Anne Pycha*

Phonological and morphological roles modulate the perception of consonant variants

https://doi.org/10.1515/lingvan-2020-0025
Received May 26, 2020; accepted October 8, 2020; published online February 8, 2021

Abstract: This study investigated how the perception of a sound is affected by its phonological and morphological roles within a word. We asked American English listeners (n = 24) to judge differences among phonetic variants of sounds [l], [n], [ɹ] in three word conditions: 1) at morpheme boundaries with a phonological process, such as [n] in *down-ed*, which triggers voicing agreement on the suffix, 2) internally without a process, such as [n] in *mound*, and 3) at morpheme boundaries alone, such as [n] in *town-ship*. We used Praat synthesis with different acoustic settings to create variants, e.g., $[n]_a$, $[n]_b$, $[n]_c$, which were spliced into a base to produce three tokens, *dow*$[n]_a$*ed*, *dow*$[n]_b$*ed*, *dow*$[n]_c$*ed*. Identical variants were used across conditions (e.g., in condition 2: *mou*$[n]_a$*d*, *mou*$[n]_b$*d*, *mou*$[n]_c$*d*). On each trial, participants heard two tokens of the same word (e.g., *dow*$[n]_a$*ed* – *dow*$[n]_b$*ed*) and rated the difference between the target sound using a sliding scale with endpoints "0% (totally identical)" and "99% (totally different)". Analysis with linear mixed-effects model revealed significant differences between ratings among all conditions, with the pattern *township* < *downed* < *mound*. These results suggest that a sound's phonological and morphological roles within a word affect how people perceive it. We evaluate this finding in light of the differing predictions made by phoneme-based theories, which incorporate phonemes as a fundamental unit, versus exemplar theories, which argue that phonological units are emergent.

Keywords: exemplar theory; phoneme-based theory; speech perception; speech sound variation

1 Introduction

Language experience – such as exposure to words and sentences – modulates the perception of speech sounds (Hume and Johnson 2001). A key finding is that listeners categorize a speech sound differently when it is embedded in a word context, compared to a non-word context (e.g., Ganong 1980). Several theories can account for this fact, including phoneme-based frameworks (e.g., Kazanina et al. 2018), which claim that phonemes are fundamental units of phonological structure, as well as exemplar or "emergence" frameworks, which argue that different units of analysis emerge under different circumstances (e.g., Bybee and Hopper 2001; Goldinger and Azuma 2003). However, not all word contexts are alike. Depending upon a word's phonological and morphological structures, a speech sound may play different roles: it may demarcate morpheme boundaries, for example, or serve as a trigger for phonological alternations. In such cases, the predictions of phoneme-based versus emergence theories diverge. In the current study, we tested these differing predictions by asking how the perception of a speech sounds changes depending upon the role it plays within a word.

When a speech sound plays a contrastive role in a word, listeners perceive it differently than they otherwise would. This was first demonstrated by Ganong (1980), who showed that listeners were more likely to categorize an ambiguous sound as /t/ in the context of __*ask*, but as /d/ in the context of __*ash*. This lexical

***Corresponding author: Anne Pycha**, Department of Linguistics, University of Wisconsin, Milwaukee, WI, USA,
E-mail: pycha@uwm.edu

effect on perception has been replicated in many subsequent studies (for review, see Pitt and Samuel 1993). Theories that emphasize the primacy of phonemes (Pierrehumbert 2002, 2016; e.g., Kazanina et al. 2018) can account for this finding because each lexical entry (such as /tæsk/) is composed of explicitly-represented phonemes (such as /t/). Therefore, activation of /tæsk/ also activates /t/. Emergence theories, including many versions of exemplar theory (e.g., Bush 2001; Bybee 2003; Bybee and Hopper 2001; Bybee and Scheibman 2008), can also account for this finding. In this framework, previously-heard instances of spoken words are stored in exemplar clouds, and newly-heard instances are categorized according to their similarity with those exemplars. Thus, upon hearing *{T/D}ask*, listeners will tend to categorize it as *task* (and accordingly, chose /t/ as their response) because it matches previously stored exemplars; they will not tend to categorize it as *dask*, because this form has no stored exemplars (see Johnson 1997).

