Wei Lai*, Lacey Wade and Meredith Tamminga

Individual differences in simultaneous perceptual compensation for coarticulatory and lexical cues

https://doi.org/10.1515/lingvan-2020-0040
Received July 24, 2020; accepted June 15, 2021; published online May 5, 2022

Abstract: Idiosyncratic perceptual compensation behaviors are considered to have a bearing on the perceptual foundation of sound change. We investigate how compensation processes driven by lexical and coarticulatory contexts simultaneously affect listeners' perception of a single segment and the individual differences in the compensation patterns. Sibilants on an /s-ʃ/ continuum were embedded into four lexical frames that differed in whether the lexical context favored /s/ or /ʃ/ perceptually and whether the vocalic context favored /s/ or not. Forty-two participants took a lexical decision task to decide whether each stimulus was a word or not. They also completed the autism-spectrum quotient questionnaire. The aggregate results of the lexical decision task show coexistence of lexically induced and coarticulatorily induced perceptual shifts in parallel. A negative correlation was found between the two kinds of perceptual shifts for individual listeners in lexical decisions, lending support to a potential trade-off between compensation magnitudes on different levels of cue integration. No correlation was found between the perceptual shifts of individuals and the results of the autism-spectrum quotient questionnaire.

Keywords: cue integration; individual difference; perceptual compensation; sibilant

1 Introduction

Speech signals are highly variable. The lack of invariant mapping between phonemes and their acoustic manifestations gives rise to a substantial amount of ambiguity in speech perception. To resolve this ambiguity, listeners may attend to multiple types of perceptual cues, meaning "any information that systematically influences listeners' perception of a contrast" (Schertz and Clare 2019: 2). While the most obvious of these are the cues in the acoustic signal that are primarily associated with a phonological contrast (such as voice onset time for stop voicing), listeners can also make use of cues that are external to the contrast-associated signals but are informative of the contrast's distribution in a language. For example, listeners may take into consideration the influence of contrast-external information such as lexical (Ganong 1980), phonotactic (Pitt 1998), and coarticulatory (Mann and Repp 1980) contexts, as well as cues to gestural (McGurk and MacDonald 1976) and social (Hay et al. 2006) contexts. The process where listeners make perceptual adjustments to account for the latter, external kind of cues in speech perception is called *perceptual compensation*. For example, a listener might use their knowledge of the English lexicon to infer that a sound ambiguous between /t/ and /d/ is more likely to be /d/ when embedded in *?ice* (so that it forms the real word *dice*) but more likely to be /t/ when embedded in *?ype* (so that it forms the real word *type*) (Ganong 1980).

Zellou (2017) summarizes the purposes of perceptual compensation as twofold: to predict the upcoming speech stream and improve processing efficiency (e.g., Fowler 1984), and to make linguistic decisions about what was said (e.g., Mann and Repp 1980). While these two functions of compensation are not mutually

***Corresponding author: Wei Lai**, Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, USA, E-mail: wei.lai@vanderbilt.edu
Lacey Wade and Meredith Tamminga, Department of Linguistics, University of Pennsylvania, Philadelphia, PA, USA, E-mail: laceya@sas.upenn.edu (L. Wade), tamminga@ling.upenn.edu (M. Tamminga)

exclusive, in this study, we will focus on the latter by examining to what extent the presence versus absence of a contextual cue changes listeners' phoneme categorization responses. In real-world listening, listeners have access to many different kinds of contextual information at once. In this paper, we investigate how listeners respond when two kinds of contextual information relevant to the identification of a single segment – lexical information, and coarticulatory information – are available simultaneously. The *type/dice* example just given is referred to as *lexical compensation* (stemming from the "Ganong effect", Ganong 1980) because it relies on lexical knowledge. *Coarticulatory compensation*, on the other hand, is when listeners make use of their knowledge about how certain phonological environments give rise to coarticulation that predictably alters the acoustic properties of segments. For example, Mann and Repp (1980) show that a segment ambiguous between /s/ and /ʃ/ is more likely to be perceived as /s/ when preceded /u/ than when preceded by /a/: listeners know that a following rounded vowel will lower a sibilant's frequency (making a true /s/ sounds more /ʃ/-like) and can therefore "factor out" this coarticulatory influence before identifying the segment.

While having multiple contextual cues may seem helpful for listeners who need to decode the variability in the speech stream in order to identify what was said, different contextual cues are not guaranteed to point the listener toward the same conclusion. This raises the question of how listeners integrate conflicting external cues. Even when separate cues bias the listener in the same direction, it is unclear whether that confers an advantage in identification, or is just a source of redundancy. Finally, it is not a given that all listeners resolve these kinds of cue integration problems in the same way, as previous studies have observed individual differences in perceptual compensation and in the integration of segment-intrinsic cues (Schertz and Clare 2019; Yu and Zellou 2019). Because perceptual compensation responses to different kinds of context have mostly been studied in isolation, we know little about integrating simultaneous external cues. What do listeners, on average, do when both lexical and coarticulatory cues are present simultaneously? And do individual listeners differ in whether they rely more on lexical or coarticulatory cues? We investigate these questions using a lexical decision experiment in which ambiguous fricatives are embedded in contexts that provide both lexical and coarticulatory cues to the ambiguous segment's identity, with the two cue types being consistent in some conditions and conflicting in others.

