Title
Listeners can use coarticulation cues to predict an upcoming novel word

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Moore, Charlotte E
Bergelson, Elika

Publication Date
2021

Peer reviewed

Listeners can use coarticulation cues to predict an upcoming novel word

Charlotte Moore (charlotte.moore@duke.edu)
Elika Bergelson (elika.bergelson@duke.edu)
Department of Psychology and Neuroscience, 417 Chapel Drive
Durham, NC 27710 USA

Abstract

During lexical access, listeners turn unfolding phonetic input into words. We tested how participants interpret words that aren't in their lexicon, either due to their coarticulation cues or because they label a novel object. In a 2-picture Visual World study, 57 adults saw a familiar object and an unfamiliar object, while hearing sentences directing their gaze to the target in 3 conditions: with a familiar word ("crib"), a novel word ("crig"), or a familiar word with coarticulation cueing a novel word ("cri(g)b"). When coarticulation cues matched the novel word ("cri(g)b"), participants looked more at the unfamiliar object than when the cues matched the familiar word, suggesting lexical competition can include a novel word under appropriate circumstances. When hearing a novel word (e.g. "crig"), participants showed two patterns: Roughly half looked more at the unfamiliar object, as expected, while the rest surprisingly looked more at the familiar object. We discuss the interaction of mutual exclusivity, phonetic similarity, and coarticulation cues in driving lexical access.

Keywords: Coarticulation, Visual World Paradigm, Mutual Exclusivity, Speech Perception

Introduction

To achieve the speeds necessary for typical speech, articulators maximize efficiency by moving into position for upcoming sounds while earlier sounds are still being articulated. This preparation leads to a phonetic blending throughout an utterance, referred to as coarticulation (Fowler, 1981). Coarticulation cues can provide early acoustic signals to upcoming speech sounds before those sounds have been fully articulated.

Across eyetracking studies, findings suggest that coarticulation cues facilitate speech perception: participants find the target object more quickly when they hear useful coarticulation cues compared to when they hear no cues (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Salverda, Kleinschmidt, & Tanenhaus, 2014). Similarly, misleading coarticulation cues can create lexical competition between the cued word and the realized one, delaying target looking relative to controls. For example, in Dahan, Magnuson, Tanenhaus, & Hogan (2001), participants were prompted to find an object (e.g. a rod) with edited tokens where coarticulation cues in the vowel were either correct (e.g. *"ro(d)d"*) or misleading. When participants heard *"ro(k)d"*, where the vowel had coarticulation cues for the word "rock" but the final phoneme matched "rod," they were slower to find the target compared to trials where cues were consistent (*"ro(d)d"*). This effect occurred whether or not a rock was present as a competitor image during the trial. In contrast, coarticulation cues that matched a novel word like in *"ro(p)d"* did not slow participants down. These findings suggest that the delay in target looking was caused by lexical competition between two known words; since *rop* was not a word participants knew, it did not create competition and thus did not influence participants' target-looking. However, since this study only used familiar referents, it leaves open the possibility that referential context (i.e. the presence of novel objects) might interact with coarticulation cues to drive lexical competition, even for novel words. We explore this possibility below.

Studies with adults complement a robust literature looking at speech comprehension during development. Studies with infants often test their ability to perceive larger, phonemic changes (Von Holzen & Bergmann, 2018). For example, when hearing one-phoneme mispronunciations, infants look less to a target object than when the label is correctly pronounced (Mani & Plunkett, 2010). Even more relevantly, toddlers show gradient responses to mispronunciations: 1-feature changes disrupt comprehension less than 2- or 3-feature changes. Crucially, this gradient response only emerges when a plausible referent (e.g. a novel object) for a novel word (i.e. a mispronounced familiar word) is present (White & Morgan, 2008; White, Morgan, Wier, & Brown, 2004). This suggests that the greater pragmatic context plays a role even in early speech comprehension.