In addition to contrast, however, speech sounds play other important roles within a word. Consider an American English example such as *downed* [dɑʷnd] (e.g., *the missile downed the fighter jet*). In addition to contrast, the speech sound [n] demarcates the boundary of the root morpheme *down*, and it also serves as the trigger for the phonological process of voicing agreement, according to which the past-tense suffix must be realized as a voiced [d], and not voiceless [t] (*[dɑʷnt] is an ungrammatical realization of the past tense). However, in a superficially similar word such as *mound* [mɑʷnd], [n] plays neither of these roles. It does not demarcate a boundary, and it does not serve as a trigger for agreement (*mount* [mɑʷnt] is perfectly grammatical). A similar distinction holds for the sound [ɹ] in words like *lured* versus *curd*, and for the sound [l] in words like *filled* versus *guild*.

If language experience modulates the perception of speech sounds, we can broadly hypothesize that listeners should perceive [n] differently in words like *downed* versus *mound*. The specific direction of this difference, however, will depend upon the framework we adopt. A phoneme-based theory predicts that listeners should treat [n] more categorically when it occurs in a word like *downed*, compared to a word like *mound*. This is because the [n] in *downed*, as a trigger for voicing agreement, behaves similarly to other sounds, such as the [ɹ] in *lured* and the [l] in *filled*, and therefore belongs to an "equivalence class" that unites the behavior of physically different speech events (Pierrehumbert 2016: 37). The same logic holds for variants of [n] in *downed* that may be physically different, such as $[n]_a$, $[n]_b$, or $[n]_c$; they all trigger voicing agreement and therefore participate in the equivalence class. In *mound*, however, [n] does not trigger voicing agreement, and variants such as $[n]_a$, $[n]_b$, $[n]_c$ do not participate in this equivalence class. If listeners' experiences with equivalence classes shape perception, then, we would hypothesize listeners should treat [n] more categorically – and therefore be less sensitive to surface differences in variants of [n] – in *downed* contexts, compared to *mound* contexts.

Emergence theories, on the other hand, do not make this prediction. For one thing, processes such as voicing agreement have no status in these frameworks, so there is nothing special about cases in which [n] acts as a trigger. Furthermore, "speech sound [n]" is not a fundamental unit of analysis. Instead, as listeners generalize across stored exemplars, a chunk like [n] may *emerge* as a category – if it does so, however, it will compete for activation with other, larger chunks, and the parameters of this competition change from one situation to the next (Goldinger and Azuma 2003; Vitevitch and Luce 1999; Walsh et al. 2010). For example, the signal [dɑʷnd] contains two strong competitors mitigating against the emergence of [n] as a category, namely the chunk [dɑʷnd] (corresponding to the whole word *downed*), and the chunk [dɑʷn] (corresponding to the root *down*). Meanwhile, the signal [mɑʷnd] only contains one strong competitor to the chunk [n], namely the chunk [mɑʷnd] (corresponding to *mound*). The chunk [mɑʷn] is unlikely to activate, because it has no exemplar clouds (**moun* isn't a word) and because it shares no meaning with *mound*. The same logic holds for similar pairs, such as *lured* versus *curd* and *filled* versus *guild* – in this last case, even though the chunk [gɪl] does have exemplar clouds (*gill* happens to be a word), it is unlikely to activate in response to [gɪld] because it shares no meaning. Thus, all things being equal, [n] is more likely to emerge as a category in *mound*, compared to *downed*. Emergence theories therefore predict a pattern that is opposite to that of phoneme-based theories: listeners should treat [n] more categorically – and therefore be less sensitive to surface differences in variants of [n] – in *mound* contexts, compared to *downed* contexts.

The current study tests these predictions by asking American English listeners to judge differences among phonetic variants of the sounds [n], [l], [ɹ] in words where they participate in voicing agreement (e.g., *downed*, *lured*, *filled*) versus in words where they do not (e.g., *mound*, *curd*, *guild*). On each trial, participants heard two tokens, such as *dow*$[n]_a$ *ed* ~ *dow*$[n]_b$ *ed* or *mou*$[n]_a$*d* ~ *mou*$[n]_b$ *d*, and rated the difference between the target sounds, such as [n]. To create variants, we used different synthesized voices, a choice which allowed us to focus on reactions to lower-level, non-allophonic acoustic changes. As we have outlined, phoneme-based theories predict that listeners should be less sensitive to these surface differences in the context of words like *downed*, while emergence theories predict that they should be less sensitive to them in the context of words like *mound*.

