Auditory Processing Difficulties Influence on Perceptual Learning

Department of Psychological Science
Katherine M. Fiallo and Alexia Galati, Ph.D.

Abstract

Understanding spoken language is a complex process that comes to most people easily, yet past research demonstrates that some people face difficulties with auditory processing. Listeners integrate auditory, visual, and other cues to understand speech. However, people with auditory processing difficulties rely less on visual cues, such as lip-reading. We investigated how listeners' reported difficulties in perceiving auditory information interact with their use of visual cues. This was a replication and extension of a study by Kraljic, Samuel, and Brennan (2008). Participants viewed a speaker pronouncing some words in one of four conditions, depending on which phoneme was changed (?S or ?SH) and whether the speaker held a pen-in-hand (characteristic) or had a pen-in-mouth (incidental) while producing these words. We assessed each participant's phonemic

Katherine M. Fiallo

Katherine M. Fiallo is an undergraduate student at the University of North Carolina at Charlotte, graduating in Spring 2022. She is majoring in Psychology and Philosophy. Her interests include cognitive science, biological psychology, and early childhood development. She has conducted research in perceptual learning based speech perception. Katherine aspires to enter a clinical psychology graduate program next year and continuing to develop her research skills in psychological science.

Alexia Galati, Ph.D.

Dr. Alexia Galati is an Assistant Professor in Psychological Science at the University of North Carolina at Charlotte, where she directs the Cognition in Interaction Lab. Dr. Galati's research focuses on how language users keep track of their conversational partners' perspective, and how they adapt their linguistic and non-linguistic behavior to coordinate with one another. She uses a variety of methods, including behavioral experiments, discourse analysis, eye-tracking, mouse-tracking, perceptual paradigms, and virtual reality technology. Dr. Galati holds a B.A. (Honors) in Psychology from Stanford University, a M.A. in the Social Sciences from the University of Chicago, and a Ph.D. in Experimental Psychology from Stony Brook University. Before joining UNC Charlotte, she was a postdoctoral researcher at the University of Cyprus and a Marie Skłodowska-Curie fellow at UC Merced.

boundaries in a category-identification task, in which participants categorized ambiguous sounds as being more S-like or SH-like. We measured participants' perceptual learning—the restructuring of their phonemic boundaries—based on their categorization of these ambiguous sounds. Afterward, participants were surveyed on their language background and their auditory processing difficulties. We did not replicate the findings on the effect of visual context on perceptual learning from the original study by Kraljic et al. (2008). Contrary to our predictions, we also did not find an effect of individual differences in auditory processing on perceptual learning and no interaction with visual context. Given that the effect of visual context has been replicated before, our results are inconclusive. The effect of individual differences on perceptual learning should be explored further to gain insight into the underpinnings of perceptual learning and how to improve speech perception for those with auditory difficulties.

Keywords: speech perception, perceptual learning, auditory processing difficulties, phonetic adjustment

Auditory Processing Difficulties and Perceptual Learning

During speech perception, listeners actively interpret spoken speech by combining various cues with previous knowledge. Listeners map acoustic signals onto linguistic elements to understand other speakers (Diehl et al., 2004). Speech perception, though seemingly effortless is an extremely complex process, as it requires listeners to integrate their knowledge of linguistic categories and the variable input produced by speakers. This challenge of mapping speaker variability onto one's mental representation of linguistic categories is often known as the *lack of invariance* problem. The lack of invariance problem arises from the variability in the acoustic signal produced both within and between speakers (Magnuson et al., 2020). Even the same speaker has considerable variability in their speech due to contextual factors (e.g., their speaking rate, the noise in the environment), making flexibility in processing speech sounds essential. Comprehending speech, given these variations, is challenging and involves using several external cues to perceive words accurately (Branigan et al., 2000).

Factors reducing ambiguity during speech perception

Speech variability can be overcome through several means. Listeners adapt their boundaries of speech sounds based on experiences with other speakers. They use lexical, syntactic, and visual information from the speaker to create a representation of the speaker (Branigan et al., 2000). Given the evidence that listeners adjust to speech variations dynamically (Norris et al., 2003), the representations must be flexible. Goldinger's work (1996) suggests part of this flexibility comes from the integration of the episodic memory traces of past speakers to overcome speaker variability. People have an episodic memory for speakers that encompasses voice details and possibly other aspects of speech (Goldinger, 1996).

Visual context is also a factor that helps reduce ambiguity during speech perception. One well known example that demonstrates the role of visual context is the McGurk effect (McGurk & MacDonald, 1976). In McGurk & Macdonald's experiment, those who saw a visual stimulus that differed from the auditory stimulus, in terms of the speech sound presented, would respond incorrectly. For example, if participants heard "gaga" but saw speakers physically create "baba," they typically perceived "dada", integrating the two sources of information. This integration of information suggests that visual evidence about the speaker's articulation influences speech perception (McGurk & MacDonald, 1976).

There are several theories of speech perception, with some emphasizing the articulatory gestures involved in speech production. One such theory is Liberman's motor theory of speech perception (1967), which characterizes speech production as the result of causal links between phonemes, neuromotor commands, muscle contractions, vocal tract shapes, and acoustic signals. In this view, every phoneme corresponds to a muscle contraction. Additionally, in this view, speech perception operates as speech production does except "backwards," with each segment of language corresponding to a neural signal. People can decode acoustic signals by recalling the articulatory events seen (Liberman, 1967). This view can handle well how listeners perceive variable speech, given that perception relies on the articulatory gestures of speech rather than the speech signal.

Consistent with theories emphasizing the role of observed articulatory gestures on speech perception, other studies highlight the importance of visual information, particularly lip-reading, (Dodd et al., 2008; Woodhouse et al, 2009). Dodd and colleagues (2008) explored how incongruences with audiovisual stimuli can impact the speech perception of people with phonological processing disorders and speech difficulties. The researchers found that lip-reading and heard speech are combined into an articulatory code, regardless of one's own ability to produce the sound. Not only is the combination of audiovisual information important, but research suggests that the auditory cortex is activated simply by viewing a video of someone speaking without any sound (Woodhouse et al., 2009).