In what follows, we explore how adults respond when coarticulation cues in a familiar word point to a novel word while a possible referent for that novel word is visually available. In a context where a novel label is expected, listeners might attend to coarticulation cues that do not match any known words. Do listeners need to know a word already for its coarticulation cues to interfere with word recognition? In order to answer this question, we presented participants with unknown objects alongside familiar ones, and manipulated coarticulation cues. Germane to this design is the assumption that new words go with new referents, i.e. mutual exclusivity.

Mutual Exclusivity

Mutual exclusivity (Carey & Bartlett, 1978; Markman & Wachtel, 1988) refers to learners' preference to map novel words onto novel objects, as opposed to mapping novel words onto objects whose names are already known. Older children use mutual exclusivity more reliably than younger children to match novel objects with novel labels, with the rate of use stabilizing in early childhood (Lewis, Cristiano, Lake, Kwan, & Frank, 2020). Studies with adults do not generally employ the same simple-choice mutual exclusivity tasks used with young children. However, cross-situational learning studies suggest that adults *do* use mutual exclusivity, and learn more novel words in conditions where they know the names of more distractor objects (Hendrickson & Perfors, 2019) or when a one-to-one mapping is maintained between objects and labels (Yurovsky & Yu, 2008).

Mutual exclusivity is less robust in some contexts (e.g. multilingual learners use it less; Byers-Heinlein & Werker (2009)). Further, mutual exclusivity cannot constrain learning in all contexts or it would bar learners from acquiring typical language features like synonyms or taxonomic relations. While most studies of mutual exclusivity use novel words that aren't neighbors of common nouns (e.g. *dax*), phonetic similarity may play a role. For instance, White, Yee, Blumstein, & Morgan (2013) showed that adults do not always use mutual exclusivity when perceiving newly-learned words. In an artificial language, they find that adults accepted one-feature mispronunciations of newly-learned words as correct even when a plausible alternate referent was available (e.g. they selected the learned object *zad* when they heard *vad*). Adults were most likely to accept an incorrect label when they had only heard the new word once, suggesting that newly learned wordforms are particularly fragile. Here we extend this line of research beyond newly-learned words by examining whether adults use mutual exclusivity when considering novel words alongside highly similar early-learned nouns, i.e. words whose representations are ostensibly well-engrained in the lexicon.

While under-studied with regards to mutual exclusivity, phonological similarity may be particularly relevant for real-time word comprehension. Indeed, speech is notoriously variable in its realization (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), requiring listeners to perceive incoming speech adaptively (Kleinschmidt & Jaeger, 2015). Phonological similarity affects speech comprehension generally; words that share cohorts or rhymes create competition with one another (Allopenna, Magnuson, & Tanenhaus, 1998). In the context of mutual exclusivity, a novel word that is highly similar to a known word may be difficult to learn because it may be perceived as a speech error. Put another way, if my friend mentions her "new dresh," it is more likely that she is discussing the known object *dress* than a novel object *dresh*. However, the calculus of this situation changes if I see my friend holding an unknown object.

By testing familiar words and novel words that differ only in their final consonant, we can observe how adults perceive a novel word in a context where both a novel word and a familiar word are plausible. This allows us to test the relative contributions of phonetic cues and referential context to speech comprehension.

The Present Study

Our study has two aims. The first (and primary) aim is to determine what happens while listeners perceive coarticulation cues in a context that makes a novel word plausible. Do listeners take the greater context into account when understanding speech and consider any legal sound combination, or are listeners constrained by words they already know while attending to sub-phonemic cues like coarticulation? We assume that adults will look at familiar objects when hearing familiar labels (e.g. a photo of a crib when hearing "crib"), and novel objects when hearing novel labels (e.g. a photo of a new object when hearing "crig") (Markman & Wachtel, 1988; Markman, Wasow, & Hansen, 2003). Here we ask what they do with misleading stimuli where the *vowel* cues a novel word, but the *final consonant* is consistent with a familiar one (e.g. *cri(g)b*). The answer to this question will tell us whether phonetic cues (i.e. coarticulation) interface with pragmatic cues during lexical competition and selection.

