

JOURNAL ARTICLE

The voice of experience: Causal inference in phonotactic adaptation

Thomas Denby and Matthew Goldrick

Department of Linguistics, Northwestern University, Evanston, IL, US

Corresponding author: Thomas Denby (tdenby1@gmail.com)

Successfully grappling with widespread linguistic variation requires listeners to adapt to systematic variation in the environment while discarding incidental variation, based on listeners' prior experience. We examine the role of prior experience in phonotactic learning. Talkers who differ in their language background are more likely to vary in their phonotactic grammars than talkers who share a language variety. This predicts stronger adaptation to novel phonotactics when listeners are exposed to multiple talkers from different versus shared language backgrounds. We tested this by exposing listeners to two talkers, each of whom exhibited a different phonotactic constraint, in a recognition memory task. In Experiment 1, English listeners exposed to talkers differing in language background (English versus French) showed a greater degree of adaptation relative to cases where the talkers shared a language background (English or French). Experiment 2 found similar results when English listeners were exposed to talkers from different, non-native language backgrounds (Hindi versus Hungarian), suggesting that listeners make fine-grained distinctions between different non-native language phonotactics. These results suggest that phonotactic adaptation is flexible, but constrained by the fine-grained causal inferences listeners draw from their prior experience.

Keywords: Phonotactics; phonotactic learning; rational inference; adult language learning; accent adaptation; generalization

1. Introduction

In our day-to-day lives we encounter an enormous amount of linguistic variation. Individual speakers, for example, widely vary in their vowel productions (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995). Successfully navigating such widespread variation requires us to quickly and effectively adapt. Based on what we are currently experiencing in a given context, we must update our expectations to better match future input we will encounter in that context. This process of adaptation enables better prediction, which allows us to more efficiently process future events.¹ In the case of phonetics, we must adapt to novel speakers, dialects, languages, and other task-relevant properties that serve to distinguish different contexts. Such flexibility is critical to our ability to accurately perceive speech from different speakers and in different environments. The type of variation we encounter is not random, however—it is highly structured, with individual speakers, dialects, languages, and contexts all varying in different ways and to different degrees (Kleinschmidt & Jaeger, 2015). Properly attributing the source of variation to its underlying cause is critical to successful adaptation. To do so, speakers must use their

¹ We differentiate shorter-term *adaptation* from longer-term *learning* primarily based on the different time courses for each process.

prior experience with variation as a guide, making causal inferences about the source of variation. Doing so allows speakers to adapt to systematic variation for the task at hand, while ignoring variation incidental to the current task (Liu & Jaeger, 2018; Samuel, Brennan, & Kraljic, 2008).

For example, if someone hears a talker consistently produce an idiosyncratic [s] that sounds unusually like [ʃ] (e.g., *shick* instead of *sick*), adapting to that specific individual's [s] productions will be advantageous for perceiving that individual's speech in the future, as it is a stable property of the individual speaker. This is a form of systematic variation, guided by the listener's past experience with individual phonetic variation (e.g., Kraljic & Samuel, 2007).² If, on the other hand, the speaker happens to have a pen in their mouth while talking, the listener can infer that the source of the idiosyncratic [s] production may be due to an incidental factor: the obstruction from the pen. This incidental variation is unlikely to be predictive of the speaker's future speech in other contexts (i.e., when they do not have a pen in their mouth); as such, listeners are less likely to adapt under these conditions (Liu & Jaeger, 2018; Samuel et al., 2008). Critically, listeners do not completely disregard all causally ambiguous input (i.e., idiosyncratic productions when the talker has a pen in their mouth). Instead, they hold it in memory, as it may be predictive of future input in similar contexts (i.e., future productions when the talker has a pen in their mouth) or it may prove to be predictive after further disambiguating evidence (i.e., the talker produces the same idiosyncratic productions without a pen in their mouth; Kraljic & Samuel, 2011; Liu & Jaeger, 2018). In other words, adaptation requires listeners to properly attribute variation to its underlying source for the given task.

In this paper, we focus on the role of systematic versus incidental variation in adaptation to novel phonotactic constraints. Phonotactics—constraints on the possible sequences and positions of sounds within words and syllables—differ widely between languages, but much less so between individual speakers of a single language variety. English, for example, allows voiced plosives (i.e., [b], [d], and [g]) in syllable-final position; Dutch, on the other hand, does not allow voiced plosives in syllable-final position. While such phonotactic differences between speakers of Dutch and English are systematic, encountering two English speakers who differ in this way is unlikely. There are communicative constraints against widespread phonotactic variation between speakers within language varieties; if individual speakers differed in this way, it would lead to unreliable cues to word and syllable boundaries, resulting in frequent errors in lexical access (Pierrehumbert, 2001).

The underlying structure of phonotactic variation, and speakers' previous experience with this variation, likely plays a role in the ways speakers adapt to novel phonotactic constraints. Research over the past 20 years has found that speakers quickly adapt to novel phonotactic constraints (e.g., “[s, ʃ, f] are restricted to onset position, while [p,t,k] are restricted to coda position”) in both speech production (e.g., speech error patterns; Dell, Reed, Adams, & Meyer, 2000) and perception (e.g., memory error patterns; Bernard, 2015).

We explore the hypothesis that phonotactic adaptation is constrained by the types of causal inferences speakers make about the source of phonotactic variation. These causal inferences are based on speakers' prior experience with phonotactic variation: Speakers of different languages systematically differ, often quite drastically, in their phonotactics; while speakers of the same language varieties are unlikely to vary in this way. As such, we predict that when learners encounter such variation between speakers of a single language

² Note that these adaptation effects are also influenced by bottom-up acoustic similarity (Eisner & McQueen, 2005; Reinisch & Holt, 2014), interacting with prior experience to determine degree of adaptation (Kleinschmidt & Jaeger, 2015).

variety, they will infer it is incidental, rather than systematic. In other words, they will not attribute the source of the variation as being a durable, context-independent trait of the talker. This hypothesis predicts that when speakers are exposed to multiple talkers with distinct phonotactic grammars, either in perception or production, they will show a high degree of adaptation if those talkers clearly differ in their language background (e.g., one native Hindi talker and one native English talker), and a low degree of adaptation if they do not (e.g., two native English talkers). Indeed, the only previous study to examine adaptation to individual talkers who share a language variety (e.g., “Talker A doesn’t end their syllables in /f/; Talker B doesn’t end their syllables in /n/”) found that speakers did not adapt under these conditions (as assessed in production using a speeded repetition task; Onishi, Chambers, & Fisher, 2002).

Here, these predictions are tested using a perceptual phonotactic adaptation paradigm (Bernard, 2015; Denby, Schecter, Arn, Dimov, & Goldrick, 2018). Participants are exposed to two talkers, each of whom differs in their phonotactic grammar (e.g., “for Talker A, [s, ʃ, f] are restricted to onset position; for Talker B while [p,t,k] are restricted to coda position”). Crucially, in some conditions the talkers differ in their language backgrounds; in other conditions, the talkers share a language background (in other words, listeners should detect a shared or different ‘accent’ between the two talkers). In Experiment 1, English listeners are exposed to native French and English talkers exhibiting different phonotactic constraints. The results of Experiment 1A show that, as predicted, the highest degree of adaptation generally occurred when talkers differed in their language backgrounds, and the lowest degree occurred when both talkers were native speakers. Surprisingly, listeners adapted to a moderate degree when both talkers were *non-native* (i.e., two French talkers). This finding is replicated in Experiment 1B, which controls for phonetic differences between talkers and the difficulty of the learning task, as well as Experiment 2. This result may be due to listeners’ higher confidence that two native talkers are definitely speaking the same language, and less certainty about the shared versus distinct background of two non-native talkers.

In Experiment 2, we investigate the structure of listener knowledge of non-native phonotactic variation. English listeners are exposed to talkers of two non-native languages (Hindi and Hungarian). If listeners make distinctions within non-native phonotactic grammars, they should adapt when talkers differ in their language backgrounds. If, on the other hand listeners only distinguish between native versus non-native phonotactics, without further distinctions between non-native phonotactics, they will infer both non-native speakers share a single phonotactic grammar and show a small degree of adaptation. Results suggested listeners were sensitive to distinctions within non-native languages: Listeners adapted to a high degree when talkers differed in their language backgrounds, regardless of whether one of them was native (e.g., English talker versus Hindi talker) or not (e.g., Hungarian talker versus Hindi talker).

Together, these experiments aim to extend theories of the role of causal inference in adaptation into the domain of phonotactics and shed light on the mechanisms underlying the speed and flexibility of phonotactic adaptation.

2. Background

2.1. Phonotactics

Knowing the phonotactics of a language entails knowing real words in that language (e.g., English *flick*), as well as what constitutes possible words (*frick*), and what constitutes impossible words (*fnick*; Chomsky & Halle, 1965). This knowledge guides perception in profound ways, as it eliminates some options as possible words but not others. For example, Massaro and Cohen (1983) find that the same token, ambiguous between [r] and [l], is

perceived differently based on the legality of the phonotactic context in which it's heard: in the [t?i] context, it's more often perceived as [r]; in the [s?i] context, it's more often perceived as [l]. Phonotactics also influences word segmentation (McQueen, 1998), and a number of other perceptual processes (for a review, see Goldrick, 2011).

Given the importance of phonotactics in perceptual processes, an efficient learner should quickly adapt to novel phonotactic constraints to better guide perception in the future. Indeed, listeners can quickly learn artificial phonotactic constraints in experimental settings (e.g., Bernard, 2015, 2017; Denby et al., 2018; Onishi et al., 2002; Richtsmeier, 2011; Steele, Denby, Chan, & Goldrick, 2015). Bernard (2015), for example, exposed participants to a series of spoken syllables exhibiting an experimental constraint (e.g., [p] cannot appear in coda; [f] cannot appear in onset). Participants were asked after each syllable whether they had heard that syllable earlier in the experiment (no feedback was provided). After a number of repetitions of the exposure set, a handful of novel syllables were presented, half of which followed the constraint and half of which violated the constraint. Participants were more likely to false alarm on novel syllables that followed the constraint than those that violated it, suggesting participants were utilizing the novel constraint to make memory judgments.