Visual context also shapes listeners' attributions about the speakers' auditory signals, allowing for more flexibility in perceptual adjustments. Kraljic and colleagues (2008) explored how the source of auditory signals—based on visual context—influenced listeners' flexibility in speech perception. The researchers measured participants' perceptual learning: the way people restructure their phonemic boundaries to better understand speakers' variation. The present study is a partial replication of the study by Kraljic and colleagues (2008) study, with a focus on their manipulation of audiovisual context.

Kraljic and colleagues (2008) tested perceptual learning by using words containing the phonemes [s] (s sound) and [ʃ] (sh sound). All [s] and [ʃ] words were recorded and had two forms, one with normal pronunciation and another with the [s] or [ʃ] sound shifted towards an ambiguous [~sʃ]. Participants completed a lexical-decision task (i.e., judging whether these tokens were words or nonwords) to expose them to the typical or atypical pronunciations. For the lexical decision task, participants were assigned to one of two conditions, the audiovisual or audio-only. We will only be discussing the audiovisual condition since it pertains to the proposed study. In that condition, participants viewed videos of the same female speaker pronouncing these tokens. The speaker either had a pen in her hand or her mouth. Participants viewed one of these two versions of the stimuli (i.e., a between-participants manipulation), allowing the researchers to examine whether listeners were sensitive to external attributions for varied speech. The *incidental group* always saw the pen-in-mouth visual on atypical tokens (with the [~sʃ] sound) and the c*haracteristic* group always saw the pen-in-hand visual on atypical tokens. All atypical tokens (those with [~sʃ] sound) were in the first half of the list. Kraljic and colleagues (2008) predicted listeners would experience perceptual learning in the characteristic group—when atypical tokens could be attributed to the speaker's idiolect (i.e., the speaker's particular way of speaking)—but not in the incidental group—when atypical tokens could be attributed to an external cause.

Consistent with these predictions, Kraljic et al. (2008) found that listeners only showed perceptual learning in the characteristic group. When exposed to a visual cue that would explain the mispronunciation, listeners appeared to assume that the variation was a result of the circumstance. Kraljic and colleagues concluded that listeners ignored incidental, atypical pronunciations (with the pen-in-mouth), possibly due to sorting the incidental causes separately from other tokens (produced without the pen). These findings challenge Liberman's motor theory, as listeners were able to map speakers' phonemes in a typical manner, even with obstructions to the articulatory gestures.

In a more recent study, Liu and Jaeger (2018) set out to explore an alternative explanation for the finding of Kraljic and colleagues (2008). The researchers suggested that when listeners are uncertain of the cause of unusual word pronunciation, they use causal inferences, thus potentially blocking their shifting of phonemic boundaries. Liu and Jaeger (2018) examined if boundary shifts are blocked under uncertainty and whether this uncertainty is maintained after viewing unambiguous stimuli (pen-in-hand). The study by Kraljic and colleagues (2008) only measured perceptual learning of atypical pronunciations when they were followed by typical pronunciations of filler words and nonwords. Liu and Jaeger (2018) argued that beliefs can be updated after exposure to audiovisual stimuli that suggest a causal characteristic (pen-in-hand) since listeners remain uncertain of the cause of atypical pronunciations. Their results supported their hypothesis: participants shifted their boundaries when the characteristic condition (pen-in-hand) was followed by the incidental condition (pen-in-mouth). Participants showed causal reasoning when faced with the characteristic (pen-in-hand) stimuli and maintained those inferences later when exposed to disambiguating evidence. People are aware that the atypical pronunciation may have an incidental cause, even when not explicitly seen. These results indicate that causal reasoning prevents the restructuring of phonemic boundaries to stop preemptively attributing atypical pronunciations as a characteristic of the speaker. Liu and Jaeger expanded on the original Kraljic et al. (2008) study, suggesting listeners maintain some perceptual evidence from past experience which can be used to adapt to disambiguating evidence.

The role of individual differences

In our study, we further probe the integration of visual information with auditory speech, while considering individual differences. Our study is a partial replication of Kraljic et al. (2008) study with an additional focus on individual differences as they relate to auditory processing. We included the audiovisual conditions from the study of Kraljic and colleagues (2008) (excluding their audio-only condition), given our focus on the visual components' effect on perceptual learning. Other departures from Kraljic et al (2008) were modeled after Liu and Jaeger's (2018) study: we administered the study online and reduced the number of stimuli participants experienced during the lexical-decision task. The online aspect was shown to produce perceptual learning in a previous web-based study (Kleinschmidt & Jaeger, 2012).

Additionally, using less stimuli has been successful in eliciting perceptual learning in past experiments (Kleinschmidt & Jaeger, 2011; Vroomen et al., 2007). In the present study, we utilize these techniques and add additonal factors to examine.

In addition to these changes to the Kraljic and colleagues' (2008) study, the present study focuses on the role of individual differences in speech perception. People experience speech perception in different ways. How people process auditory information is one of many factors that can influence speech perception. Central auditory processing is how listeners perceive auditory information; it involves the central auditory nervous system as well as other neurological processes to create auditory perceptions (American Speech-Language-Hearing Association, 2005). A growing body of research demonstrates that some people experience auditory processing issues, affecting the ways in which they perceive speech (Chermak, 1997; American Speech-Language-Hearing Association, 2005). One tool to identify individuals with auditory processing difficulties is the (modified) Amsterdam Inventory for Auditory Disability [(m)AIAD] (Meijer, 2003), a self-report questionnaire developed to determine auditory disabilities (Bamiou et al., 2015). Individual auditory processing differences are important to better understand how people overcome the lack of invariance problem when listening to spoken speech, especially as it relates to visual cues (e.g., lip-reading). Auditory processing differences provide insight into the effect of lip-reading as auditory processing difficulties are correlated with impaired lip-reading abilities (Dodd et al., 2008; Woodhouse et al., 2009). Since lip-reading aids speech perception of ambiguous pronunciations, those with auditory processing difficulties rely more on auditory information as visual cues can cause confusion (Bellis & Ferre, 1999). That reliance on auditory information suggests that those with more auditory processing difficulties may be more likely to attribute phonological ambiguity as a characteristic of the speaker, rather than incidental, regardless of the visual information presented. This guides the hypotheses of the present study.

