	0.12	-0.36	0.11
moBlock:Condition.Test.PenM	0.04	0.12	-0.17	0.28
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.03	0.07	-0.16	0.10
moCondition.Test.Audio:Condition.Test.PenM	-0.10	0.06	-0.22	0.02
moBlock:moCondition.Test.Audio	0.03	0.01	0.01	0.06
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.26	0.22	-0.17	0.68
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.07	0.12	-0.31	0.16
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.03	0.02	-0.01	0.07
moBlock:moCondition.Test.Audio:Condition.Test.PenM	-0.01	0.02	-0.04	0.03
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.02	0.04	-0.10	0.06

Table 24: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1c	Pen location Mouth -> fewer <i>ashi</i> -responses	-0.49	0.34	-1.04	0.06	12.4	0.926	
1c	Pen effect increases for more <i>ashi</i> -like acoustic input	-0.10	0.06	-0.20	0.00	19.2	0.950	*
1c	Pen effect increases for visually <i>ashi</i> -biased input	-1.01	0.49	-1.83	-0.24	57	0.983	*
1c	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.07	0.12	-0.27	0.12	2.8	0.737	

Table 25: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1c	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.29	0.05	0.22	0.37	> 3999	1.000	*
1c	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	0.42	0.28	-0.03	0.88	15.7	0.940	
1c	Acoustic and visual bias effects are independent	-0.03	0.07	-0.16	0.10	36.1	0.973	

Table 26: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
1c	Pen effect is stable over blocks	0.03	0.08	-0.12	0.21	96.8	0.990	
1c	Acoustic effect is stable over blocks	0.03	0.01	0.01	0.06	1.5	0.598	*
1c	Visual bias effect is stable over blocks	-0.05	0.08	-0.22	0.11	86.6	0.989	

6 Experiments 2 and 2b

We conducted two experiments that were designed to test whether listeners need to see the effects of the pen on the articulators *during the production of the fricative*, or whether the presence of a pen during the production of a fricative is sufficient to cause the effects observed in Experiments 1a-c. To this end, Experiments 2 and 2b introduced a black box that occluded the talker’s mouth during the articulation of the fricative (see main text for details). The two experiments were identical except that Experiment 2b introduced an additional task intended to assure that participants paid attention to the videos (rather than just the audio). This change in procedure from Experiment 2 to 2b was intended to address an alternative explanation for the results of Experiment 2. As reported below, Experiment 2 did not find the effects of pen location that were present in Experiments 1a-c. This raised the question as to whether participants in Experiment 2 failed to pay attention to the video (unlike in Experiments 1a-c).

On the one hand, the catch task employed in Experiment 2 (same as in Experiments 1a-c: participants were asked to press “B” when they saw a white dot in the video) would seem to reject that possibility: as in Experiments 1a-c, participants that pressed “B” in the absence of a white dot were excluded from the analysis of Experiment 2. However, since none of the trials actually contained a white dot, this means that participants that did not pay any attention to the video and never pressed “B” would fail to be excluded from the analysis of Experiment 2. In Experiment 2b, we thus replaced the catch task from Experiments 1a-c & 2, and instead asked Experiments to press SPACE when the pen was in the mouth. As in Experiment 2, participants in Experiment 2b also had to press “X” or “M” to indicate whether they heard *asi* or *ashi*.

6.1 Procedure

Experiment 2 employed the exact same tasks as Experiment 1a-c. Experiment 2b, however, introduced an additional task. Participants were asked to press SPACE whenever the pen was in the talker’s mouth. After pressing SPACE, participants had to answer whether they heard *asi* or *ashi*, as in all preceding experiments.

6.2 Exclusions

Figure 3 summarized the exclusions for Experiment 2 and 2b. Of note is the high exclusion rate for Experiment 2b, for which almost one fourth of all participants stated after the experiment that they had not been wearing head phones. This deterioration of data quality for experiments conducted over Mechanical Turk post 2020 was also observed in other experiments conducted in our lab during the same time period. Like in the present study, data quality tended to decrease in particular for the later studies of a series of experiments. One possible explanation is our recruitment criterion, which only allowed participants to see the experiment if they had *not* participated in any previous experiment of the series. This successively reduces the participant pool, potentially making experiments in a series increasingly more vulnerable to less cooperative participants. This motivated our switch to the Prolific

crowdsourcing platform for subsequent experiments. Experiment 2b was the last experiment we conducted over Mechanical Turk.