To help disentangle the effects of morpheme boundaries versus voicing agreement, we also included a third word context (*township*, *fearful*, *illness*), in which the sounds [n], [l], [ɹ] occur at a root morpheme boundary (i.e., *town*, *fear*, *ill*), but do not serve as a trigger for voicing agreement (i.e., there is no requirement for the following sound to agree in voicing, and forms like *[taʊnʒɪp] are not grammatical). Phoneme-based theories predict that words like *township* and *mound* should pattern together, because [n] does not participate in voicing agreement in either context. Emergence theories, on the other hand, predict that words like *township* and *downed* should pattern together, because in both contexts, [n] competes with multiple emergent constituents (e.g., [tɑʷn], [tɑʷnʃɪp]) rather than just one.

2 Method

The experiment collected listeners' judgments of the sounds [n], [ɹ], [l] when they were embedded into different word contexts. An overview of the design, using [n] as an example, is in Table 1.

As seen in Table 1, we used three word contexts, which differed according to whether the target sound played only a contrastive role only (*mound*), a morphological and phonological role (*downed*), or a morphological role alone (*township*). As a baseline, we also used an isolated condition, in which the target sound occurred on its own without any word context.

We synthesized three variants of the target sound, $[n]_a$, $[n]_b$, $[n]_c$, and spliced these variants into the base words. On each trial, participants heard two tokens of the same word (e.g., *dow*$[n]_a$*ed* – *dow*$[n]_b$*ed*) and rated the difference between the two variants of the target sound using a sliding scale with endpoints "0% (totally identical)" and "99% (totally different)".

2.1 Target word selection

We used the Clearpond database (Marian et al. 2012) to search for English words that ended in the clusters [nd], [ɹd], or [ld]. This search returned complex words ending with the past-tense suffix, such as *downed*, as well as simple words ending with [d], such as *mound*. We also searched for words that ended in one of the target sounds plus a sequence representing a derivational suffix such as *-ship*, *-ful*, *-ness*, etc. For example, searches for final sequences [nʃɪp], [ɹʃɪp], [lʃɪp] returned words such as *township*, *starship*, *battleship*. To help ensure that the target sound would occur in coda position, we searched only suffixes that began with a consonant.

Table 1: Overview of experimental design.

Roles	Example	Tokens		
Contrast only	*mound*	[maʊ$[n]_a$ d]	[maʊ$[n]_b$ d]	[maʊ$[n]_c$ d]
Morphology + Phonology	*downed*	[daʊ$[n]_a$ d]	[daʊ$[n]_b$ d]	[daʊ$[n]_c$ d]
Morphology	*township*	[taʊ $[n]_a$ ʃɪp]	[taʊ $[n]_b$ ʃɪp]	[taʊ $[n]_c$ ʃɪp]
Isolated	*n*	$[n]_a$	$[n]_b$	$[n]_c$

From the pool of candidate words, we removed proper nouns (*England*), words with irregular alternations (*told*), and words that contained more than two morphemes (*ownership*). We also removed words that were potentially taboo (*killed*), or that formed homonyms across word types (*soared* and *sword*). Finally, we removed words that contained multiple instances of the target sound, such as *contained*, as well as words with orthographic doubling, such as *stunned*.

From the remaining set of words, we selected 54 contrast-only words, 54 morphology + phonology words, and 54 morphology words, for a total of 162. Each condition contained an equal number of words with targets [n], [ɹ], [l]. The mean log frequency was 0.596 for contrast-only, 0.596 for morphology + phonology words, and 0.562 for morphology words. The full set of stimuli is listed in the Appendix.

Note that words like *township* present a couple of confounds. Ideally, all stimulus words would have contained the same number of syllables and the target sound would have always occurred in the same syllabic position followed by the sound [d]. However, constraints on the English lexicon made this impossible. The past-tense *-ed* and plural *-s* are the only English suffixes that adjoin to existing syllable codas and do not increase overall syllable count; all other suffixes, such as *-ship*, *-ful*, *-ness*, form a new syllable. Furthermore, the only other suffix that begins with [d] is *-dom*, for which Clearpond returned only potential two candidates from the entire lexicon, *boredom* and *stardom*. Note also that relatively few English monomorphemic words end in the clusters [nd], [ɹd], or [ld]; in order to have a sufficient number of stimuli, we were obliged to include words with one, two, and three syllables (*mound*, *astound*, *correspond*). In interpreting the results, we must bear these confounds in mind.