2 Background

2.1 Coexisting compensation processes for coarticulatory and lexical cues

Although coarticulatory and lexical cues coexist in real-world listening, coarticulatory compensation and lexical compensation are usually studied separately, with different stimuli and tasks. Segmental frames used in the same coarticulatory compensation study typically differ in the segment(s) adjacent to the critical phoneme but maintain equivalent lexical statuses that are independent of the perception of the target phoneme (for word frames, see Beddor 2009; Best et al. 1981; Fitch et al. 1980; for nonword frames, see Fowler 1984, 2006; Mann and Repp 1980; Yu and Lee 2014). In contrast, the segmental frames adopted in the same lexical compensation study usually have maximally identical segments near the target phoneme, while their lexical statuses vary with the perceived phoneme, and this variation works in opposite directions between frames (e.g., Ganong 1980; Stewart and Ota 2008). These designs appropriately isolate from confounding factors whichever compensation effect is under investigation, but they leave open the question of how different compensation mechanisms interact when multiple contextual factors are simultaneously perceptually relevant, as in real-world listening.

Moreover, different experimental paradigms may be more typically used to investigate one kind of compensation or another. There is some overlap in the use of phoneme identification (Ganong 1980; Mann and Repp 1980) and eye-tracking (Kingston et al. 2016; Mitterer et al. 2013). Other than that, lexical compensation has been often examined with lexical decision tasks (Stewart and Ota 2008), while coarticulatory compensation has been often examined with sound discrimination tasks (Stephens and Holt 2003; Yu and Lee 2014;

Zellou 2017). These factors make coarticulatory and lexical effects obtained in different studies not necessarily comparable.

One reason that the relationship between coarticulatory compensation and lexical compensation is interesting but complicated is that they may operate on different timescales and processing levels. It is generally thought that coarticulatory compensation comes into play when listeners parse blended acoustic streams into prelexical units. It is disputed whether the knowledge garnered from acoustic signals that listeners parse is gestural (Fowler 1986; Liberman and Mattingly 1985), featural (Gow 2003), or spectral (Diehl et al. 2004), and to what extent it is physiologically based (Liberman and Mattingly 1985) or language-specific (Beddor and Krakow 1999). Meanwhile, compensation for lexical cues depends on higher-order language-specific knowledge and is relevant up until the full lexical item has been accessed. It is intuitive to conceive of the process of phoneme identification (shaped by coarticulatorily compensation) as feeding forward onto lexical representations. However, speech perception models diverge in their assumptions about whether higher-level lexical information extends downward into prelexical perceptual mechanisms. An *autonomous* framework posits a unidirectional flow of activation where, for example, the lexical representations do not affect processing in the prelexical representations that feed up to them (e.g., Norris et al. 2000). In contrast, an *interactive* model conceives of bidirectional flows of information between different levels of activation nodes, which allows lexical information to reshape the perception of prelexical units (McClelland and Elman 1986). Although we will not resolve these debates here, the irresolution regarding how different compensation processes are integrated into a united architecture provides an impetus for more empirical investigations on perceptual compensation for multiple cues.

Experimental studies on the cooperation of compensation processes usually focus on coexisting audio-visual cues and lexical-semantic contexts (e.g., Barutchu et al. 2008; Brancazio 2004; Sams et al. 1998). Regarding the interplay between coarticulatory and lexical compensation, a series of experimental studies (Magnuson et al. 2003; McClelland et al. 2006; Samuel and Pitt 2003) have shown that lexical information about the identity of one speech sound could feed downward into perceptual mechanisms, triggering contrastive perception of neighboring speech sounds (but see McQueen et al. 2006, 2009 for alternative interpretations). For example, Magnuson et al. (2003) found that stimuli on a /t-k/ continuum were more likely to be perceived as /k/ following an ambiguous fricative in an /s/-biased word frame (Christma_ **c**apes) and /t/ following the same ambiguous fricative in an /ʃ/-biased word frame (fooli_ **t**apes), suggesting that the output of lexical compensation can be used as a source of coarticulatory compensation in the perception of a following phoneme. While these results have lent some support to the possibility that lexical-level information can penetrate prelexical categorization mechanisms, the interaction between processes is realized through chains of decisions about different words and segments that are partly dependent on one another. In this paper, we ask what happens if two compensation effects are simultaneously brought to bear on the *same* segment. In other words, how do listeners integrate coarticulatory and lexical cues that either align or conflict in signaling an intended phonemic category?