Method

Study overview and predictions

We examined how individual differences in auditory processing affect perceptual learning of speech, replicating and extending the study by Kraljic and colleagues (2008). Following that study, we created four conditions, based on which phoneme was changed (?S signifying [s] or ?SH signifying [ʃ]) and on whether listeners viewed the pen-in-hand (characteristic) visual or the pen-in-mouth (incidental) version of these words. Participants were randomly assigned one of four conditions: ?S-characteristic, ?S-incidental, ?SH-characteristic, or ?SH-incidental (Appendix A). Participants experienced a set of 100 total stimuli. 50% of the stimuli were words, 20 of which contained the critical phonemes [s] or [ʃ]. The remainder of the stimuli were nonwords.

During the lexical-decision task, participants were exposed to audiovisual stimuli and were asked to indicate if the sound was a word or a nonword. After the lexical-decision task, participants completed a category-identification task in which they categorized six syllables into two categories, S and SH. Finally, participants answered questions regarding their experiences processing auditory information by completing the (modified) Amsterdam Inventory for Auditory Disability questionnaire. The responses from the (m)AIAD determined participants' individual auditory processing difficulties.

Based on previous work on utilizing lip-reading to distinguish ambiguous sounds and links between auditory processing difficulties and poor lip-reading abilities, we hypothesized that those with more auditory processing difficulties would be more likely to attribute phonological ambiguity as a characteristic of the speaker in both the incidental and characteristic conditions. Insofar as those with auditory rely less on lip-reading, we expected that they would be less likely to take the pen-in-mouth into account and experience more perceptual learning (shifting of sound categories) in both conditions. In contrast, we expected that those with fewer auditory processing difficulties would perform similar to the findings by Kraljic et al. (2008): only showing perceptual learning in the characteristic conditions. Therefore, as auditory processing difficulties increase, we expected that participants would show more perceptual learning in the incidental condition.

Participants

Fifty-four participants from UNC Charlotte were recruited through the subject pool of the Department of Psychological Science, managed by SONA systems, to participate for course credit. They were at least 18 years old and identified as English speakers. Participants were instructed to complete the experiment in a quiet room with headphones. Three participants who took a significantly longer time to complete the experiment (> 90 mins) were excluded. The average completion time for the remainder of the sample was 25.08 minutes. Similarly, we excluded four participants who scored lower than 75% accuracy on the lexical-decision task, as low scores may indicate inattentiveness, following Kraljic and Samuel (2005). We initially planned to exclude participants who did not use headphones. However, because a substantial proportion of participants did not use headphones (61.70%), we did not exclude these participants. We will return to these points in the discussion. Finally, we had planned to exclude any participants who failed to respond as instructed on an attention check question; however, no participant failed this attention check. Therefore, analyses were based on the data of 47 participants.

Among these 47 participants, there were 22 female, 24 male, and one nonbinary participant. Their mean age was 19.57 (*SD* = 2.00; range 18-28).

Procedure

Participants completed this experiment online, without supervision by the researchers. Participants were randomly assigned to one of four conditions: ?S-characteristic, ?SH-characteristic, ?S-incidental, and ?SH-incidental. Participants were not told that some words would have an ambiguous sound. To administer the experiment, we used the Psytoolkit software (Stoet, 2010; 2017).

Participants were first given instructions on how to complete the lexical-decision task. Instructions stated that participants would view 100 different audiovisual stimuli and would respond by pressing a button to categorize the sound as a “nonword” (F) or “word” (J).

After completing the lexical-decision task, participants viewed instructions for the category-identification task, which asked participants to categorize syllables as containing “S” or “SH.” Participants were informed that they would hear vowel-consonant-vowel syllables (e.g., “a[ʃ]i” and “a[s]i”) and were instructed to quickly press the button, “S” (F key) or “SH” (J key), that best corresponded to what they heard. Participants heard the same six syllables 10 times each.

After completing the category-identification task, participants answered some “Post Experiment Questions” regarding their experience completing the main tasks, such as if they wore headphones. This included an attention check question regarding the perceived gender identity of the speaker, man or woman (Liu & Jaeger, 2018).

Next, participants were presented with the questions from the (m)AIAD. They responded to 28 questions based on their experience, clicking on one of the four responses to each question (i.e., almost never (0), occasionally (1), frequently (2), almost always (3)).

Finally, participants responded to questions about their sociolinguistic background. Before completing the study, participants were debriefed about the purpose of the study—measuring how auditory processing difficulties influence the integration of visual information with audio stimuli.

Materials

Exposure: Lexical-Decision Task

Word and Nonword Stimuli. The original list of stimuli created by Kraljic and Samuel (2005) had 100 words and 100 nonwords, all spoken by a single speaker. There were 20 critical [s]-words (e.g., episode) and 20 critical [sh]-words (e.g., beneficial). For each critical word, a second audio stimulus was created, replacing the [s] or [ʃ] sound with an ambiguous [~sʃ] (a mixture of /s/ and /ʃ/ sounds). The remaining 60 tokens were filler-words, not containing either phoneme, [s] or [ʃ]. All audio was paired with visual stimuli of a woman pronouncing the words. Due to a loss of files from Kraljic et al. (2008), we used videos created for a replication by Babel (2016), who kindly provided them. Each video featured a female speaker with a pen, either in her mouth or in her hand. The original audio stimuli from Kraljic et al. (2008) were paired with a Babel (2016) video for all conditions.

Stimulus Lists. Unlike Kraljic et al. (2008), who used 2 lists of 120 stimuli, we created four lists of 100 stimuli each. In each list, half the items were words (n= 50), and the other half were nonwords (n=50). Of the 50 words, 20 were "critical words" and 30 were filler words. Filler words and nonwords were randomly selected from the original Kraljic et al. stimuli and used in all four lists (Appendix B). Additionally, half the items were paired with pen-in-hand videos, and the other half were paired with pen-in-mouth. Consistent with Kraljic et al. (2008), [s] and [ʃ] words were presented in the first half of the stimuli, randomly inserted amongst words and nonwords.