Table 27: Summary of exclusions.

Experiment	Exclude Participant.Reason	n	Percent
CISP-2	Catch trials	1	1.6%
CISP-2	No headphones	1	1.6%
CISP-2	none	58	90.6%
CISP-2	Reaction time	2	3.1%
CISP-2	Swapped response keys	2	3.1%
CISP-2b	No headphones	14	21.9%
CISP-2b	none	48	75%
CISP-2b	Reaction time	2	3.1%

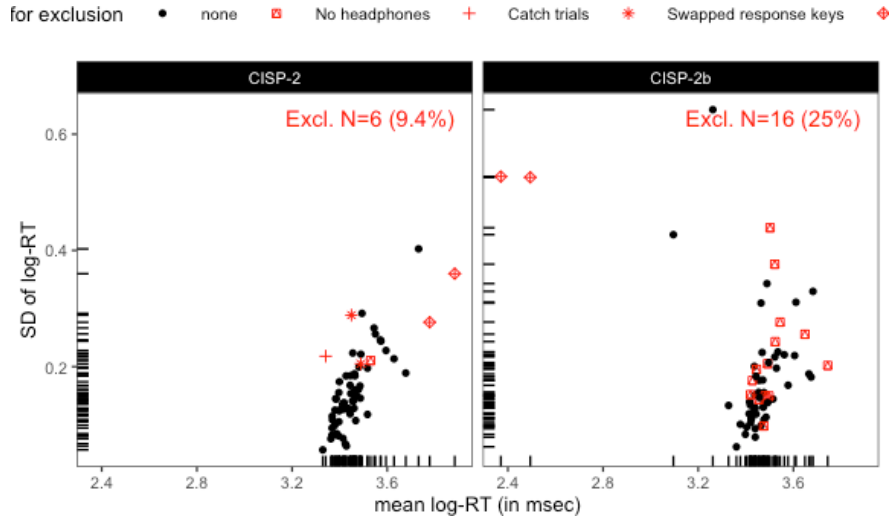


Figure 3: Mean and SD of log-transformed reaction times for participants in Experiments 2, 2b.

6.3 Analyses - Experiment 2

Table 29: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$
2	Pen location Mouth -> fewer <i>ashi</i> -responses	0.00	0.21	-0.35	0.34	1	0.490
2	Pen effect increases for more <i>ashi</i> -like acoustic input	0.02	0.05	-0.06	0.10	0.6	0.386
2	Pen effect increases for visually <i>ashi</i> -biased input	-0.18	0.27	-0.62	0.26	2.9	0.744
2	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.06	0.08	-0.19	0.08	3.1	0.754

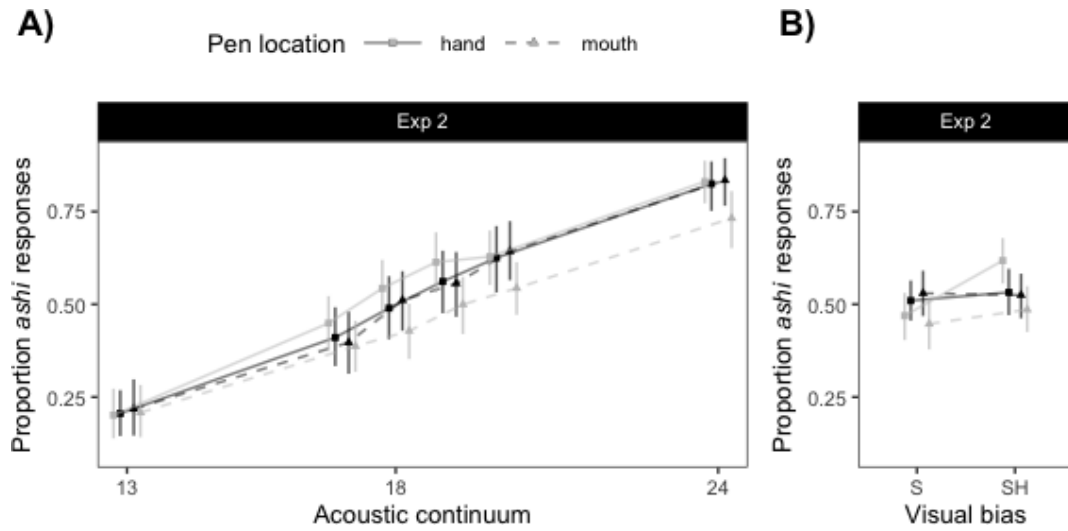


Figure 4: (ref:exp2-plot)

Table 28: Results of Bayesian mixed-effect logistic regression for Experiment 2.