The second (auxiliary) aim of this study is to examine the relationship between mutual exclusivity and the phonological similarity of a novel word to a known word. Mutual exclusivity may only function as a word-learning heuristic when the novel label is sufficiently different from known words in the context. If this is the case, we would expect to find that listeners do not map a novel label like "crig" to a novel object when a crib is also an available referent; instead, listeners may assume they misheard the speaker or that the word was mispronounced. On the other hand, if listeners' mutual exclusivity bias does not interact with a lower-level phenomenon like phonetic similarity, then we would expect to find that listeners use the mutual exclusivity bias at very high rates, even when the novel words presented are highly similar to known words. In this case we expect to see listeners looking significantly to e.g. a spiky dog toy when they hear "crig."

To achieve these aims, we present an eyetracking study which simultaneously tackles questions of both coarticulation processing and mutual exclusivity. Taken together, the results highlight interactions between phonetic, lexical, and pragmatic considerations. While our main goal was to seek evidence that coarticulation perception interacts with the pragmatic context, this work also inadvertently highlighted some limits to the mutual exclusivity bias in adults.

Methods

Participants

We collected data in two samples. Sample 1 (N = 20 participants) was a convenience sample of adults affiliated with a private university in the United States. The second sample (N = 37) was recruited through this university's undergraduate research program. Both consisted of fluent English speakers. Four participants had native languages other than English. Their exclusion did not change the results; they are retained in the analyses below. Race data was not collected.

Design

We use a modified Visual World paradigm akin to Looking While Listening (Allopenna, Magnuson, & Tanenhaus, 1998; Fernald, Zangl, Portillo, & Marchman, 2008). This study was a control for a study with infants, thus the design is geared towards younger participants, i.e we provide two images instead of four, and there is no overt selection response.

After 4 warm-up trials, participants saw 24 experimental trials. On each, they saw two images and heard one sentence. Sentences came from one of three conditions, which differed only at the target word. In the **match** condition, participants heard a label for the familiar object (e.g. “crib”). In the **mismatch** condition, participants again heard a label for the familiar object, but the coarticulation cues on the vowel did not match the expected final consonant and instead matched a novel word (e.g. the vowel /I/ in “crib” cued an upcoming /g/). The actual final consonant matched the familiar word (e.g. “cri(g)b”). In the **unfamiliar label** condition, participants heard a novel word which differed from the familiar word only in its final consonant (e.g. “crig”). Each trial contained a familiar object and an unfamiliar object. Each pair of images was always presented together, making 8 unique pairs of images. All pairs were presented three times throughout the study, once with the audio prompt for each of the conditions listed above. See Table 1 and Figure 1.

Stimuli

Each yoked pair contained one photograph of a familiar, early-learned object (e.g. a crib) and one of an unfamiliar object with no well-known name (e.g. a spiky dog toy). All images were superimposed on a 500x500 pixel gray background. Auditory stimuli were created in Praat (Boersma, 2001). A female speaker produced all stimuli sentences. For unfamiliar tokens, stimuli were used as recorded. To create the match condition, we took the familiar word and spliced together the onset and vowel of one token with the final consonant of another token of the same word (e.g. *cri(b)b*). Stimuli were spliced at the closest zero-crossing to the closure of the stop consonant. For the mismatch condition, we spliced the onset and vowel of the novel word and the final consonant of the familiar word (e.g. *cri(g)b*). By splicing both the match and mismatch conditions, we could be sure that differences between the two conditions were not due to the splice itself. Stimuli are available at bit.ly/36Ijy54.

Procedure

Participants came to the lab and signed the consent form. They were then asked about their language background using a 3-question survey to determine when they learned English and if they spoke any additional languages. Participants in Sample 1 then proceeded to the eyetracking component. During data collection of Sample 1, we noticed that some participants were looking to the familiar object even when they heard the unfamiliar label. To ensure that this was not a result of a misunderstanding about the nature of the task, we

Table 1: The 8 stimuli words included in the study by condition. The sound in parentheses indicates the sound being coarticulated in the vowel.