2.2 Individual differences in compensation for multiple contexts

Although cue integration in perception in the aggregate may reflect community perception norms, individuals may differ substantially in whether and how much they tend to rely on a specific cue. At least two lines of research have shed light on individual differences in perceptual compensation. One line of these studies reports evidence of "partial" compensation in the perception of coarticulatory vowel nasalization (Beddor and Krakow 1999; Fowler and Brown 2000) and anticipatory vowel-to-vowel coarticulation (Beddor et al. 2002). They found that some listeners attribute only some of the coarticulatory derivations – whereas others attribute all of them – to the source. The other line of studies examines individual compensation behaviors as a way of investigating the link between production and perception (e.g., Beddor and Krakow 1999; Yu 2019; Zellou 2017). These studies found that speakers who exhibit more coarticulation in production also compensate more for coarticulatory contexts in perception (but see Grosvald 2009; Kataoka 2011 for null effects). Still, the individual-differences studies on perceptual compensation mostly focus on a single coarticulatory cue. Little is known about the individual differences in perceptual compensation for *multiple* contextual cues.

Individual listeners' relative dependence on multiple cues in speech perception has been substantially examined by a separate line of studies on cue weighting (see Schertz and Clare 2019 for a review). Differing from contextual cues examined in perceptual compensation studies, cues examined in the cue weighting literature are acoustic manifestations intrinsic rather than extrinsic to the phonological contrast, such as F0 and voice onset time as cues to stop voicing (Shultz et al. 2012), spectral and temporal cues to vowel identity (Flege et al. 1997), and F2 locus and F3 locus as cues to glides (Ainsworth and Paliwal 1984), among others. The cue weighting literature features a trade-off relationship between individual listeners' reliance on different cues to the same contrast. In the perception of stop voicing, for example, listeners who attend more to the voice onset time are found to be affected less by F0 cues (Coetzee et al. 2018; Lee et al. 2013). Similarly, listeners whose tone perception varies substantially with phonation cues are less likely to attend to F0 and vowel formants (Kuang and Cui 2018). These results raise the possibility that there could be a similar trade-off relationship between perceptual compensation for different kinds of contextual cues at the individual level.

Although this trade-off hypothesis has not been investigated directly within the behavior of individual listeners, it has been suggested in previous work on perceptual compensation for external cues. Yu (2013) observed that across separate studies on coarticulatory compensation (Yu 2010), phonotactic compensation (Yu et al. 2011), and lexical compensation (Stewart and Ota 2008), the magnitude of perceptual compensation bears different correlations with listeners' Autism-Spectrum Quotient scores (AQ, Baron-Cohen et al. 2001), a self-reported measure of subclinical phenotypic similarity to autism spectrum conditions. In particular, AQ was found to correlate negatively with how much listeners compensate for lexical (Stewart and Ota 2008) and phonotactic (Yu et al. 2011) contexts; meanwhile, it correlates positively with how much listeners compensate for coarticulatory context and talker vocal track differences (Yu 2010). To reconcile these seemingly disparate results, Yu proposed that listeners who compensate more for phonotactic and lexical cues may compensate less for coarticulatory cues, which he suggested stems from individual differences in holistic versus non-holistic processing styles. Since a nonholistic processing style is characterized by heightened attention to detail and difficulty in higher-order processing, he hypothesizes that listeners with high AQ attend more to the acoustic variation of sibilants with coarticulatory contexts and vocal tracts but less to lexical and phonotactic contexts, and vice versa.

While this proposal was motivated by individual differences in processing style, the basic premise of a trade-off between compensation types has not itself been empirically established, given the lack of studies on the relationship between compensation for contextual cues that we have described. In particular, such a trade-off has not been investigated within the *same* group of listeners and on the same stimuli/segments. The second research question we investigate, therefore, is whether individual reliance on lexical cues is negatively correlated with reliance on coarticulatory cues, suggesting that individuals tend to rely primarily on one cue or the other.

3 The present experiment

The present experiment introduces coexisting perceptual biases induced by lexical and coarticulatory contexts that work either in conjunction or in opposite directions, with the goal of probing the perceptual processes involved in these behaviors. As we outlined in Sections 1 and 2, it is our goal to answer the following two questions: First, how do listeners integrate simultaneous lexical and coarticulatory cues that are conflicting or consistent? When the two cues align, are they independently useful in identifying the ambiguous signal, and how are multiple cues reconciled when they conflict? And second, is there a negative correlation between the magnitude of compensation for lexical and coarticulatory cues on an individual basis?

To answer these questions, we use a lexical decision task to examine perception of ambiguous sounds between /s/ and /ʃ/. The ambiguous fricatives are embedded into segmental frames with coexisting lexical and coarticulatory contexts that independently induce either /s/-favoring or /ʃ/-favoring perceptual biases. This design enables us to explore the perceptual behaviors of the same participants when lexical and coarticulatory cues are consistent or conflict, as well as to determine how listener behaviors vary in their dependence on different cues in a single task.