Multiple findings with this paradigm confirm that this recognition memory task (used in this paper) taps into abstract, phonological learning mechanisms. It is sensitive to syllable structure, generalizing beyond the simple position of constrained segments (Bernard, 2015). Denby et al. (2018) show that learning in the task is sensitive to the diversity of contexts in which the forms appear (i.e., type frequency), inconsistent with simple exemplar accounts (e.g., Goldinger, 1998) that model recognition memory by aggregate activation of stored traces (i.e., token frequency).

2.2. Adaptation and variation

Perceptual adaptation is guided by the presence of variation in environment and the structure of this variation. In the perception of faces, for example, there is substantial variation in facial features across individuals. The structure of this variation impacts adaptation; learners adapt differently to novel face shapes that are similar versus dissimilar to faces they have previously experienced (e.g., Little & Apicella, 2016).

In the context of speech, adaptation is motivated by the huge amount of inter- and intra-speaker phonetic variation speakers have previously encountered (e.g., Hillenbrand et al., 1995). Listeners adapt to such variation: Nygaard and Pisoni (1998), for example, found that English listeners more accurately recognized words and sentences in noise for familiar talkers, suggesting they learn idiosyncratic features of that talker's speech and use that knowledge to guide perception of that talker in the future. Critically, talker variation is structured by higher-level factors (e.g., social structures; see Drager, 2010, for a review). These factors guide adaptation to novel speakers (see Kleinschmidt, 2018, for a review). For example, for native English listeners, exposure to Spanish-accented talkers improves recognition accuracy for novel Spanish-accented talkers, especially for words including Spanish vowels that are less characteristic of English (Sidaras, Alexander, & Nygaard, 2009).

Listeners can also use past experience with phonetic variation to make causal inferences about the source of variation they encounter. Kraljic, Samuel, and Brennan (2008) exposed English listeners to a talker producing ambiguous [s~ʃ] productions; previous work had shown that learners will adapt to such ambiguous productions. In one condition, listeners heard the ambiguous productions in an exposure phase with a video depicting the talker with a pen in their mouth, providing an incidental source for the acoustic variation. In a second condition, listeners were exposed to the same talker, but with a video depicting

the talker holding the pen in their hand—suggesting the acoustic variation reflected an idiosyncratic property of the talker. Consistent with a causal inference process, listeners showed significant adaptation in the pen in the hand but not in the pen in the mouth condition (see also Kraljic & Samuel, 2011; Liu & Jaeger, 2018).

Recent work in the rational learner framework has characterized results such as these by viewing adaptation as a process of uncovering the underlying structure that generates observable events and inferring causal relations that help to explain those events (Qian, Jaeger, & Aslin, 2012). Within this framework, the structured variability that forms the basis of our experience with language is encoded via a hierarchical indexical structure (Kleinschmidt & Jaeger, 2015; Pajak, Fine, Kleinschmidt, & Jaeger, 2016). For example, listeners could model structured phonetic variation by including different languages at the top of the hierarchy (e.g., Hindi, French), with native versus non-native accents one step below, followed by dialects within the native accent and sociolinguistic groupings (e.g., gender), with individual speakers at the bottom.

2.3. Phonotactic adaptation and variation

While the rational learner framework has generally been applied to talker adaptation, it makes markedly different predictions for phonotactic learning. This is because unlike talker variation, phonotactic variation is greatest across languages and smallest across individuals.³ We hypothesize that listeners will therefore assume that they should build separate models for speakers of different languages, while speakers within a dialect will be assumed to fall under a single model.

Consistent with the rational learner framework, previous studies suggest that participants treat language data encountered in an experimental context as a separate ‘lab language,’ distinct from the one used outside the lab. Warker (2013) exposed English-speaking participants to language data exhibiting novel, complex phonotactic constraints. As in previous studies of speech errors, these complex constraints required sleep-based consolidation to acquire and so did not influence speech error production until the second experimental session (see Anderson & Dell, 2018, for a review and meta-analysis). Warker (2013) found that the length of time between the first and second session did not significantly impact performance. Second session performance after one day was not significantly different than second session performance occurring a full week later. Participants retained their knowledge of the experimental constraints, despite the huge amount of conflicting evidence participants received from English in the intervening week between experimental sessions. The lack of sensitivity to intervening English experience suggests listeners may treat an artificial ‘lab language’ as a distinct language.

Our experiments apply the rational learner framework to the learning of multiple ‘lab languages.’ Based on listeners’ prior experience, we predict a greater degree of adaptation to talker-specific phonotactics when the talkers differ in their language background, and less adaptation when talkers have the same background. There could be several sources for such prior experience. Listeners may only require occasional, incidental exposure to non-native phonotactics (from either speakers of different languages, or accented speakers of their native language) to learn that different languages can have different phonotactic constraints. Many listeners would naturally come across such speech in their daily lives in an industrialized society such as the United States (Mechanical Turk workers, which is the population we sampled from, also have higher rates of education than the general

³ While phonotactics clearly varies across dialects, the extent of this variation is unclear. Staum Casasanto (2008) provides evidence that listeners’ processing of phonotactic variants is affected by speaker dialect, suggesting that phonotactics can vary across dialects. Quantifying this variation and examining its implications for phonotactic adaptation is a key area for future work.

U.S. population; Levay, Freese, & Druckman, 2016). Alternatively, listeners may require a high degree of exposure, such as having spent time learning a non-native language, or proficiency in two or more languages. To address this question, we analyze the self-reported language backgrounds of our listeners.

In Experiments 1A and B, we test the prediction that listeners will show more robust adaptation when two talkers differ in their language backgrounds, but not when they share a language background. Note that this distinction is predicted only if listeners can detect that the two talkers differ in the language background in the first place. As such, we predict that the degree of adaptation will be a function of how much evidence listeners have that the two talkers differ in language background. We examine this by manipulating the strength of the phonetic cue to language background. (Alternatively, if listeners can readily detect the background of non-native speakers based on weak cues, cue strength will not strongly impact performance.) Motivated by unexpected findings in Experiment 1A, Experiment 1B replicates several conditions, controlling for potential phonetic differences between talkers and the difficulty of the learning task. In Experiment 2, we use phonotactic adaptation to explore the structure of listeners' models of non-native phonotactics. Do listeners maintain models of only a native versus non-native grammar, or do they make distinctions between non-native languages?

3. Experiment 1A

In an artificial language paradigm, we expose English listeners to second-order constraints that require tracking talker information (e.g., Talker A's codas are restricted to [s, ʃ, f]; Talker B's codas are restricted to [p, t, k]), while manipulating the language background of the two talkers. The experiment contains four conditions: In the Native Shared condition, both talkers are native English speakers; in the Non-Native Shared condition, both talkers are French speakers; in the Weak Different and Strong Different conditions, one talker is a French speaker, while the other is a native English speaker. Each participant is exposed to a single pair of talkers in a between-participant design. The Weak Different and Strong Different conditions are distinguished by the strength of the acoustic cue to the French talker's language background: In the Strong condition, the French talker produces a vowel uncharacteristic of English (front rounded [y]); in the Weak condition, the French speaker produces the more English-like back rounded [u] vowel. Note that both the front rounded [y] and back rounded [u] French vowels are perceptually assimilated to [u] by native English listeners (Levy, 2009); that said, in both the Weak and Strong conditions, there are a number of cues to talker language background, as there are many phonetic differences between French and English beyond [y] relevant for the stimuli in this study. First, while French and English [i] are acoustically similar (Strange et al., 2007), French [u] is produced with a lower F2 (i.e., further back) than English [u], although this difference is likely not as large as that between French [y] and English [u] (Flege, 1987). Second, voicing distinctions for French plosives differ from those in English: French voiceless plosives are short-lag and unaspirated (i.e., short voice onset time), rather than long-lag (long voice onset time) and aspirated, as in English; and French voiced plosives are frequently pre-voiced (negative voice onset time) rather than short-lag, as in English (Caramazza & Yeni-Komshian, 1974). Third, coronal consonants—particularly plosives such as [t]—tend to be produced further forward in the mouth (i.e., as dental stops) in French than in English (Dart, 1998).

If listeners adapt based on their prior experience with phonotactic variation and talker language background, they should adapt to a greater degree in the Different conditions,

Table 1: Summary of conditions in Experiment 1A, along with experimental speaker language background, gender, and stimulus vowels. *English vowels* are denoted with italics; French vowels are denoted with underlines. Note that virtually all listeners were native English speakers.

	Native Shared	Non-Native Shared	Weak Different	Strong Different
Speaker Language Background	Shared <i>English</i>	Shared French	<i>English</i> versus French	<i>English</i> versus French
English versus French Vowels	<i>[i, u]</i>	<u>[i, y]</u>	<i>[i, u]</i>	<i>[i, u]</i>
	<i>[i, u]</i>	<u>[i, y]</u>	<u>[i, u]</u>	<u>[i, {u/y}*]</u>
Speaker Gender	Different	Different	Same	Same
Predicted Degree of Adaptation	Low	Low	Moderate	High

* N.B. For Strong Different condition, French syllables included [y] in familiarization syllables and [u] in generalization syllables.

since talkers who differ in language background are more likely to differ in their phonotactic grammars. Among the Different conditions, the two talkers are phonetically less distinct in the Weak condition; as such, listeners have less evidence that the two speakers do not share a language background. Thus we predict a greater degree of adaptation for the Strong Different condition than the Weak Different condition. In both Shared conditions, talkers do not differ in language background; listeners' prior experience should suggest that the talkers are unlikely to differ in their phonotactic grammars. As such, we predict the smallest degree of adaptation in these conditions. (See **Table 1** for summary of conditions and predictions.)

Participants are tested using a continuous recognition memory task (Bernard, 2015, 2017; Denby et al., 2018; Steele et al., 2015), in which they are auditorily presented with a series of syllables and asked whether they have previously heard each syllable within the experiment. Participants are first exposed to multiple repetitions of a set of *familiarization* syllables, all of which follow the phonotactic constraint (e.g., Speaker A says *fut*; Speaker B says *puf*). After the first four repetitions to the familiarization syllables, listeners hear nine more repetitions of the entire set of familiarization syllables, but now with a handful of novel *generalization* syllables mixed in. Half of these are legal (i.e., follow the phonotactic constraint), while the other half are illegal (i.e., violate the constraint; for example, Speaker A saying *tish*; Speaker B saying *tuk*). If listeners are tracking the constraint, generalization syllables that follow the constraint should seem more familiar than those that do not; as such, participants should be more likely to incorrectly believe they had previously heard legal generalization syllables. For example, a participant might hear Speaker A say *fut*, *kit*, *sik*, *tup*, etc., multiple times during familiarization. If that participant is tracking the constraint, during generalization they may believe they had previously heard *tut*, since syllables with similar phonotactic patterns (i.e., voiceless stops in coda position) appeared in familiarization. In contrast, participants should be unlikely to false alarm (i.e., incorrectly respond "yes") to *tus*, however, since no syllables spoken by Talker A in familiarization contained coda fricatives.