Category-Identification Task

Participants were asked to categorize six different syllables—all with consonances ranging on an S-SH continuum—into two categories, S or SH. The continuum had six points with steps from [s]-like to [ʃ]-like. All six syllables fit into each of the six points on the continuum, with the [s]-like stimulus being "asi" and the [ʃ]-like stimulus being "ashi" (Appendix C). Participants were instructed to categorize these syllables as containing an S or SH sound by quickly pressing a button ("S" (F key) or "SH" (J key)). Participants heard the six syllables ten times in a randomized order.

To measure perceptual learning, following Kraljic and colleagues (2008), we computed the absolute difference between "SH" responses and "S" responses in the category-identification task (|% "SH" responses - % "S" responses|).

(modified) Amsterdam Inventory for Auditory Disability and Handicap

The (m)AIAD (Meijer 2003) is a self-report questionnaire used to determine auditory disabilities. It contains 28 questions and a Likert response scale (see Appendix D), allowing participants to judge how often they experience specific auditory difficulties in their daily life. Questions concern five categories of auditory perception: distinction of sounds, auditory localization, intelligibility in noise, intelligibility in quiet, and detection of sound (Appendix E). Responses were on a Likert scale of 0-3 based on four response options: almost never (0), occasionally (1), frequently (2), almost always (3). Total scores were determined by adding the scores from all 28 questions for each participant. The higher the participant's score, the fewer auditory difficulties they are attributed to have. We presented the (m)AIAD as an online survey to participants.

Language Background and Experience Questionnaire

The Language Background and Experience questionnaire includes 44 questions (all taken from Cox & Goldrick's OSF files (2021)). The questionnaire involves several response methods, including multiple-choice, Likert scales, and open-ended responses. It includes questions regarding participants' fluency in the English language as well as any other language(s) the participant may know. There are also a few general demographic questions (e.g., gender and age) as well as more specific questions that relate to language, such as where the participant lived while learning the language and how often they speak the language. The questionnaire also includes questions regarding participants' hearing loss, speech or language impairments, or visual impairments–identical to those in the original questionnaire. We chose to include these questions from the original questionnaire to ensure that participants' perceptual learning differences are not due to these factors.

Questions Regarding Experience Completing the Experiment

Participants were asked eight questions regarding the main tasks, identical to those used in the Liu and Jaeger (2018) study. These questions ensured that participants were attentive during the study, did not experience technical issues, and completed the study in a quiet room with headphones.

Results

Analysis Plan

Our goal was to examine the effect of visual context, phonemic manipulation, and individual on the perceptual learning effect. Toward that end, our primary confirmatory analysis plan involved building a linear regression model with predictors for visual context (incidental: "pen-in-mouth;" characteristic: "pen-in-hand"), phonemic manipulation (S-word context, as in "episode" vs. SH-word context, as in "beneficial") (See Appendix B), and individual differences auditory processing (entered as the centered and scaled (m)AIAD score). The regression model also included the interaction between visual context and individual differences in auditory processing, since we had hypothesized the size of the perceptual effect could differ across auditory processing difficulties. The perceptual learning effect was defined as the absolute difference between “SH” and “S” responses) in the category-identification task for each participant. To evaluate the statistical significance of the effect of each predictor, we used p = .05 as the criterion level, as it is common in the social and behavioral sciences.

Below, we report these planned analyses, as well as an exploratory analysis we conducted to establish the effect of wearing headphones, after establishing through the “Questions Regarding Experience Completing the Experiment” that the majority of participants did not wear headphones as instructed. We also conducted a set of exploratory analyses aimed at confirming that participants were responding as expected in the category identification task, selecting “SH” more frequently when the stimuli were more SH-like: these analyses involved the step in the continuum of the stimulus as a predictor of the individual trial choices (SH or S).

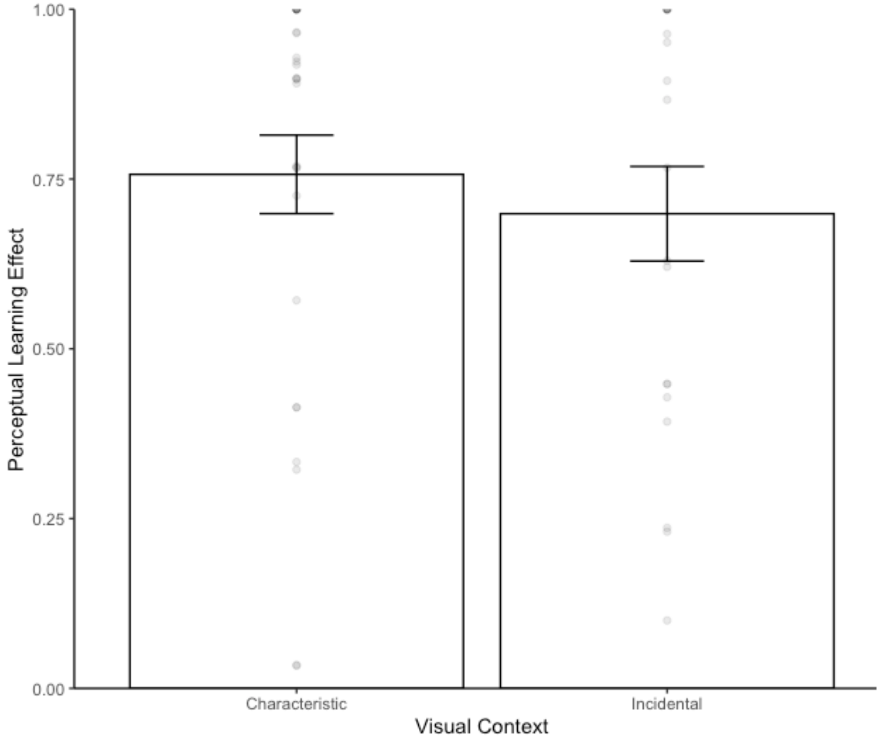
Finally, we provide some descriptive statistics about participants’ individual differences in auditory processing on the (m)AIAD questionnaire, and about participants’ language background, including their fluency in English and other languages spoken.