Experimental conditions Eight steps of ambiguous sibilants from an /s-ʃ/ continuum were each spliced into four auditory contextual frames, namely, /əˈ_ɛnd/ (as in *ascend*), /əˈ_eɪmd/ (as in *ashamed*), /əˈ_um/ (as in *assume*), and /əˈ_ʊɹ/ (as in *assure*). These four frames fit in a design of two lexical by two coarticulatory conditions, as displayed in Table 1. The two frames in the top row both have a rounded vowel (/u/ or /ʊ/) after the target sibilant. This provides an /s/-favoring coarticulatory context, as listeners are expected to compensate for the lowered sibilant frequency by allocating it to the lip-rounding of the following vowel and would therefore be more likely to perceive an /s/. In contrast, the other two frames in the bottom row are coarticulatorily unbiased. Regarding lexical conditions, the two frames in the left column are lexically /s/-favoring because positing an /s/ sound results in a real word for them (*assume*, *ascend*) while positing /ʃ/ results in a nonword. Similarly, the two frames in the right column are lexically /ʃ/-favoring since positing /ʃ/ would make a word (*assure*, *ashamed*) while positing /s/ would not.

Stimuli The stimuli were read by a female speaker with a US Midland dialect and were recorded in a sound-attenuated booth with a Yeti microphone at a sampling rate of 44,100 Hz. She was instructed to read all of the four frames with a falling pitch and with stress on the second syllable. Then, she was instructed to pronounce the nonword sequences /əˈseɪ/ and /əˈʃeɪ/ in a similar manner, which represents her [s] and [ʃ] pronunciations in a lexically and coarticulatorily unbiased context. The two unbiased sibilants were then cut out and blended together in Praat with different proportions to generate an eight-step sibilant continuum. They range from 15%[ʃ]85%[s] to 85%[ʃ]15%[s], with a difference of 10%[ʃ] increase and 10%[s] decrease between adjacent steps. Each of the eight synthesized sibilants were then spliced into each of the four contextual frames with their original sibilant cut out, resulting in 32 stimulus items. Finally, all of these stimuli were normalized to 70 dB.

Subjects Fifty-one subjects were recruited from the subject pool of the University of Pennsylvania to participate in exchange for course credit. The data from nine subjects were excluded on the basis of unexpected performance in the lexical decision task: namely, they identified 98–100% of the stimuli with the frame of *ascend* as nonwords in the lexical decision task. The responses of the remaining 42 participants were used for further analysis in this paper. They are 36 women and six men aged 18–22 years old.

Procedure All the subjects completed a lexical decision task following the instructions below: "For each sound you hear, please decide as quickly and accurately as you can whether the sound is a Nonword (a nonsense word in English) or a Word (a real English word). Press the *Z* key using your left index finger for Nonword. Press the *M* key using your right index finger for Word." The tasks was implemented in the experimental presentation software Ibex (Drummond 2017) using the PennController experiment toolkit (Zehr and Schwarz 2018). Each participant responded to eight sibilants varying along the /s-ʃ/ continuum in each of the four lexical frames with seven repetitions in a single block (8 steps × 4 frames × 7 repetitions = 224 items), with the order of these stimuli randomized for each participant.[1]

Table 1: Experimental design of two lexical by two coarticulatory conditions.

	lexically /s/-favoring	**lexically /ʃ/-favoring**
coarticulatorily /s/-favoring	/əˈ_um/ *assume*	/əˈ_ʊɹ/ *assure*
coarticulatorily unbiased	/əˈ_ɛnd/ *ascend*	/əˈ_əɪmd/ *ashamed*

1 The participants also completed a sound discrimination task and an AQ questionnaire in the same experimental session. However, post hoc identification of design issues in the discrimination experiment make it difficult to interpret, while a problem in response recording on the AQ questionnaire made it impossible to compare fairly with previous studies. Therefore, we focus exclusively on the lexical decision results in this paper.

Predictions We address our first research question by examining the difference in responses between different lexical and coarticulatory conditions in the aggregate. If participants rely on multiple cues simultaneously, as we predict, we would expect an enhancement in the perceptual bias when cues align and a diminution in the perceptual bias when cues contradict. In this experiment, we would expect to see more /s/-equivalent responses for *assume* than for any other category, because both coarticulatory and lexical cues bias the listener toward hearing /s/. Similarly, we would expect to see responses across the continuum for items with conflicting cues intermediate of those with only /s/-biasing or only /ʃ/-biasing cues, such that the word *assure* (lexically /ʃ/-favoring, coarticulatorily /s/-favoring), would have fewer /s/ responses than *ascend* (lexically /s/-favoring, coarticulatorily unbiased) and fewer /ʃ/ responses than *ashamed* (lexically /ʃ/-favoring, coarticulatorily unbiased). In predicting that participants will make use of multiple cues simultaneously, a further question arises as to whether multiple cues have an independent, additive effect, or whether they interact such that consistent cues provide more or less of an effect than the sum of the individual effects alone. This would be indicated by an interaction between coarticulatory and lexical frames in the statistical modeling. We predict that both cues contribute independently and that their effects are additive, such that the effect of consistent cues will be roughly the sum of the independent effects of lexical and coarticulatory cues.