Note that speaker gender was also manipulated across conditions: In the Shared conditions, speakers differed in gender, while in the Different condition, speakers shared a gender. Much like accent, gender conveys sociolinguistic differences between speakers (e.g., Oh, 2011). This served as a control on phonetic and social distance between talkers in each condition: While talkers in the Different conditions were distinguished by their accent, talkers in the Shared conditions were distinguished by their gender. As such in each condition the two talkers differed along social and phonetic lines, either by gender or accent.

Power analyses, based on the results of an initial pilot study (see Appendix A), were run to approximate the number of participants required. The design and analysis of the experiment—including predictions, number of participants, stimulus design, and model structure—were defined before data collection in a pre-registration on the Open Science Foundation (OSF) platform (<https://osf.io/dbcqx/>). Stimuli, experimental lists, data, the listener language background questionnaire, and analysis files for all experiments can be found on the OSF (<https://osf.io/a6pju/>).

3.1. Participants

Based on the power analysis, 256 participants, split evenly between the four conditions (64 per condition), were required. However, participants had to pass a set of experimental criteria (see Data Analysis section) to ensure that they were adequately attending to the task. As such, participants were iteratively recruited until there were 64 participants who passed the criteria in each condition. A total of 455 participants were recruited through Amazon Mechanical Turk (AMT; Buhrmester, Kwang, & Gosling, 2011); of these, 260 (57.1%) passed the criteria. This passing rate was similar to previous studies using this paradigm over AMT (see Steele et al., 2015 and Denby et al., 2018). The task is quite difficult; participants performing substantially below our inclusion criteria performed at chance on the task (see Appendix B for full breakdown and discussion of participant passing rates). Due to limitations within our online framework and AMT, four participants who passed the criteria were exposed to a unique experimental list that a previous participant had been exposed to. Three of these participants were excluded; one such participant was included, however, in the Weak Different condition, as one unique experimental list did not have a participant due to experimenter error. Participants were required to have U.S. IP addresses, and were fluent speakers of English; 98.4% of participants who passed the criteria self-identified as native speakers of English, while two participants self-identified as speaking a non-North American dialect of English. 99.2% of participants who passed the criteria had no speech or hearing impairments. (Note that model results were qualitatively identical when non-native and hearing- and speech-impaired participants were excluded from the analysis.)

3.2. Stimuli

Stimuli were recorded in a soundproof booth at a 44.1 kHz sampling rate, and normalized to 60 dB SPL. Four talkers recorded stimuli: a female native English speaker; a male native English speaker; a female native French speaker; and a male native French speaker. Talkers produced syllables from orthographic representations of syllables on a monitor; orthography reflected the language background of the speaker. Both French talkers were multilingual, but were instructed to produce the syllables as though they were French, rather than English, words. Syllables were presented in a random order.

Stimuli consisted of consonant-vowel-consonant syllables with voiceless stops [p, t, k] and voiceless fricatives [f, s, ʃ] as onsets. Vowels for the English speaker are either [i] or [u]. For the French speakers, vowels are [i] and [u] in the Weak Different condition; in the Strong Different condition, vowels in familiarization syllables were [i] and [y]; for generalization syllables vowels were [i] and [u]. We confirmed that the strength manipulation was reflected in the stimuli through examination of F1 and F2 at vowel midpoints (see Appendix C). The result was a total of 108 possible syllables (6 onset consonants * 3 vowels * 6 coda consonants) recorded by French speakers, and 72 possible syllables (6 onsets * 2 vowels * 6 codas) recorded by English speakers (as English speakers only produced [u] and not [y]). Participants were exposed to 72 unique syllables in each condition.

3.3. Procedure

Participants were asked to fill out a demographic form that included information about their language background, geographic areas in which they had previously resided, whether they were a native or non-native speaker of English, and whether they had any hearing or language impairments. Participants were free to opt out of answering any questions.

To ensure listeners had a working audio set-up and basic fluency with English, an audio pre-test was administered in which listeners identified two English words spoken by a talker not involved in the rest of the experiment by typing the words with their keyboards.

Participants performed a recognition memory task. The question “Have you heard this before?” was on the screen for the entire experiment. On each trial, an auditory stimulus was presented. Participants answered the question by clicking a “Yes” or “No” button on the screen. After each click, there was a 500 ms interstimulus interval before the following stimulus played. The “Yes” and “No” buttons disappeared from the screen until the stimulus completed playing. Participants had unlimited time to answer the question, and no feedback was provided. There were no breaks in between experimental blocks; blocks were not demarcated in any way to the participant.

3.4. Design

Stimuli were split in half, into generalization and familiarization syllables (36 each), by onset-vowel pairs, and counter-balanced across participants. For example, Participant A hears the onset-vowel pair [tu_] in familiarization syllables (e.g., *toof*) and the onset-vowel pair [ti_] in novel generalization syllables (e.g., *teef*); the converse pattern holds for Participant B (e.g., [ti_] in familiarization; [tu_] in generalization). Onset-vowel pairs were [ti], [hi], [su], [pi], [ku], [fu] for one pattern, and [tu], [hu], [si], [pu], [ki], [fi] for the other. Among the 36 familiarization syllables, half (18) will end in fricatives, while half will end in stops. These subsets of 18 syllables will each be repeated by a different talker, such that a given talker will only repeat syllables ending in either fricatives or stops. Thus, during familiarization participants will be exposed to a phonotactic constraint linking manner in coda position (fricative versus stop coda) and speaker. Which talker produces which set is counterbalanced across participants. Among the 36 generalization syllables, each speaker produces half (18) of the set. Among this subset, half (9) follow the constraint established in the familiarization set, and half violate this constraint (i.e., both speakers say novel generalization syllables that end in both stops and fricatives).

The first four blocks of the experiment make up the familiarization phase. In each block, participants are exposed to the 36 familiarization syllables (half said by each speaker) in random order. In the generalization phase, there are nine further randomly ordered repetitions of the familiarization set, but each repetition is now intermixed with four generalization syllables. This results in a total of 504 trials (36 familiarization syllables * 13 blocks + 36 generalization syllables).

In both of the Shared conditions, the two talkers have different genders (e.g., male English and female English talker in Native Shared). In the Different conditions, however, talkers have the same gender; talker gender was counter-balanced across participant (e.g., Participant A hears a female French talker and a female English talker; Participant B hears a male French talker and a male English talker).

3.5. Data analysis

Following previous work (Denby et al., 2017; Steele et al., 2015), participants had to pass a set of criteria to ensure that they were adequately attending to the task: As in previous studies, during the generalization phase (blocks 5–13), participants must correctly accept

at least 90% of the syllables they had previously heard, and correctly reject at least 10% of the novel generalization syllables that they had not heard, regardless of whether the syllable is phonotactically conforming or not. (Note that loosening the criteria to include a greater number of participants does not qualitatively alter the results; see Appendix B). Participants who did not pass these criteria were excluded from the analysis.

Generalization data was analyzed using logistic mixed-effects regressions with maximal effects structures (Barr, Levy, Scheepers, & Tily, 2013). The dependent measure was the rate at which participants false alarmed (e.g., incorrectly responded “yes” to novel syllables). Fixed effects for the model consisted of *legality* and three contrast-coded terms: *language difference*, in which the Shared and Different conditions were contrasted; *strength*, in which the Weak and Strong Different conditions are contrasted; and *accent*, in which the two Shared conditions are contrasted. In addition, an interaction term was included between legality and each of the contrast-coded terms. Random effects included random intercepts and random slopes by legality for both participants and items (where ‘item’ was defined as individual tokens spoken by specific talkers; e.g., French male talker’s [tif]). Finally, a likelihood ratio test, between models with and without each contrast term as a fixed effect, was included to test for statistical significance (Barr et al., 2013).

We measure the degree of adaptation using the size of the *legality advantage*: the “yes” response rates to legal generalization syllables minus the “yes” response rate to illegal syllables. Our account predicts that listeners adapt when their prior experience suggests that the two talkers are likely to have different phonotactic grammars. This should yield an interaction between legality and the language difference terms, such that the legality advantage is larger in the Different conditions (i.e., when talkers differ in their language backgrounds) than in the Shared conditions. Further, as adaptation requires that listeners recognize the talkers as having different language backgrounds, we predict the legality advantage will be larger when the cue to language background is stronger (i.e., more robust adaptation in the Strong Different condition than the Weak Different condition), as shown by an interaction between legality and cue strength. Finally, listener behavior should not change between the two Shared conditions depending on whether the talkers are native or non-native speakers. In both Shared conditions (i.e., two French talkers or two English talkers) the talkers share a language background, and listeners should therefore infer they share a phonotactic grammar. As such, we predict no interaction between the legality and accent contrast term.

3.6. Results

A 95% confidence interval (CI) for each analysis of mean values was estimated using a bootstrap method, in which the distribution of a statistic is estimated by repeatedly resampling (1,000 times) from the observed data (with replacement).

Participants correctly accepted a mean of 91.0% of familiarization syllables (CI [90.6%, 91.3%]); participants falsely recognized (i.e., incorrectly responded “yes” to) 55.9% of novel generalization syllables (CI [53.2%, 58.6%]). The crucial measure, however, was the difference in the rate of false recognitions for legal versus illegal syllables, and whether this ‘legality advantage’ was modulated by talker language background. The mean legality advantage across participants was 12.5% (CI [10.5%, 14.7%]), replicating previous results showing that listeners show higher false recognition rates on novel legal syllables (i.e., syllables following constraints they’ve been previously exposed to) than novel illegal syllables. Moreover, the legality advantage is modulated by language background—as can be seen in **Figure 1**, the legality advantage is small in the Native Shared condition, and moderate in the other three conditions.