Category-Identification Task Results

Overall, the mean perceptual learning effect (absolute difference between “SH” and “S” response) in the category-identification task was .73 (*SD* = .30), which suggests a large difference in the proportions of “SH” and “S” responses. Participants identified the ambiguous stimuli as “SH” most of the time (the proportion of “SH” responses for the characteristic condition: *M*= .84, *SD* = .23; incidental: *M* = .85, *SD* = .16). In an open response question, many participants expressed that the stimuli in the category-identification task seemed more “SH” like. As Figure 1 shows, participants’ perceptual learning effect was numerically larger in the characteristic condition (pen-in-hand; *M* = .76, *SD*= .30) than the incidental (pen-in-mouth; *M* = .70, *SD* = .31) condition, consistent with Kraljic and colleagues (2008).

Figure 1

Perceptual Learning Effect Across the Two Conditions of (characteristic vs. incidental).



Note. Error bars represent the standard error of the mean and points represent the scores of individual participants.

However, this effect of visual context was not statistically significant (see results of the linear regression model in Table 1). The interaction between visual context and individual differences, which was of theoretical interest, was also not significant: there was no evidence that the perceptual learning effect depended on the participant's auditory processing differences. As illustrated in Table 1, none of these predictors had a significant effect on perceptual learning.

Table 1

Results of Linear Regression Model on the Perceptual Learning Effect

	B	*SE*	*t*	*p-value*
(Intercept)	0.69	0.09	7.84	<.0001
Visual context	-0.07	0.09	-0.72	0.48
Phonemic manipulation	0.11	0.10	1.08	0.29
(m)AIAD (scaled and centered)	0.003	0.06	0.05	0.96
Visual context * (m)AIAD	-0.05	0.10	-0.48	0.63

Note. Visual context and phonemic manipulation as categorical predictors, and (m)AIAD is a continuous predictor.

Exploratory Analysis: Examining the Effect of Headphone Use

Since many participants did not wear headphones (62%), we conducted a follow-up exploratory analysis examining if headphone usage interacted with perceptual learning. We added "headphones" as a categorical predictor to the linear regression model described above. This factor did not have a significant effect on perceptual learning (B = -.029, SE = .162, t= -.18, p = .86) and did not change the effects of the remaining factors in the model (p > .46). This suggests that headphone usage did not drive the null effect of visual context on perceptual learning.

Exploratory Analysis: Examining the Effect of Stimulus Ambiguity (Step in Continuum)

We also conducted exploratory analyses on phoneme choices ("SH" or "S") on individual trials as predicted by visual context, phonemic manipulation, and the stimulus step in the continuum. We built two mixed logistic regression models with S-choice (as shown in Table 2) or SH-choice (as shown in Table 3) as the binary dependent variable (with values 1 vs. 0 indicating the presence and absence of each choice). Visual context, phonemic manipulation, and step in the continuum were modeled as fixed effects and participants as a random effect. The random effect structure included random intercepts and random slopes for steps in the continuum (which was a within-participants factor)[01].

Table 2

"SH"-Choice Mixed-effects Model

	χ^2	*df*	*p-value*
(Intercept)	9.38	1	0.002
Visual Context	0.085	1	0.77
Phonemic Manipulation	0.27	1	0.60
Continuum level	22.51	5	<.0001

[01]The syntax for the linear mixed effect model in R for these models was: glmer(PhonemeChoice/~VisualContext + Phoneme + continuumlevel + (1 +continuumlevel | participantCatID), data=CatIDData, family = "binomial", control=glmer-Control(optimizer="nloptwrap", calc.derivs=FALSE))

Table 3

"S"-Choice mixed-effects model

	χ^2	*df*	*p-value*
(Intercept)	9.32	1	0.002
Visual Context	0.09	1	0.76
Phonemic Manipulation	0.28	1	0.60
Continuum level	22.65	5	<.0001

Visual context (for SH choices: $\chi^2(1) = .09$, $p = .77$; for S choices: χ^2 $(1) = .09$, $p = .76$) and phonemic manipulation (for SH choices: $\chi^2(1) = .27$, $p = .60$; for S choices: χ^2 $(1) = .28$, $p = .60$) were not significant predictors of phoneme choice. However, the continuum step was (for SH choices: $\chi^2(5) = 22.51$, $p < .001$; for S choices: $\chi^2(5) = 22.65$, $p < .001$]. As Figures 2 and 3 show, participants were more likely to respond "SH", which is consistent with the stimuli being more SH-like as the steps increased.