We address our second research question by fitting a mixed-effects logistic model to obtain the estimates of lexical context and coarticulatory context for each participant, which allows us to evaluate whether the two factors are negatively correlated with one another for individual listeners. We predict that individual participants will exhibit a trade-off such that those who rely more on lexical cues rely less on coarticulatory cues, and vice versa. We therefore expect to see a negative correlation between lexical bias and coarticulatory bias for both tasks. Other possible patterns that would not support this prediction would be that some individuals tend to rely on both types of external cues more than others (i.e., a positive correlation between lexical and coarticulatory cue integration), or that there is no relationship between cue usage at the level of the individual.

4 Results

Analyses were conducted using the R statistical environment (R Core Team 2015). Mixed-effects logistic models were run using the lme4 package (Bates et al. 2015), and plots were created using ggplot2 (Wickham 2016). Data and analysis scripts are available at https://osf.io/kefn4.

Figure 1 shows the means and standard errors of /ʃ/-equivalent response rates (i.e., Word responses for *assure* and *ashamed*, and Nonword responses for *ascend* and *assume*) in the lexical decision task with different embedding frames. A mixed-effects logistic model was evaluated to predict /ʃ/-equivalent lexical decision responses, with LexicalContext (sum-coded; /s/-favoring: 1; /ʃ/-favoring: –1), CoarticulatoryContext (sum-coded; /s/-favoring: 1; unbiased: –1), Step of the mixed [ʃ] proportion (centered), and their three-way interaction as fixed-effect predictors, along with by-subject random slopes for Step*LexicalContext*CoarticulatoryContext. The model shows a significant Step effect ($\beta = 0.11$, $p < 0.001$), suggesting an expected higher /ʃ/-equivalent response rate as proportion of mixed [ʃ] increases, as shown in Figure 1. The model also shows a significant effect of CoarticulatoryContext ($\beta = -0.29$, $p < 0.001$), indicating that the unbiased coarticulatory contexts (indexed by the black lines) induce more /ʃ/-equivalent responses than the /s/-favoring contexts (indexed by the grey lines). The LexicalContext effect corresponds to the difference between the dashed lines and the solid lines in Figure 1 and is also significant ($\beta = -1.20$, $p < 0.001$). None of the interaction items are significant (LexicalContext*CoarticulatoryContext: $\beta = 0.01$, $p = 0.89$; Step*LexicalContext: $\beta = 0.005$, $p = 0.16$; Step*CoarticulatoryContext: $\beta = -0.005$, $p = 0.11$; Step*LexicalContext*CoarticulatoryContext: $\beta = 0.005$, $p = 0.09$). In other words, we do not find evidence that the coarticulatory bias is stronger in one lexical condition or the other, or that the number of /ʃ/-equivalent responses increases more rapidly with Step in one condition than in the others.

We then extracted the by-participant regression coefficients for LexicalContext and CoarticulatoryContext from the previous mixed-effects logistic model to investigate whether lexical and coarticulatory estimates negatively correlate with each other on an individual basis. The model fit of each participant was plotted in Figure 2, with lexical effect corresponding to the *x*-axis and coarticulatory effect corresponding to the *y*-axis.

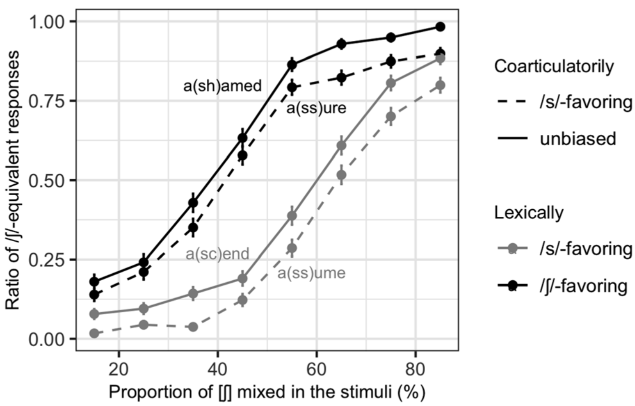

Figure 1: The means and standard errors of /ʃ/-equivalent responses in the lexical decision task.

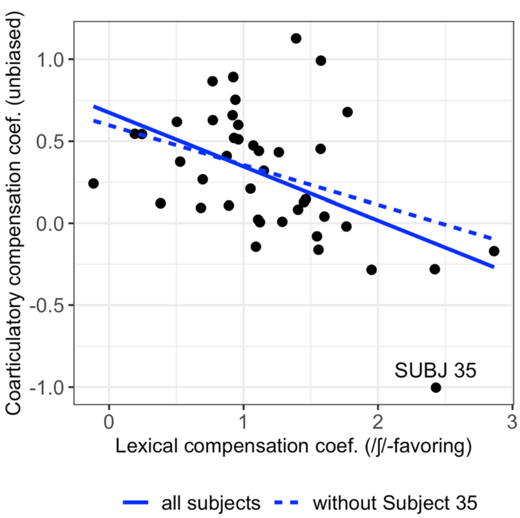

Figure 2: By-participant model estimates of lexical context and coarticulatory context in the lexical decision task. Each point in the graph represents a unique participant.

