Individual differences in the use of top-down versus bottom-up cues to resolve phonetic ambiguity

Anne Marie Crinnion[1], **Christopher C. Heffner**[2], **Emily B. Myers**[1]

[1] University of Connecticut

[2] University at Buffalo

Abstract

How listeners weight a wide variety of information to interpret ambiguities in the speech signal is a question of interest in speech perception, particularly when understanding how listeners process speech in the context of phrases or sentences. Dominant views of cue use for language comprehension posit that listeners integrate multiple sources of information to interpret ambiguities in the speech signal. Here, we study how semantic context, sentence rate, and vowel length all influence identification of word-final stops. We find that while at the group level all sources of information appear to influence how listeners interpret ambiguities in speech, at the level of the individual listener, we observe systematic differences in cue reliance, such that some individual listeners favor certain cues (e.g., speech rate and vowel length) to the exclusion of others (e.g., semantic context). While listeners exhibit a range of cue preferences, across participants we find a negative relationship between individuals' weighting of semantic and acoustic-phonetic (sentence rate, vowel length) cues. Additionally, we find that these weightings are stable within individuals over period of one month. Taken as a whole, these findings suggest that theories of cue integration and speech processing may fail to capture the rich individual differences that exist between listeners, which could arise due to mechanistic differences between individuals in speech perception.

Introduction

Natural speech occurs in rich contextual settings, providing listeners with many sources of information that can aid in interpreting the acoustic signal. These cues include properties of the speech signal (e.g., vowel formants, voice onset time), as well as cues from semantic context, syntactic content, and even the rate of the signal itself (examples of the use of various cues to interpret the acoustic signal include: Borsky et al., 1998; Bosker et al., 2020;

Broderick et al., 2018; Bushong & Jaeger, 2019; Fox & Blumstein, 2016; Heffner et al., 2013; Jesse, 2021; Kaufeld et al., 2020; Morrill et al., 2015; Özyürek et al., 2007; Szostak & Pitt, 2013). Listeners are faced with the challenge of simultaneously integrating multiple cues to phonetic identity arising from distinct levels of the language hierarchy. Of interest is whether listeners tend to rely on different cues in a similar manner or whether some individuals tend to rely more on one cue versus the other. In order to test this question, it is necessary to assess individual differences in processing in a way that establishes the extent to which meaningful differences are reliable and stable within individuals over time. Importantly, assessing individual differences in cue reliance allows us to take an important step towards understand the question of how listeners resolve ambiguities in speech, by testing whether they may be not just one answer to the question but rather multiple.

Studies of contextual influences on speech perception typically take a group-level approach where the influence of one aspect of language processing (e.g., lexical knowledge) on another (e.g., acoustic-phonetic processing) is analyzed across a sample of participants (e.g., Ganong, 1980; Getz & Toscano, 2019; Gow et al., 2008). At a group level, most studies show that listeners are able to leverage context and acoustic detail in order to resolve phonetic ambiguity in the input. More recent work, however, suggests that *individual listeners* vary in the types of cues they tend to use to identify spoken words (Giovannone & Theodore, 2021a, 2023; Ishida et al., 2016; Kaufeld et al., 2020). For instance, Kaufeld and colleagues (2020) used a sentence listening task to examine how listeners relied on different types of contexts to interpret phonetic ambiguities in speech. More specifically, listeners heard sentences in Dutch with vowel-length minimal pair final items (e.g., *as* and *aas*). Sentences contained three different cues to the final word: (1) morphosyntactic gender markers (e.g., *de* vs. *het*), (2) sentence rate cues (fast vs. slow), and (3) vowel length along a 5-step continuum. Thus, each sentence contained three potentially conflicting cues to the identity of the final word. Data from the sentence listening task in the aggregate suggest that listeners make use of all available information, resulting in additive effects of morphosyntactic gender markers, rate, and vowel length on final word identification. However, Kaufeld and colleagues found that individual listeners tended to rely more on either "knowledge-based" information (morphosyntactic gender markers) or "signal-based" acoustic-phonetic information (rate and vowel length). These individual differences in response patterns were corroborated by simultaneously collected eye-tracking data. Individual differences in cue use likely did not reflect just a decision-threshold strategy carried out after the fact; eye tracking data reveals differences in looks as early as 200ms after the onset of the target word, where listeners who used acoustic-phonetic information looked more to the target item that was consistent with acoustic-phonetic cues and listeners who used morphosyntactic information looked more to the morphosyntactically valid target. These findings thus suggest that listeners vary in their cue weighting during online processing of speech.

What remains unknown is whether these tendencies to rely more on a certain cue reflect a consistent individual difference in speech processing. That is, do they reflect a preference for a cue in the moment, or a tendency to use that same cue consistently over time? Addressing this question requires multisession testing to assess the reliability of these observations. A parallel question was addressed in work from Giovannone and Theodore

(2023), where the stability of individual differences in the use of lexical and phonetic cues in single-word processing was tested over two sessions separated by at least two weeks. Participants performed three different tasks that demonstrate lexical effects on acoustic-phonetic perception. On each of the three tasks, listeners reliably differed in the degree to which they relied on lexical as opposed to phonetic information. Therefore, at least on those specific single-word listening tasks, listeners tend to reliably differ in the use of lower-level and higher-level sources of information.

In order to make claims about how listeners use different types of cues, researchers need to know whether a measure they are using to make those claims is one that will *reliably* demonstrate those individual differences (e.g. Blott et al., 2023; Hintz et al., 2020; Staub, 2021). It is tempting to assume that any behavioral paradigm that shows individual differences in performance is useful for, say, predicting an outcome measure. However, this assumption is not always an accurate one. When an individual difference metric is unreliable, this fact has real consequences; the reliability of a measure effectively acts as a cap on the extent to which it correlates with other metrics (Spearman, 1904) meaning that unreliable measures effectively endanger the likelihood that a study can give useful information. All told, it is clear that some minimal criteria need to be established to ensure that the paradigm is both reliable as well as valid (see Hedge et al., 2018). Reliability refers to the stability of responses over time or across items: showing that a measure is reliable suggests that the pattern is not entirely idiosyncratic to the testing session and shows some consistency within the individual. The validity of a measurement is more difficult to establish. It involves showing that the measure indeed taps the construct that it is intended to measure, often through comparison against an established standard, or using disparate materials.