Figure 2

Proportion of "SH" Choices Across the Six Steps of the Continuum

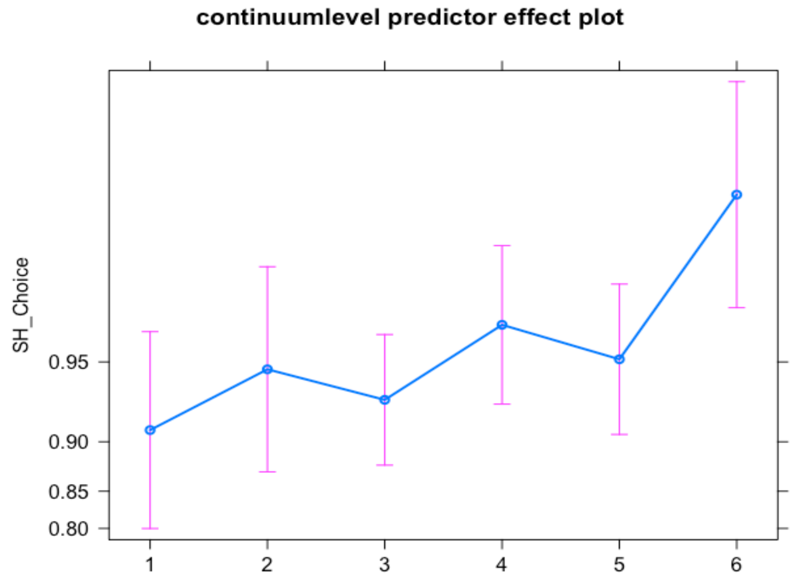
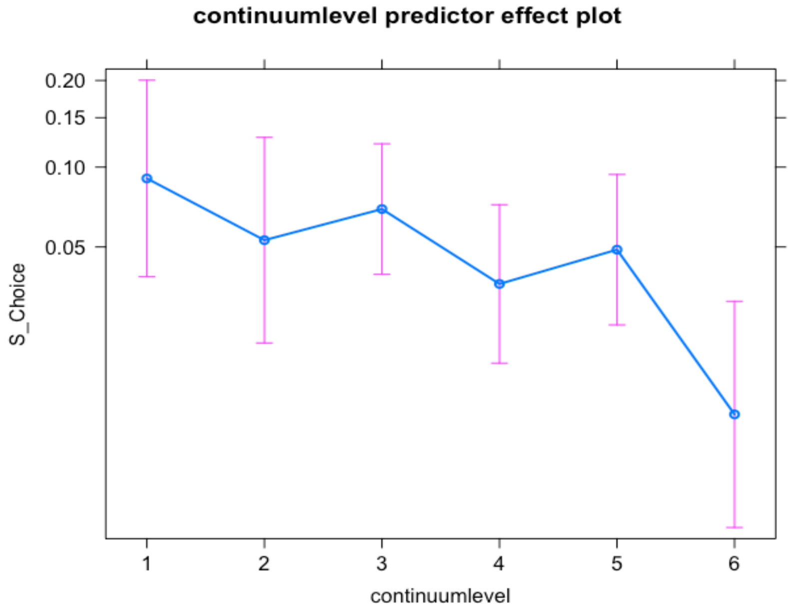


Figure 3

"S" Choices Across the Six Steps of the Continuum



Summary of Results of Category-Identification Task

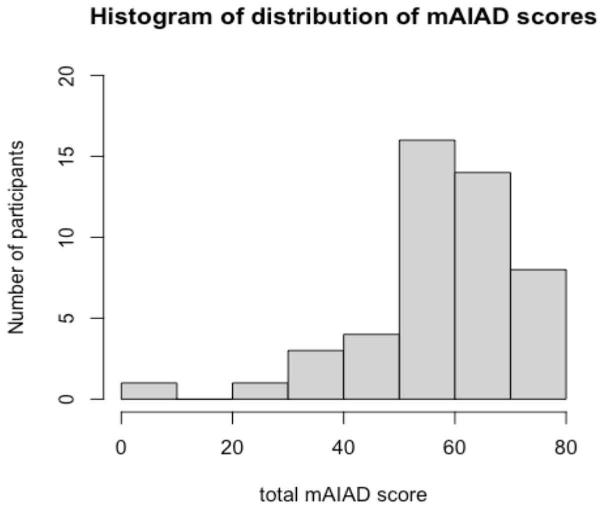
In contrast to Kraljic and colleagues (2008), we found no effect of visual context on perceptual learning, which was computed based on the aggregated proportions of "SH" and "S" responses. We also found no effect of visual context on the individual selections of either phoneme in more sensitive analyses that modeled participants as random effects. We did not find an effect of individual differences or interaction of individual differences with visual context.

Sample Descriptives: Individual Differences in Auditory Processing

In terms of individual differences in auditory processing, participants had a mean (m)AIAD score of 57.19 (*SD* = 13.47; range = 0-78, median = 58). Figure 3 presents the distribution of these scores. As shown, one participant scored 0, indicating extreme auditory processing difficulties (or else responding inauthentically). We did not exclude this participant from the analysis because they did not fail any of the other exclusion criteria we had specified in advance.

Figure 3

Histogram of (m)AIAD Scores in this Sample



Sample Descriptives: Language Background

Using responses to the Language Background questionnaire, we computed participants' mean fluency in English. This was based on four questions asking about ability in reading, writing, speaking, and listening (rated on a 0-10 scale). On average, participants scored 8.64 (*SD* = 1.35) across the four items. For listening, which is most relevant to this study: 8.81 (*SD* = 1.56).

The majority of participants first learned English in North Carolina (*N* = 29, 61.70%). 15 participants learned English in a state besides North Carolina, and the remaining three learned English outside of the USA (Australia, Jamaica, and Nigeria). Twenty-one participants knew a second language, and four of them a third language.

Discussion

In contrast to Kraljic and colleagues (2008), we found no effect of visual context on perceptual learning. Additionally, we found no effect of visual context on the individual phoneme selections during the category-identification task. We hypothesized that those with more auditory processing difficulties would demonstrate more perceptual learning in the incidental condition (pen-in-hand). Our findings did not support our hypothesis, as individual differences in auditory processing showed no effect on perceptual learning and did not interact with visual context.

Despite findings by Kraljic and colleagues (2008) and the successful replication of those findings by Liu and Jaeger (2018), we did not find a significant effect of visual context on perceptual learning. This is surprising as Liu and Jaeger (2018) replicated Kraljic and colleagues (2008) results successfully with an online experiment. The failure to replicate previous findings could be due to a small sample size or technical reasons, which we describe below.

As noted, our findings did not support our hypothesis that individual differences moderate perceptual learning. There was no significant difference between those who scored higher on the (m)AIAD and those who scored low, in that (m)AIAD was not a significant predictor of perceptual learning. This is unsurprising since the effect of visual context on perceptual learning was not detected. Another reason for these findings could be insufficient variation in individual differences in this sample since most participants scored high on the (m)AIAD.

Because the study was conducted fully online, it was difficult to control the environmental conditions under which participants completed the experiment. Despite previous research showing perceptual learning can still be produced via online studies (Kleinschmidt & Jaeger, 2012), we did not find an effect of visual context on perceptual learning. Participants may have been inattentive while completing the study, and as a result, may have missed the variation in how speakers produced phonemes during the lexical-decision task or missed the differences in "S" and "SH" sounds in the category-identification task.