The model estimates of lexical context for individual participants vary from 0 to 3, suggesting that individuals almost always show more /ʃ/-equivalent responses in lexically /ʃ/-favoring frames. The estimates of the coarticulatory context vary roughly from –0.25 to 1, except for Subject 35 who has an estimation of –1 for coarticulatory compensation, which we cannot account for. This means that most participants show more /ʃ/-equivalent responses in front of unrounded vowels (/e ɛ/) than rounded vowels (/u ʊ/), and participants who do not obey this pattern do not seem to deviate much from a null effect (in other words, we think it is plausible that the small negative coefficients represent noise around a lack of coarticulatory compensation for those individuals). Moreover, there appears to be a negative correlation between individuals' coefficients of lexical compensation and those of coarticulatory compensation. A Pearson's correlation test reveals that this correlation is significant ($r = -0.49$, $t\ (40) = -3.53$, $p = 0.001$). This negative correlation still holds after the responses of Subject 35 are removed ($r = -0.39$, $t\ (39) = -2.66$, $p = 0.01$). This means that in the word identification task, participants who showed larger shifts in response to /ʃ/-favoring lexical contexts were inclined to show smaller shifts in response to unbiased coarticulatory contexts, and vice versa.

5 Discussion

In this experiment, we evaluated whether listeners simultaneously compensate for coarticulatory and lexical cues that exert contradictory or consistent influences on the identification of a single segment, and whether this compensation behavior varies across individual listeners.

The first research question we ask is how listeners, on average, integrate simultaneous lexical and coarticulatory cues that are either conflicting or consistent. Do lexical decision responses reflect the perceptual biases induced by both of the cues, or does a single cue provide sufficient information for listeners to rely on? If listeners make use of both cues, we would expect combined cues that exert the same directional bias to elicit a larger effect than either cue alone, and combined cues that exert opposite directional biases to elicit a smaller apparent effect than either of the single cues. This is consistent with the aggregate pattern we observed in Figure 1, which exhibit a clear two-by-two stratification in the likelihood of /ʃ/-equivalent responses depending on whether the lexical cue favors the perception of /s/ or /ʃ/ and whether the coarticulatory cue is /s/-favoring or unbiased. Specifically, an additive effect of consistent cues is exemplified by the fact that the word frame *assume* (lexically and coarticulatorily /s/-favoring) induces more /s/ responses than *ascend* (lexically /s/-favoring and coarticulatorily unbiased). Similarly, a diminishing effect of contradictory cues is evidenced by the fact that *assure* (lexically /ʃ/-favoring but coarticulatorily /s/-favoring) induces fewer /ʃ/ responses than *ashamed* (lexically /ʃ/-favoring and coarticulatorily unbiased) and fewer /s/ responses than *ascend* (lexically /s/-favoring and coarticulatorily unbiased). This result is not consistent with the alternative possibility that one cue is sufficient and other cues are simply a source of redundancy, which would otherwise predict no difference between *ascend* which only has one /s/-biasing cue and *assume*, which has two.

A further question arises as to whether the multiple influencing cues have an independent, additive effect, or whether they interact such that consistent cues provide more or less of an effect than the sum of the individual effects alone. As a response to this question, our statistical model shows an insignificant interaction between CoarticulatoryContext and LexicalContext ($\beta = 0.01$, $p = 0.89$), lending support to an additive rather than an interactive relationship between the two compensation effects. This finding is also roughly visually apparent in Figure 1, where the gap between *ashamed* and *assure* and that between *ascend* and *assume* have somewhat similar breadths, indicating that the magnitude of coarticulatory compensation is stable across different lexical conditions. Likewise, the difference between *ashamed* and *ascend* is similar to the difference between *assure* and *assume*, indicating that the magnitude of lexical compensation is stable across coarticulatory contexts. However, this finding is not consistent with similar studies evaluating a different set of contextual cues. In a perceptual learning study, Ullas et al. (2020) found that the learning effect caused by combined cues in conjunction is equivalent in size to that caused by audiovisual cues alone. However, with coexisting lexical and coarticulatory cues, we find an additive effect of different speech compensation mechanisms. It remains unclear whether the discrepancy should be attributed to the different natures of the tasks or the cues involved in the two studies. Future investigations are needed to answer this question.