Match	Mismatch	Unfam.
crib	cri(g)b	crig
foot	foo(p)t	foop
bed	be(g)d	beg
block	blo(p)ck	blop
egg	e(b)gg	ebb
dog	do(b)g	dob
hand	ha(N)d	hang
sock	so(p)ck	sop

Figure 1: Two example images presented as in a trial with the corresponding three auditory prompts. Prompts were heard aloud and are written here for demonstrative purposes.

collected an additional sample where we included more information about the study. In Sample 2, participants were read a script that stated that this study was designed as a control for an infant study, and reminded them that infants hear many words they may not know and that some of those words may sound very similar to words they already know.

Next, all participants were seated in front of a 1240x1028 pixel display connected to an Eyelink 1000 Plus eyetracker and calibrated using 5-point calibration. The eyetracking study itself contained 4 warm-up trials with a single image on the screen (e.g. a cracker) and a prompt to look at the image (e.g. “look at the cracker!”). After the warm-up trials, the test trials began. Each test trial featured two images on the screen, one of a familiar object and one of an unfamiliar object (See Figure 1). Participants first saw both images in silence for 2500ms, then heard one of the three corresponding prompts (coarticulation match, coarticulation mismatch, or unfamiliar label). Looking was recorded throughout the trial. After the experiment, participants were thanked and when applicable provided with course credit.

Data Analysis plan

Our first question was whether mismatching coarticulation cues (e.g. *cri(g)b*) led listeners to look more at the unfamiliar object relative to match trials. To answer this, we fit growth

Table 2: Fixed effects of the growth curve analysis. 'Int' represents the intercept term. Terms labeled 'ot' represent orthogonal polynomials, with number indicating the order, e.g. ot1 is the linear term. Degrees of freedom (df) are approximate and estimated using Satterthwaite's method.

Fixed Effs	Est.	SD	t =	df	p value
Int.	-1.36	0.16	-8.35	114	<0.001 ***
ot1	-5.31	0.30	-17.90	114	<0.001 ***
ot2	0.07	0.21	0.35	114	0.729
ot3	0.39	0.15	2.63	114	0.01 *
ot4	0.35	0.11	3.11	114	0.002 **
Cond	-0.23	0.05	-5.06	114	<0.001 ***
Sample	-0.01	0.09	-0.11	114	0.916
ot1:Cond.	-1.00	0.30	-3.37	114	0.001 **
ot2:Cond.	0.58	0.21	2.82	114	0.006 **
ot3:Cond.	0.44	0.15	2.94	114	0.004 **
ot4:Cond.	-0.25	0.11	-2.19	114	0.03 *

curves for just the mismatch and match conditions. Our second question was based on a surprising discovery in our initial visualization of the data in Sample 1 (N = 20). Namely, we found that counter to our predictions, not all participants readily looked at the unfamiliar object when hearing the unfamiliar word. In an exploratory analysis, we first categorized participants into two "listener types" based on their behavior on unfamiliar label trials. Next, we fit growth curves to their performance on these trials, and added listener type as a fixed effect to a base model to confirm that listener type helped explain variance in behavior. Finally, we returned to our original model but added listener type to look at any differences in performance on the mismatch and match conditions based on performance in the unfamiliar label condition.

For the growth curve analyses, we follow Mirman (2014). We conduct pseudo-logistic growth curve analyses using weighted empirical logits. We analyzed participants' gaze to the *unfamiliar object* (the distractor image) in the match and mismatch conditions from 200-1500ms after the target word's onset in each of 8 trials per condition. To accomplish this, each participant's gaze data was divided into 20ms bins. For each participant, we summed looks (0 or 1) for each of the 20 ms bins across items, resulting in a condition-level value (with possible values ranging from 0-8) for each bin. These condition-level sums were then used to create the empirical logits and weightings used in the model.