Another limitation is that we had a small sample size—smaller than our target of 140. Our sample size of 47 may have been too small to detect an effect. Other limitations may have arisen from the data collection methods used on the (m)AIAD. Participants may have not been able to accurately assess themselves on the self-report questionnaires to determine auditory processing difficulties [(m)AIAD].

Despite these limitations, we were able to collect rich information about the participants' language background, which can be used by future researchers interested in examining the relationship between perceptual learning and sociolinguistic background.

Future research can use a larger sample size, which may allow for more robust and conclusive results. Other improvements include using a more controlled experimental setting to ensure participants' attentiveness. More attention checks can also be included to ensure participants' attentiveness. Future research can replace the self-reporting method of measuring auditory processing difficulties, so that participants' auditory processing difficulties are assessed through a task (e.g., having participants report what they hear following audio stimuli).

Another future direction could involve examining other sources of individual differences, beyond those concerning auditory processing. For example, dialectal differences among participants can be explored further. Since S and SH could be allophones in some words in different dialects (e.g., "street" and "shtreet"), exposure to such dialectal variation could impact how perceptual learning for these phonemes is affected by visual context (i.e., by attributions about the source of variability in the speaker). Future research can ask participants about their dialect or use a task that reveals information about dialectical differences.

Given that the effect of visual context has been replicated before, our results are inconclusive. We did not find evidence for such an effect, but as noted this study had some limitations. The role of individual differences in auditory processing on perceptual learning should be further explored, as this can lead to auditory disabilities accommodation development.

APPENDIX A

Table A1
Distribution of items in ?S-characteristic condition

	Pen-in-hand	Pen-in-mouth
Words	10 **[sh]**: **normal** version	
	10 **[s]**: **atypical** version	
	10 filler words	20 filler words
Non-words	20 nonwords	30 nonwords

Table A2
Distribution of items in ?S-incidental condition

	Pen-in-hand	Pen-in-mouth
Words	10 **[sh]**: **normal** version	
		10 **[s]**: **atypical** version
	15 filler words	15 filler words
Non-words	25 nonwords	25 nonwords

APPENDIX A

Table A3
Distribution of items in ?SH-characteristic condition

	Pen-in-hand	Pen-in-mouth
Words	10 **[s]**: **normal** version	
	10 **[sh]**: **atypical** version	
	10 filler words	20 filler words
Non-words	20 nonwords	30 nonwords

Table A4
Distribution of items in ?SH-incidental condition

	Pen-in-hand	Pen-in-mouth
Words	10 **[s]**: **normal** version	
	10 **[sh]**: **atypical** version	
	15 filler words	15 filler words
Non-words	25 nonwords	25 nonwords

APPENDIX B

Table B1

Filler Words/ Nonwords and Four Stimulus List

Filler Words	***Filler Nonwords***	perkum
negate	kradomet	emhoutic
lethal	Ithomel	bimobel
tutorial	lirthy	alnadiro
blueberry	pirugalo	aknid
keyboard	mowery	bikanian
continually	bimikay	ryligal
panic	niritaly	ibirak
marina	anolipa	marody
eighty	rakil	nowim
liability	rikmaral	admunker
lobbying	tilegkalo	rumatik
membrane	kermimer	bamtel
lingering	gondimually	loubel
ironic	hilder	kloumidiger
platonic	tamical	namuery
nightmare	gerbualo	rengimer
directory	bawaseet	aigi
inhabit	nomemtoly	lilgrai
pilgrim	itempider	durkuwomt
outnumber	hintarber	
laminate	kegimel	
burglary	pogunemd	
document	wonontic	
gullible	neramgory	
honeymoon	mikid	
hurdle	rawamtee	
worldly	onple	
turbulence	waiper	
melancholy	gairelom	
undermine	Indalier	

Note. 40 Filler Words and 60 Filler Nonwords (all used in each of the four lists)

APPENDIX B

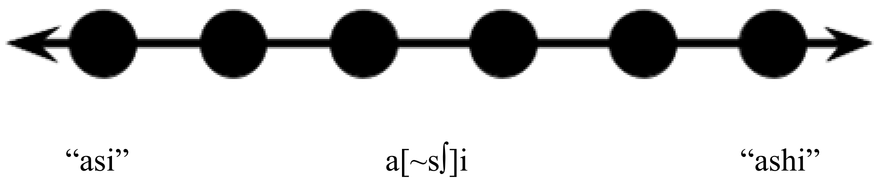
Table B2

Level of Phonemic Manipulation: Dark Grey Representing S-word Phonemic Manipulation and Light Gray Representing SH-word Phonemic Manipulation

?S-characteristic	?S-incidental	?SH-characteristic	?SH-incidental
10 [s]→[~sʃ] words	***10 [s]→[~sʃ] words***	***10 [ʃ]→[~sʃ] words***	***10 [ʃ]→[~sʃ] words***
literacy	episode	initial	ambition
medicine	arkansa	machinery	beneficial
obscene	coliseum	negotiate	brochure
parasite	compensate	official	commercial
peninsula	democracy	parachute	crucial
personal	dinosaur	pediatrician	efficient
pregnancy	embassy	publisher	flourishing
reconcile	eraser	reassure	glacier
rehearsal	hallucinate	refreshing	graduation
tennessee	legacy	vacation	impatient
10 [ʃ] words	***10 [ʃ] words***	***10 [s] words***	***10 [s] words***
initial	ambition	literacy	episode
machinery	beneficial	medicine	arkansa
negotiate	brochure	obscene	coliseum
official	commercial	parasite	compensate
parachute	crucial	peninsula	democracy
pediatrician	efficient	personal	dinosaur
publisher	flourishing	pregnancy	embassy
reassure	glacier	reconcile	eraser
refreshing	graduation	rehearsal	hallucinate
vacation	impatient	tennessee	legacy