Further, our study also raises a broader question as to how coarticulatory compensation fits into the architecture connecting integration mechanisms of contrast-intrinsic and contrast-extrinsic cues, and what the interplay between these mechanisms looks like at different levels of this architecture. The results found by considering the individual differences shed some light on the interplay between integration mechanisms. In particular we find a negative correlation between individual listeners' perceptual biases induced by lexical cues and by coarticulatory cues from their lexical decision responses. This provides evidence for a trade-off relationship at the level of the individual listener between coarticulatory compensation and lexical compensation, which parallels the trade-off between contrast-intrinsic cues established in the cue weighting literature.

The idea of a trade-off between compensation mechanisms was first proposed by Yu (2013) as an attempt to reconcile discrepancies between existing findings of different correlations between AQ and compensation processes at different levels. In this study, we evaluated individual compensatory trade-offs directly, instead of via their joint association with AQ as a proxy measure for processing style. With the same group of listeners and the same stimuli, we did find evidence for an individual-level trade-off between compensatory cue reliance, lending empirical support to one component of Yu's (2013) proposal. Although the cognitive parameters underpinning individual perception patterns remain unclear, we do see in this study that seemingly distinct perceptual mechanisms may be interrelated at the level of the individual listener. Our finding therefore contributes further to our previous knowledge of the reliability of individual phonetic behaviors across tasks (Yu and Lee 2014), occasions (Wade et al. 2021), and modalities (Arnon 2020). The idea that speech perception (and production) processes of the same individual may systematically correlate with one another is well known. However, the specific relationships between separate processes induced with different cues, tasks, and timescales have remain largely unexplored. Work on these many dimensions of individual differences holds promise for our understanding of the processes of speech perception.

Acknowledgments: This paper is based on work supported by the National Science Foundation under Grant Number 1627972, "Cognitive characteristics of the leaders of language change". The ideas in this paper have been improved by feedback from audiences at the 5th Workshop of Sound Change. Special thanks go to Pam Beddor and Alan Yu for their helpful feedback.
Research funding: This work was supported by the Division of Behavioral and Cognitive Sciences (grant number NSF #1627972).

References

Ainsworth, William & Kuldip Paliwal. 1984. Correlation between the production and perception of the English glides /w, r, l, j/. *Journal of Phonetics* 12(3). 237–243.
Arnon, Inbal. 2020. Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods* 52(1). 68–81.
Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin & Emma Clubley. 2001. The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males, females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31. 5–17.
Barutchu, Ayla, Sheila G. Crewther, Patricia Kiely, Melanie J. Murphy & David P. Crewther. 2008. When /b/ill with /g/ill becomes /d/ill: Evidence for a lexical effect in audiovisual speech perception. *European Journal of Cognitive Psychology* 20(1). 1–11.
Bates, Douglas, Martin Mächler, Bolker Ben & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
Beddor, Patrice Speeter. 2009. A coarticulatory path to sound change. *Language* 85. 785–821.
Beddor, Patrice Speeter, James D. Harnsberger & Stephanie Lindemann. 2002. Language-specific patterns of vowel-to-vowel coarticulation: Acoustic structures and their perceptual correlates. *Journal of Phonetics* 30(4). 591–627.
Beddor, Patrice Speeter & Rena Arens Krakow. 1999. Perception of coarticulatory nasalization by speakers of English and Thai: Evidence for partial compensation. *The Journal of the Acoustical Society of America* 106(5). 2868–2887.
Best, Catherine T., Barbara Morrongiello & Rick Robson. 1981. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics* 29(3). 191–211.
Brancazio, Lawrence. 2004. Lexical influences in audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance* 30(3). 445–463.
Coetzee, Andries W., Patrice Speeter Beddor, Kerby Shedden, Will Styler & Daan Wissing. 2018. Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics* 66. 185–216.
Diehl, Randy L., Andrew J. Lotto & Lori L. Holt. 2004. Speech perception. *Annual Review of Psychology* 55. 149–179.
Drummond, Alex. 2017. Ibex: Internet based experiments. https://github.com/addrummond/ ibexfarm (accessed 30 April 2020).
Fitch, Hollis L., Donna M. Erickson, Terry Halwes & Alvin M. Liberman. 1980. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics* 27(4). 343–350.
Flege, James Emil, Ocke-Schwen Bohn & Sunyoung Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics* 25(4). 437–470.
Fowler, Carol A. 1984. Segmentation of coarticulated speech in perception. *Perception & Psychophysics* 36(4). 359–368.
Fowler, Carol A. 1986. An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics* 14(1). 3–28.
Fowler, Carol A. 2006. Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics* 68(2). 161–177.
Fowler, Carol A. & Julie M. Brown. 2000. Perceptual parsing of acoustic consequences of velum lowering from information for vowels. *Perception & Psychophysics* 62(1). 21–32.
Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1). 110–125.
Gow, David W. 2003. Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics* 65(4). 575–590.

Grosvald, Michael. 2009. Interspeaker variation in the extent and perception of long-distance vowelto- vowel coarticulation. *Journal of Phonetics* 37(2). 173–188.