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.58	0.25	-2.05	-1.11
Condition.Test.OriginalLabelSH	0.12	0.27	-0.41	0.68
Condition.Test.PenM	-0.04	0.28	-0.60	0.51
Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.03	0.35	-0.72	0.65
moBlock	-0.03	0.14	-0.33	0.25
moCondition.Test.Audio	0.35	0.03	0.30	0.41
moBlock:Condition.Test.OriginalLabelSH	-0.05	0.08	-0.21	0.13
moBlock:Condition.Test.PenM	0.02	0.10	-0.16	0.22
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	-0.03	0.05	-0.12	0.06
moCondition.Test.Audio:Condition.Test.PenM	0.02	0.05	-0.08	0.11
moBlock:moCondition.Test.Audio	0.00	0.05	-0.10	0.10
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	-0.03	0.14	-0.32	0.27
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.08	-0.22	0.10
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.01	0.02	-0.03	0.04
moBlock:moCondition.Test.Audio:Condition.Test.PenM	0.02	0.02	-0.01	0.05
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.05	-0.10	0.10

Table 30: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
2	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.35	0.03	0.31	0.40	> 3999	1.000	*
2	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	0.05	0.20	-0.27	0.39	1.4	0.582	
2	Acoustic and visual bias effects are independent	-0.03	0.05	-0.12	0.06	52.8	0.981	

Table 31: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$
2	Pen effect is stable over blocks	0.08	0.10	-0.13	0.28	57.8	0.983
2	Acoustic effect is stable over blocks	0.00	0.05	-0.10	0.10	53.4	0.982
2	Visual bias effect is stable over blocks	-0.03	0.09	-0.22	0.16	85.5	0.988

6.4 Comparison of Experiment 2 against Experiment 1c

Figure 5 show the predictions for the effect of pen location in both Experiments 2 and 1c, based on a model fit to the data from both experiments, while allowing Experiment to interact with all other predictors.

