

NORTHWESTERN UNIVERSITY

The Voice of Experience: Causal Inference in Phonotactic Adaptation

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Linguistics

By Thomas Denby

EVANSTON, ILLINOIS

June 2019

Abstract

Successfully grappling with the widespread linguistic variation of daily life requires speakers to adapt to systematic variation in the environment while discarding incidental variation, based on their prior experience. In the case of phonotactics, speakers' prior experience is that talkers who differ in their language background are likely to vary in their phonotactic grammars, while talkers who share a language variety are unlikely to do so. As such, we predict that when speakers are exposed to multiple talkers whose phonotactics vary, and those talkers differ in their language background, listeners will infer the variation is systematic and adapt. Conversely, if the talkers share a language background, listeners will infer the variation is incidental, and not adapt.

In Study 1, we tested this prediction in a perception experiment, by exposing listeners to two talkers, each of whom exhibited a different phonotactic constraint, in a recognition memory task. In Experiment 1, when listeners were exposed to talkers who differed in their language background (1 English vs. 1 French talker), they showed a high degree of adaptation; when the talkers shared a language background (2 English or 2 French talkers), listeners showed a low-to-moderate degree of adaptation. In Experiment 2, we examined the granularity of listener knowledge of variation in non-native phonotactics by including a novel condition with two non-native talkers (1 Hindi vs. 1 Hungarian talker). Listeners showed a high degree of adaptation even when both talkers were non-native speakers with different language backgrounds, suggesting that listeners make distinctions between different non-native language phonotactics.

In Study 2, we examine the role of causal inference in speech production. Recent work suggests adaptation in production may differ from perception, as production may utilize simple associative learning mechanisms that may not take high-level indexical features into account. We explore this question using a modified tongue twister paradigm, in which participants repeat syllable sequences from two model talkers, with each talker exhibiting a different phonotactic constraint. Mirroring Study 1, model talkers either shared a non-native language background; shared a native language background; or differed in their language background (the languages backgrounds in question were German and English). In addition, a control condition was included following previous tongue twister experiments, in which the phonotactic constraint was conditioned on the identity of the adjacent vowel. Results were largely inconclusive—there was some evidence of increased adaptation when participants were exposed to model talkers with different language backgrounds, but the effect was inconsistent. In addition, no effect was found in the control condition.

Together, these results suggest that phonotactic adaptation is flexible, but constrained by the causal inferences listeners draw from their prior experience, particularly in perception.

Acknowledgements

It's hard to look back over the last six years and adequately put into words, or even fully account for, the profound help my friends, family, and colleagues have given me to get to this point. I'll do my best.

First and foremost, I'd like to thank my advisor and mentor, Matt Goldrick. It's been an honor to work with you, and I felt so profoundly lucky so often in my graduate career to have your expert guidance, calm demeanor, and infinite wisdom in my corner. You're a model of what a good mentor, teacher, and researcher should be, and I feel so thankful to have had that model throughout grad school. I'll miss our meetings, spent untangling knotty ideas and making obscure Simpsons references. I would not be the researcher and thinker I am without your help, nor would this dissertation have been possible. Thank you for everything. (And sorry about all those deadlines I missed...)

Thank you Ann Bradlow and Jennifer Cole, my committee members, for your expertise and critical eye. You each consistently challenged my assumptions in a way that made me think more deeply about my work. Your experience and acumen made the dissertation stronger, and your encouragement and positivity buoyed my spirits along the way.

Thank you to the National Science Foundation, for their generous funding and support (grant #1728173).

Thank you to the members of SoundLab, past and present, who heard me talk about this work in its various forms for 5 years and provided invaluable feedback along the way

(and listened to me complain about hand-coding speech errors every week). Special thanks to lab members Emily Cibelli, Erin Gustafson, Nicole Mirea, and Maria Gavino in particular for your advice, support, and friendship. Thank you to all of my fellow Ling grad students over the years, for their camaraderie and support.

Thank you to Chun Chan, whose technical wizardry and ability to solve difficult logistical problems made huge portions of this project possible. Thank you also to the many brilliant linguists who gave me advice and ideas for novel conditions along the way, especially Melissa Baese-Berk and Gary Dell.

Thank you to Kat Hall, my cohort-mate, confidant, and best bud. Taking this journey with you has been one of the great joys of my life. You were my rock throughout graduate school, and hanging out with you never failed to fill me with joy and appreciation for what a strong, brilliant, and silly person you are. To Sean, thank you for your companionship, your warmth, and your endless goodness. You and Kat (and Abbey and Miles!) are my Chicago family, and I love you.

Thank you to my other cohort-mate Nayoun Kim, for being the perfect office buddy and always supporting me in the toughest times. I'll always think back on our giggly, punch-drunk days in Swift 309 spent agonizing over our dissertations with fondness.

Thank you to my many Chicago friends for keeping me sane, healthy, and happy over the last six years. To Laura, John, and Taylor, for being my favoritest drinking buddies and generally the most refreshing, comforting, wonderful human beings in Chicago. To Libby, for being with me through all the ups and downs (#sadnesstwinsforever). To the

founding members of TatLPD (Emily Cibelli, Angela Cooper, Erin Gustafson, Cindy Blanco, Julie Matsubara), I'll miss our jam sessions.

Thank you to Brianna Kaufman and Alex Kapelman, my long-distance best bros, who supported me and let me blow off steam countless times on the phone. You are the truest friends I could ever ask for.

Thank you to all four of my parental units. To Mom and Dad, my words will do no justice to what you have given me, but here goes. Thank you for always being there for me, always supporting me and my education, always lending me an empathetic ear. Thank you for stoking my intellectual curiosity, for teaching me how to write and think critically, for pushing me to do more, and for picking me up when I fall. Most of all, thank you for kindness, and for your love. To Janet and Susan, thank you for being the most wonderful, supportive, and affectionate step parents I could ever ask for. I am filled with gratitude that you both came into my life, and I know I'm smarter, kinder, and more open because of it.

Finally, thank you to my partner Cassandra Rose, who has been by my side for every up and down of this project (and there were many). You spent endless hours sitting on the couch with me while we both clacked away on our laptops, supporting my work, listening to me kvetch, and boosting my spirits with frequent bouts of extreme silliness. Your patience and your goodness and your kindness are unlimited. You are the best person I know, and I am so lucky to have you in my life. I love you and I like you.