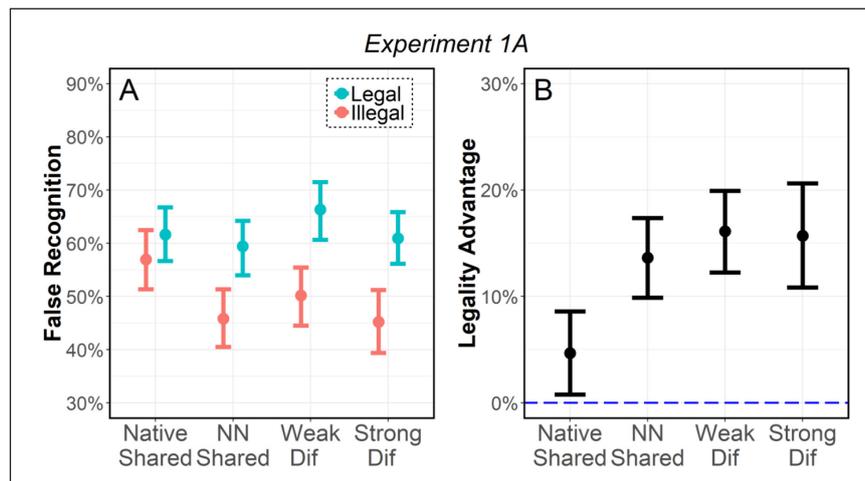


Figure 1: A: False recognition rates for legal and illegal generalization syllables in Experiment 1A. **B:** Legality advantage (false recognition rate on legal generalization syllables minus false recognition rate on illegal generalization syllables) for Experiment 1A. In both panels, error bars reflect bootstrapped 95% confidence intervals.

Our analysis showed a significant main effect of legality ($\beta = 0.64$, $SE \beta = 0.06$, $\chi^2(1) = 73.6$, $p < .0001$), as listeners were more likely to falsely recognize legal syllables over illegal syllables. In addition, there was a significant interaction of legality with the language difference contrast term ($\beta = 0.72$, $SE \beta = 0.20$, $\chi^2(1) = 12.6$, $p < .001$), as listeners showed a greater legality advantage in the Different conditions than in the Shared conditions. Legality also interacted with accent ($\beta = 0.45$, $SE \beta = 0.14$, $\chi^2(1) = 9.8$, $p < 0.01$), but not strength ($\beta = -0.05$, $SE \beta = 0.14$, $\chi^2(1) = 0.14$, $p = .70$). In other words, the legality advantage was greater in the Non-Native Shared condition than the Native Shared condition, but was not different across the Strong and Weak Different conditions. (Note: For all experiments, full model results including random effects structure can be found on <https://osf.io/rdez4/>.)

3.7. Experiment 1A discussion

Experiment 1A exposed listeners to talker-specific phonotactic constraints while modulating the language background of talkers. Listeners were able to successfully adapt within each condition, acquiring talker-specific constraints. Moreover, this adaptation was modulated by the language background of the talkers: Listeners showed a modest degree of adaptation when exposed to two talkers with a shared native language background (Native Shared condition), and a greater degree of adaptation if either or both talkers had a non-native language background (Strong Different, Weak Different, and Non-Native Shared). There was no difference in adaptation based on the strength of the cue to language background (Strong Different versus Weak Different), suggesting that even with the weaker cue to the non-native language background of talkers (i.e., the French [u] vowel, rather than [y]), listeners are confident of the non-native language background of the talker.

Counter to our predictions, however, adaptation was affected by language background even when both talkers shared a language: There was a greater degree of adaptation when both talkers shared a non-native language background than when they shared a native language background. Perhaps more surprising, adaptation was equally robust when talkers shared a non-native language background as when their language backgrounds *differed* (i.e., one native and one non-native talker). It is possible that any inclusion of talkers with a non-native language background increases listener confidence that talkers are speaking two different languages. This may be because of the asymmetry in listener

knowledge of their native phonetics versus non-native phonetics: Due to listeners' extensive knowledge of their native language, when they encounter two native speakers they are likely confident that those two speakers share a language (even when they are both speaking an artificial, non-native language). When listeners encounter two talkers with a shared non-native language background, on the other hand, they may be less confident that these talkers share a language background, given their relative paucity of experience with non-native (in this case, French) phonetics.

If the asymmetry in listener knowledge between native and non-native phonetics is driving the difference between the two Shared conditions, however, this asymmetry should also result in the greatest degree of adaptation for the Different conditions, which was not the case. That is, listener confidence of having encountered multiple languages should be highest when one of those languages is a native language.

There were two limitations of Experiment 1A that may have affected adaptation. First, the productions of the two French talkers showed markedly different pitch contours, with the female French speaker sometimes producing syllables with a flat pitch, but other times producing syllables with dramatic rises in pitch (particularly for syllables ending in fricatives). The male French speaker, on the other hand, much more consistently produced syllables with a flat pitch. This difference may have been salient enough that listeners inferred that the two talkers did not share a language background, increasing the legality advantage in the Non-Native Shared condition. The differences in pitch contour may have been due to differences during recording, or the different backgrounds of the two talkers. The male French speaker was 23 years old, and had lived in the United States for less than a year. He was from Paris, and self-identified as speaking a standard dialect of French. The female French speaker was 41 years old, had lived in the United States for 13 years, was from south of France, and identified as speaking a non-standard dialect of French.

To address this limitation, in a follow-up experiment replicating three of the four conditions in Experiment 1A (see below), we recorded a novel female French speaker, whose language background was more similar to that of the male French speaker, and who was instructed to imitate the male speaker's productions to ensure phonetic similarity across speakers. As such, we predict a lower degree of adaptation in the Non-Native Shared condition in Experiment 1B than in 1A. Note that the Native Shared condition was not replicated, as it was the only condition that could not have been affected by the aberrant pitch contours of the female French speaker (since it included only English speakers).

A second limitation of Experiment 1A was that in the Strong Different condition, listeners were exposed to familiarization syllables that included the uncharacteristic French [y] vowel; generalization syllables, however, had the French [u] vowel. This was intended to provide a more direct comparison across the Strong and Weak conditions by ensuring that generalization sets were identical across conditions. However, this design may have also attenuated adaptation in the Strong Different condition, given that [u] is a weaker cue to talker language background. Moreover, it increased the phonetic distance between familiarization and generalization sets, as listeners encountered a novel French vowel in the generalization set that was not present in familiarization syllables. Low false recognition rates for syllables in the Strong Different condition spoken by a French talker and containing [u] (38.1%) reflected this. This is lower than false recognition for French syllables containing [i] (63.7%) in the Strong Different condition, as well as French syllables containing [u] in the Weak Different condition (60.0%).

To address this limitation, in the Strong Different condition of Experiment 1B, familiarization and generalization syllables spoken by French talkers contained matching vowels. If the increased phonetic distance in the Strong Different condition depressed the

legality effect for that condition, we would predict a greater degree of adaptation for the Strong Different condition in Experiment 1B than in 1A.

4. Experiment 1B

In Experiment 1A, participants unexpectedly adapted to the same degree in the Non-Native Shared as they did in the Different conditions. In Experiment 1B, three of the conditions from Experiment 1A were replicated (Strong Different, Weak Different, and Non-Native Shared) while two limitations of the previous experiment were addressed that may have caused the unexpected results.

4.1. Participants

As in Experiment 1A, participants were iteratively recruited from AMT until there were 64 participants in each of the three conditions who passed the experimental inclusion criteria. A total of 418 participants were recruited, of which 192 (46.4%) passed the criteria. Of the speakers who passed the criteria, 98.9% identified as native English speakers. All participants identified as having no speech or hearing impairments. No participant identified as speaking a non-American dialect of English. (Note that model results were qualitatively similar when non-native participants were excluded from the analysis.)

4.2. Stimuli

Stimuli from three of the four talkers were identical to that in Experiment 1A; however, stimuli from a novel female French speaker were recorded to replace the stimuli of the female French speaker from Experiment 1A. In a soundproof booth, the novel female French speaker heard each of the male French speaker's productions in random order over headphones. After the male speaker's production was played, she was instructed to imitate it; each syllable was also provided in French orthography, and appeared on a monitor after the audio had finished playing.

The novel female French speaker was 24 years old, grew up in the southwest of France, lived in Paris as an adult, and had lived in the United States for less than a year at the time of recording. She self-identified as speaking a standard dialect of French as an adult, despite having grown up speaking a non-standard dialect (Southwestern French).

In the Strong Different condition in Experiment 1B, vowels spoken by French speakers in both familiarization and generalization syllables were always [i] or [y]. (This differed from the Strong Different condition in Experiment 1A, in which French speakers used [i] and [y] in familiarization syllables, but [i] and [u] in generalization syllables.) Acoustic analysis confirmed the strength manipulation (see Appendix C). Stimuli were otherwise identical to those in Experiment 1A.

4.3. Data analysis

Significance was assessed using a logistic mixed-effects regression identical to that in Experiment 1A, with the exception of a fixed effect for accent, which was not included (there was no Native Shared condition in Experiment 1B). The model had fixed effects of legality, language difference (i.e., Non-Native Shared versus both Different conditions) and strength (i.e., Weak versus Strong Different conditions). An interaction term was included between legality and both contrast-coded terms; random effects included random intercepts and random slopes by legality for both participants and items.

We predict a significant difference between the Different conditions and the Non-Native Shared condition, as shown by an interaction between legality and the language difference terms. We also predict a significant interaction between legality and the strength contrast term.

4.4. Results

Participants correctly accepted a mean of 90.3% of familiarization syllables (CI [89.7%, 90.9%]); participants falsely recognized (i.e., incorrectly responded “yes” to) 60.2% of novel generalization syllables (CI [57.3%, 63.1%]). The mean legality advantage across participants was 19.4% (CI [17.1%, 21.8%]). Critically, the legality advantage is modulated by language background—as can be seen in **Figure 2**, the legality advantage is moderate in the Non-Native Shared condition, and large in the Different conditions.

A logistic mixed effects regression found a significant main effect of legality ($\beta = 1.02$, $SE \beta = 0.07$, $\chi^2(1) = 127.6$, $p < .0001$), as well as language difference ($\beta = -1.06$, $SE \beta = 0.19$, $\chi^2(1) = 28.82$, $p < .0001$), as listeners were more likely to falsely recognize legal syllables, as well as syllables in the Non-Native Shared condition. In addition, there was a significant interaction of legality with the language difference contrast term ($\beta = 0.57$, $SE \beta = 0.18$, $\chi^2(1) = 9.81$, $p < .01$), as listeners showed a greater legality advantage in the Different conditions than in the Shared condition. Legality did not interact with strength ($\beta = -0.02$, $SE \beta = 0.15$, $\chi^2(1) = 0.03$, $p = .87$), as the legality advantage was not significantly different across the Strong and Weak Different conditions.