As an instance of the importance of reliability in speech perception research, in prior work from our group (Heffner et al., 2022), we tested the reliability of a set of five behavioral paradigms designed to test aspects of speech perception, testing participants twice, and, when possible, with distinct sets of stimuli (i.e., different talkers, different phonetic contrasts). Some effects that are robust and replicable at a group level (e.g., lexically-guided perceptual learning for speech) have poor test—re-test reliability, whereas others (e.g., a measure of accent adaptation) show excellent reliability across sessions and materials. Some of these measures were then used with more confidence in a study of individual differences in speech learning and adaptation (Heffner & Myers, 2021). In the current study, we aimed to establish whether individual differences in cue preference were stable over time, in a context where multiple cues were available to resolve ambiguities in spoken sentences. This study provides a first step towards establishing the reliability of this measure.

While there thus has been work showing reliable individual differences in cue use at the single word level (e.g., Giovannone & Theodore, 2023), recent neural data suggests that processing of single words differs considerably from processing connected speech (and arguably, from processing speech at even just the sentence level; Gaston et al., 2023). While sentence listening tasks are not meant to be entirely representative of natural speech processing, they come closer than single word paradigms in helping us understand how

listeners process speech in everyday life. We thus sought to explore (a) how listeners make use of multiple, sometimes competing cues, (b) whether and how individual listeners vary in the types of cues they use to resolve ambiguities in speech, and (c) whether the cues individual listeners tend to rely on are stable over time.

To do this, we adapted the sentence listening task from Kaufeld and colleagues and tested listeners twice over the span of a month. Listeners heard sentences and then were asked to identify the final word in a two-alternative forced choice design. Those final words were always minimal pairs that differed in word-final stop voicing (e.g., "cup" vs. "cub"). We manipulated three different cues to the identity of these sentence-final minimal pairs, two that arise from details of the acoustic signal itself (sentence rate and vowel length), and one "knowledge-based cue" (semantic context). Of these cues, two can also be said to arise from the surrounding context (semantic information and speech rate information) whereas one is embedded in the acoustics of the target word itself (vowel length). The selection of these specific cues is in fact an oversimplification of the amount of information available to the listener to perceive speech. There are, for example, many different acoustic cues that are known to influence speech perception, some more than others (Crinnion et al., 2020; McMurray & Jongman, 2011). Based on previous work suggesting that listeners may differ in the dominance of cues at different levels of processing (e.g., Giovannone & Theodore, 2021a, 2023; Ishida et al., 2016; Kaufeld et al., 2020), we think it is valuable to pursue this question of cue use by simplifying the dimensions on which we test cue use.

The vowel length preceding a word-final stop serves as a strong cue to voicing (longer vowels are more associated with voiced final stops; Denes, 1955). Additionally, we manipulated the rate of the preceding sentence, as in the work of Kaufeld and colleagues (2020). As acoustic context effects, particularly those involving rate, are often defined in a variety of ways (see Stilp, 2020), we define contextual speech rate as the entirety of the sentence preceding the target word (and hence excluding the target word). The rate of preceding words impacts how listeners interpret later-occurring acoustic-phonetic information (Heffner et al., 2013, 2017; Morrill et al., 2015; Toscano & McMurray, 2015), such as vowel length. These effects work in a contrastive fashion. If the preceding speech rate is fast, listeners will perceive the final vowel as relatively long, meaning that they will be more likely to interpret the final stop as voiced. Conversely, if the preceding speech rate is slow, listeners will perceive the final vowel as relatively short, meaning that they will be more likely to interpret the final stop as voiceless. Finally, we modulated the semantic context of the sentences (e.g., *The lion/plate sat near the cu[?]*). Semantic context has been shown in numerous studies to bias acoustic-phonetic perception towards more-plausible parses (Borsky et al., 1998; Bushong & Jaeger, 2019; Getz & Toscano, 2019; Jesse, 2021). While we would predict, then, that the presence of sentence rate and semantic information may similarly influence how listeners interpret ambiguities in speech to the findings from Kaufeld and colleagues (2020), we aim to test this relationship by studying these different context effects both within and across participants over multiple sessions.

The use of sentence rate in addition to just the vowel length manipulation of the target word allowed us to examine global context integration at both the semantic and the acoustic-phonetic level. Both semantic and sentence rate cues occur prior to the target word, which

eliminated the possibility that listeners focus processing solely on the target word. The combination of the cues, therefore, set us up to examine how listeners use semantic vs. acoustic-phonetic cues at the sentence level and to ask whether and how individuals differed in cue weighting.

Methods

Participants

52 participants were recruited in order to reach a sample size of 50 usable participants (22 female, 26 male, 2 unknown; age range: 18–34, mean age: 29 years) on the first day of data collection. We selected this sample size because a post-hoc power analysis on data from Kaufeld and colleagues (2020) suggests that a sample size of 4 participants is necessary to detect the rate effects and syntactic effects they observe. However, because we were interested in individual differences, we aimed for 50 usable participants on the first day of testing, which we reasoned would give us a reasonable sample with complete data on two separate days (due to attrition). We ended with 34 usable subjects with complete data on two days (see below for further description). A power analysis done use the *pwr* (Champely et al., 2017) package in R (R Core team, 2022) demonstrated that with at least 28 participants, we were powered to detect correlations of 0.5 and higher (within and across cue weights at the level of the individual). Generally, correlation coefficients of greater than .7 or .8 are usually considered to be adequate levels of reliability for many applications (Hedge et al., 2018; Nunnally, 1978; Parsons et al., 2019). As such, it was not a priority to achieve power sufficient to detect correlations much weaker than .7; even if a correlation of, say, .4 were statistically significant, it would indicate that the measure is not a meaningful individual difference metric.

Participants were paid $6 for completing the task each time, a rate consistent with Connecticut minimum wage at the time of running. Only participants who were 18–34, native speakers of North American English, and who reported normal/corrected-to-normal vision and normal hearing were recruited for this study.

Participants' data were excluded if they failed a headphone check twice (Woods et al., 2017; N = 2). The 50 subjects with included data on day one were invited back one month later to complete the task again. 35 participants returned on day two (range of time in between sessions: 31–38 days), with one participant's data excluded due to failing the headphone check, leaving 34 usable subjects with data on two separate days of testing.