Table of Contents

ABSTRACT.....	2
ACKNOWLEDGEMENTS	4
TABLE OF CONTENTS.....	7
LIST OF FIGURES.....	11
LIST OF TABLES	13
1. INTRODUCTION	15
2. STUDY 1.....	22
2.1. INTRODUCTION.....	22
2.2. BACKGROUND.....	27
2.2.1. <i>Phonotactics</i>	27
2.2.2. <i>Adaptation and Variation</i>	29
2.2.3. <i>Phonotactic Adaptation and Variation</i>	31
2.3. EXPERIMENT 1A.....	33
2.3.1. <i>Participants</i>	37
2.3.2. <i>Stimuli</i>	38
2.3.3. <i>Procedure</i>	39
2.3.4. <i>Design</i>	40
2.3.5. <i>Data Analysis</i>	41
2.3.6. <i>Results</i>	43

2.3.7. Experiment 1A Discussion.....	45
2.4. EXPERIMENT 1B	48
2.4.1. Participants.....	49
2.4.2. Stimuli.....	49
2.4.3. Data Analysis	50
2.4.4. Results.....	51
2.4.5. Discussion.....	52
2.4.6. Experiment 1 conclusion.....	54
2.5. EXPERIMENT 2	55
2.5.1. Participants.....	57
2.5.2. Stimuli.....	57
2.5.3. Procedure	59
2.5.4. Design.....	59
2.5.5. Data Analysis	60
2.5.6. Results.....	62
2.5.7. Discussion.....	63
2.6. LISTENER LANGUAGE BACKGROUND ANALYSIS.....	65
2.7. GENERAL DISCUSSION.....	68
2.7.1. Phonotactics and L2 Acquisition.....	70
2.7.2. Accent Detection.....	72
2.8. CONCLUSION	73
3. STUDY 2.....	74
3.1. INTRODUCTION.....	74

3.2.	BACKGROUND	77
3.2.1.	<i>Phonotactic learning in speech production</i>	77
3.2.2.	<i>Causal inference in phonotactic adaptation in production</i>	79
3.2.3.	<i>Current Study</i>	81
3.3.	METHODS	84
3.3.1.	<i>Participants</i>	84
3.3.2.	<i>Materials</i>	85
3.3.3.	<i>Procedure</i>	88
3.3.4.	<i>Analysis</i>	89
3.4.	RESULTS	90
3.4.1.	<i>Discussion</i>	102
3.4.2.	<i>Production-perception dynamics and phonotactic adaptation</i>	107
3.5.	CONCLUSION	108
4.	CONCLUSION	110
4.1.	STUDY 1	111
4.2.	STUDY 2	115
4.3.	FUTURE DIRECTIONS	117
4.4.	CONCLUSIONS	122
5.	APPENDIX	123
5.1.	APPENDIX A – STUDY 1 VOWEL ACOUSTICS ANALYSIS	123
5.2.	APPENDIX B – STUDY 1 PILOT STUDY	125
5.3.	APPENDIX C – STUDY 1 POWER ANALYSIS	127
5.4.	APPENDIX D – STUDY 1 PASSING RATES	128

5.5.	APPENDIX E – STUDY 1 MODEL RESULTS.....	134
5.6.	APPENDIX F – TONGUE TWISTER SAMPLE SIZE ANALYSIS	136
5.7.	APPENDIX G – STUDY 2 MODEL RESULTS.....	138
5.7.1.	<i>Within-condition models; all sessions; [m]-[n] swaps included.....</i>	<i>138</i>
5.7.2.	<i>Within-condition models; [m]-[n] swaps excluded</i>	<i>139</i>
5.7.3.	<i>Within-condition models; session 1 excluded; [m]-[n] swaps excluded</i>	<i>140</i>
6.	REFERENCES.....	142

List of Figures

Figure 2.1. 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.	44
Figure 2.2. 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.	51
Figure 2.3. 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.	62
Figure 3.1. Overall error rates by participant. Colors reflect experimental condition.	92
Figure 3.2. Percentage of errors that maintain their syllable position for constrained vs. unconstrained consonants, broken down by session and condition. All data visualized, including [m]-[n] swaps. Error bars reflect 95% confidence interval over participant; however, note that participants contributed different numbers of errors to each bar. 96	
Figure 3.3. Percentage of errors that maintain their syllable position for constrained vs. unconstrained consonants, broken down by session and condition. [m]-[n] swaps were excluded. Error bars reflect 95% confidence interval over participant; however, note that participants contributed different numbers of errors to each bar. In the Vowel condition, one participant made the majority of the errors but had a much lower mean than other participants, resulting in a CI that does not overlap with the mean for Session 1, unconstrained errors.....	99
Figure 5.1. Scatterplot of female French talkers' first and second vowel formants. Each point is a vowel, with vowel identity indicated by the appropriate IPA symbol. Color indicates talker differences.....	124
Figure 5.2. Scatterplot of hit rate (%yes on familiar items) in generalization phase by legality advantage (false alarm rate for legal items minus false alarm rate on illegal items) for all experiments in Study 1. Each dot represents a single participant; colors represent whether participants passed or failed criteria. Lines represent Loess regression; shading represents 95% confidence interval	130
Figure 5.3. Legality advantage for Experiment 1A, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.	132
Figure 5.4. Legality advantage for Experiment 1B, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.	132

Figure 5.5. Legality advantage for Experiment 2, with hit rate criterion lowered to 75%.
Error bars reflect bootstrapped 95% confidence interval.133

List of Tables

Table 2.1. Summary of conditions in Experiment 1A, along with experimental speaker language background, gender, and stimulus vowels. Note that virtually all listeners were native English speakers.....	35
Table 2.2. Summary of conditions in Experiment 2, along with experimental speaker language background, gender, and stimulus vowels. Note that virtually all listeners were native English speakers.....	56
Table 3.1. Summary of conditions, along with model talker language background, vowels, gender, and degree of adaptation.....	84
Table 3.2. Number of errors and error rates by condition.....	92
Table 3.3. Target/error matrix for all consonants and conditions. Targets are columns and errors are rows. The No Target column refers to errors that were made on extra syllables (i.e., when a participant produced more than 4 syllables in a single twister). Gradient color-coding reflects the number of errors for a given target/error combination.	94
Table 3.4. Between condition comparison; all sessions; [m]-[n] switches included.	97
Table 3.5. Between condition comparison; session 1 excluded; [m]-[n] switches included	98
Table 3.6. Different condition; session 1 excluded; [m]-[n] switches excluded.....	100
Table 3.7. Between condition comparison; all sessions; [m]-[n] switches excluded.....	101
Table 3.8. Between condition comparison; session 1 excluded; [m]-[n] switches excluded.	101
Table 5.1. Passing rates for each condition and experiment in Study 1.....	129
Table 5.2. Experiment 1A Fixed Effects.	134
Table 5.3. Experiment 1A random effects.	134
Table 5.4. Experiment 1B fixed effects.....	134
Table 5.5. Experiment 1B random effects.....	135
Table 5.6. Experiment 2 fixed effects.....	135
Table 5.7. Experiment 2 random effects.....	135
Table 5.8 Number of items analyzed (words or syllables), errors, error rates, and constraint order for previous tongue twister experiments.	137
Table 5.9. Different condition; all sessions; [m]-[n] swaps included.....	138
Table 5.10. Native Shared condition; all sessions; [m]-[n] swaps included.	138
Table 5.11. Non-Native Shared condition; all sessions; [m]-[n] swaps included.	139
Table 5.12. Vowel condition; all sessions; [m]-[n] swaps included.....	139
Table 5.13. Different condition; all sessions; [m]-[n] swaps excluded.	139
Table 5.14. Native Shared condition; all sessions; [m]-[n] swaps excluded.....	140
Table 5.15. Non-Native Shared condition; all sessions; [m]-[n] swaps excluded.....	140
Table 5.16. Vowel condition; all sessions; [m]-[n] swaps excluded.....	140
Table 5.17. Different condition; session 1 excluded; [m]-[n] swaps excluded.....	141
Table 5.18. Native Shared; session 1 excluded; [m]-[n] swaps excluded.	141

Table 5.19. Non-Native Shared condition; session 1 excluded; [m]-[n] swaps excluded. ...	141
Table 5.20. Vowel condition; session 1 excluded; [m]-[n] swaps excluded.....	141

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—i.e., updating our expectations to better match future input in a given context based on what we are currently experiencing in that context, so as to better predict and more efficiently process future events.¹ In the case of phonetics, this involves adapting 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, as well as have our speech be accurately perceived by others. 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). Uncovering the underlying structure that generates distinct patterns of variation is critical to successful adaptation. To do so, speakers must use their prior experience with variation as a guide, making causal inferences about the source of variation. Doing so allows speakers to adapt to systematic and relevant variation, while ignoring incidental

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

variation not relevant for the task at hand (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). 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 and 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; Liu & Jaeger, 2018; Kraljic & Samuel, 2011). In other words, adaptation requires listeners to properly attribute variation to its underlying source for the given task.