Results

Aim 1: Coarticulation effect

To determine whether participants were slower to find the target in the mismatch condition relative to the match condition, we used the lmerTest package to fit the following model: $\hat{Y} = (ot1 + ot2 + ot3 + ot4) * Condition + Sample + ((ot1 + ot2 + ot3 + ot4)||Subj \times Condition)$ The fixed effects were Condition (match or mismatch), Sample (1 or 2) and orthogonal polynomials to model time. Time 0 on each trial corresponds to target word onset (e.g. the start of crib in "Look at the crib!"). We use four orthogonal terms to model time in keeping with Mirman (2014), which establishes that multiple terms are needed to appropriately capture the nature of eye movement trajectories for this method.

The final model also included interaction terms between Condition and each of the 4 polynomial terms. The model including these interaction terms was significantly better than one without (χ^2 = 723.73, p < .001), suggesting that looking to the unfamiliar object unfolded differently as a function of whether participants' heard *cri(b)b*, *crig*, or *cri(g)b*.

A summary of the fixed effects is in Table 2. All estimates are in logit space. The intercept estimates the log odds of looking to the unfamiliar object in the match condition for Sample 1. Sample was not a significant predictor, suggesting that our added instructions for Sample 2 did not influence looking behavior. There was a significant main effect of condition, indicating that participants looked at the unfamiliar object significantly more overall in the mismatch condition (e.g. crig when hearing *cri(g)b*). Looking at the time estimates, the significant linear term indicates that looking to the unfamiliar object decreased overall as each trial unfolded. The significant cubic and quartic terms confirm curvilinear gaze dynamics, shown (untransformed) in Figure 2.

The slopes of the two conditions were also significantly different on the polynomial terms, corresponding to differences in the speed that participants looked away from the unfamiliar object across conditions; participants were significantly faster to look away from the unfamiliar object in the match condition, and looked less at it overall. In sum, coarticulation mismatches pulled participants' looking away from the target across the entire analysis window, i.e. participants were mis-cued by the coarticulation on the vowel in mismatch trials. This in turn suggests that the presence of an unfamiliar competitor object onscreen created lexical competition between the familiar word and the cued unfamiliar word

Aim 2: Mutual Exclusivity

Given the robustness of the mutual exclusivity (ME) constraint from early childhood on, we assumed that on hearing *"crig"*, participants would look at the unfamiliar object (e.g. the spiky dog toy) rather than the crib. To our surprise, a large proportion of our adult participants appeared not to do so, instead looking mostly at the crib upon hearing *"crig"*. Thus, in an exploratory analysis, we first categorize subjects into two groups, those who exhibited this unexpected behavior (ME Eschewers), and those who behaved as expected (ME Users). Then we model their gaze using the same type of growth curves as in Aim 1, this time comparing these two groups' performance across conditions.

To quantify the two different patterns we observed across individuals, we began by aggregating the data for each participant's looking across all 8 of the unfamiliar label trials to determine what proportion of the time participants spent looking at the unfamiliar object. We used the entire length of

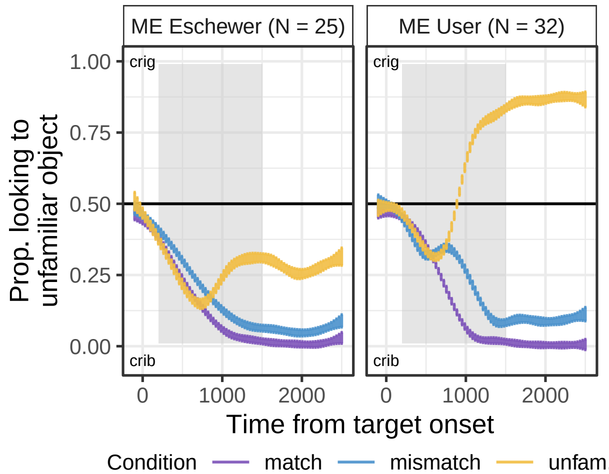

Figure 2: Looking over time to the unfamiliar object for both listener types. Horizontal line indicates chance. Shaded area indicates window of analysis for growth curve models. Width of band represents 95 percent CI.