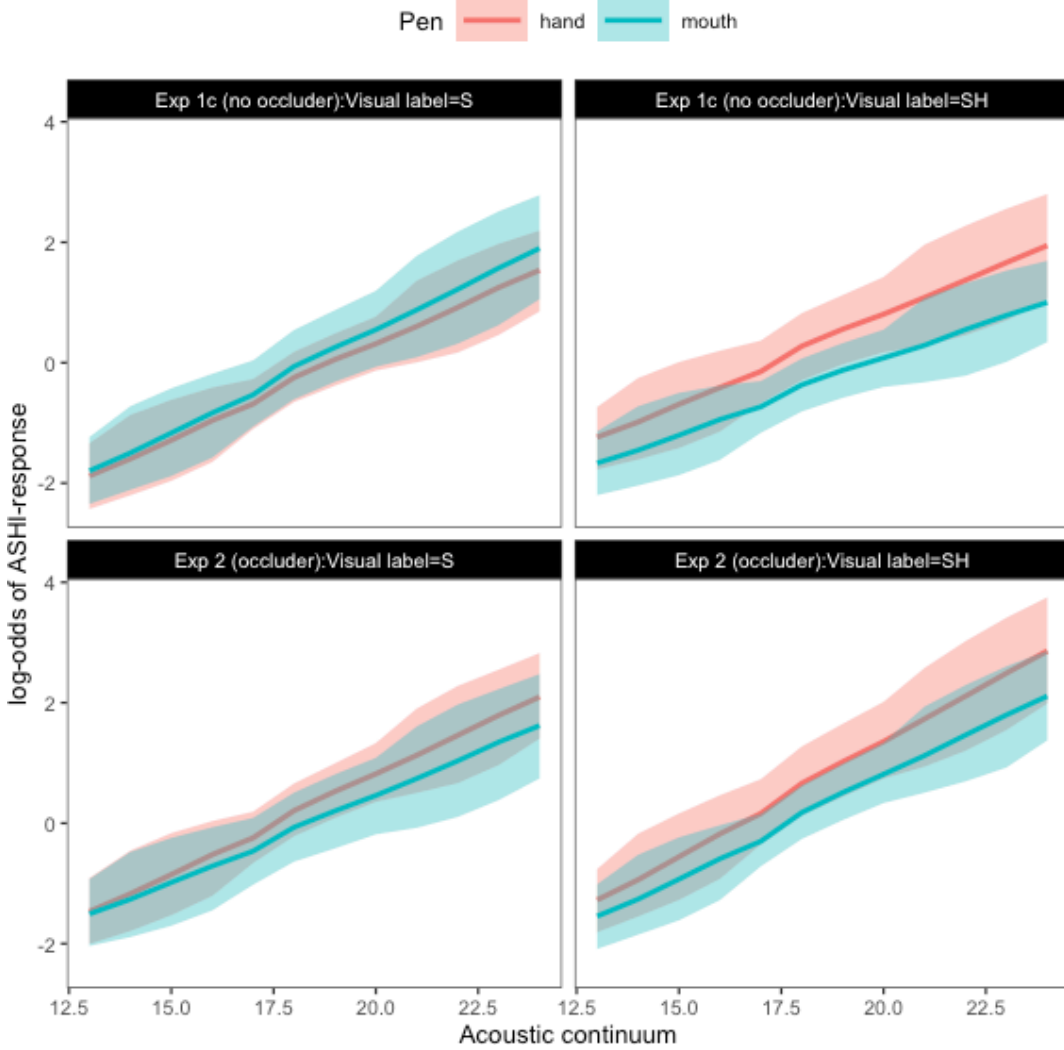


Figure 5: Comparing the effects of acoustic continuum and pen location in Experiment 2 and 1c.

If the effects of pen location in Experiments 1a-c were due to the pen visually occluding otherwise available visual cues to articulatorily relevant evidence, then both pen locations in Experiment 2 should yield *ashi*-responses similar to the pen-in-mouth condition of Experiment 1c (which used the exact same audio-visual stimuli as Experiment 2, except without a visual occluder). To test this hypothesis, we combined the data from Experiments 1c and 2, and coded both pen locations of Experiment 2 as a single new pen location (“occluded”). We then analyzed this combined data in the exact same way as Experiments 1c and 2, except that pen location now was a 3-way sliding difference-coded factor, testing (1) whether the occluded condition (Experiment 2) elicited fewer *ashi* responses than the pen-in-hand condition of Experiment 1c (as predicted by the occlusion hypothesis) and (2) whether the

pen-in-mouth condition of Experiment 1c yielded fewer *ashi* responses than the occluded condition (as predicted if occlusion does *not* explain all of the effects of pen location in Experiments 1a-c).

Overall, the results of this analysis do not provide any notable evidence for the occlusion hypothesis while providing strong evidence for the compensation hypothesis. In the middle of the acoustic continuum, *ashi*-responses were *more* likely in Experiment 2, compared to even the pen-in-hand condition of Experiment 1c ($BF = 2$). This is the opposite of what would be expected if the main effects of pen location in Experiments 1a-c were due to visual occlusion. There also was strong evidence that the pen-in-mouth condition of Experiment 1c resulted in fewer *ashi*-responses than in Experiment 2 ($BF = 73.1$).

The only way in which the visual occluder in Experiment 2 resembled the pen-in-mouth condition in Experiment 1c was in terms of their interaction with visual evidence to the articulation of “sh”: both the black rectangle in Experiment 2 and the pen-in-the-mouth in Experiment 1c reduced the effect of visually *ashi*-like stimuli, compared to the pen-in-hand condition (see third row of the two tables). These two effects were almost identical (comparison of occluder vs. pen-in-hand condition: $BF = 14.2$; comparison of pen-in-mouth vs. occluder condition: $BF = .9$). This result is not particularly surprising given that the black rectangle completely occluded any direct visual evidence of “s” vs. “sh” articulation.