In this dissertation, we will focus on the role of systematic vs. 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, only allows voiceless plosives in syllable-initial 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, as individual speakers differing in this way 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).

In this dissertation, 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 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, using a speeded repetition task (Onishi, Chambers, & Fisher, 2002).

These predictions are tested in two studies. Study 1 examines phonotactic adaptation in perception, exposing listeners 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. Adaptation is assessed via a recognition memory paradigm (e.g., Bernard, 2015; Denby, Schechter, Arn, Dimov, & Goldrick, 2018). In Experiment 1, listeners are exposed to native French and English talkers exhibiting different phonotactic constraints. We predict that when listeners are exposed to talkers with a shared language background (i.e., 2 English talkers or 2 French talkers) they will infer the talkers share a language and

therefore a single phonotactic grammar, suggesting the phonotactic variation is incidental, and show a low degree of adaptation. When listeners are exposed to talkers who differ in their language background (i.e., 1 English talker and 1 French talker), we predict listeners will infer the variation is a systematic quality of the talkers' languages, and therefore listeners will show a high degree of adaptation.

Results showed that listeners adapted to the novel constraints in all conditions, suggesting that they attended to the differing phonotactics even when the source of variation was causally ambiguous (i.e., talkers shared a language background). The highest degree of adaptation occurred, however, when talkers differed in their language backgrounds, and the lowest degree occurred when both talkers were native speakers. In other words, listeners adapted to a higher degree when the source of the variation was causally unambiguous (i.e., talkers differed in their phonotactics due to the difference in their language backgrounds). Surprisingly, listeners adapted to a moderate degree when both talkers were non-native (i.e., French talkers). This may be due to listeners' lack of knowledge of non-native languages—listeners are likely more confident judging two native talkers as speaking the same language than two non-native talkers.

How can we understand listener behavior in Experiment 1? Listeners may simply be sensitive to whether they share the language background of a talker (native) or if the talker does not share their language background (non-native). Alternatively, listeners may be sensitive to talker language backgrounds regardless of whether they themselves share a background with the talker. In the second experiment of Study 1, we investigate the structure of listener knowledge of non-native phonotactic variation. 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 vs. 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 vs. Hindi talker) or not (e.g., Hungarian talker vs. Hindi talker).

In Study 2, we examine whether speakers make causal inferences about phonotactic variation in speech production, using a modified tongue twister paradigm (Dell, et al., 2000). Participants are exposed to multiple model talkers—native English and/or native German talkers—exhibiting distinct phonotactic constraints. As in Study 1, these model talkers either differ in their language backgrounds or share a single language background; we predict a high degree of adaptation when model talkers differ in their language background, and a low degree of adaptation when model talkers share a language background. Recent evidence from phonotactic adaptation in production, however, points to a purely associative account of phonotactic adaptation in production (e.g., Anderson, Holmes, Dell, & Middleton, 2019), in which inferences about the causes of variation are not always integrated into adaptation. As such, it is possible we may find divergent results in studies 1 and 2 due to broad differences in adaptation in speech perception and production (e.g., Samuel, 2011).

Results from Study 2 were difficult to interpret given a surprisingly high number of illegal errors, across all conditions, for the consonants that were the target of the constraint. Despite this, there is some evidence of adaptation when model talkers differed in their language background, but no evidence of adaptation when model talkers shared a language background.

Together, these studies aim to extend theories of the role of causal inference in adaptation into the domain of phonotactics; explore possible differences in adaptation between perception and production; and shed light on the mechanisms underlying the speed and flexibility of phonotactic adaptation.

2. Study 1

2.1. Introduction

Listeners encounter a huge amount of variation in their day-to-day linguistic experience—for example, men, women and children show considerable variation in their vowel productions (e.g., Hillenbrand, Getty, Clark, & Wheeler, 1995)—and yet speech perception is remarkably accurate. Over the last 25 years, the importance of listener adaptation to novel talkers has come into focus: for example, when listeners are exposed to idiosyncratic productions of speech sounds, they are able to adapt their phonemic category boundaries accordingly, and do so differently for talkers with different productions (e.g., Kraljic & Samuel, 2007). This adaptation allows listeners to navigate the inter-talker variability they encounter, helping them to predict speech from that talker in the future.

The flexibility of perceptual adaptation, however, is constrained by the types of highly structured variation listeners encounter—individual talkers do not vary freely. Beyond idiosyncratic differences between individuals, talkers vary on a number of linguistic and sociolinguistic dimensions (e.g., native vs. nonnative talkers, dialect, age, race; see Drager, 2010 for review). For listeners to use this structured variation to their advantage in perception, they must identify the source of the variation, and the underlying system that generates it (e.g., Kleinschmidt, 2018; Kleinschmidt & Jaeger, 2015). In doing so, speakers must distinguish between systematic variation that is relevant for a given task, and incidental variation that is irrelevant for the task (e.g., Kraljic & Samuel, 2011; Kraljic, Samuel, & Brennan, 2008; Liu & Jaeger, 2018).

For example, imagine you are having a conversation with someone who has a bad cold. The changes to that speaker’s vocal tract from the cold (e.g., occlusion of the nasal tract) introduce distortions to the acoustic properties of their speech (e.g., Tull, Rutledge, & Larson, 1996). The acoustic distortion introduced by the cold is temporary, and not systematic for that speaker—it is not a part of the speaker’s usual state, and thus not useful for perceiving that talker’s speech in the future. If you encounter this speaker again in a week, it’s likely that this acoustic distortion will no longer be present. As such, listeners should put little weight on those productions when updating their expectations of that talker’s future speech when healthy. This constitutes one of the core challenges of speech processing: adapting to systematic and relevant variation (e.g., talker differences) that will help you better communicate in the future, while deemphasizing irrelevant and incidental variation (e.g., noisy productions)².

In this study, we examine the type of relevant vs. irrelevant variation speakers experience with regards to phonotactics—constraints on possible positions and sequences of sounds in words and syllables—and how the systematicity of variation affects adaptation. English speakers, for example, unconsciously know that *sung* [sʌŋ] is a phonologically licit structure but *ngus* [ŋʌs] is not, as [ŋ] can only appear in syllable-final position in English (e.g., Chomsky & Halle, 1965). Phonotactic constraints differ

² It should be noted that while the phonetic variation introduced by the cold is irrelevant for guiding future expectations about that particular talker, it is highly relevant for similar contexts encountered in the future: namely, when talkers have colds. As such, we might expect that listeners build a mental representation of “cold speech”, reflecting the structure of phonetic variation they experience, that allows them to better comprehend such speech in the future.

systematically between languages: unlike English, Vietnamese allows [ŋ] in onset position, as well as coda (e.g., [ŋũ], “sleep”). Such differences between languages are common: Russian allows consonant clusters such as [stv] that are not legal in English; Dutch does not allow voiced obstruents, such as [d], at the ends of words; Hawaiian not allow any consonants in coda position; and so on.