4.5. Discussion

Experiment 1B replicated the adaptation to talker-specific phonotactic constraints found in Experiment 1A, with listeners adapting in each condition. Moreover, listeners adapted to a greater degree when talkers differed in their language background (Different conditions) than when they shared a non-native language background (Shared Non-Native condition), unlike in Experiment 1A. This provides evidence that the difference in language background between talkers is critical, as opposed to the simple presence of non-native talkers.

We further predicted that the changes in stimulus design to Experiment 1B would result in (a) an increase in the legality advantage for Strong Different condition due to consistent vowels across generalization and familiarization syllables, and (b) a decline in the legality effect for the Non-Native Shared condition due to the increased phonetic similarity of the two French talkers. While the legality advantage for the Strong Different condition did increase across Experiments 1A and 1B (from a mean of 15.7% to 23.1%), a similar increase was found in the Weak Different condition (from 16.1% to 22.7%), suggesting the change in the design of the Strong Different condition was not the cause of this

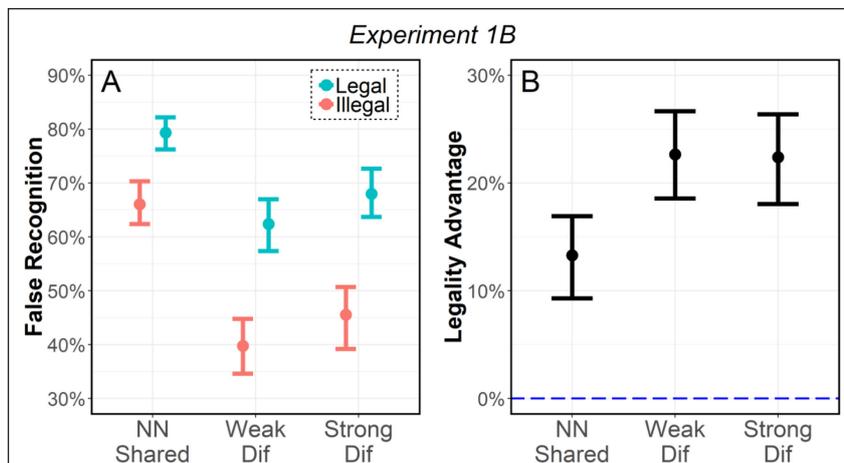


Figure 2: **A:** False recognition rates for legal and illegal generalization syllables in Experiment 1B. **B:** Legality advantage (false recognition rate on legal generalization syllables minus false recognition rate on illegal generalization syllables) for Experiment 1B. In both panels, error bars reflect bootstrapped 95% confidence intervals.

increase. In addition, for the Non-Native Shared condition, the legality effect was roughly equivalent across Experiments 1A and 1B (a mean of 13.6% in 1A and 12.7% in 1B), counter to our prediction.

The inclusion of a novel female French speaker in Experiment 1B appears to account for the increased legality advantage in the Different conditions: Listeners who heard two male speakers in the Different conditions showed a similar legality advantage across the two experiments (a mean of 20.2% in 1A and 19.3% in 1B); listeners who heard two female speakers, however, showed a substantially higher legality advantage in Experiment 1B (a mean of 25.8%) than in 1A (11.6%).

Why did the inclusion of the novel female French speaker increase adaptation in both Different conditions, without lowering adaptation in the Non-Native Shared condition (as we originally predicted)? It's possible that the female French speaker's anomalous pitch contours in Experiment 1A were distracting, shifting listener attention away from the segmental level differences between speakers. This could have reduced the distinction between the female native English and native French talkers, depressing the legality advantage in the Different conditions in 1A.

The increase of adaptation in the Different conditions, whatever the cause, suggests that the low-level phonetic properties of talkers, and the differences or similarities between talkers, affect listener inferences about talker language background. Replicating this experiment with novel talker pairs and languages (as we do below) is necessary to ensure that the pattern of adaptation found in Experiment 1 was, in fact, spurred by differences in language background, rather than the result of arbitrary individual variation.

Finally, as in Experiment 1A, there was no difference in adaptation based on the strength of the cue to language background, providing further evidence that the 'weak' cue stimuli are sufficient for listeners to detect the talker's language background. As noted above, this likely reflects the numerous phonetic differences between stimuli, above and beyond the vowel distinction.

5. Experiment 2

The results of Experiment 1, particularly the results of Experiment 1B, support our structured model of phonotactic variation, with native and non-native languages each having separate phonotactic grammars. In Experiment 2, we investigate whether listeners assign different phonotactic grammars to different non-native languages (as well as their native language). Alternatively, monolingual listeners may only assign a single phonotactic grammar to their native language, and a single phonotactic grammar to all non-native languages.

Using a similar design and recognition memory paradigm to that in Experiments 1A and B, we expose listeners to two talkers, each of whom exhibits a different novel phonotactic pattern. We conceptually replicate two conditions of Experiments 1A and B using novel stimuli, speakers, and languages (Hindi and Hungarian). In the Mixed Different condition listeners are exposed to one native English speaker, and one non-native speaker (either Hindi or Hungarian), broadly replicating the design of the Different conditions in Experiment 1. In the Non-Native Shared condition, listeners are exposed to two non-native speakers who share a language background (either two Hindi speakers or two Hungarian speakers). To address the structure of listener knowledge, we include a novel condition: In the Non-Native Different condition, listeners are exposed to two non-native speakers who differ in their language background (one Hindi speaker and one Hungarian speaker). To ensure that listeners can clearly tell talkers apart, the two talkers have different genders in each condition.

Hindi and Hungarian were chosen specifically because they are phonologically and phonetically distinct from English as well as from one another and do not have restrictions on the coda consonants used in both experiments. For instance, Hungarian has the

front-round [y] (and Hungarian speakers were instructed to produce it, similar to the French speakers in the Strong Different condition in Experiment 1), while Hindi has [u]; Hungarian produces [t] as alveolar, while Hindi has the dental [t̪]; among many other segmental and suprasegmental differences.

The within non-native distinctions hypothesis and the native versus non-native hypothesis make identical predictions in the Non-Native Shared condition—moderate adaptation, following the results of Experiment 1—and the Mixed Different condition—a high degree of adaptation. Replicating these results with novel speakers and languages should provide further evidence that talker language background affects adaptation. In the Non-Native Different condition, however, the two hypotheses make differing predictions. The within non-native distinctions hypothesis predicts a high degree of adaptation in the Non-Native Different condition—significantly higher than that in the Non-Native Shared condition—with listeners inferring that different non-native languages have different phonotactic grammars. The native versus non-native hypothesis, on the other hand, predicts a similar, moderate degree of adaptation in the Non-Native Different and Non-Native Shared conditions, as under this hypothesis listeners don't distinguish between different non-native phonotactic grammars. (See **Table 2** for summary of conditions and predictions.)

The design and analysis of the experiment—including predictions, number of participants, stimulus design, and model structure—were defined before data collection in a pre-registration on the Open Science Foundation platform (<https://osf.io/rdez4/>).

5.1. Participants

A total of 192 participants were required (64 participants for each of the three conditions). To reach 192 participants who passed the experiment criteria (see below), 441 participants were recruited, 202 of whom passed the criteria (45.8%). Ten participants who passed the criteria were exposed to an experimental list a previous participant had been exposed to and as such were excluded. As in Experiment 1, participants were required to have a U.S. IP address. Native English speakers made up 98% of participants, while one participant self-identified as speaking a non-North American dialect of English. Only a small minority (1.6%) of participants self-identified as having speech or language impairments. All model results were qualitatively identical when non-native and participants and those with impairments were excluded.

5.2. Stimuli

Stimuli were recorded in a soundproof booth at a 44.1 kHz sampling rate and normalized to 60 dB SPL. Six talkers recorded stimuli, with one male and one female speaker for three languages: English, Hungarian, and Hindi. Talkers produced disyllables from orthographic representations of disyllables on a monitor; orthography reflected the language background

Table 2: Summary of conditions in Experiment 2, along with experimental speaker language background, gender, and stimulus vowels. NN stands for non-native. Note that virtually all listeners were native English speakers.

	Non-Native Shared	Non-Native Different	Mixed Different
Language Background	Shared Hindi or Hungarian	Hindi versus Hungarian	English versus (Hindi or Hungarian)
Gender	Different	Different	Different
Predicted Degree of Adaptation	Moderate	Within NN: High Native versus NN: Moderate	High

of the speaker (a transliterated orthography was used for Hindi). All Hindi and Hungarian talkers were bilingual, but were instructed to produce stimuli as disyllables in their native language, rather than English. Disyllables were presented in a random order.

Given the added difficulty of detecting differences in talker language background between talkers of two non-native languages, stimuli consisted of disyllables rather than monosyllables to provide listeners with greater phonetic evidence of talker language background. The syllables making up the disyllabic stimuli in Experiment 2 were a subset of those used in Experiment 1. Consonants consisted of voiceless stops [p,k] and voiceless fricatives [f, ʃ]; [t] and [s], which can form complex onsets in the second syllable, were not used to ensure that each individual syllable was parsed as consonant-vowel-consonant. For English and Hindi speakers, vowels consisted of [i] and [u]; for Hungarian speakers, vowels consisted of [i] and [y]. This resulted in a total of 32 monosyllables (4 onsets * 2 vowels * 4 codas).

We created 64 disyllabic stimuli by splitting the 32 monosyllables into four groups of eight, counterbalanced for coda pattern (fricative versus stop) and onset/rhyme pattern (onset [k,f] matched with rhymes [uf, ih, uk, ip] versus onset [h,p] matched with rhymes [if, uh, ik, up]). These groups of eight are further split in two, such that each group has an even distribution of segments in each position. Each subgroup of four is crossed to create 32 disyllables (4 syllables * 4 syllables * 2 positions). Among the resulting 128 disyllables, all disyllables with gemination and reduplication are removed, and subsets were chosen such that syllables appeared an equal number of times in both positions within each group, for a total of 64 disyllables. Note that speakers put stress on the first syllable in both languages.

5.3. Procedure

The procedure was identical to that of Experiment 1.