Materials

In the cue use task, participants heard a sentence and were asked to select the final word of the sentence that they heard from two options. The two words on the screen were always minimal pairs with word-final stops (e.g., *cub* and *cup*). The position of the voiced target word option on the screen (e.g., *cub*) was consistent within trials for a given participant but counterbalanced across participants. Each sentence contained three types of cues that varied: (1) semantic information, (2) sentence rate (all words in the sentence excluding the target word), and (3) vowel duration in the final target word. Semantic information

was either voiced-biasing (e.g., lion relates to *cub*) or voiceless-biasing (e.g., plate relates to *cup*). Sentence rate was either fast or slow. Fast sentence rates should lead to a longer-sounding vowel within the target word, thereby making them voiced-biasing. Meanwhile, slow sentence rates should lead to a shorter-sounding vowel within the target word, thereby making them voiceless-biasing (Heffner et al., 2017). Likewise, vowel duration was either voiced-biasing (i.e., longer), voiceless-biasing (i.e., slower), or ambiguous.

All sentences (see Appendix A for a complete list of stimuli) were recorded by a female native English speaker in semantically congruent, incongruent, and neutral contexts (e.g., the speaker produced *The lion sat near the cub*, *The plate sat near the cub*, and *The next word you will hear is cub*). Voicing continua (e.g., target words ranging from *cub* to *cup*) were synthesized from the voiced token recorded in the neutral context. To create the continuum, the stop release burst was removed from the sound and the duration of the immediately preceding vowel was manipulated using the PSOLA manipulation in Praat (Boersma & Weenink, 2022). Target items were piloted in isolation (i.e., removed from the sentence contexts) to identify three vowel-length steps: one where responses were most voiced (62.5% of the original duration), one where responses were most voiceless (37.5% of the original duration), and one where responses were around chance between the two tokens (50% of the original duration).

Sentence contexts into which each target item was added were taken from the semantically congruent production of the sentence. For example, we took the sentence context *The lion sat near the...* from the production *The lion sat near the cub* and took *The plate sat near the...* from the production *The plate sat near the cup*. Each of these contexts had all versions of the target item added, such that all three *cub-cup* target tokens appeared in a context such as *The lion sat near the...* and *The plate sat near the...* . Two versions of each sentence context were created by manipulating the rate in the same way as the vowel length: a fast rate version at 66% of the original production and a slow rate version at 150% of the original production. Each target item (for each vowel length) was then added to each sentence context (for each rate manipulation). There were 29 total sentence contexts, with 2 semantic conditions, 3 vowel durations, and 2 rate conditions, making for a total of 348 trials during which listeners made a response indicating the final word.

Procedures

The experiment was conducted online through Gorilla (Anwyl-Irvine et al., 2020), and participants were recruited through Prolific (www.prolific.com). After providing informed consent, participants completed a headphone check (Woods et al., 2017). They then completed the cue use task (see Figure 1 and Materials above).

Four weeks after completing the tasks above, participants who did not fail the headphone check were invited to complete the exact same process as above again: consent, headphone check, and the cue use task.

Analyses

We analyzed data for participants from whom we had two days of data. For each day of data collection, we fit a separate logistic mixed effects regression model predicting the proportion

of voiced responses. This analysis was performed using the *afex* package (Singmann et al., 2015) in *R* (R Core Team, 2022), which reports model results in ANOVA-like output. The models included fixed effects for semantic bias (coded voiced-biasing +1, voiceless-biasing −1), and sentence rate (voiced-biased (fast rate) and voiceless-biased (slow rate); coded in the same way as semantic bias), vowel length step (coded numeric: 1, 2, 3; scaled), and their interactions. Random slopes and intercepts for all above mentioned fixed effects were included for participants and for items, as this was the maximal model (Barr et al., 2013), and we were interested in examining individual differences from these random effects. Models were run separately for each day in order to estimate subject-specific estimates of cue use on each day. In order to estimate individual differences, by-participant random slopes for rate bias, semantic bias, and vowel length bias were extracted from each model. Pearson correlations were calculated to assess how cue use on day one of data collection related to cue use on day two (one month later) and to assess how use of different types of cues related to use of other cues. Data and code for all analyses are available at: https://osf.io/8wmf5/?view_only=bc9e6095b45a4912824b99dcc29fcd6f.

Results

Overall cue use patterns

Figure 2 shows overall patterns of cue use on Day 1 and Day 2, demonstrating more voiced responses after voiced-biasing semantics (red vs. blue lines), after voiced-biasing sentence rates (solid vs. dashed lines), and with more voiced-like vowel length steps (x-axis). Data from Day 1 and Day 2 samples were analyzed separately before being combined to look at stability in cue use over time.

For Day 1 (N=34), our mixed effects model revealed significant effects of semantic bias ($\chi^2 = 17.55$, p < .001), vowel length bias ($\chi^2 = 46.49$, p < .001), and sentence rate bias ($\chi^2 = 31.47$, p < .001). Additionally, we found a significant sentence rate by vowel length interaction ($\chi^2 = 12.12$, p < .001), with a bigger difference between fast and slow rates at more voiced-like steps.

For Day 2 (N=34), our mixed effects model revealed significant effects of semantic bias ($\chi^2 = 12.49$, p < .001), vowel length bias ($\chi^2 = 44.47$, p < .001), and sentence rate bias ($\chi^2 = 31.03$, p < .001). We again additionally found a significant sentence rate by vowel length interaction ($\chi^2 = 7.96$, p = .005). Overall, the same patterns of results were found on both days of testing.

Reliability

Individual use of each cue (vowel length, semantic context, and rate context) was estimated by extracting the by-participant random slopes from separate models run on Day 1 and Day 2 data. Figure 3a displays these cue-use estimates for Day 1 and Day 2 for a few representative participants. To visualize the stability of cue use over time, Figure 3b shows the relationship between individual participants' use of each cue (semantic bias, vowel length bias, and rate bias) on Day 1 plotted against the use of the same cue on Day 2.

We observed a significant correlation between Day 1 and Day 2 use of semantics (r = .73, t(32) = 6.104, p <.001), vowel length (r = .85, t(32) = 9.224, p <.001), and sentence rate (r = .79, t(32) = 7.393, p <.001). These relationships suggest that patterns of cue use are stable within individuals over time.