Two individual speakers of a single language variety, however, generally do not differ in this way—sharing a phonotactic grammar is part of what it means to share a language variety. For example, encountering a native English speaker whose grammar allows [ŋ] in onset position is exceedingly unlikely. Moreover, Pierrehumbert (2001) argues that phonotactic constraints must be widely shared across talkers within a language for communication to be possible: if speakers of a single dialect systematically varied in their phonotactic grammars, phonotactic cues to word and syllable boundaries would be unreliable, leading to systematic errors in lexical access. (For example, it can be difficult to tell words apart in speech from an unfamiliar language, due in part to a lack of knowledge of that language’s phonotactic cues to word boundaries.)

As such, there is a systematic asymmetry in the degree to which phonotactic constraints vary between speakers of different languages and speakers who share a language variety. For example, an English speaker and a French speaker will vary in their phonotactic grammars much more than two English speakers who share the Inland North dialect. We examine how these systematic differences shape adaptation to novel phonotactic constraints.

Over the past two decades, laboratory paradigms have been developed that allow us to examine adaptation to variation in phonotactics. These studies suggest that talkers and listeners adapt to novel phonotactics with surprising speed: participants are able to learn arbitrary constraints (e.g., [n] *and* [f] *cannot appear in 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; Denby, Schecter, Arn, Dimov, & Goldrick, 2018).

We hypothesize that, much like the examples of perceptual adaptation discussed above, listeners adapt to relevant or systematic phonotactic variation, while ignoring variation irrelevant to the task at hand. Unlike talker-specific phonetics, however, phonotactics vary little between individual speakers within a speech community. This suggests that listeners may treat differences between individual talkers as irrelevant, and infer a single, shared phonotactic grammar for two talkers who share a dialect. Indeed, one study that investigated talker-specific phonotactic constraints (e.g., *for Fred, stops are restricted to coda position, and fricatives are unconstrained; vice versa for Barbara*) found no evidence of adaptation (Onishi, Chambers, and Fisher, 2002). Extending previous findings, we predict little adaptation will occur when speakers share a language even if both speakers are non-native (relative to the listener). In other words, whether the speakers share a native language with the listener is immaterial—as long as the speakers share a language with one another, listeners should infer the two speakers share a phonotactic grammar, and therefore not adapt.

In addition, the complementary prediction has yet to be examined. When listeners encounter phonotactic variation between two speakers who do not share a language background (e.g., a native speaker of English vs. a native speaker of French), we predict they should treat such differences as systematic based on their prior experience with phonotactic variation across languages, and infer separate phonotactic grammars for each speaker.

It is unclear what type and degree of experience with non-native languages is required to make inferences about speaker language background and phonotactic variation. It's possible that listeners only require occasional, incidental exposure to non-native phonotactics (from either speakers of different languages, or accented speakers of their native language). 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 U.S. population; Levay, Freese, and 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.

This study consists of two sets of artificial grammar experiments in which native English listeners are exposed to two talkers, each of whom exhibits a different talker-dependent phonotactic constraint. Crucially, the language background of talkers is manipulated. In Experiments 1A and B, talkers share a language native to listeners (2 English talkers); share a non-native language (2 French talkers); or do not share a language

background (1 French talker, 1 English talker). We predict that listeners will adapt only when talkers *differ* in their language background, as listeners infer from their prior experience that language-dependent variation is relevant, while within-dialect variation is incidental. To preview the results, we find differences in the *degree* of adaptation, rather than the categorical presence vs. absence of adaptation, based on shared or different talker language backgrounds. This leads us to refine our hypothesis: listeners adapt to a greater degree to variation they interpret as relatively systematic, and to a lower degree to variation they interpret as relatively incidental.

In Experiment 2, we examine the structure of listener knowledge of variation in non-native phonotactics. Listeners may be sensitive only to whether or not they share a language background with the talker, distinguishing only between the listener's native language and all non-native languages. Alternatively, listeners' prior knowledge could make distinctions between multiple non-native languages. To explore this question, listeners are exposed to non-native talkers who differ in their language backgrounds (1 Hindi talker, 1 Hungarian talker).

2.2. Background

2.2.1. 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 (first language: McQueen, 1998; second language: Weber and Cutler, 2006), and a number of other perceptual processes (e.g., Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Otake, Yonehama, Cutler, van der Lugt, 1996; Pitt & McQueen, 1998; Vitevitch & Luce, 1999; for a review, see Goldrick, 2011).

Given the importance of phonotactics in prediction 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; 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. 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.

2.2.2. *Adaptation and Variation*

The guiding role of prior experience is a general property of adaptation. In the perception of faces, for example, learners adapt differently to novel face shapes that are similar vs. dissimilar to faces they have previously experienced (e.g., Little & Apicella, 2016; Webster, Kaping, Mizokami, & Duhamel, 2004), suggesting that the type of variation learners have previously experienced affects how they adapt to novel patterns. Similarly, in the adaptation to the phonetics of individual talkers (e.g., Creel, Aslin, & Tenenhaus, 2008; Eisner & McQueen, 2005; Goldinger, 1996; Kraljic & Samuel, 2007; Nygaard & Pisoni, 1998; Pardo, 2006), adaptation is motivated by the huge amount of inter- and intra-speaker phonetic variation speakers have previously encountered (e.g., Hillenbrand et al., 1995). Nygaard and Pisoni (1998), for example, found that 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.

In addition to adaptation to variation between individual talkers, adaptation is conditioned on the structured variation introduced by higher-level (socio)linguistic factors (e.g., Kleinschmidt, 2018). A substantial body of work has suggested that listeners encode this structured variation (see Drager, 2010, for review). In turn, these factors guide adaptation to novel speakers, most clearly in the case of non-native accent adaptation (Bradlow & Bent, 2008; Reinisch & Holt, 2014; Xie & Myers, 2017). 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). This suggests listeners have models of specific non-native accents, encoded as distinct from native accents, and use this information to guide adaptation. There is also evidence that in some contexts native listeners treat non-native speech as a distinct concept, as they are able to generalize *across* non-native accents (Baese-Berk, Bradlow, & Wright, 2013).

Listeners can also use past experience with phonetic variation to make causal inferences about the source of the variation for the novel experimental talkers and contexts. Kraljic, Samuel, and Brennan (2008) exposed listeners to a talker producing ambiguous [s~ʃ] 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; in a second condition, listeners were exposed to the same talker, but with a video depicting the talker holding the pen in their hand. The pen in the mouth condition provided listeners with an incidental source for the variation: the pen in the talker's mouth was disrupting their productions. As such, listeners did not adapt to the talker's idiosyncratic [s~ʃ] boundary in that condition. Listeners who saw the video of the talker with the pen in their hand, however, inferred that the variation they were exposed to was a systematic characteristic of that talker's speech, and adapted accordingly – appropriately utilizing previous experience to constrain adaptation (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 process of uncovering the underlying structure that

generates observable events and inferring causal relations that help to explain those events (Qian, Jaeger, and Aslin, 2012). Such models of how the world works allow generalization to novel situations—prior experience can guide expectations about what will be encountered, especially in similar contexts, and help make sense of novel experiences. Prior experience can also constrain adaptation, when the novel context is dissimilar to those we have experienced previously. This rigidity is an important feature of the system, as total plasticity would require inefficiently building an entirely novel model for every novel context. 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, Kleinshmidt, & 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 vs. 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.