5.4. Design

Stimuli were split in half, into generalization and familiarization disyllables (32 each), by onset-vowel/coda pairings, and counter-balanced across participants. For example, in disyllables that include syllables ending in codas [k,f], Participant A hears the onset-vowel pair [fu_] (e.g., *fookpeek*) in familiarization disyllables, and the onset-vowel pair [fi_] in novel generalization disyllables (e.g., *feekpook*). Participant B hears the converse pattern (e.g., *feekpook* in familiarization; *fookpeek* in generalization). Among the 32 familiarization disyllables, syllables in half (16) end in fricatives, while syllables in the other half end in stops.

The sets were split in half again into subsets of eight, such that each syllable only appears once in each position (e.g., *fi* appears once as the first syllable and once as the second syllable). To decrease the overall confusability of the sets, participants hear each speaker produce only one subset of eight in familiarization (although twice as often; see below), while the other matching subset is withheld. As in Experiment 1, each speaker repeats familiarization disyllables that end in a different coda pattern (e.g., Speaker A ends their syllables in stops; Speaker B in fricatives); which talker produces which set is counterbalanced across participants.

Among the 32 generalization syllables, each speaker produces half (16) of the set. Among this subset, half (16) follow the constraint established in the familiarization set, and half violate this constraint (i.e., both speakers say novel generalization disyllables that end in both stops and fricatives).

The first two blocks of the experiment consists of the familiarization phase. In each block, participants are exposed to two repetitions of each of the 16 familiarization disyllables (half said by each speaker) in random order, for a total of 32 tokens per block.

Pilot testing suggested that due to the increased similarity of tokens in this experiment, two repetitions of each disyllable per block were required to ensure adequate levels of recognition performance on the familiarization tokens. In the generalization phase, these randomized sets of 32 tokens are repeated in eight further blocks; each generalization block also includes four intermixed generalization disyllables. This results in a total of 352 trials (16 familiarization disyllables * 2 repetitions/block * 10 blocks + 32 generalization syllables).

5.5. Data analysis

As in Experiment 1, participants had to pass a set of criteria to ensure that they adequately attended to the task. To achieve similar overall passing to those in Experiment 1, given the increased confusability of the familiarization set in Experiment 2, the criteria for performance were slightly lowered: Participants had to correctly accept at least 85% of familiar items (as opposed to 90% in Experiment 1). As in Experiment 1, participants had to correctly reject at least 10% of the novel generalization items that they had not heard. However, the first two criteria could result in a participant passing who, for example, correctly accepted familiarization items 85% of the time, but who correctly rejected generalization items only 10% of the time (i.e., false alarmed 90% of the time on novel words). That participant would be responding “yes” more often on generalization syllables (90%) than on familiarization syllables (85%), suggesting they were unable to sufficiently differentiate between familiar and novel items. Therefore we added a third criteria to catch such cases: Participants could not incorrectly respond “yes” on generalization items (i.e., false alarm) more often than they correctly responded “yes” to familiarization items.

Generalization data was analyzed using a logistic mixed-effects regression. Fixed effects included *legality* and two contrast-coded terms: *language difference*, in which the Non-Native Shared condition was contrasted with the two Different conditions; and *non-native language background*, in which the Non-Native Shared and Non-Native Different conditions contrasted with the Mixed Different condition. Furthermore, the model included an interaction term between legality and each of the contrast-coded terms. The random effects structure included random intercepts and random slopes of legality by both participants and items.

The within non-native hypothesis predicts a significant difference between the Non-Native Shared and the two Different conditions, as indicated by the interaction term between legality and the non-native term; the native versus non-native hypothesis does not predict such a difference. Such a difference would indicate that listeners showed a larger legality advantage in the Different conditions, despite one of these conditions including speakers of two different non-native languages.

5.6. Results

Participants correctly accepted 89.3% of familiarization disyllables (CI [88.7%, 90.0%]) and falsely recognized 69.8% of generalization syllables. The mean legality advantage was 14.0% (CI [11.3%, 16.5%]). Crucially, as shown in **Figure 3**, the difference in language background modulates the legality advantage: Similar to Experiment 1B, the legality advantage is moderate in the Non-Native Shared condition, and large in both Different conditions.

The results from the logistic mixed effects regression show a main effect of legality ($\beta = 0.76$, $SE \beta = 0.08$, $\chi^2(1) = 62.83$, $p < .0001$), showing that listeners were more likely to false alarm on legal disyllables. The interaction between legality and language difference was also significant ($\beta = 0.71$, $SE \beta = 0.21$, $\chi^2(1) = 11.1$, $p < .001$), as listeners showed a greater legality advantage on the Different conditions than in the Shared condition.

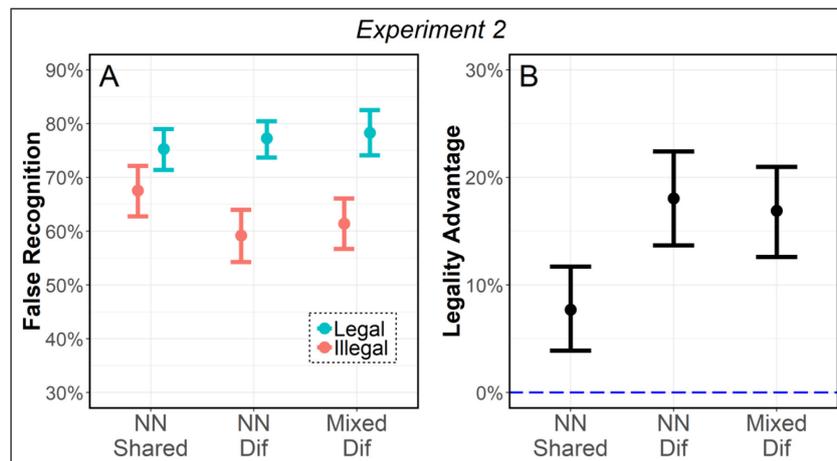


Figure 3: **A:** False recognition rates for legal and illegal generalization syllables in Experiment 2. **B:** Legality advantage (false recognition rate on legal generalization syllables minus false recognition rate on illegal generalization syllables) for Experiment 2. In both panels, error bars reflect bootstrapped 95% confidence interval.

Legality did not interact with the non-native background term ($\beta = -0.04$, $SE \beta = 0.21$, $\chi^2(1) = 0.03$, $p = 0.87$), as listeners did not show a larger legality advantage in the two Non-Native conditions versus the Mixed condition. (For full model results, see <https://osf.io/rdez4/>.)

5.7. Discussion

Listeners in Experiment 2 adapted to talker-specific constraints in each condition. This replicates findings from Experiments 1A and B using novel talkers, languages, and stimulus design, providing further evidence that listeners can adapt to talker-specific constraints. As in Experiments 1A and B, the degree of adaptation was modulated by the language background of the talkers: Listeners showed a high degree of adaptation when talkers differed in the language background (Mixed Different and Non-Native Different conditions), and a low-to-moderate degree of adaptation when talkers shared a language background (Non-Native Shared).

Listeners adapted at a similar rate in both Different conditions, regardless of whether they were exposed to one Hindi and one Hungarian talker (Non-Native Different) or one English and one Hindi/Hungarian talker (Mixed Different). This suggests that listeners make distinctions between different non-native phonotactic grammars, and assign different phonotactic grammars to different non-native languages. In other words, if the phonetics of two languages are perceptibly different—regardless of whether they are native or non-native languages—listeners can infer that those languages have separate phonotactic grammars.

While Experiment 2 replicated the relatively higher legality advantage in Different versus Shared conditions found in Experiment 1, the overall legality advantages are lower in Experiment 2 (e.g., in Experiment 1B the mean legality advantage in the Different conditions is 22.5%; in Experiment 2 it's 17.5%). To the extent that these differences in effect sizes between experiments are meaningful, it is likely due to differences in the designs of the two experiments. In Experiment 2, the stimulus set was much more confusable than in Experiment 1. This likely caused the relatively high overall false recognition rate (57.7% in Experiment 1; 69.8% in Experiment 2). This also may have lowered the legality advantage, as participants may have begun to hit a ceiling on false recognition rates for legal syllables.

6. Performance across experiments

6.1. Overall effect of different language backgrounds

Pooling the results across all experiments, which includes 640 participants who passed the experimental criteria, 14 talker pairs, four languages, and two experimental designs, the 384 participants in all Different conditions have a mean legality advantage of 18.7% (CI [17.0%, 20.5%]), nearly twice as high as the 256 participants in all Shared conditions (9.9%, CI [7.9%, 11.8%]); see **Figure 4**). This suggests that listeners do indeed adapt to a higher degree when listeners differ in their language backgrounds, due to the causally unambiguous source of phonotactic variation in those conditions.

Alternative explanations for the difference in legality advantage between Shared and Different conditions—such as idiosyncrasies of particular talker combinations—seem unlikely, given the relatively large difference between conditions, general consistency of the overall effect, and variety of talkers, talker pairs, and languages across experiments and conditions. That said, the results from Experiment 1A serve as an exception to this pattern, with listeners in the Non-Native Shared condition adapting at a similar rate to those in the Different conditions. Replacing a talker with idiosyncratic productions with a different talker in Experiment 1B resulted in a higher rate of adaptation in the Different conditions, possibly because listeners interpreted the two female talkers in 1A as sharing a language background. This suggests that individual talker characteristics can have a large effect on listener adaptation. In addition, other predictions, like those between Weak and Strong Different conditions were not met. As such, this phenomenon requires replication and further investigation—in particular, listeners should be exposed to a greater number of talker pairs, given possible listener sensitivity to fine-grained phonetic differences between talkers. The Weak/Strong distinction also relied on a relatively narrow phonetic difference (a single vowel difference of [u] versus [y]); it's possible that other more salient phonetic differences may result in a stronger effect.