Individual patterns of cue use

Of interest was how cue use pattern within an individual. One hypothesis is that listeners might differ in how much they considered the preceding context (semantic context and rate context) compared to concentrating on local details of the target itself (vowel length). Another hypothesis is that listeners might instead differ in prioritizing top-down cues from semantics (the semantic context) compared to those that originate in details of the acoustic signal (rate context and vowel length). To assess whether individuals tended to use one cue over the other, we looked for the relationship between the use of different cues within each day. As shown in Figure 3c, we observed negative relationships between semantic context cues and both acoustic-phonetic cues (rate, vowel length). This suggests that individuals who tend to rely more heavily on cues from semantics downweight both cues originating from details of the acoustic-phonetic signal (rate and vowel length). Specifically, semantic bias was negatively correlated with vowel length bias (day 1: r = −.62, t(32) = −4.468, p <.001; day 2: r = −.58, t(32) = −4.005, p <.001) and with rate (day 1: r = −.71, t(32) = −5.681, p <.001; day 2: r = −.50, t(32) = −3.242, p =.003). In contrast, the two cues arising from acoustic-phonetic information tended to correlate with one another: vowel length bias was strongly and positively correlated with rate (day 1: r = .96, t(32) = 18.472, p <.001; day 2: r = .98, t(32) = 26.338, p <.001). Taken as a whole, these relationships suggest that individuals tend to either use top-down cues (semantics) or cues originating from the acoustic signal (rate, vowel length).

Discussion

Overall, we found that listeners as a group tend to use all available information to resolve ambiguities in speech. However, at the level of the individual, some listeners instead relied more on semantic information or on acoustic-phonetic information (speech rate and vowel length). Individual cue use profiles were consistent over time, with listeners similarly weighting acoustic-phonetic and semantic cues across two sessions at least one month apart. This consistency is noteworthy, suggesting that listeners' implicit or explicit strategies for weighting different cues in the input are consistent over time.

Individual patterns of semantic and acoustic-phonetic cue use mirror findings with other combinations of cues (e.g., morphosyntactic and acoustic-phonetic, lexical and acoustic-phonetic; Giovannone & Theodore, 2021b, 2023; Kaufeld et al., 2020). Our results extend previous findings in two critical ways. Like Giovannone and Theodore (2023), we find reliability in cue use over time, and like Kaufeld et al. (2020), we find differences in cue weighting at the level of sentence processing. However, the current study combines insights from both previous studies: we observe stability in cue use at the level of sentence processing. Taken together, these studies suggest that individual listeners tend to weight different sources of information differently. It should be noted that despite tendencies to

observe a trade-off relationship between context (semantic) and acoustic-phonetic cues, most listeners used some combination of both cues (i.e., very few listeners had zero effect of a specific source of information), and that the relative strength of these cues is remarkably stable over a month. These cues, then, when considered in combination, may prove to be potent individual difference measures that could be applied to examine correlations with other individual difference metrics.

Having established cue reliability at the sentence level (an arguably more naturalistic listening environment than single-word paradigms) and replicated the pattern of individual differences in work from Kaufeld and colleagues, we hope to shed new light on the issue of whether these individual differences reflect true profiles of individuals or task-based strategies. The eyetracking data from Kaufeld et al. (2020) suggests that these differences reflect differences in online processing, not simply differences in post-perceptual decision processes. Whether these processing differences are reflective of a listener's general profile or are induced by the task remains an open question. Giovannone and Theodore (2023) find little cross-task consistency in reliance on lexical and acoustic-phonetic cues, though they argue that the tasks they use may be tapping different constructs. Future work should explore how cue weighting varies (or shows stability) across tasks.

Of interest is which cue preferences tend to pattern together. Listeners might plausibly adopt a strategy in the current study to ignore context since the target word was always sentence final. If some listeners were "context users" and others "context ignorers", we might expect that some listeners would only use vowel length (and essentially not use any context information to bias interpretation). We observe, however, that the use of rate tends to pattern with the use of vowel length, suggesting distinct weightings of top-down (semantic) vs. bottom-up (acoustic-phonetic) information.

The finding that people who use speech rate cues also tend to use vowel length cues supports the idea that these two cues may not be independent, but rather that the way in which listeners use temporal cues like vowel length reflect compensation for contextual cues, like sentence rate (Toscano & McMurray, 2015). These cues can be viewed as asymmetrically contingent on one another. Under one contingency, a listener who is influenced by sentence rate uses this rate information to interpret the relative length of the vowel in the final word. Under this contingency, it is hard to imagine a listener using speech rate information while being insensitive to the vowel length of the target; the rate information is useful only with regard to the interpretation of the vowel duration. However, it is possible to imagine a world in which a listener uses vowel duration but not sentence rate. That is, a listener could conceivably treat the sentence listening task as a "target word listening" task by disregarding the sentence rate information and concentrating only on the length of the vowel in the target word. However, this pattern was not attested in our data, and would have been unlikely given that listeners have been shown to be sensitive to the rate of talkers they are not even attending to (Bosker, Sjerps, et al., 2020).

One potential limitation of this work is the fact that we only test a subset of cues (e.g., there are many acoustic cues that listeners may use). While we argue that more studies should explore distinct profiles of cue weightings at the level of the individual, the nature of many

of these types of experiments (including this one) necessitates a dimensionality reduction in the space of available information. Individuals may vary in their use of cues at a given level (e.g., Idemaru et al., 2012). However, we know that listeners as a whole tend to use secondary cues when primary cues are unavailable (e.g., Zhang et al., 2021). Thus, while this data tested only certain cues (and hence may have missed more fine grained nuanced in the cue use patterns of individual listeners), the fact that (a) we (and others) find strong influence of the cues we tested on average and (b) listeners reliably differ in their use of these cues highlights that this type of work provides motivation for characterizing individual patterns of cue use.

To better elucidate mechanisms of language processing, we need to understand how listeners use information across levels of language processing, by studying questions such as whether listeners who rely more on semantic cues also tend use syntactic or lexical cues more. Even within the set of cues we use here, looking at whether individuals' cue weightings hold across talkers and phonetic contrasts, for example, would provide some evidence to whether these findings were more representative of broader traits of individuals. Because we present a reliable task that measures reliance on acoustic-phonetic and semantic cues, we provide a strong first step in assess individual differences under other situations. While studying cross-task relationships may prove difficult due to uncertainty about underlying processes involved in a given task, it is important to assess whether stable effects on one task mirror those on another, or whether observed individual differences reflect specific (stable) strategies that participants adopt (Giovannone & Theodore, 2023; Heffner et al., 2022).