We selected fifty-four filler words, half of which contained initial clusters with a target sound (*snow*, *preach*, *blemish*), and half of which contained final clusters with a target sound (*present*, *disturb*, *gulp*). Each group contained equal numbers of the targets [n], [l], and [ɹ]. To be used in practice trials, we also selected six words with target sound [m] (*small*, *smooth*, *named*, *ashamed*, *roomful*, *bottomless*).

2.2 Stimulus creation

All stimuli were created in Praat (Boersma and Weenink 2018). We did not use natural speech because it is possible that word context affects production: speakers may plausibly produce the [n] differently in *downed* versus *mound*. In addition, natural speech varies idiosyncratically from one talker to the next. In order to focus solely on the influence of word context on perception, and to help ensure that our results can be replicated, the experiment used synthesized speech.

We used Praat's SpeechSynthesizer feature, which produces outputs with different pre-specified voice settings, using American English as the target language. To create each base token, we entered its orthographic representation (e.g., *mound*) and chose the voice variant "Female 2". We used a text grid to demarcate the beginning and end of the target sound.

To create variants of the target sound, we varied the artificial voice in which they were synthesized, which allowed us to create different tokens of [n], [ɹ], [l] without explicitly manipulating other acoustic characteristics – such as duration or pitch – that might lead listeners to believe that an incorrect allophone had been produced for a particular word. Variants were synthesized so as to match the base token's vowel context. Thus for example, for words like *downed*, *mound*, *township*, we created three variants of [n] by entering the orthographic representation of a word containing the same vowel + [n] sequence (e.g., *crown*), and choosing the three female voice variants "Female 2", "Annie", and "Steph". We choose these three different voice settings, from among the many others in Praat, because they impressionistically sounded the most different from one another. For each of these three tokens of [kɹaʊn], we cut out the [n] portion, creating variants we will refer to as $[n]_{a,}$ $[n]_{b,}$ $[n]_{c.}$

Other variants were produced in a similar manner; for example, for *guild*, *filled*, *illness*, we created three variants of [l] by synthesizing *pill* with "Female 2", "Annie", and "Steph", and cutting out the [l] portion to create $[l]_{a,}$ $[l]_{b,}$ $[l]_{c.}$ In total, we synthesized thirty-four context words ([n] in 15 vowel contexts + [ɹ] in 7 vowel contexts + [l] in 12 vowel contexts), each of which was synthesized with the same three voice settings, ultimately producing a total of 102 variants.

The final step was to cross-splice each variant into the appropriate base, producing three versions of each word as depicted in Table 1.

The procedure for constructing filler and practice stimuli was the same as for targets, except that we used different voice variants than for targets.

2.3 Lists

We created nine trial lists, which included two types of trials. On "different" trials, there were two different variants of the same word. On "same" trials, which were included so that participants did not create an expectation of difference, there were two identical variants of the same word.

The 162 target words were divided into three primary sets, X, Y, Z, of 54 words each, evenly distributed across the three word types and the target sounds ([n], [ɹ], [l]). Each primary set was then further divided into three lists, e.g., X1, X2, X3, such that each list contained two trials (one "same", one "different") per target word. For example, the word *mound* was in set X. Lists X1, X2, X3 all included *mou*$[n]_a$*d* ~ *mou*$[n]_a$*d*. Furthermore, List X1 included *mou*$[n]_a$*d* ~ *mou*$[n]_b$*d*, List X2 included *mou*$[n]_a$*d* ~ *mou*$[n]_c$*d*, and List X3 included *mou*$[n]_b$*d* ~ *mou*$[n]_c$*d*. Other lists were constructed in similar fashion. The three "different" trial types were distributed evenly within each list. Each list contained a total of 108 trials with target words (54 targets × 2 trial types).

For the isolation condition, every participant heard the same set of trials. Specifically, for each of the 34 vowel-context words, participants heard the corresponding target sound in one "same" trial and three different trials. For example, for the [n] created from *crown*, listeners heard $[n]_a$ ~ $[n]_a$, $[n]_a$ ~ $[n]_b$, $[n]_a$ ~ $[n]_c$, $[n]_b$ ~ $[n]_c$. Thus, each list contained a total of 136 trials with isolated vowels (34 vowel-contexts x 4 trial types).