Table 32: Effects of pen location: occlusion vs. hand

Exp	Hypothesis	$\hat{\beta}$	SE	CI_l	CI_u	BF	$p_{posterior}$	
2 vs. 1c	Pen location Mouth -> fewer ASHI-responses	0.14	0.33	-0.40	0.66	0.5	0.345	
2 vs. 1c	Pen effect increases for more ASHI-like acoustic input	-0.04	0.05	-0.12	0.04	3.8	0.790	
2 vs. 1c	Pen effect increases for visually ASHI-biased input	-0.74	0.50	-1.59	0.07	14.2	0.934	
2 vs. 1c	Pen effect increases even more when acoustic and visual input is ASHI-biased	-0.02	0.06	-0.12	0.09	1.6	0.610	

Table 33: Effects of pen location: mouth vs. occlusion

Exp	Hypothesis	$\hat{\beta}$	SE	CI_l	CI_u	BF	$p_{posterior}$	
2 vs. 1c	Pen location Mouth -> fewer ASHI-responses	-0.71	0.33	-1.25	-0.19	73.1	0.986	*
2 vs. 1c	Pen effect increases for more ASHI-like acoustic input	-0.04	0.04	-0.11	0.03	5.1	0.836	
2 vs. 1c	Pen effect increases for visually ASHI-biased input	0.02	0.40	-0.65	0.67	0.9	0.484	
2 vs. 1c	Pen effect increases even more when acoustic and visual input is ASHI-biased	-0.07	0.08	-0.21	0.06	4.3	0.811	

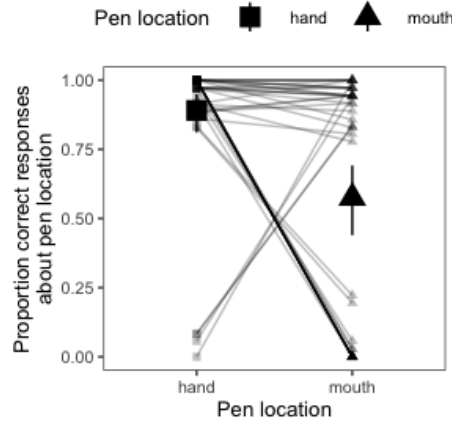
6.5 Analysis - Experiment 2b

We first analyzed participants’ accuracy in detecting when the pen was in talker’s mouth, in which case participants were supposed to press SPACE. Then we turn to the analysis of participants’ categorization responses.

6.5.1 SPACE bar presses

Participants’ overall accuracy was above chance (73.1%). Overall, this suggests that participants could detect the pen presence above chance, while also suggesting that participants experienced the task to be difficult.

When the pen was in the talker’s mouth, participants were only slightly above chance (57.5% accurate). When the pen was in the talker’s hand, participants were much more accurate (89%). This asymmetry is not particularly surprising: correct detection of the pen in the mouth required a key press, whereas correct detection that the pen was not in the mouth (but rather in the hand) required no action at all.



A closer look at participants’ SPACE responses reveals that some participants almost never pressed the SPACE bar and some always pressed the SPACE bar, regardless of whether the pen was in the mouth (the lines forming the “X” in the figure). For example, 17 (16%) participants pressed SPACE on 90% or more of all trials or did not press SPACE on 90% or more of the trials (all but one of these participants, never pressed SPACE). If those participants are excluded from the analysis, performance improved to 86.1% accuracy, with statistically indistinguishable performance when the pen was in the mouth vs. hand.

We note that, in total, about half of the participants in Experiment 2b did not follow instructions—either by not wearing headphones for the experiment or by never/always pressing SPACE. As mentioned under Exclusions, we have encountered similar issues with data quality over Mechanical Turk over recent years whenever we conducted series of experiments (for which any subsequent experiment only recruits from the pool of participants that have *not* taken any previous experiments in the series). While we see no reason why the data from cooperative participants in Experiment 2b should not be analyzed, the fact that only about 50% of all participants seem to meet this criterion limits the conclusions that can be drawn from this experiment.

6.5.2 Categorization responses

Next, we turn to participants categorization responses—i.e., whether participants heard *asi* or *ashi*. Figure 6 summarizes participants’ categorization responses. For this figure, we did not exclude the 17 participants that never or always pressed SPACE. Separate analyses—not summarized here—did, however, show that the results are unchanged when those participants are excluded. Compared to Experiment 2, participants in Experiment 2b exhibited noticeably weaker effects of the acoustic continuum (more shallow slope in Figure 6a). This suggests a trade-off introduced by the secondary task in Experiment 2b: either because participants focused more on the visual input, or because the categorization response was delivered *after* the response to the visual stimulus (SPACE press), participants’ responses were less affected by the acoustic continuum. Critically, however, Experiment 2b replicates all effects found in Experiment 2.