It is important to note that findings of aggregate additive effects of semantic context, sentence rate, and vowel length support the idea that higher-level language processing (e.g., lexical, semantic) can influence acoustic-phonetic perception. On the other hand, however, patterns of cue additivity at the level of the group do not necessarily arise because every listener aggregates information from all cues equally. Some listeners rely more heavily on cues from semantic context, whereas others favor cues from vowel length and speech rate. Furthermore, evidence for influence of context on neural encoding of low-level acoustic-phonetic information suggests that the effect of context on perception is not uniform: context may only influence encoding at perceptually more ambiguous acoustic-phonetic tokens (Sarrett et al., 2020). This nuanced pattern of influence across levels of language processing supports the idea that different types of information might be weighted differently depending on the language context (e.g., situations where a given cue is less determinate) but also depending on the individual (Martin, 2016). Theoretically, these findings challenge ideas about a 'normative' listener (McMurray et al., 2023) and highlight the need for theory building that can account for this data (e.g., see Rueckl, 2016 for discussion of this idea in terms of reading). Existing models of phonetic categories and speech perception more broadly, for example, may be able to account for this range of findings with parameters that could be adjusted for the individual. Or, it may be the case that findings about individual differences truly suggest different processing mechanisms (as opposed to just reweightings) that require distinct models.

With these data, we hope to highlight the importance of understanding participant-level effects in conjunction with presenting aggregate patterns. If group-level patterns of cue

additivity only arise because the sample is made up of equal numbers of participants weighting one cue over another and vice versa, then theories of cue integration and speech perception are severely underserved.

Funding

Appendix A

Minimal Pair	Voiced-Biasing Context	Voiceless-Biasing Context
bag-back	The hotel doorman dealt with my...	The chiropractor dealt with my...
bead-beet	The vegetable farmer loved to talk about her favorite type of...	The jewelry maker loved to talk about her favorite type of...
bed-bet	The kind housekeeper made a very large...	The risky gambler made a very large...
bleed-bleat	The wound began to...	The goat began to...
bug-buck	I checked my program and found another...	I checked my wallet and found another...
cab-cap	The traveler tried to find a...	The ball player tried to find a...
card-cart	The poker star picked out a...	The food shopper picked out a...
code-coat	The programmer wanted to show me her new...	The snowboarder wanted to show me her new...
cob-cop	I spotted the corn on the...	I spotted the badge on the...
cub-cup	The lion sat near the...	The plate sat near the...
fad-fat	The hipster wanted to keep the...	The sous chef wanted to keep the...
grade-grate	His exam scores went into the...	His wastewater went into the...
grand-grant	The pianist got another...	The scientist got another...
hog-hawk	I looked in the mud to see the...	I looked in the tree to see the...
lab-lap	The scientist sat in her...	The baby boy sat in her...
league-leak	The bowler dropped out of the...	The water dropped out of the...
maid-mate	The rich socialite called out to his...	The Australian called out to his...
pad-pat	The assistant gave him a...	The caretaker gave him a...
pig-pick	The old farmer lost his...	The guitarist lost his...
pod-pot	The pea is in the...	The soil is in the...
pub-pup	The bar owner walked over to her...	The dog owner walked over to her...
rag-rack	The dirt was on the...	The shoe was on the...
seed-seat	The gardener took a...	The viewer took a...
side-site	Her mortal enemy is on the bad...	Her naughty teenager is on the bad...
tag-tack	The expensive scarf had a large...	The bulletin board had a large...
toad-tote	In the pond, I saw a cute...	In the mall, I saw a cute...

Minimal Pair	Voiced-Biasing Context	Voiceless-Biasing Context
trade-trait	John's clever swap was his best...	John's clever wit was his best...
ward-wart	The hospital had a...	The ugly frog had a...
wig-wick	The bald actor had a...	The big candle had a...

References

Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, & Evershed JK (2020). Gorilla in our midst: An online behavioral experiment builder. Behavior Research Methods, 52(1), 388–407. [PubMed: 31016684]

Barr DJ, Levy R, Scheepers C, & Tily HJ (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278. 10.1016/j.jml.2012.11.001

Blott LM, Gowenlock AE, Kievit R, Nation K, & Rodd JM (2023). Studying individual differences in language comprehension: The challenges of item-level variability and well-matched control conditions. Journal of Cognition, 6(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10487189/

Boersma Paul & Weenink David. (2022). Praat: Doing phonetics by computer [Computer program]. http://www.praat.org/

Borsky S, Tuller B, & Shapiro LP (1998). "How to milk a coat:" The effects of semantic and acoustic information on phoneme categorization. The Journal of the Acoustical Society of America, 103(5), 2670–2676. 10.1121/1.422787 [PubMed: 9604360]

Bosker HR, Peeters D, & Holler J (2020). How visual cues to speech rate influence speech perception. Quarterly Journal of Experimental Psychology, 73(10), 1523–1536. 10.1177/1747021820914564

Bosker HR, Sjerps MJ, & Reinisch E (2020). Temporal contrast effects in human speech perception are immune to selective attention. Scientific Reports, 10(1), Article 1. 10.1038/s41598-020-62613-8

Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, & Lalor EC (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. Current Biology, 28(5), 803–809.e3. 10.1016/j.cub.2018.01.080 [PubMed: 29478856]

Bushong W, & Jaeger TF (2019). Dynamic re-weighting of acoustic and contextual cues in spoken word recognition. The Journal of the Acoustical Society of America, 146(2), EL135–EL140. [PubMed: 31472578]

Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A, Ford C, Volcic R, & De Rosario H (2017). pwr: Basic functions for power analysis.