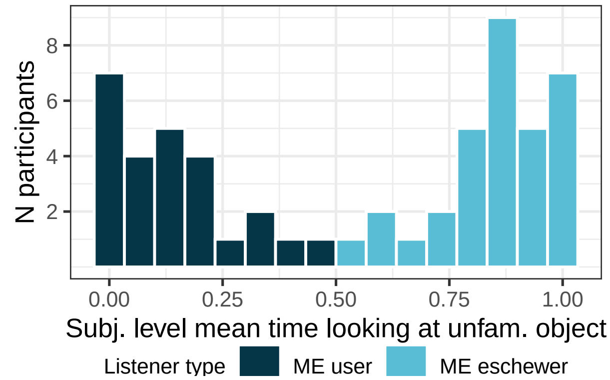

Figure 3: Histogram of unfamiliar label trial responses. X-axis shows subject-level mean proportion of time spent looking at the unfamiliar object 200-4000ms post-target word onset (full trial length). Higher proportions represent greater use of mutual exclusivity. Color indicates how participants were sorted in growth curve models.

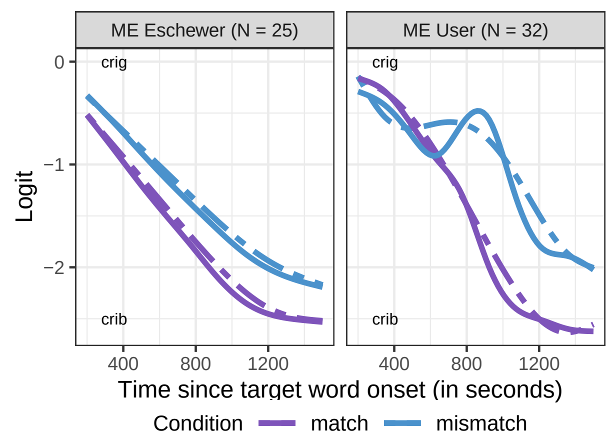

Figure 4: Comparison of the transformed raw data in logit space (solid lines) and the predicted values from the growth curve model which contains listener type as a predictor (dotted lines).

the trial after the target word was uttered (from 200ms after target word onset to 4000ms after). We use the full window rather than making assumptions about the timeline of looking behavior given this unexpected result.

Participants who spent more than 50% of the trial after the target word's onset looking at the unfamiliar object on unfamiliar label trials were categorized as ME Users (N = 32). Participants who spent less than 50% of the time looking at the unfamiliar object were labeled ME Eschewers (N = 25).[1] Neither group showed a preference for the unfamiliar object before the target word was spoken (50.4% for ME Eschewers and 49.8% for ME Users). See Figure 3.

After establishing the two distinct groups, we conducted an additional growth curve analysis predicting looking behavior in only the unfamiliar label condition using the same approach described above, comparing a baseline model to one with listener type (ME User vs. Eschewer).

The model with listener type as a fixed effect fit the data significantly better than one without it (χ^2 = 15.12, $p < .001$). The final model was $\hat{Y} = (ot1 + ot2 + ot3 + ot4) + Sample + ListenerType + ((ot1 + ot2 + ot3 + ot4)|Subj)$. This suggests that in line with our holistic trial-level split, ME Users and Eschewers diverged in their dynamic responses to unfamiliar labels; we explore possible explanations in the Discussion. To see if this difference held beyond the unfamiliar trials, we next modeled the coarticulation effect in the match and mismatch conditions using listener type as a predictor.

Listener type to predict coarticulation processing

We extended the model in Aim 1, adding listener type as a fixed effect. This significantly improved model fit (χ^2 = 12.98, p = < .001), suggesting that in addition to differing in their responses in the unfamiliar label trials, the two listener-types had divergent looking trajectories in the two conditions where coarticulation cues were manipulated as well. Figure 4 highlights this difference: ME Users exhibited a characteristic 'bump' (from ~500 to 1000ms post-target onset) in the mismatch condition wherein they began looking at the unfamiliar object more when they heard the coarticulation cues in the vowel, then returned their gaze to the familiar object upon hearing the final consonant. In the ME Eschewers, this 'bump' is missing, suggesting that the coarticulation cues on the vowel did not create the same effect for these participants. Instead, ME Eschewers reduced their looking monotonically.