Table 35: Effects of pen location.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$
2b	Pen location Mouth -> fewer <i>ashi</i> -responses	0.16	0.27	-0.27	0.59	0.4	0.277
2b	Pen effect increases for more <i>ashi</i> -like acoustic input	-0.02	0.07	-0.12	0.08	1.8	0.648
2b	Pen effect increases for visually <i>ashi</i> -biased input	0.07	0.27	-0.38	0.52	0.7	0.405
2b	Pen effect increases even more when acoustic and visual input is <i>ashi</i> -biased	-0.06	0.07	-0.18	0.05	4.1	0.803

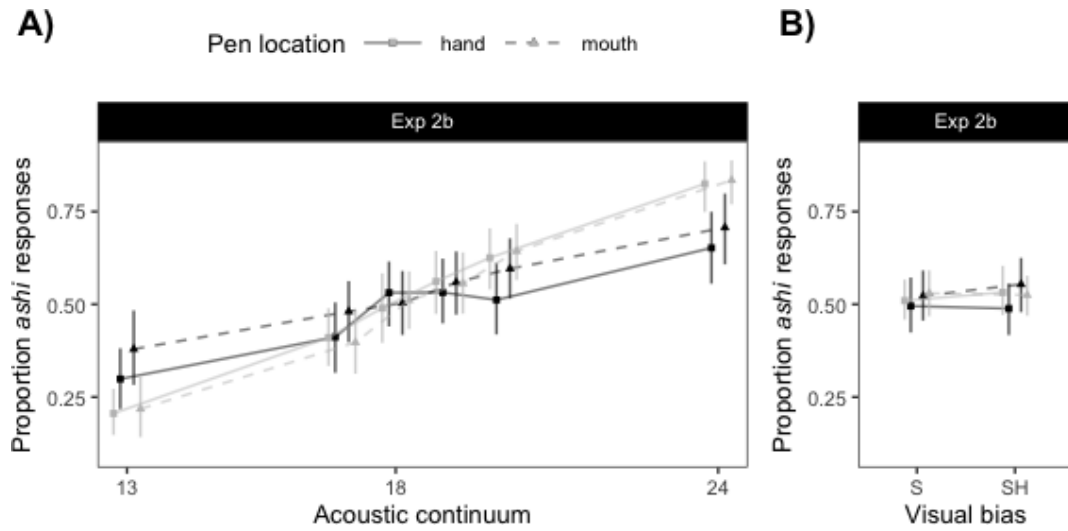


Figure 6: Summary of participants' responses in Experiments 2b, depending on pen location and acoustic continuum step (Panel A) or visual bias (Panel B). For comparison, the results from Experiment 2 are shown in the background. The two experiments were identical except for the secondary task introduced in Experiment 2b.

Table 34: Results of Bayesian mixed-effect logistic regression for Experiment 2b.

Covariate	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	-1.06	0.29	-1.63	-0.50
Condition.Test.OriginalLabelSH	0.10	0.27	-0.46	0.62
Condition.Test.PenM	0.21	0.32	-0.40	0.85
Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.22	0.32	-0.42	0.84
moBlock	0.02	0.18	-0.36	0.39
moCondition.Test.Audio	0.18	0.06	0.07	0.30
moBlock:Condition.Test.OriginalLabelSH	-0.04	0.09	-0.22	0.13
moBlock:Condition.Test.PenM	0.04	0.09	-0.14	0.22
moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.03	0.05	-0.07	0.12
moCondition.Test.Audio:Condition.Test.PenM	-0.02	0.07	-0.14	0.10
moBlock:moCondition.Test.Audio	0.00	0.03	-0.06	0.07
moBlock:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.03	0.19	-0.33	0.41
moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.07	-0.20	0.08
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH	0.02	0.02	-0.01	0.05
moBlock:moCondition.Test.Audio:Condition.Test.PenM	-0.01	0.01	-0.03	0.02
moBlock:moCondition.Test.Audio:Condition.Test.OriginalLabelSH:Condition.Test.PenM	0.00	0.03	-0.06	0.07

Table 36: Effects of acoustic continuum and visual bias.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{posterior}$	
2b	Acoustic continuum more <i>ashi</i> -like -> more <i>ashi</i> -responses	0.18	0.06	0.10	0.27	189.5	0.995	*
2b	Visual <i>ashi</i> -bias -> more <i>ashi</i> -responses	0.17	0.21	-0.18	0.50	3.9	0.795	
2b	Acoustic and visual bias effects are independent	0.03	0.05	-0.07	0.12	57.4	0.983	

Table 37: Changes across blocks.