Crinnion AM, Malmskog B, & Toscano JC (2020). A graph-theoretic approach to identifying acoustic cues for speech sound categorization. Psychonomic Bulletin & Review. 10.3758/s13423-020-01748-1

Denes P (1955). Effect of Duration on the Perception of Voicing. The Journal of the Acoustical Society of America, 27(4), 761–764. 10.1121/1.1908020

Fox NP, & Blumstein SE (2016). Top-down effects of syntactic sentential context on phonetic processing. Journal of Experimental Psychology: Human Perception and Performance, 42(5), 730–741. 10.1037/a0039965 [PubMed: 26689310]

Ganong WF (1980). Phonetic categorization in auditory word perception. Journal of Experimental Psychology: Human Perception and Performance, 6(1), 110–125. 10.1037/0096-1523.6.1.110 [PubMed: 6444985]

Gaston P, Brodbeck C, Phillips C, & Lau E (2023). Auditory Word Comprehension Is Less Incremental in Isolated Words. Neurobiology of Language, 4(1), 29–52. 10.1162/nol_a_00084 [PubMed: 37229141]

Getz LM, & Toscano JC (2019). Electrophysiological Evidence for Top-Down Lexical Influences on Early Speech Perception. Psychological Science, 30(6), 830–841. 10.1177/0956797619841813 [PubMed: 31018103]

Giovannone N, & Theodore RM (2021a). Individual Differences in Lexical Contributions to Speech Perception. Journal of Speech, Language, and Hearing Research : JSLHR, 64(3), 707–724. 10.1044/2020_JSLHR-20-00283 [PubMed: 33606960]

Giovannone N, & Theodore RM (2021b). Individual Differences in the Use of Acoustic-Phonetic Versus Lexical Cues for Speech Perception. Frontiers in Communication, 6, 120.

Giovannone N, & Theodore RM (2023). Do individual differences in lexical reliance reflect states or traits? Cognition, 232, 105320. 10.1016/j.cognition.2022.105320 [PubMed: 36442381]

Gow DW, Segawa JA, Ahlfors SP, & Lin F-H (2008). Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. NeuroImage, 43(3), 614–623. 10.1016/j.neuroimage.2008.07.027 [PubMed: 18703146]

Hedge C, Powell G, & Sumner P (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. Behavior Research Methods, 50, 1166–1186. [PubMed: 28726177]

Heffner CC, Dilley LC, McAuley JD, & Pitt MA (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. Language and Cognitive Processes, 28(9), 1275–1302. 10.1080/01690965.2012.672229

Heffner CC, Fuhrmeister P, Luthra S, Mechtenberg H, Saltzman D, & Myers EB (2022). Reliability and validity for perceptual flexibility in speech. Brain and Language, 226, 105070. 10.1016/j.bandl.2021.105070 [PubMed: 35026449]

Heffner CC, & Myers EB (2021). Individual Differences in Phonetic Plasticity Across Native and Nonnative Contexts. Journal of Speech, Language, and Hearing Research, 64(10), 3720–3733. 10.1044/2021_JSLHR-21-00004

Heffner CC, Newman RS, & Idsardi WJ (2017). Support for context effects on segmentation and segments depends on the context. Attention, Perception, & Psychophysics, 79(3), 964–988. 10.3758/s13414-016-1274-5

Hintz F, Dijkhuis M, van 't Hoff V, McQueen JM, & Meyer AS (2020). A behavioural dataset for studying individual differences in language skills. Scientific Data, 7(1), 429. [PubMed: 33293542]

Idemaru K, Holt LL, & Seltman H (2012). Individual differences in cue weights are stable across time: The case of Japanese stop lengths. The Journal of the Acoustical Society of America, 132(6), 3950–3964. 10.1121/1.4765076 [PubMed: 23231125]

Ishida M, Samuel AG, & Arai T (2016). Some people are "more lexical" than others. Cognition, 151, 68–75. [PubMed: 26986746]

Jesse A (2021). Sentence context guides phonetic retuning to speaker idiosyncrasies. Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(1), 184–194. 10.1037/xlm0000805 [PubMed: 31855000]

Kaufeld G, Ravenschlag A, Meyer AS, Martin AE, & Bosker HR (2020). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 46(3), 549–562. 10.1037/xlm0000744 [PubMed: 31343252]

Martin AE (2016). Language Processing as Cue Integration: Grounding the Psychology of Language in Perception and Neurophysiology. Frontiers in Psychology, 7. 10.3389/fpsyg.2016.00120

McMurray B, Baxelbaum KS, Colby S, & Tomblin JB (2023). Understanding language processing in variable populations on their own terms: Towards a functionalist psycholinguistics of individual differences, development, and disorders. Applied Psycholinguistics, 1–28.

McMurray B, & Jongman A (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. Psychological Review, 118(2), 219–246. 10.1037/a0022325 [PubMed: 21417542]

Morrill T, Baese-Berk M, Heffner C, & Dilley L (2015). Interactions between distal speech rate, linguistic knowledge, and speech environment. Psychonomic Bulletin & Review, 22(5), 1451–1457. 10.3758/s13423-015-0820-9 [PubMed: 25794478]

Özyürek A, Willems RM, Kita S, & Hagoort P (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. Journal of Cognitive Neuroscience, 19(4), 605–616. [PubMed: 17381252]

Rueckl JG (2016). Toward a theory of variation in the organization of the word reading system. Scientific Studies of Reading, 20(1), 86–97. 10.1080/10888438.2015.1103741 [PubMed: 26997862]

Sarrett ME, McMurray B, & Kapnoula EC (2020). Dynamic EEG analysis during language comprehension reveals interactive cascades between perceptual processing and sentential expectations. Brain and Language, 211, 104875. 10.1016/j.bandl.2020.104875 [PubMed: 33086178]

Singmann H, Bolker B, Westfall J, Aust F, & Ben-Shachar MS (2015). afex: Analysis of factorial experiments. R Package Version 0.13–145.

Spearman C (1904). The proof and measurement of association between two things. The American Journal of Psychology, 15(1), 72–101. 10.2307/1412159

Staub A (2021). How reliable are individual differences in eye movements in reading? Journal of Memory and Language, 116, 104190. 10.1016/j.jml.2020.104190

Szostak CM, & Pitt MA (2013). The prolonged influence of subsequent context on spoken word recognition. Attention, Perception, & Psychophysics, 75(7), 1533–1546. 10.3758/s13414-013-0492-3

Toscano JC, & McMurray B (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. Language, Cognition and Neuroscience, 30(5), 529–543. 10.1080/23273798.2014.946427 [PubMed: 25780801]

Woods KJ, Siegel MH, Traer J, & McDermott JH (2017). Headphone screening to facilitate web-based auditory experiments. Attention, Perception, & Psychophysics, 79(7), 2064–2072.