[1]These distinct patterns were already present in the first half of the study and became stronger in the second half of trials, especially for the ME Users. See bit.ly/36Ijy54.

Discussion

We asked whether the referential context changes how coarticulation cues are used in speech comprehension. Previous work suggests that competition from within the lexicon is a prerequisite for coarticulation competition effects (Dahan, Magnuson, Tanenhaus, & Hogan, 2001). Instead, we found that when listeners were provided with a plausible referent for an unfamiliar word, they were slower and less accurate at finding the target on mismatch trials compared to match trials. This suggests that listeners were not limited to the words in their lexicon when predicting the identity of a word from its vowel. Even ME Eschewers were sensitive to coarticulation mismatches, though the difference between match and mismatch trials was less pronounced for these participants. These findings provide evidence that coarticulation perception does not merely create competition between known words in the lexicon. Instead, coarticulation cues that correspond to novel words seem to guide behavior when a plausible referent for the novel word is available. This type of pragmatic inference where participants take context into account during speech comprehension may explain why we find a robust effect of coarticulation between our match and mismatch conditions where Dahan et al. (2001) did not.

While the divergence between the match and mismatch conditions is consistent with a lexical competition account, a possible alternative explanation is that upon hearing the mismatch stimuli (e.g. cri(g)b), participants were simply confused about the target. We find this explanation unlikely, given the systematic looking trajectories that participants provided across conditions; confusion would likely result in more random looking behavior than what our results have captured. That said, a lexical competition account would be bolstered by a replication of Dahan et al (2001), to allow us to compare our findings to a condition with multiple familiar objects onscreen.

While we intended to examine the interaction of coarticulation cues with pragmatic considerations, along the way we unexpectedly discovered a limit to the mutual exclusivity bias. Participants fell into two distinct behavior patterns when faced with a novel word that was very similar to a known word. 56% of participants showed the pattern we would expect based on mutual exclusivity, i.e. looking to the unfamiliar object when hearing the unfamiliar word. The other 44% showed the opposite pattern, looking at the familiar object for most of each trial even when they heard an unfamiliar word. Why might this be?

One possibility is that participants could not *detect* differences in final consonants across conditions. If this were the case, we would expect the timecourse of looking behavior in all three conditions to be equivalent; this is not what we find. Instead, all participants looked at the familiar object (e.g the crib) least when they heard e.g. "crig." Furthermore, participants looked less at the target on mismatch trials than on match trials, suggesting they detected fine-grained phonetic distinctions. It is therefore highly unlikely that half of our participants were unable to perceive the differences between each condition.

A Bayesian explanation in the vein of Kleinschmidt & Jaeger (2015) is that participants correctly heard each prompt, but because the prior probability of the familiar word was so much higher than a novel word, the winning candidate was the familiar word. Under this view, participants either interpreted the unfamiliar words as a mispronunciation of the familiar word instead of as a novel word, or assumed they misheard the prompt. These are plausible explanations for the data, since the unfamiliar words differed only in the final phoneme. Further, the mismatch condition likely added uncertainty, as those trials sounded unusual due to their conflicting cues. This explanation also helps us account for the emergence of two divergent approaches. ME Eschewers may have stronger priors to expect familiar words or less clear speech, whereas ME Users may have weaker priors in those domains.

To probe this possibility further, we took advantage of our randomized trial orders. Order 1 began with a familiar word while Order 2 began with an unfamiliar word. 20/29 Order 1 participants were ME Users, while only 12/28 Order 2 participants were ME Users. This suggests that the first trial participants saw may have biased them towards expecting familiar or unfamiliar words, in effect shifting their priors.