The same 54 filler words occurred on every list. There were two trials (one "same", one "different") for each filler word. Thus, each list contained a total of 108 trials with fillers (54 filler words x 2 trial types). The same practice trials also occurred on every list. There were ten trials for isolated [m] sounds and twelve trials for [m] words, for a total of 22 practice trials (eleven "same", eleven "different").

In sum, there were nine different lists, which each contained a total of 374 trials (108 target words, 136 target sounds in isolation, 108 fillers, and 22 practice items).

2.4 Procedure

Participants completed the experiment, which was built using E-Prime 3.0 software, in a quiet laboratory environment, seated at an individual carrel equipped with a computer and mouse, wearing high-quality headphones. Participants received verbal instructions from the researcher, which were reiterated on the computer screen.

Each participant was randomly assigned to one of the nine lists. Participants first completed the practice trials, then proceeded to the remainder of the experiment. Trials were blocked by target sound, with [n] stimuli in one block, [ɹ] stimuli in another block, and [l] stimuli in another block. The order of these blocks was randomized for each participant. Within each block, trials were further grouped such that all isolation stimuli occurred first, followed by one-syllable words, two-syllable words, and three-syllable words. Within each grouping, trials were randomized for each participant.

On each trial, participants heard two tokens of the same word (e.g., *dow*$[n]_a$*ed* ~ *dow*$[n]_b$*ed*) or isolated sound (e.g., $[n]_a$ ~ $[n]_b$) and were asked to "Rate how different the sounds are." Simultaneously, the screen displayed printed information, indicating the current target and word, e.g., "Target sound: "n", Current word: downed". This ensured that participants attended to the appropriate sound, and that they interpreted the context word correctly. The screen also displayed a sliding scale with endpoints "0% (totally identical)" and

"99% (totally different)". Participants used the mouse to manipulate the scale and indicate their judgment of how different the target sounds were.

2.5 Participants

Participants ($n = 24$) were native speakers of American English, between the ages of 18 and 32, with no known history of speech or language impairments. In return for their participation, they received either extra credit in a linguistics course, or small cash compensation.

3 Results

Mean difference ratings for each type of word context are displayed in Table 2. Because our hypothesis focused on how listeners responded to surface variation (and not to lack of variation), we present and analyze only "different" trials. For reference, the mean difference ratings for the "same" trials were: Contrast-only (*mound*) = 4.25 (10.41), Morphology+Phonology (*downed*) = 5.52 (14.53), Morphology (*township*) = 3.54 (8.37), Isolated = 8.08 (16.60).

To analyze these results, we used a linear mixed-effects model as implemented in the R function lme() from the package nlme. The outcome variable was the difference rating (a value from 0 to 99). The predictor value was the word type, and we used treatment coding with the Isolated condition as the baseline. Random intercepts were included for participant, item, comparison type between variants (a~b, a~c, b~c), target sound ([n], [ɹ], [l]), and syllable count (1, 2, 3). Models with random slopes failed to converge. Results of this analysis are in Table 3.

These results show that all three word conditions differed significantly from the Isolation baseline. Post-hoc comparisons using Tukey's test also revealed significant differences between the Contrast-only and Morphology + Phonology conditions ($\beta = 5.81$, std. error = 1.83, $z = 3.18$, $p < 0.008$), the Contrast-only and Morphology conditions ($\beta = -15.79$, std. error = 1.83, $z = -8.64$, $p < 0.001$), and the Morphology + Phonology and Morphology conditions ($\beta = -9.99$, std. error = 1.83, $z = -5.46$, $p < 0.001$).

Because the use of a sliding scale can vary from one individual to the next, we also calculated participant-specific z-scores. The mean difference ratings for the four conditions were: Contrast-only (*mound*) = 0.41,

Table 2: Difference ratings (means, standard deviations) given to surface variants of [n], [l], [ɹ] on "different" trials.

Roles	Difference rating
Contrast-only (*mound*)	36.49 (33.19)
Morphology + Phonology (*downed*)	30.69 (31.10)
Morphology (*township*)	20.70 (26.28)
Isolated	53.11 (30.67)

Table 3: Results of linear fixed-effects model on difference ratings, with Isolation as baseline.