2.2.3. Phonotactic Adaptation and Variation

If listeners build a hierarchical indexical structure based on the variation they encounter, such a structure should differ for levels of linguistic structure that exhibit different patterns of structured variation. Given that phonotactic variation is greatest

across languages and smallest across individuals³, we hypothesize that listeners will assume that they should build separate models for speakers of different languages, while they will assume that speakers within a dialect should be assigned to the same model. This hypothesis predicts that listeners should show a greater degree of adaptation to talker-specific phonotactics when the talkers differ in their language background.

This prediction is consistent with evidence from previous phonotactic adaptation studies, in which listeners are exposed to artificial languages with non-native phonotactic constraints, and appear to quickly learn these novel constraints, suggesting that listeners assume they have encountered a non-native “laboratory language”. Indeed, participants are able to maintain separate models for English and a learned laboratory language. In a production task, Warker (2013) exposed participants to complex, “second-order” phonotactic constraints, in which the possible positions of a phoneme depend on features of surrounding phonemes (e.g., [æ] *cannot be followed by* [s] *and cannot be preceded by* [f]; the reverse constraint is true for [ɪ]). Participants required two experimental sessions on separate days to acquire the second-order constraints (see also Warker & Dell, 2006). When they returned to the experiment after a week, they 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 (e.g., that

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.

[æ] can be followed by [s] in words like *pass*). This suggests listeners may treat artificial languages as non-native languages, even when the exposure talker has a native language background. As a whole, these results are consistent with the hypothesis that learners treat what they learn in the lab as a distinct language.

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 cue to language background. 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 vs. non-native grammar, or do they make distinctions between non-native languages?

2.3. Experiment 1A

In an artificial language paradigm, we expose participants 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 (see Appendix A for acoustic analysis of French vowels). 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]. First, while French and English [i] are acoustically similar (Strange, Weber, Levy, Shafiro, Hisago, and Nishi, 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).

Table 2.1. Summary of conditions in Experiment 1A, along with experimental speaker language background, gender, and stimulus vowels. Note that virtually all listeners were native English speakers.

	<i>Native Shared</i>	<i>Non-Native Shared</i>	<i>Weak Different</i>	<i>Strong Different</i>
Speaker Language Background	Shared English	Shared French	English vs. French	English vs. French
English vs. French Vowels	[i, u] [i, u]	[i, y] [i, y]	[i, u] [i, u]	[i, u] [i, {u/y}*]
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.

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, 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 will unlikely differ in their phonotactic grammars. As such, we predict the smallest degree of adaptation in these conditions.

Participants are tested using a continuous recognition memory task (Bernard, 2015, 2017; Denby, et al., 2017; 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 4 repetitions to the familiarization syllables, listeners hear 9 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.

An initial pilot study (see Appendix B) was run to approximate the number of participants needed for requisite statistical power (see Appendix C for details of power analysis). 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 (osf.io/dbcqx/).

2.3.1. *Participants*

Based on a power analysis (see Appendix C for details), 256 participants, split evenly between the 4 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, and 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; see Appendix D for full breakdown of participant passing rates). Due to limitations within our online framework and AMT, 4 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 2 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.)

2.3.2. Stimuli

Stimuli were recorded in a soundproof booth at a 44.1 kHz sampling rate, and normalized to 60 dB SPL. 4 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]. 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.

2.3.3. *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 2 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 500ms 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.

2.3.4. *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 vs. 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 4 blocks of the experiment consists of 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 9 further randomly ordered repetitions of the familiarization set, but each repetition is now intermixed with 4 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).

2.3.5. 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 (note that loosening the criteria to include a greater number of participants does not qualitatively alter the results; see Appendix D). Participants who did not pass these criteria were excluded from the

analysis. In addition, for participants who attempted to complete the experiment multiple times within a single session, only the first attempt was included.⁴

Generalization data was analyzed using logistic mixed-effects regressions with maximal effects structures (Barr, Levy, Scheepers, and 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.⁵ Finally, a likelihood ratio test, between models with and without each contrast term as a fixed effect, was included to test for statistical significance.

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

⁴ There were a number of reasons why participants would attempt the experiment multiple times (e.g., if they accidentally closed their browser window halfway through).

⁵ Note that random intercepts and random slopes for participants and items by legality was the maximal effects structure. Items were defined as individual tokens spoken by specific talkers (e.g., French male talker’s [tif]), rather than more abstract “phonological” syllables (e.g., /tif/ spoken by all talkers). As such, participants in different conditions were exposed to a different subset of syllables, and contrast terms could not be included as random slopes for items.

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. 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.

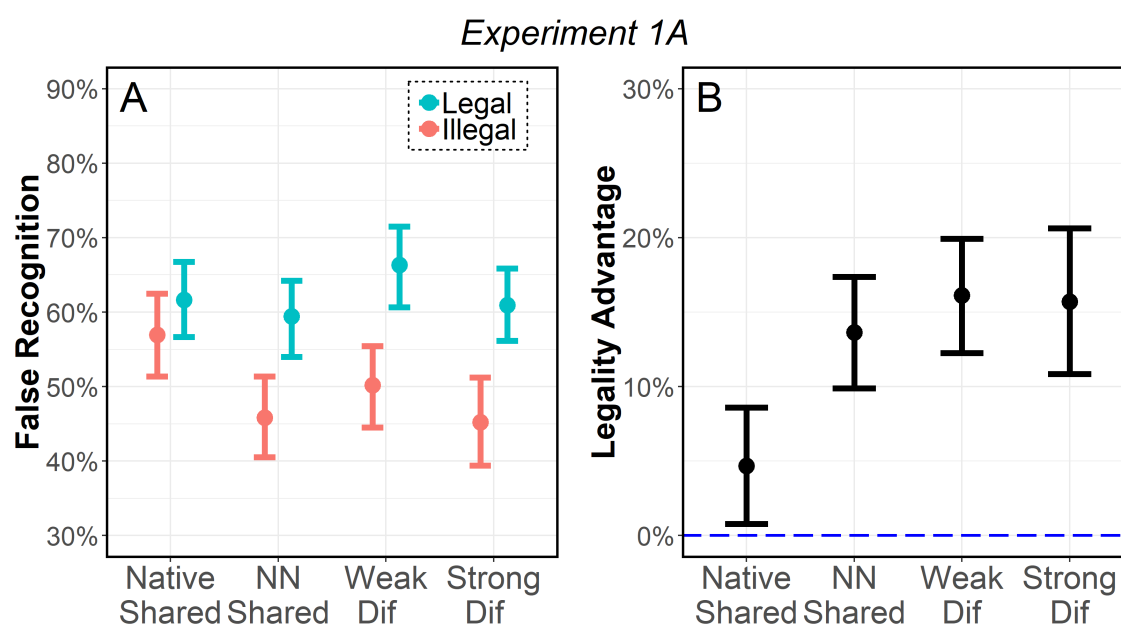
2.3.6. Results

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 from the observed data (with replacement). Distributions for means across participants were estimated with 1,000 replicates, sampling across means within each participant.