6.2. Listener language background analysis

Results from Experiments 1 and 2 strongly suggest that previous experience with non-native languages, and the phonotactic variation that different languages exhibit, constrain listeners' adaptation to novel non-native phonotactics. *How much* experience with

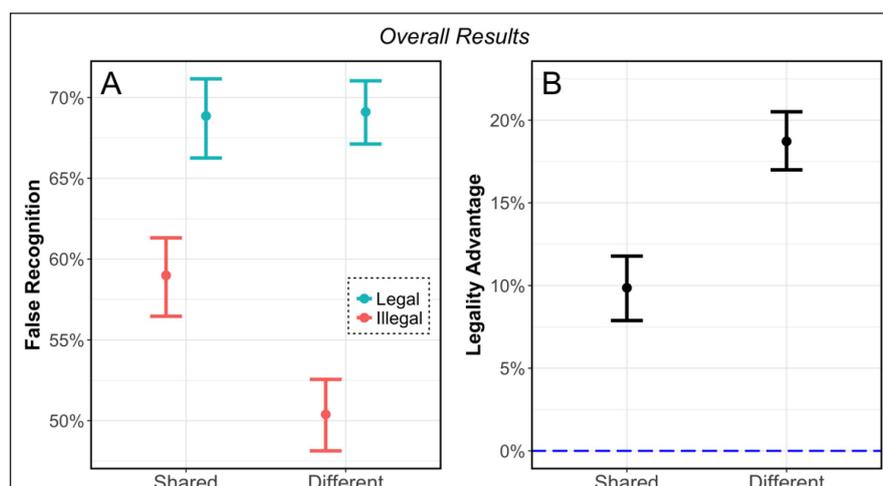


Figure 4: **A:** False recognition rates for legal and illegal generalization syllables in Different and Shared conditions, pooled across all experiments. **B:** Legality advantage (false recognition rate on legal generalization syllables minus false recognition rate on illegal generalization syllables) for all Different and Shared conditions. In both panels, error bars reflect bootstrapped 95% confidence interval.

non-native languages is necessary to make such inferences? It's possible that the threshold is quite low, with monolingual speakers able to make such inferences through their daily exposure to non-native languages. Alternatively, multilingual speakers may more readily make these inferences based on their past experience learning languages.

Participants reported their language backgrounds on a questionnaire before taking the experiment (see the OSF site for the full questionnaire). Participants reported the languages they know, their age of acquisition, and the length of time speaking those languages. Any participants with an age of acquisition of five years old or earlier were classified as having early second language (L2) experience. A minority (33.8%, $N = 216$) participants reported speaking at least one language other than English, while 7.7% (49) had early L2 experience. Among participants in Experiment 1, in which participants were exposed to French speakers, 6.7% (30) reported knowing some amount of French. Among the participants in Experiment 2, who were exposed to Hindi and Hungarian speakers, none reported knowing Hindi or Hungarian.

Mixed-effects regressions were used to assess differences based on L2 experience. Separate models were run for early L2 experience, any L2 experience, and French L2 experience. For early L2 and any L2 experience, data was pooled over both experiments; For French L2 experience, only data from Experiment 1 was included. None of these models showed an overall effect of L2 experience, nor did they show an effect of L2 experience on the legality advantage.

These results suggest that a relatively small degree of exposure to non-native phonotactics is required to make inferences about talkers' phonotactic grammars based on their language background. This is not surprising given listeners' sensitivity to non-native phonotactics, even in infants as young as nine months old (Mattys & Jusczyk, 2001). Listeners also take into account talker phonotactics when judging speaker accentedness. When listeners hear speech in the speaker's L2, sequences that are legal in the speaker's native language (L1) are deemed less accented than sequences illegal in the speaker's L1 (Park, 2013). In other words, monolingual listeners are highly sensitive to non-native phonotactics, and are likely to attend to such non-native patterns when they appear in the input, even if that input is relatively limited. However, it's also possible our self-reported measures of L2 proficiency are unreliable (Tomoschuk, Ferreira, & Gollan, 2018), and we do not have a large sample of listeners with L2 experience (especially early L2 experience). Future work should investigate the relationship between listener language background and inferences about phonotactic variation directly, by comparing bilingual and monolingual populations.

7. General discussion

Previous research within the perceptual adaptation literature has frequently been couched within the rational learner framework (e.g., Liu & Jaeger, 2018): Listeners use their past experience to model the underlying structure that generates variation in speech forms, and make causal inferences based on this structure when exposed to novel input. In the case of phonotactics, there is massive variation between the phonotactic systems of distinct languages, and relatively little variation within a single dialect. In the case of phonotactics, then, the rational learner framework predicts that listeners leverage this past experience with phonotactic variation to make causal inferences about novel phonotactic constraints during adaptation. Across three experiments, this prediction was largely confirmed. Experiments 1A and B found a high degree of adaptation to novel phonotactic constraints by English listeners when talkers differed in language background (French versus English). In contrast, adaptation was moderate when talkers shared a language background different from that of the listeners (two French talkers) and low, in Experiment

1A, when talkers shared the listeners' native language background (two English talkers). Experiment 2 showed that the high degree of adaptation in cases where two speakers differ in their language background generalizes to non-native languages (Hindi versus Hungarian) as well. This pattern of results supports the hypothesis that listeners make distinctions between non-native phonotactic grammars, and use this information when making inferences about whether or not talkers shared a phonotactic grammar.

While learning was stronger when talkers differed in language background, learning also occurred when talkers shared a language background (c.f. Onishi et al., 2002). Note that learning was significantly weaker than the adaptation observed with different language backgrounds. This suggests that listeners' assumptions that same-language talkers should have the same phonotactic constraints can be overridden (to a degree) by bottom-up phonotactic patterns in the novel input. Additionally, among the shared conditions, the Non-Native Shared condition showed a greater degree of adaptation relative to the Native Shared condition. This is likely due to the asymmetry between listener knowledge of non-native versus native phonetics. Native English listeners have less knowledge of the French phonetic system than the English one; therefore while the listeners may perceive two French speakers as phonetically similar, listeners won't be as confident as they are for two English speakers.

7.1. Phonotactics and L2 acquisition

Why is phonotactic adaptation so rapid, robust, and flexible? In these experiments, listeners were able to simultaneously adapt to two distinct, complex (i.e., second-order) phonotactic constraints within a single short experimental session, showing sensitivity to different non-native languages and even individual speakers. One possibility is that phonotactics are a critical tool in the earliest stages of L2 acquisition. This may be because phonotactic constraints guide speech perception by limiting the number of lexical and phonological candidates listeners have to consider (e.g., listeners perceive ambiguous sounds as the option that results in a legal, rather than illegal, sequence; Massaro & Cohen, 1983). This may be particularly important when speech perception is less accurate in the early stages of acquisition. Phonotactics also act as an important cue in word segmentation (McQueen, 1998), which in turn is a precursor to lexical acquisition. Indeed, some evidence suggests that adult listeners learn novel L2 words with high phonotactic probability more easily than those with low probability (Storkel, Armbrüster, & Hogan, 2010).

The learners in these experiments adapted to subset phonotactics, where English(-like) sounds are more constrained than they are in English. This appears to be fairly easy relative to other types of adaptation (e.g., acquiring perceptual distinctions between two L2 sound categories that assimilate into a single L1 category; Best, McRoberts, & Goodell, 2001). Adapting to a novel language's subset phonotactics may serve as a cognitively inexpensive adjustment that aids listeners with some of the most important early tasks of acquisition: speech perception, word segmentation, and lexical acquisition. If this is the case, we would predict that learners who are able to more successfully adapt in phonotactic adaptation experiments would also be more successful in early L2 acquisition.

7.2. Language background detection

Listeners are capable of perceiving remarkably fine-grained differences in language background (Atagi & Bent, 2013). However, in cases where listeners are exposed to single words with subset phonotactics, as in the current experiments, this is a harder task (Park, 2013). How were listeners able to do this, particularly in Experiment 2, in which they were tasked with distinguishing speakers from two non-native language backgrounds? The task in our experiment may be easy to perform; listeners only must

(implicitly) decide whether two speakers have the same or distinct language backgrounds. Furthermore, listeners are not forced to make this decision on each single item (as in Park, 2013); they are able to build their representation of the speaker's language background over the course of the experiment. Better understanding how listeners make decisions about language background is a key area for future work.

7.3. Future directions

We hypothesized that learners are recruiting their L2 acquisition faculties in phonotactic adaptation, and treating the laboratory exposure as a novel language. According to our account, this motivates adaptation, as learners are able to separate their experience with their native language from the novel experimental input, and therefore rapidly adapt to novel constraints. If learners believe they are being exposed to speech from their native language, however, they should be less likely to adapt, as they have extremely strong priors about their native language phonotactic constraints from a lifetime of experience. This prediction could be tested by exposing learners to selected speech from their native language. For example, real word stimuli could be presented in sentential context, or accompanied by corresponding pictures. If rapid phonotactic adaptation is part of the process of L2 acquisition we would predict low rates of adaptation in such cases.

Dialect differences may present another case where we expect low rates of adaptation. While there is some evidence that dialects can differ in phonotactic constraints (Staum Cassanto, 2008), the high degree of lexical and phonological overlap between dialects presumably leads to small differences in phonotactic structure. A key question for future work is to develop more precise estimates of listeners' prior knowledge of differences between dialects versus language; the resulting quantitative predictions would provide a novel test of our account.

Another question that arises from this research involves the nature of explicit listener knowledge about talkers. First, future studies in this paradigm should employ a post-experiment survey probing listeners' explicit beliefs about the language background of the talkers they were exposed to. Beyond probing listener knowledge, however, is the additional question of how presenting explicit information about talkers as part of the experimental design would impact adaptation. In speech perception research, modifying listener expectations about talker characteristics such as dialect, even in subtle ways, can have important consequences for speech perception (e.g., Hay & Drager, 2010). Even manipulating the number of speakers that listeners expect to hear can affect processing of the same linguistic input (Magnuson & Nusbaum, 2007). In the current studies, we hypothesized that adaptation is induced by the learner's belief that they are being exposed to different languages. This belief about talkers' language backgrounds comes from phonetics alone, as participants are given no explicit information about the experimental talkers whatsoever. If we presented explicit information that speakers differed in their language backgrounds (e.g., "Barbara grew up speaking English in Ohio, while Béla grew up speaking Hungarian in Budapest") we might expect it to strengthen participants' confidence that the talkers do not share a language background, and thus boost adaptation.

Alternatively, presenting explicit information that talkers share a language background may dampen adaptation. In some conditions, it may be the case that listeners are already fully confident in their beliefs about talkers' language backgrounds, suggesting that their confidence could not be increased further by top-down information. In the Different conditions, for example, listeners may have already been fully confident that talkers differed in those cases, as modulating the 'non-nativeness' of the phonetic vowel cues in the Strong versus Weak Different conditions did not change the degree of adaptation. In

this case, explicit information that talkers differed in their language backgrounds may not affect adaptation. Information that talkers share a language background, however, may decrease adaptation. In the Shared conditions, listeners appeared to show varying degrees of confidence about the language backgrounds of talkers. We might expect in the Non-Native Shared conditions, for example, that if listeners are explicitly told the two speakers share a language background, the degree of adaptation might decrease. In this same context, pushing participants in the reverse direction, by giving them information that the two talkers *differ* in their language backgrounds, might override the phonetic similarities of the two talkers and increase adaptation.