Predictor	Value	Std. error	DF	*t*	*p*
Contrast-only (*mound*)	−16.35	2.09	1341	−7.83	<0.001
Morphology + Phonology (*downed*)	−22.16	2.09	1341	−10.62	<0.001
Morphology (*township*)	−32.15	2.09	1341	−15.40	<0.001

Morphology + Phonology (*downed*) = 0.21, Morphology (*township*) = –0.12, Isolated (*n*) = 1.00. A model using z-scores as the outcome variable produced results comparable to those in Table 3.

4 Discussion

Our results provide some qualified support for phoneme-based theories, but also raise further questions. To begin, difference ratings for phonetic variants of [n], [ɹ], [l] were highest in the Isolation condition, and significantly lower in all three word conditions. This is consistent with previous work demonstrating that lexical entries affect perception of individual speech sounds (e.g., Ganong 1980; Gow et al. 2008; Samuel 2001), and suggests that our experimental task functioned as intended.

Within the word contexts, difference ratings were significantly lower when the target sound participated in voicing agreement (*downed*) compared to when it did not (*mound*). This result suggests that participants perceived the [n] in *downed* words more categorically, and is consistent with the predictions of phoneme-based theories, which grant special status to a sound when it participates in an equivalence class.

The results for *township* words present a puzzle, however. If the phonological role of [n] in *downed* were the primary factor affecting listeners' perceptual judgments, we would expect *township* and *mound* to pattern similarly, because both contexts lack voicing agreement. If, on the other hand, the morphological role of [n] in *downed* were the primary factor, we would expect *township* and *downed* to pattern similarly, because both contexts contain a morpheme boundary. In fact, neither scenario was borne out: difference ratings in the morphology condition (*township*) were significantly lower than both the contrast-only (*mound*) and the morphology + phonology conditions (*downed*). Neither phoneme-based nor emergence theories predict this outcome.

As noted earlier, words like *township* contained unavoidable confounds which may be responsible for this result. In *township* words, the target sounds occupied coda-final position, while in words like *mound* and *downed*, they occupied coda non-final position. If listeners' judgments depend on syllable position, this confound could have affected our results. Also in *township* words, the target sounds were followed by a variety of other sounds (e.g., [ʃ], [f], [s]), while in *mound* and *downed* words, the targets were always followed by [d]. If listeners' judgments were affected by the heterogeneity of segmental environments in *township* words, this confound could have affected our results. Note, though, that our technique for constructing stimuli, using sounds that were synthesized and cross-spliced from non-coarticulated environments, should have mitigated any effect.

A potentially more interesting explanation concerns frequency of the root morpheme. Post-hoc analysis revealed that the mean log frequency of roots in the morphology + phonology condition (e.g., *down* in *downed*, *lure* in *lured*, *fill* in *filled*, etc.) was 1.26 (0.61) while that of roots in the morphology condition (e.g., *town* in *township*, *fear* in *fearful*, *ill* in *illness*, etc.) was 1.59 (0.87). If listeners are less sensitive to variation when it occurs at the boundary of a higher-frequency root, compared to a lower-frequency root, this could potentially explain the current results. Such a scenario would provide support for emergence theories, which explicitly incorporate word frequency as a core factor affecting phonological structure (e.g., Bybee and Hopper 2001), although it might also be accommodated within those phoneme-based theories that incorporate word exemplars (Pierrehumbert 2002). An interesting challenge would be to reconcile this exemplar-based scenario with the notion of equivalence classes, which was supported by our results for *mound* versus *downed* contexts. Future research could pursue these challenges using words with different morpho-phonological structures and controlled root frequencies.

Future research could also explore alternative methods for creating surface variants. Here, we manipulated synthesized voice settings, such that the difference between e.g. $[n]_a$ and $[n]_b$ amounted to a difference in talker source. Although we chose this method specifically in order to avoid creating variants that might be perceived as allophonic, it admittedly created a somewhat artificial listening situation, and represents just a first step into the investigation of this issue.

This study has shown that listeners' sensitivity to surface variation in speech sounds crucially depends upon the phonological and morphological roles played by those sounds within a word. Our results move beyond the basic distinction of word versus non-word contexts, and offer a new demonstration of the idea that language experience affects perception of speech sounds.