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 vs. 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 2.1, the legality advantage is modest in the *Native Shared* condition, and relatively large in the other three conditions.

Figure 2.1. 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 ($\beta = 0.64$, s.e. $\beta = 0.06$, $\chi^2(1) = 73.6$, $p < .0001$), 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 ($\beta = 0.72$, s.e. $\beta = 0.20$, $\chi^2(1) = 12.6$, $p < .001$), as listeners showed a greater legality advantage in the *Different* conditions than in the *Shared* conditions. Legality also interacted with accent ($\beta = 0.45$, s.e. $\beta = 0.14$, $\chi^2(1) = 9.8$, $p < 0.01$), but not strength ($\beta = -0.05$, s.e. $\beta = 0.14$, $\chi^2(1) = 0.14$, $p = .70$). 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. (See full model results in Appendix E.)

2.3.7. 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 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* vs. *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 vs. 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, to the point that listeners may have inferred that the two talkers did not share a language background, increasing the legality advantage in the *Non-Native Shared* condition. This 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 3 of the 4 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.

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 meant the generalization sets were identical across *Strong* and *Weak* conditions, allowing for a

more direct comparison. However, it 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.

2.4. Experiment 1B

In Experiment 1A, participants adapted to an unexpectedly high degree in the *Non-Native Shared* condition relative to 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.

2.4.1. *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. No participants were excluded due to exposure to previously seen experimental lists. 98.9% of participants who passed the criteria 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.)

2.4.2. *Stimuli*

Stimuli from 3 of the 4 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.) Stimuli were otherwise identical to those in Experiment 1A.

2.4.3. Data Analysis

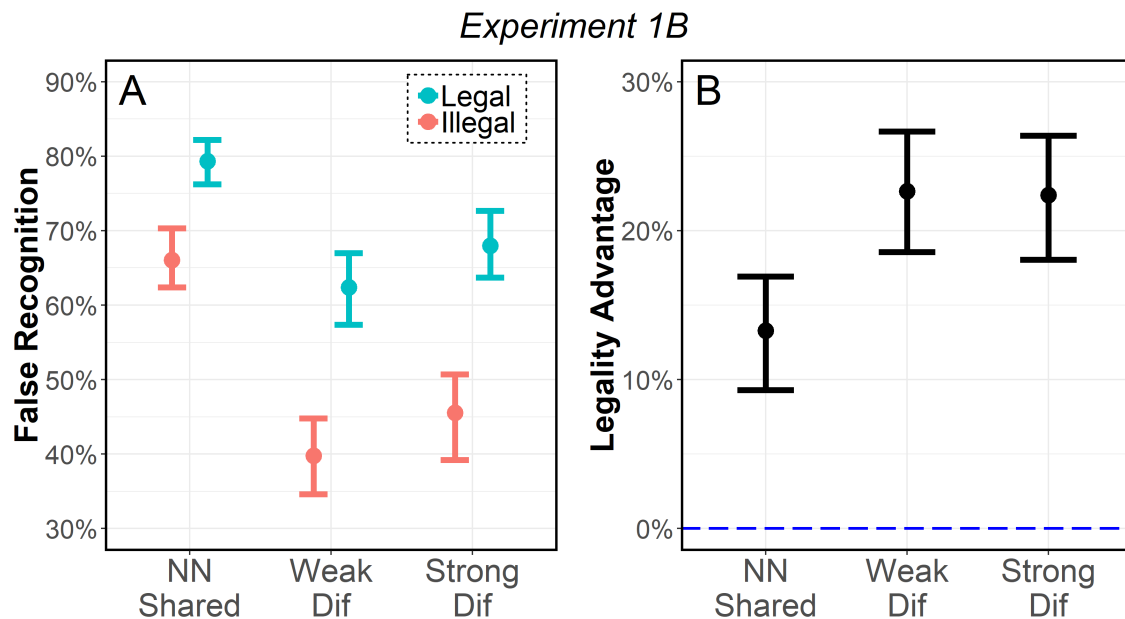
Significance was assessed using a logistic mixed-effects regression identical to that in Experiment 1A, with the exception of a fixed effect for accent, which was not included (there was no *Native Shared* condition in Experiment 1B). The model had fixed effects of legality, language difference (i.e., *Non-Native Shared* vs. both *Different* conditions) and strength (i.e., *Weak* vs. *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.

2.4.4. 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 2.2, the legality advantage is moderate in the *Non-Native Shared* condition, and large in the *Different* conditions (the *Native Shared* condition from Experiment 1A was included for reference).

Figure 2.2. 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.



A logistic mixed effects regression found a significant main effect of legality ($\beta = 1.02$, s.e. $\beta = 0.07$, $\chi^2(1) = 127.6$, $p < .0001$), as well as language difference ($\beta = -1.06$, s.e. $\beta =$

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 *Different* conditions. In addition, there was a significant interaction of legality with the language difference contrast term ($\beta = 0.57$, s.e. $\beta = 0.18$, $\chi^2(1) = 9.81$, $p < .01$), as listeners showed a greater legality advantage in the *Different* conditions than in the *Shared* condition. Legality did not interact with strength ($\beta = -0.02$, s.e. $\beta = 0.15$, $\chi^2(1) = 0.03$, $p = .87$), as the legality advantage was not significantly different across the *Strong* and *Weak Different* conditions.

2.4.5. 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 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%; see Appendix D for talker means in each condition and experiment).

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? One possibility is that the original female French speaker in Experiment 1A was not sufficiently phonetically distinct from the female English speaker, perhaps due to the French speaker's extended time in the United States. If this were the case, listeners would have inferred that they shared a language background, resulting in decreased rates of adaptation. Note that this differed from our original prediction: that the female French speaker in Experiment 1A was not similar enough to the male French speaker, resulting in higher than expected adaptation in the *Non-Native Shared* condition (we saw similar

adaptation for this condition in Experiments 1B). A second possibility is that the female French speaker's anomalous pitch contours were distracting, shifting listener attention away from the segmental level differences between speakers, and towards the prosodic differences.

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 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 is sufficient for listeners to detect the talker's language background.

2.4.6. *Experiment 1 conclusion*

Experiments 1A and B provide evidence that listeners' previous experience with phonotactic variation—that speakers with different language backgrounds exhibit a large degree of phonotactic variation, while speakers within a speech community do not—constrains adaptation to novel phonotactic constraints. Further, the results suggest that listeners have a structured model of phonotactic variation, with native and non-native languages each having separate phonotactic grammars. The specificity of listener

knowledge of non-native phonotactics, however, is unclear. In Experiment 2, we investigate whether listeners assign different phonotactic grammars to different non-native languages (as well as their native language). Alternatively, listeners may only assign a single phonotactic grammar to their native language,⁶ and a single phonotactic grammar to all non-native languages.

2.5. Experiment 2

In Experiment 2, 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 2 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 2 Hindi speakers or 2 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).

⁶ In the case of multilingual speakers, one phonotactic grammar to each of their native languages.

Table 2.2. 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.