Exp	Hypothesis	$\hat{\beta}$	SE	CI _l	CI _u	BF	$p_{\text{posterior}}$
2b	Pen effect is stable over blocks	0.02	0.10	-0.17	0.21	79.6	0.988
2b	Acoustic effect is stable over blocks	0.00	0.03	-0.06	0.07	100.1	0.990
2b	Visual bias effect is stable over blocks	0.01	0.09	-0.17	0.20	88.3	0.989

References

- Aust, Frederik, and Marius Barth. 2024. *papaja: Prepare Reproducible APA Journal Articles with R Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>.
- Bache, Stefan Milton, and Hadley Wickham. 2022. *Magrittr: A Forward-Pipe Operator for r*. <https://CRAN.R-project.org/package=magrittr>.
- Barth, Marius. 2023. *tinylabls: Lightweight Variable Labels*. <https://cran.r-project.org/package=tinylabls>.
- Bürkner, Paul-Christian. 2017. “brms: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software* 80 (1): 1–28. <https://doi.org/10.18637/jss.v080.i01>.
- . 2018. “Advanced Bayesian Multilevel Modeling with the R Package brms.” *The R Journal* 10 (1): 395–411. <https://doi.org/10.32614/RJ-2018-017>.
- . 2021. “Bayesian Item Response Modeling in R with brms and Stan.” *Journal of Statistical Software* 100 (5): 1–54. <https://doi.org/10.18637/jss.v100.i05>.
- Clark, Michael. 2024. *Lazerhawk: Miscellaneous Functions Mostly Inspired by Synthwave*. <https://github.com/m-clark/lazerhawk>.
- Eddelbuettel, Dirk, Romain Francois, JJ Allaire, Kevin Ushey, Qiang Kou, Nathan Russell, Iñaki Ucar, Doug Bates, and John Chambers. 2024. *Rcpp: Seamless r and c++ Integration*. <https://CRAN.R-project.org/package=Rcpp>.
- Fowler, Carol A. 2006. “Compensation for Coarticulation Reflects Gesture Perception, Not Spectral Contrast.” *Perception & Psychophysics* 68: 161–77.
- Gabry, Jonah, Rok Češnovar, Andrew Johnson, and Steve Bronder. 2024. *Cmdstanr: R Interface to 'CmdStan'*. <https://mc-stan.org/cmdstanr/>.
- Grolemund, Garrett, and Hadley Wickham. 2011. “Dates and Times Made Easy with lubridate.” *Journal of Statistical Software* 40 (3): 1–25. <https://www.jstatsoft.org/v40/i03/>.
- Liao, Yongfu. 2019. *Linguisticsdown: Easy Linguistics Document Writing with r Markdown*. <https://CRAN.R-project.org/package=linguisticsdown>.
- Liu, Linda, and T Florian Jaeger. 2018. “Inferring Causes During Speech Perception.” *Cognition* 174: 55–70.
- Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. <https://CRAN.R-project.org/package=tibble>.
- R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Stan Development Team. 2020. “StanHeaders: Headers for the R Interface to Stan.” <https://mc-stan.org/>.
- . 2024. “RStan: The R Interface to Stan.” <https://mc-stan.org/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2023a. *Forcats: Tools for Working with Categorical Variables (Factors)*. <https://CRAN.R-project.org/package=forcats>.
- . 2023b. *Stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, and Lionel Henry. 2023. *Purrr: Functional Programming Tools*. <https://CRAN.R-project.org/package=purrr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.

- Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>.
- Wilke, Claus O., and Brenton M. Wiernik. 2022. *Ggtext: Improved Text Rendering Support for 'Ggplot2'*. <https://CRAN.R-project.org/package=ggtext>.
- Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.