Hay, Jennifer, Paul Warren & Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4). 458–484.

Kataoka, Reiko. 2011. *Phonetic and cognitive bases of sound change*. Berkeley: University of California Dissertation.

Kingston, John, Joshua Levy, Amanda Rysling & Adrian Staub. 2016. Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology: Human Perception and Performance* 42(12). 1969–1988.

Kuang, Jianjing & Aletheia Cui. 2018. Relative cue weighting in production and perception of an ongoing sound change in Southern Yi. *Journal of Phonetics* 71. 194–214.

Lee, Hyunjung, Stephen Politzer-Ahles & Jongman Allard. 2013. Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics* 41(2). 117–132.

Liberman, Alvin M. & Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition* 21(1). 1–36.

Magnuson, James S., Bob McMurray, Michael K. Tanenhaus & Richard N. Aslin. 2003. Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science* 27(2). 285–298.

Mann, Virginia A. & Bruno H. Repp. 1980. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics* 28(3). 213–228.

McClelland, James L. & Jeffrey L. Elman. 1986. The trace model of speech perception. *Cognitive Psychology* 18(1). 1–86.

McClelland, James L., Daniel Mirman & Lori L. Holt. 2006. Are there interactive processes in speech perception? *Trends in Cognitive Sciences* 10(8). 363–369.

McGurk, Harry & John MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264(5588). 746–748.

McQueen, James M., Alexandra Jesse & Dennis Norris. 2009. No lexical-prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language* 61(1). 1–18.

McQueen, James M., Dennis Norris & Anne Cutler. 2006. Are there really interactive processes in speech perception? *Trends in Cognitive Sciences* 10(12). 533.

Mitterer, Holger, Sahyang Kim & Taehong Cho. 2013. Compensation for complete assimilation in speech perception: The case of Korean labial-to-velar assimilation. *Journal of Memory and Language* 69(1). 59–83.

Norris, Dennis, James M. McQueen & Anne Cutler. 2000. Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences* 23. 299–325.

Pitt, Mark A. 1998. Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics* 60(6). 941–951.

R Core Team. 2015. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sams, Mikko, Petri Manninen, Veikko Surakka, Pia Helin & Riitta Kättö. 1998. McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context. *Speech Communication* 26(1–2). 75–87.

Samuel, Arthur G. & Mark A. Pitt. 2003. Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory and Language* 48(2). 416–434.

Schertz, Jessamyn & Emily Clare. 2019. Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews: Cognitive Science* 11. e1521.

Shultz, Amanda A., Alexander L. Francis & Fernando Llanos. 2012. Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America* 132(2). EL95–EL101.

Stephens, Joseph D. W. & Lori L. Holt. 2003. Preceding phonetic context affects perception of nonspeech (L). *The Journal of the Acoustical Society of America* 114(6). 3036–3039.

Stewart, Mary E. & Mitsuhiko Ota. 2008. Lexical effects on speech perception in individuals with “autistic” traits. *Cognition* 109. 157–162.

Ullas, Shruti, Elia Formisano, Eisner Frank & Anne Cutler. 2020. Audiovisual and lexical cues do not additively enhance perceptual adaptation. *Psychonomic Bulletin & Review* 27(4). 707–715.

Wade, Lacey, Wei Lai & Tamminga Meredith. 2021. The reliability of individual differences in vot imitation. *Language and Speech* 64(3). 576–593.

Wickham, Hadley. 2016. *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. Available at: http://ggplot2.org.

Yu, Alan C. L. 2010. Perceptual compensation is correlated with individuals’ “autistic” traits: Implications for models of sound change. *PLoS One* 5(8). e11950.

Yu, Alan C. L. 2013. Individual differences in socio-cognitive processing and the actuation of sound change. In Alan C. L. Yu (ed.), *Origins of sound change: Approaches to phonologization*, chap. 10, 201–227. Oxford: Oxford University Press.

Yu, Alan C. L. 2019. On the nature of the perception-production link: Individual variability in English sibilant-vowel coarticulation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1). 1–29.

Yu, Alan C. L., Julian Grove, Martina Martinović & Sonderegger Morgan. 2011. Effects of working memory capacity and “autistic” traits on phonotactic effects in speech perception. *International Congress of the Phonetic Sciences* 17. 2236–2239.

Yu, Alan C. L. & Hyunjung Lee. 2014. The stability of perceptual compensation for coarticulation within and across individuals: A cross-validation study. *Journal of the Acoustical Society of America* 136(1). 382–388.
Yu, Alan C. L. & Georgia Zellou. 2019. Individual differences in language processing: Phonology. *Annual Review of Linguistics* 5. 131–150.
Zehr, Jeremy & Florian Schwarz. 2018. PennController for internet based experiments (Ibex). https://www.pcibdex.net/ (accessed 30 April 2020).
Zellou, Georgia. 2017. Individual differences in the production of nasal coarticulation and perceptual compensation. *Journal of Phonetics* 61. 13–29.