Zhang X, Wu YC, & Holt LL (2021). The Learning Signal in Perceptual Tuning of Speech: Bottom Up Versus Top-Down Information. Cognitive Science, 45(3), e12947. [PubMed: 33682208]
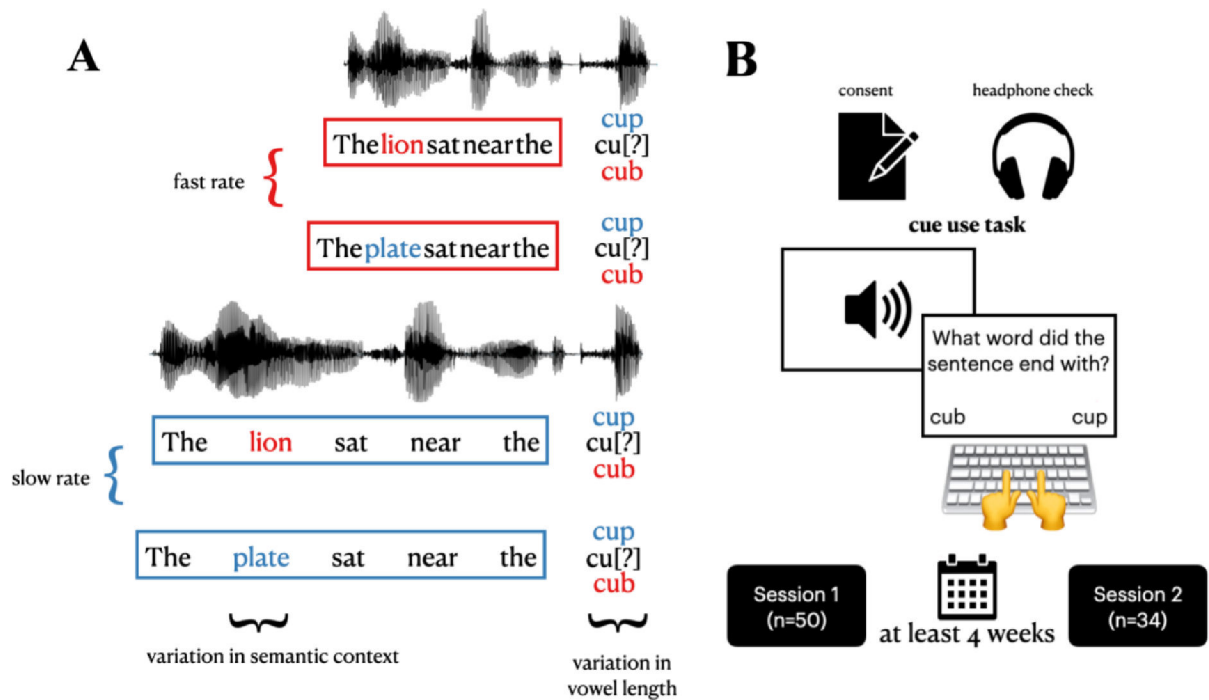
Figure 1.
A. Schematic of cue use task. Sentences contained sentence rate cues, semantic context cues and target word vowel length cues to final target word identity. Sentence rates were either fast (voiced-biasing) or slow (voiceless-biasing). Semantic context cues were either voiced-biasing (e.g., lion relates to *cub*) or voiceless-biasing. Vowel lengths on the final target word were either long (voiced-biasing), short (voiceless-biasing) or of an ambiguous length (vowel lengths were determined in separate pilot testing). *B.* Schematic of two-session experiment. After giving informed consent and completing a headphone check, participants completed the cue use task. Participants were invited back and completed the same set of tasks at least four weeks later (with 34 participants completing both days of experimentation).
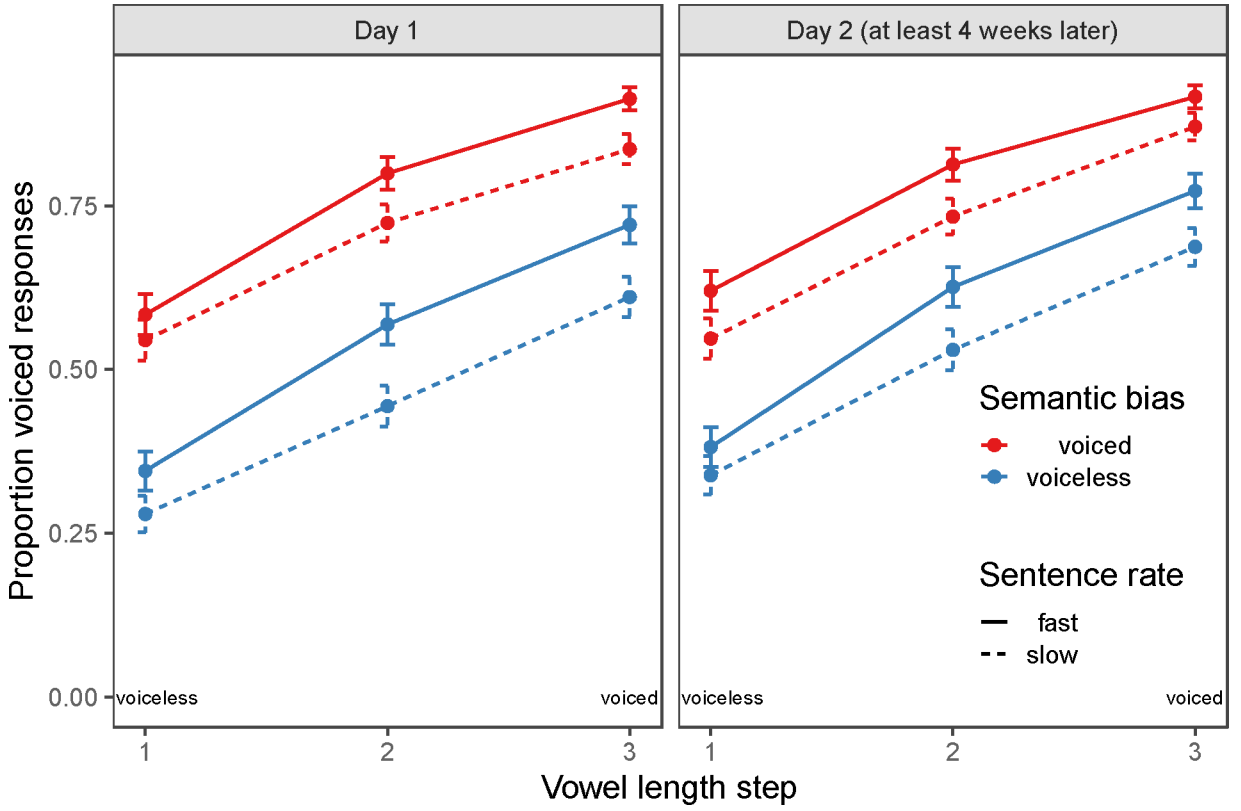
Figure 2.