Because this bifurcation of our participants was unexpected, these results require confirmatory follow-up beyond the second sample reported here. It would also be informative to vary the size of the phonetic difference between the match and unfamiliar conditions to see if the number of ME Eschewers scales with the size of the difference. Another fruitful avenue would be to query what features of the listener predict ME using/eschewing behavior (e.g. frequency of exposure to new words, frequency of exposure to unclear or accented speech, or personality factors like agreeableness.) Another important next step would be to add an overt response to our design, asking participants to click on the correct object in each trial. This would provide discrete evidence of the interpretation participants committed to, complementing the gradient gaze data.

Taken together, our results add evidence that the comprehension of sub-phonemic cues is subject to pragmatic factors outside of language. This further supports the interconnectedness across levels of linguistic analysis, from global contextual effects to low-level phonetic cues. Further, we provide one of the first direct tests of mutual exclusivity in adults, showing that adults use mutual exclusivity to varying degrees when labels are sufficiently similar to known words.

Conclusion

Coarticulation cues are a fundamental component of speech perception, used to distinguish words in the lexicon from one another early in the speech stream. This study provides new evidence that coarticulation cues can also be used to distinguish known words from novel ones when a plausible novel referent is provided, and that this effect is present whether listeners map a novel word to an unfamiliar referent or not.

We also provided surprising evidence that individual differences play a large role in how participants perceive unfamiliar words that sound highly similar to known words. Taken together, these results show that coarticulation perception creates more than lexical competition - understanding this seemingly low-level phonetic cue involves a complex interplay between the context, the acoustics of the speech stream, and the possible intentions of the speaker.

References

10 Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. http://doi.org/10.1006/jmla.1997.2558

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *The Journal of the Acoustical Society of America*, *133*(4), 2350–2366. http://doi.org/10.1121/1.4794366

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot. Int.*, *5*(9/10), 341–345.

Byers-Heinlein, K., & Werker, J. F. (2009). Monolingual, bilingual, trilingual: Infants' language experience influences the development of a word-learning heuristic. *Developmental Science*, *12*(5), 815–823. http://doi.org/10.1111/j.1467-7687.2009.00902.x

Carey, S., & Bartlett, E. (1978). Acquiring a Single New Word.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, *16*(5-6), 507–534. http://doi.org/10.1080/01690960143000074

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In *Developmental psycholinguistics* (pp. 97–135). Amsterdam, Netherlands: John Benjamins Publishing Company. http://doi.org/10.1075/lald.44.06fer

Fowler, C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech & Hearing Research*, *24*(1), 127–139. http://doi.org/10.1044/jshr.2401.127

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition*, *189*, 11–22. http://doi.org/10.1016/j.cognition.2019.03.005

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203. http://doi.org/10.1037/a0038695

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. http://doi.org/10.18637/jss.v082.i13

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. http://doi.org/10.1016/j.cognition.2020.104191

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461. http://doi.org/10.1037/h0020279

Mani, N., & Plunkett, K. (2010). Twelve-Month-Olds Know Their Cups From Their Keps and Tups. *Infancy*, *15*(5), 445–470. http://doi.org/10.1111/j.1532-7078.2009.00027.x

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. http://doi.org/10.1016/0010-0285(88)90017-5

Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, *47*(3), 241–275. http://doi.org/10.1016/S0010-0285(03)00034-3

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. CRC Press.

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. http://doi.org/10.1016/j.jml.2013.11.002

Von Holzen, K., & Bergmann, C. (2018). A Meta-Analysis of Infants' Mispronunciation Sensitivity Development. In *Annual Conference of the Cognitive Science Society. Cognitive Science Society (US)* (p. 6).

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. http://doi.org/10.1016/j.jml.2008.03.001

White, K. S., Morgan, J. L., Wier, L., & Brown. (2004). When is a dar a car ? Effects of mispronunciation and referential context on sound-meaning mappings.

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, *68*(4), 362–378. http://doi.org/10.1016/j.jml.2013.01.003

Yurovsky, D., & Yu, C. (2008). Mutual Exclusivity in Cross-Situational Statistical Learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *30*(30), 7.