Acknowledgments: I thank associate editor Georgia Zellou, as well as an anonymous reviewer, for helpful comments that greatly improved the paper. I also thank audiences at Northwestern University, the meeting of the Berkeley Linguistics Society (2020), and the conference of the Association for Laboratory Phonology (2020).

Appendix

Stimulus words used in the experiment.

Target	Morphology + Phonology	Contrast-only	Morphology
[l]	boiled	weld	pollster
	sailed	wield	malware
	crawled	scold	frailty
	healed	guild	coolness
	smiled	yield	ailment
	ruled	mold	railway
	nailed	mild	gentleness
	muffled	shield	normalcy
	doubled	fold	concealment
	detailed	bald	rivalry
	stumbled	wild	annulment
	crippled	field	novelty
	strangled	gold	battleship
	revealed	cuckold	cruelty
	traveled	scaffold	royalty
	troubled	behold	specialty
	scrambled	ahold	settlement
	canceled	emerald	jewelry
[n]	toned	fend	kinship
	zoned	bland	township
	churned	mound	mournful
	yearned	gland	spoonful
	coined	strand	stainless
	pawned	astound	inward
	dined	stipend	spineless
	shined	rescind	brainless
	downed	amend	downward
	leaned	ascend	onward
	loaned	almond	sinful
	rained	commend	painless
	warned	descend	manhood
	joined	dividend	runway
	burned	vagabond	painful
	signed	correspond	womanhood
	learned	apprehend	weaponry
	ironed	comprehend	certainty
[ɹ]	bared	curd	boredom
	floored	fjord	fearful
	paired	shard	wireless

(continued)

Target	Morphology + Phonology	Contrast-only	Morphology
	aired	lard	starship
	toured	turd	fearless
	lured	nerd	cheerful
	squared	herd	careless
	stared	ward	careful
	poured	toward	awareness
	dared	yard	ownership
	scored	concord	tenderness
	matured	discord	membership
	obscured	discard	powerless
	impaired	accord	leadership
	explored	absurd	scholarship
	endured	award	powerful
	adored	aboard	neighborhood
	secured	leotard	wonderful

References

Boersma, Paul & David Weenink. 2018. Praat: Doing phonetics by computer (Version 6.0.24). Retrieved from: www.praat.org.

Bush, Nathan. 2001. Frequency effects and word-boundary palatalization in English. In Joan Bybee & Paul J. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 255–280. Amsterdam: John Benjamins Publishing.

Bybee, Joan. 2003. *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, Joan & Joanne Scheibman. 2008. The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics* 37(4). 575–596.

Bybee, Joan & Paul Hopper (eds.). 2001. *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins Publishing.

Ganong, William F. 1980. Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance* 6(1). 110–125.

Goldinger, Stephen D. & Tamiko Azuma. 2003. Puzzle-solving science: The quixotic quest for units in speech perception. *Journal of Phonetics* 31(3). 305–320.

Gow, David W., Jennifer A. Segawa, Seppo P. Ahlfors & Fa-Hsuan Lin. 2008. Lexical influences on speech perception: A Granger causality analysis of MEG and EEG source estimates. *NeuroImage* 43(3). 614–623.

Hume, Elizabeth & Keith Johnson (eds.). 2001. *The role of speech perception in phonology*. Leiden: Brill. Retrieved from: https://brill.com/view/title/23297.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Keith Johnson & John W. Mullenix (eds.), *Talker variability in speech processing*, 145–165. Cambridge, MA: Academic Press.

Kazanina, Nina, Jeffrey S. Bowers & William Idsardi. 2018. Phonemes: Lexical access and beyond. *Psychonomic Bulletin & Review* 25(2). 560–585.

Marian, Viorica, James Bartolotti, Sarah Chabal & Shook Anthony. 2012. CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE* 7(8). e43230.

Pierrehumbert, Janet B. 2002. Word-specific phonetics. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology*, vol. 7, 101–139. Berlin: Mouton de Gruyter.

Pierrehumbert, Janet B. 2016. Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics* 2. 33–52.

Pitt, Mark A. & Arthur G. Samuel. 1993. An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance* 19(4). 699–725.

Samuel, Arthur G. 2001. Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science* 12(4). 348–351.

Vitevitch, Michael S. & Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language* 40(3). 374–408.

Walsh, Michael, Bernd Möbius, Travis Wade & Hinrich Schütze. 2010. Multilevel exemplar theory. *Cognitive Science* 34(4). 537–582.