Overall response data from the cue use task. The x-axis represents vowel length step (modulated by vowel length) ranging from most voiceless to most voiced. The y-axis shows the proportion of voiced responses (e.g., *cub*, as opposed to *cup*). Color represents the semantic bias of the sentence, with red representing voiced-biasing contexts (e.g., *The lion sat ...*, where lion relates more to cub than to cup) and blue representing voiceless-biasing contexts (e.g., *The plate sat...*, where plate relates more to cup than to cub. Line type indicates the contextual rate of the sentence, with solid lines representing a fast rate (voiced-biasing). The panel on the left shows data from the first day of testing, and the panel on the right shows data from the second day of testing (each day N = 34 participants).
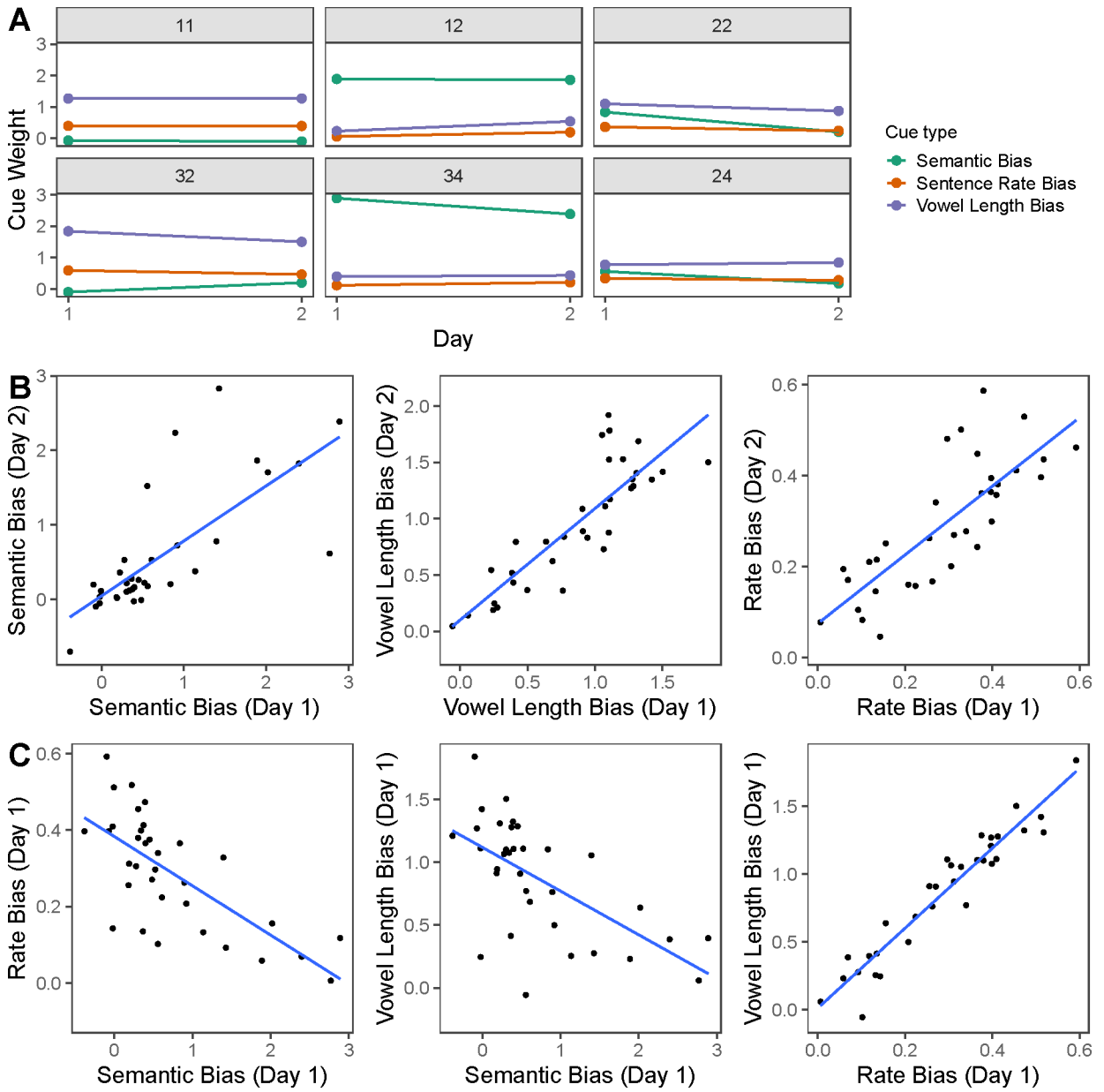
Figure 3.

A. Individual cue weights across two days of testing (separated by at least 4 weeks) from six sample participants. The left-most column displays participants who weight acoustic-phonetic cues more heavily (participants 11 & 32). The middle column displays participants who weight semantic cues more heavily (participants 12 & 34). The right-most column displays participants with relatively equal weighting across cues. Note that across all sample participants, we observe relatively consistent cue weighting across test sessions. *B.* Panel displaying relationship between cue use across test sessions (separated by at least four weeks). The x-axis represents the by-subject slope estimate for a given cue on the first day of testing and the y-axis represents the slope on the second day of testing. Each point represents an individual subject. *C.* Panel displaying relationship between different cues

on the first day of testing. The x- and y-axes represent by-subject slope estimates for the respective cue, with individual points representing individual subjects. Note that we observe an inverse relationship between the use of semantic and acoustic-phonetic cues (rate, vowel length), suggesting that individuals tend to weight one cue more heavily.