Finally, we could ask what other domains this link between talker language background and underlying grammar extend to. Speakers experience variation at every level of linguistic representation. Many of these domains may hold a similar structure in variation to phonotactics—a high degree of variation between talkers of different language varieties, and a low degree of variation between talkers of the same language variety. As such, we would expect the same principles of causal inference to apply. For example, this inference may extend to learning of novel or unlikely syntactic or morphological structures. In the case of artificial language paradigms (e.g., Schumacher, Pierrehumbert, & Lashell, 2014), if stimuli are presented by non-native talkers, it may boost adaptation. For adaptation in native language contexts (e.g., Jaeger & Snider, 2013), if stimuli are presented by talkers of different dialects it may also increase adaptation, as syntax may vary to a greater degree between speakers of different dialects (e.g., Labov, 1969) than it does between individuals within a speech community.

8. Conclusion

In three experiments, we have shown that listeners use their prior experience with phonotactic variation—that languages vary in their phonotactics much more than individual speakers of the same dialect—to guide their adaptation to novel phonotactic constraints. Listeners evaluate the underlying structure generating phonotactic variation, and exhibit a large degree of adaptation to systematic sources of phonotactic variation (i.e., listeners who differ in their language background), and a smaller degree of adaptation to incidental sources of variation (i.e., listeners who share a language background). This effect extends to differences between different non-native languages. Together, these results illuminate a core linguistic ability: appropriately adapting to our dynamic language environment based on our prior experience.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** Study 1 Power analysis. DOI: <https://doi.org/10.5334/labphon.267.s1>
- **Appendix B.** The impact of changes to participant inclusion criteria on the results of study 1. DOI: <https://doi.org/10.5334/labphon.267.s2>
- **Appendix C.** Acoustic properties of French vowels from Experiments 1A and 1B. DOI: <https://doi.org/10.5334/labphon.267.s3>

Acknowledgements

Thanks to Ann Bradlow, Jennifer Cole, Melissa Baese-Berk, Chun Chan, and members of the SoundLab for helpful comments and discussion.

Funding Information

Supported by National Science Foundation Grant BCS 1728173.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Both authors contributed to the study design. Testing, data collection, and analysis were performed by the first author. The first author drafted the manuscript; both authors revised and approved the final version for submission.

References

- Anderson, N. D., & Dell, G. S. (2018). The role of consolidation in learning context-dependent phonotactic patterns in speech and digital sequence production. *Proceedings of the National Academy of Sciences*, *115*, 3617–3622. DOI: <https://doi.org/10.1073/pnas.1721107115>
- Atagi, E., & Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics*, *41*, 509–519. DOI: <https://doi.org/10.1016/j.wocn.2013.09.003>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *44*, 255–278. DOI: <https://doi.org/10.1016/j.jml.2012.11.001>
- Bernard, A. (2015). An onset is an onset: Evidence from abstraction of newly-learned phonotactic constraints. *Journal of Memory and Language*, *78*, 18–32. DOI: <https://doi.org/10.1016/j.jml.2014.09.001>
- Bernard, A. (2017). Novel phonotactic learning: Tracking syllable-position and co-occurrence constraints. *Journal of Memory and Language*, *96*, 138–154. DOI: <https://doi.org/10.1016/j.jml.2017.05.006>
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794. DOI: <https://doi.org/10.1121/1.1332378>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. DOI: <https://doi.org/10.1177/1745691610393980>
- Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, *2*, 239–245. DOI: [https://doi.org/10.1016/S0095-4470\(19\)31274-4](https://doi.org/10.1016/S0095-4470(19)31274-4)
- Chomsky, N., & Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics*, *1*(2), 97–138. DOI: <https://doi.org/10.1017/S0022226700001134>
- Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of Phonetics*, *26*, 71–94. DOI: <https://doi.org/10.1006/jpho.1997.0060>
- Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1355–1367. DOI: <https://doi.org/10.1037/0278-7393.26.6.1355>
- Denby, T., Schecter, J., Arn, S., Dimov, S., & Goldrick, M. (2018). Contextual variability and exemplar strength in phonotactic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 280–294. DOI: <https://doi.org/10.1037/xlm0000465>
- Drager, K. (2010). Sociophonetic variation in speech perception. *Linguistics and Language Compass*, *4*(7), 473–480. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00210.x>
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*, 224–238. DOI: <https://doi.org/10.3758/BF03206487>

- Flege, J. E. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65. DOI: [https://doi.org/10.1016/S0095-4470\(19\)30537-6](https://doi.org/10.1016/S0095-4470(19)30537-6)
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279. DOI: <https://doi.org/10.1037/0033-295X.105.2.251>
- Goldrick, M. (2011). Utilizing psychological realism to advance phonological theory. In J. Goldsmith, J. Riggle, & A. Yu (Eds.) *The handbook of phonological theory*, 2nd edition (pp. 631–660). Oxford: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444343069.ch19>
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, *48*(4), 1689–1699. DOI: <https://doi.org/10.1515/ling.2010.027>
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, *97*(5), 3099–3111. DOI: <https://doi.org/10.1121/1.411872>
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition*, *127*, 57–83. DOI: <https://doi.org/10.1016/j.cognition.2012.10.013>
- Kleinschmidt, D. F. (2018). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, *34*(1), 43–68. DOI: <https://doi.org/10.1080/23273798.2018.1500698>
- Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *148*–203. DOI: <https://doi.org/10.1037/a0038695>
- Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15. DOI: <https://doi.org/10.1016/j.jml.2006.07.010>
- Kraljic, T., & Samuel, A. G. (2011). Perceptual learning evidence for contextually-specific representations. *Cognition*, *121*(3), 459–465. DOI: <https://doi.org/10.1016/j.cognition.2011.08.015>
- Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332–338. DOI: <https://doi.org/10.1111/j.1467-9280.2008.02090.x>
- Labov, W. (1969). *The study of nonstandard English*. Urbana, IL: National Council of Teachers of English. DOI: <https://doi.org/10.1016/B978-0-12-754850-0.50014-3>
- Levy, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of Mechanical Turk samples. *Sage Open*, *6*(1). DOI: <https://doi.org/10.1177/2158244016636433>
- Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults’ French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670–82. DOI: <https://doi.org/10.1121/1.3224715>
- Little, A. C., & Apicella, C. L. (2016). Face adaptation in an isolated population of African hunter-gatherers: Exposure influences perception of other-ethnicity faces more than own-ethnicity faces. *Psychonomic Bulletin and Review*, *23*(2), 439–444. DOI: <https://doi.org/10.3758/s13423-015-0919-z>
- Liu, L., & Jaeger, T. F. (2018). Inferring causes during speech perception. *Cognition*, *174*(January), 55–70. DOI: <https://doi.org/10.1016/j.cognition.2018.01.003>
- Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 391–409. DOI: <https://doi.org/10.1037/0096-1523.33.2.391>

- Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, 34(4), 338–348. DOI: <https://doi.org/10.3758/BF03203046>
- Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78(2), 91–121. DOI: [https://doi.org/10.1016/S0010-0277\(00\)00109-8](https://doi.org/10.1016/S0010-0277(00)00109-8)
- McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, 39(1), 21–46. DOI: <https://doi.org/10.1006/jmla.1998.2568>
- Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60(3), 355–376. DOI: <https://doi.org/10.3758/BF03206860>
- Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39, 59–67. DOI: <https://doi.org/10.1016/j.wocn.2010.11.002>
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83(1), 13–23. DOI: [https://doi.org/10.1016/S0010-0277\(01\)00165-2](https://doi.org/10.1016/S0010-0277(01)00165-2)
- Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning additional languages as hierarchical probabilistic inference: Insights from first language processing. *Language Learning*, 66(4), 900–944. DOI: <https://doi.org/10.1111/lang.12168>
- Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, 41(2), 78–87. DOI: <https://doi.org/10.1016/j.wocn.2012.11.001>
- Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, 16(5–6), 691–698. DOI: <https://doi.org/10.1080/01690960143000218>
- Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to represent a multi-context environment: More than detecting changes. *Frontiers in Psychology*, 3, 228. DOI: <https://doi.org/10.3389/fpsyg.2012.00228>
- Reinisch, E., & Holt, L. L. (2014). Lexically guided phonetic retuning of foreign-accented speech and its generalization. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 539. DOI: <https://doi.org/10.1037/a0034409>
- Richtsmeier, P. T. (2011). Word-types, not word-tokens, facilitate extraction of phonotactic sequences by adults. *Laboratory Phonology*, 2(1), 157–183. DOI: <https://doi.org/10.1515/labphon.2011.005>
- Sidasaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *The Journal of the Acoustical Society of America*, 125(5), 3306. DOI: <https://doi.org/10.1121/1.3101452>
- Staum Casasanto, L. (2008). Does Social Information Influence Sentence Processing? *Proceedings of the Annual Conference of the Cognitive Science Society*, 30, 799–804.
- Steele, A., Denby, T., Chan, C., & Goldrick, M. (2015). Learning non-native phonotactic constraints over the web. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*.
- Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2010). Differentiating phonotactic probability and neighborhood density in adult word learning, 49(6), 1175–1192. DOI: [https://doi.org/10.1044/1092-4388\(2006/085\)](https://doi.org/10.1044/1092-4388(2006/085))
- Strange, W. (2007). Cross-language phonetic similarity of vowels. In O. S. Bohn & M. J. Munro (Eds.) *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 35–55). Amsterdam: John Benjamins. DOI: <https://doi.org/10.1075/llt.17.08str>
- Tomoschuk, B., Ferreira, V. S., & Gollan, T. H. (2018). When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition*, 1–21. DOI: <https://doi.org/10.1017/S1366728918000421>

Warker, J. A. (2013). Investigating the retention and time course of phonotactic constraint learning from production experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(1), 96–109. DOI: <https://doi.org/10.1037/a0028648>

How to cite this article: Denby, T., & Goldrick, M. 2021 The voice of experience: Causal inference in phonotactic adaptation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12(1): 5, pp. 1–28, DOI: <https://doi.org/10.5334/labphon.267>

Submitted: 09 March 2020 **Accepted:** 25 November 2020 **Published:** 24 February 2021

Copyright: © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 