APPENDIX C

“S” - “SH” Continuum



Each point represents one of the six steps, each having an accompanying audio stimulus. Stimuli closer to the left sound more like “asi” whereas those on the right sound more like “ashi”

APPENDIX D

(modified) Amsterdam Inventory For Auditory Disabilities

1. Can you understand a shop assistant in a crowded shop?
2. Can you carry on a conversation with someone in a quiet room?
3. Do you immediately hear from what direction a car is approaching when you are outside?
4. Can you hear cars passing by?
5. Do you recognize members of your family by their voices?
6. Can you recognize melodies in music or songs?
7. Can you carry on a conversation with someone during a crowded meeting?
8. Can you carry on a telephone conversation in a quiet room?
9. Can you hear from what corner of a lecture room someone is asking a question during a meeting?
10. Can you hear somebody approaching from behind?
11. Do you recognize a presenter on TV by his/her voice?
12. Can you understand the text that's being sung?
13. Can you easily carry on a conversation with somebody in a bus or car?
14. Can you understand the presenter of the news on TV?
15. Do you immediately look in the right direction when somebody calls you in the street?
16. Can you hear noises in the household, like running water, vacuuming, a washing machine?
17. Can you discriminate between the sound of a car and a bus?
19. Can you follow a conversation between a few people during dinner?
20. Can you understand the presenter of the news on the radio?
21. Can you hear from what corner of a room someone is talking to you being in a quiet house?
22. Can you hear the door-bell at home?
23. Can you distinguish between male and female voices?
24. Can you hear rhythm in music or songs?
25. Can you carry on a conversation with someone in a busy street?
26. Can you distinguish intonations and voice inflection in people's voices?
27. Do you hear from what direction a car horn is coming?
28. Do you hear birds singing outside?
29. Can you recognize and distinguish different musical instruments?

Excluded items:

18. Do you experience that music is too loud for you, while others around don't complain about the loudness?
30. Do you miss parts of music while listening to music or song?

REFERENCES

American Speech-Language-Hearing Association. (2005). (Central) auditory processing disorders.

Babel, M. (2016). Replication of Kraljic T, Samuel AG, Brennan SE (2008, PS 19(4)). Retrieved from osf.io/pj5hb

Bellis, T. J., & Ferre, J. M. (1999). Multidimensional approach to the differential diagnosis of central auditory processing disorders in children. Journal of the American academy of audiology, 10, 319-328. http://dx.doi.org/10.1055/s-0042-1748503

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. Cognition, 75(2), B13-B25. http://dx.doi.org/10.1016/S0010-0277(99)00081-5

Cox, L., & Goldrick, M. (2021, April 2). Adult sensitivity to phonetic detail in word learning. Retrieved from osf.io/7vxpd

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech perception. Annual review of psycholo-gy, 55, 149-179. http://dx.doi.org/10.1146/annurev.psych.55.090902.142028

Dodd, B., Mcintosh, B., Erdener, D., & Burnham, D. (2008). Perception of the auditory-visual illusion in speech perception by children with phonological disorders. Clinical Linguis-tics & Phonetics, 22(1), 69-82. http://dx.doi.org/10.1080/02699200701660100

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(5), 1166-1183. http://dx.doi.org/10.1037/0278-7393.22.5.1166

Kleinschmidt, D. F., & Jaeger, T. (2012). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. Proceed-ings of the Annual Meeting of the Cognitive Science Society, 34. Retrieved from https://escholarship.org/uc/item/9v45w71z

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? Cognitive Psychology, 51, 141-178.

REFERENCES

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts. Psy-cho-logical Science, 19(4), 332-338. http://dx.doi.org/10.1111/j.1467-9280.2008.02090.x.

Kramer, S. E., Kapteyn, T. S., Festen, J. M., & Tobi, H. (1995). Factors in subjective hearing disability. Audiology, 34(6), 311-320. http://dx.doi.org/10.3109/00206099509071921

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Percep-tion of the speech code. Psychological Review, 74(6), 431-461.

Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. Cognition, 174, 55-70. http://dx.doi.org/10.1016/j.cognition.2018.01.003.

Magnuson, J. S., Nusbaum, H. C., Akahane-Yamada, R., & Saltzman, D. (2020). Talker famili-ar-ity and the accommodation of talker variability. Attention, Perception, & Psychophys-ics, 83, 1842-1860. http://dx.doi.org/10.3758/s13414-020-02203-y

Meijer, A. G., Wit, H. P., TenVergert, E. M., Albers, F. W., & Muller Kobold, J. E. (2003). Re-liability and validity of the (modified) Amsterdam Inventory for Auditory Disability and Handicap. International Journal of Audiology, 42(4), 220—226. https://doi.org/10.3109/14992020309101317

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748. http://dx.doi.org/10.1038/264746a0

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. Cognitive Psy-chology, 47(2), 204-238. http://dx.doi.org/10.1016/s0010-0285(03)00006-9.

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. Behavior Research Methods, 42(4), 1096-1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. Teaching of Psychology, 44(1), 24-31. https://doi.org/10.1177/0098628316677643

REFERENCES

Woodhouse, L., Hickson, L., & Dodd, B. (2009). Review of visual speech perception by hearing and hearing-impaired people: Clinical implications. International Journal of Language & Communication Disorders, 44(3), 253-270. http://dx.doi.org/10.1080/13682820802090281

ISSUE 2

UNIVERSITY OF NORTH CAROLINA AT CHARLOTTE
UNDERGRADUATE RESEARCH JOURNAL

The University of North Carolina at Charlotte Undergraduate Reseach Journal was established through the Office of Undergraduate Research. The publication of the journal is a year-long process that is possible through the efforts of volunteer reviewers, contributing student authors, and editors.

UNCCURJ | ISSUE 2 | https://our.charlotte.edu

OFFICE OF UNDERGRADUATE RESEARCH UNC CHARLOTTE

The Office of Undergraduate Research
Atkins 237 (Area 49)
9201 University City Blvd. |Charlotte, NC 28223
undergradresearch@uncc.edu | (704) 687-5316