	<i>Non-Native Shared</i>	<i>Mixed Different</i>	<i>Non-Native Different</i>
Language Background	Shared Hindi or Hungarian	English vs. (Hindi or Hungarian)	Hindi vs. Hungarian
Gender	Different	Different	Different
Predicted Degree of Adaptation	Moderate	High	<i>Within NN</i> : High <i>Native vs. NN</i> : Moderate

The within non-native distinctions hypothesis and the native vs. 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 within non-native distinctions hypothesis predicts a high degree of adaptation, with listeners inferring that different non-native languages have different phonotactic grammars. The native vs. 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.

2.5.1. *Participants*

192 participants were required (64 participants for each of the 3 conditions). To reach 192 participants who passed the experiment criteria (see below), 441 participants were recruited, 202 of whom passed the criteria (45.8%). 10 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. 98% of participants self-identified as native speakers of English, while 1 participant self-identified as speaking a non-North American dialect of English. 98.4% of participant self-identified as having no speech or language impairments. All model results were qualitatively identical when non-native and participants and those with impairments were excluded.

2.5.2. *Stimuli*

Stimuli were recorded in a soundproof booth at a 44.1 kHz sampling rate and normalized to 60 dB SPL. 6 talkers recorded stimuli, with 1 male and 1 female speaker for 3 languages: English, Hungarian, and Hindi. Talkers produced disyllables from orthographic representations of disyllables on a monitor; orthography reflected the language background 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).

64 disyllabic stimuli were created by splitting the 32 monosyllables into 4 groups of 8, counterbalanced for coda pattern (fricative vs. stop) and onset/rhyme pattern (onset [k,f] matched with rhymes [uf, ih, uk, ip] vs. onset [h,p] matched with rhymes [if, uh, ik, up]). These groups of 8 are further split in two, such that each group has an even distribution of segments in each position. Each subgroup of 4 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.

2.5.3. Procedure

The procedure was identical to that of Experiment 1.

2.5.4. 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 8, such that each syllable only appears once in each position (e.g., *fif* 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 8 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 2 blocks of the experiment consists of the familiarization phase. In each block, participants are exposed to 2 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 study, 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 8 further blocks; each generalization block also includes 4 intermixed generalization disyllables. This results in a total of 352 trials (16 familiarization disyllables * 2 repetitions/block * 10 blocks + 32 generalization syllables).

To ensure that listeners can clearly tell talkers apart, the two talkers have different genders in each condition.

2.5.5. 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. Finally, a third criterion was included: participants had to correctly reject novel generalization items at least as often as they rejected familiarization items (e.g., if a participant correctly accepted familiarization items only 85% of the time, the participant had to correctly reject generalization syllables at least 15% of the time). This ensured participants who were unable to sufficiently differentiate familiar and novel items were not included in the analysis.

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.

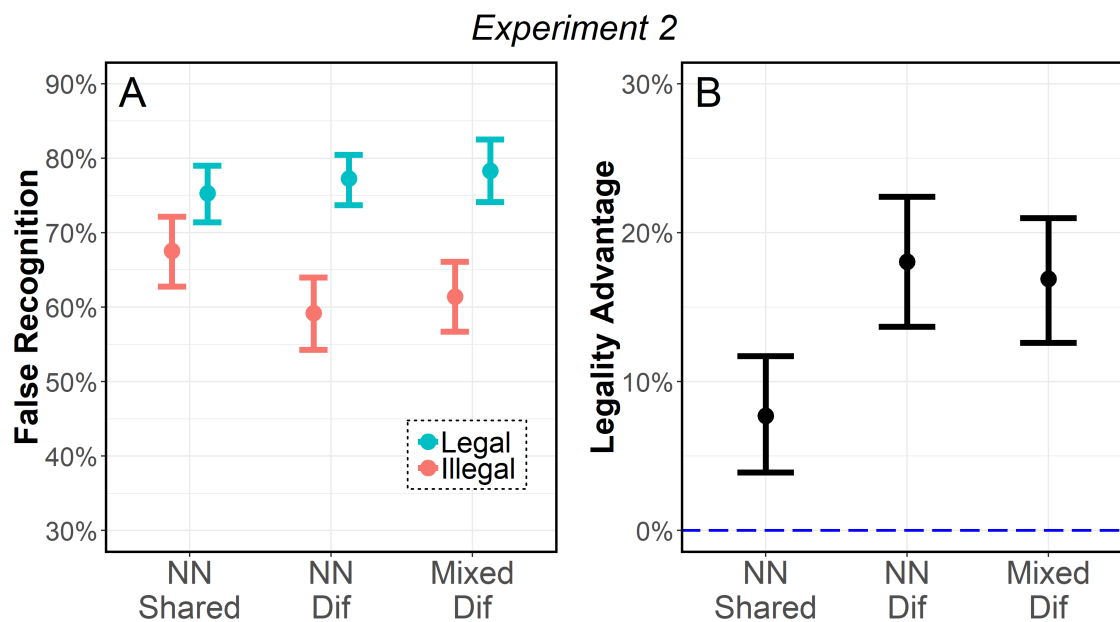
The within non-native hypothesis predicts a significant difference between the *Non-native Shared* and the two *Different* conditions, as indicated by the interaction term between legality and the non-native term; the native vs. non-native hypothesis does not predict such a difference. Such a difference 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.

2.5.6. Results

Participants correctly accepted 89.3% of familiarization disyllables (CI [88.7%, 90.0%]) and falsely recognized 69.8% of generalization. The mean legality advantage was 14.0% (CI [11.3%, 16.5%]). Crucially, as shown in Figure 2.3, 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.

Figure 2.3. 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.



The results from the logistic mixed effects regression show a main effect of legality ($\beta = 0.76$, s.e. $\beta = 0.08$, $\chi^2(1) = 62.83$, $p < .0001$), showing that listeners were more likely to false alarm on legal disyllables. The interaction between legality and language difference was also significant ($\beta = 0.71$, s.e. $\beta = 0.21$, $\chi^2(1) = 11.1$, $p < .001$), as listeners showed a greater legality advantage on the *Different* conditions than in the *Shared* condition. Legality did not interact with the non-native background term ($\beta = -0.04$, s.e. $\beta = 0.21$, $\chi^2(1) = 0.03$, $p = 0.87$), as listeners did not show a larger legality advantage in the two *Non-native* conditions vs. the *Mixed* condition.

2.5.7. 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.

How can we understand the similar levels of adaptation in the *Non-Native Different* and *Mixed Different* conditions? These results are consistent with the hypothesis that participants are treating input they receive in phonotactic learning experiments as a non-native “lab language”. Participants are exposed to nonsense words in semantically meaningless contexts, and likely infer such languages are not native. The speed with which participants adapt to novel phonotactic constraints (within one to two sessions) also suggests they are learning a novel phonotactic grammar, rather than adjusting one with which they have a lifetime of experience. This is further supported by evidence in production that speakers maintain experimental constraints for at least a week, and possibly longer, despite intervening evidence to the contrary outside of the experiment (Warker, 2013). In previous perception experiments, listeners learn rapidly despite being exposed to a single native talker. In other words, it is likely that learners treat any speaker in the experimental context as being “non-native”, even if that speaker has phonetics consistent with English. Thus the *Mixed Different* condition has two “non-native” languages, despite one of them having English phonetics. As such we don’t expect differences between the *Mixed Different* and *Non-native Different* conditions, given listeners may infer they both

have speakers of two different non-native languages. Note that this is compatible with the moderate adaptation found in the *Non-native Shared* condition in Experiment 1. We argue that this is due to the asymmetry in knowledge of native vs. non-native phonetics. In other words, listeners are likely to know that two native speakers of English share a language background; they are less likely to know that two French speakers do.

While Experiment 2 replicated the relatively higher legality advantage in *Different* vs. *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.

2.6. 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 listener's adaptation to novel non-native phonotactics. Listeners infer that speakers of different non-native languages are unlikely to share a phonotactic grammar, as

opposed to speakers sharing a non-native language background. But *how much* experience with 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. 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 5 years old or earlier were classified as having early second language (L2) experience. 33.8% of participants (216 total) 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.

Overall, participants who had any L2 experience showed a mean legality advantage (i.e., false recognition rates for legal stimuli minus false recognition rates for illegal stimuli) of 15.8% (CI [13.6%, 18.1%]), and an overall false recognition rate, regardless of legality, of 61.0% (CI [58.2%, 63.6%]); participants with no L2 experience showed a legality advantage of 14.7% (CI [13.1%, 16.4%]) and a false recognition rate of 61.5% (CI [59.5%, 63.6%]). Participants with early L2 experience, on the other hand, showed an 18.9% legality advantage (CI [14.6%, 23.0%]) and 57.3% false recognition rate (CI [51.5%, 62.7%]); participants with late or no L2 experience showed a 14.8% legality advantage (CI

[13.4%, 16.2%]) and 61.7% false recognition rate (CI [60.0%, 63.3%]). Listeners with French experience in Experiments 1A and B showed a 13.1% legality advantage (CI [6.3%, 16.5%]), while listeners with no French experience in Experiments 1A and B showed a 15.6% legality advantage (CI [13.1%, 16.6%]).

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. These models did not show an effect of L2 experience on the legality advantage, as there were no significant interactions between the legality and L2 experience terms. Moreover, there were no significant main effects (i.e., overall differences in false recognition rates) based on L2 experience.

Based on these results, it does not appear that participants with L2 experience made stronger inferences about the phonotactic grammars of speakers based on the speaker's language backgrounds. This suggests that a relatively small degree of exposure to non-native phonotactics is required to make inferences about talker's phonotactic grammars based on their language background, as listeners without extensive L2 experience do so. This is not surprising given listeners' sensitivity to non-native phonotactics, even in infants as young as 9-months-old (Mattys and 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 speakers' native language (L1) are deemed less accented than sequences illegal in the speakers' 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. That said, there were a number of limitations to the current analysis of listener language background: first, listeners self-reported their own language background, and were not asked to assess their own proficiency (although fluency self-assessments may not be entirely reliable; Tomoschuk, Ferreira, & Gollan, 2018). This means there was likely a wide range of proficiency levels within listeners with L2 experience. Second, there was a relatively small number of listeners with L2 experience (especially early L2 experience), which may have resulted in too small of a sample to draw definitive conclusions. Future work should investigate the relationship between listener language background and inferences about phonotactic variation directly, by comparing bilingual and monolingual populations.

2.7. General Discussion

Recent results from the perceptual adaptation literature (e.g., Liu and Jaeger, 2018) suggest that listeners use their past experience to uncover 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. Our prediction that listeners would leverage this past experience with phonotactic variation to make inferences about novel phonotactic constraints during

adaptation was largely confirmed in three experiments. Experiments 1A and B found a high degree of adaptation to novel phonotactic constraints by English listeners when talkers differed in language background (French vs. English). In contrast, adaptation was moderate when talkers shared a non-native language background (two French talkers) or relatively low when talkers shared a 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 vs. 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, although further replication is necessary to confirm that this is reliable across a range of talkers.

While learning was stronger when talkers differed in language background, learning also occurred when talkers shared a language background. We argue that the more confident listeners are that two talkers differ in their language background, the greater adaptation will be; conversely, when listeners are confident that the talkers share a language background, adaptation will be lower. When listeners are exposed to two native English talkers, for example, we predicted they would be highly confident that the two talkers share a language background (and therefore a phonotactic grammar), and would not adapt. Surprisingly, listeners were able to adapt in this condition, in contrast to previous work reporting null results (Onishi et al., 2002). Listener adaptation was lowest in this condition, however, suggesting that listeners had a strong prior belief that the two English talkers would have similar phonotactics relative to other pairs of speakers. It

appears the evidence that the two talkers shared a language background, in the form of their similar phonetics, wasn't enough to entirely overcome the evidence to the contrary. This includes the different phonotactic patterns exhibited by each talker; the fact that listeners were listening to non-native words; and that a single participant never heard both talkers produce the same word.

The effect of confidence concerning talker language background is more acute when two non-native talkers share the same language background, as they did in the *Non-Native Shared* condition. Here, we found a greater degree of adaptation relative to the *Native Shared* condition. This is likely due to the asymmetry between listener knowledge of non-native vs. 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.

2.7.1. *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.⁷ 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 and 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).

For the learner, adapting to subset phonotactics 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; e.g., Best, McRoberts, and Goodell, 2001). In other words, when the listener is faced with the overwhelming prospect of acquiring and understanding an unfamiliar language, adapting to the 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

⁷ This may be particularly true for the type of subset phonotactic constraints listeners were exposed to in these experiments (i.e., restrictions on sequences that are legal in the listener's L1), which do not present the same perceptual issues as superset phonotactics (i.e., difficult to perceive sequences that are illegal in the listener's L1, such as [dl] in word onset for English listeners; see Dupoux, et al., 1999).

are able to more successfully adapt in phonotactic adaptation experiments would also be more successful in early L2 acquisition.

It's also possible these adjustments, at least in perception, are short-lived, which may explain the speed with which they are made: early in acquisition, learners may need to readjust with each exposure to the non-native language. Finally, there is the question of why listeners can learn *multiple* non-native phonotactic grammars so quickly. Of course, we know this must be possible in the long term, since multilingual speakers have access to three or more distinct phonotactic grammars. Furthermore, this may be a particularly important skill in contexts with language contact occurring between multiple languages.

2.7.2. *Accent Detection*

Listeners are capable of making remarkably fine-grained distinctions in non-native accent detection (Atagi and 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 between two non-native accents? One possibility is the type of comparison listeners had to make in the current experiments was easier than those in previous studies. In many accent detection studies, listeners are given an auditory free classification task, in which they compare a large number of different accented speakers at once on a gradient two-dimensional scale (e.g., Atagi and Bent, 2013). While the high number of different accents increases the difficulty of this task, listeners are allowed

to take as long as they want and re-listen to audio samples of each speaker. Other studies use a two-alternative forced choice discrimination task, in which listeners are presented with a native and non-native speaker producing the same item and must distinguish between them (e.g., Park, 2013). In the current study, the listener's task is less demanding than these previous studies: listeners only must (implicitly) decide whether two speakers have different accents, not which one is the native vs. non-native speaker. Moreover, only 2 languages at a time are involved. Finally, listeners do not have to decide with each item which speaker is native and which is non-native; they are able to build their representation of the speaker's language background over the course of the experiment.

2.8. Conclusion

In three experiments, we have shown that listeners use their prior experience with phonotactic variation—that language 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 adaption 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.

3. Study 2

3.1. Introduction

Speech is produced with a huge amount of variation between talkers and contexts. To effectively comprehend their interlocutors, speakers must continually adapt to novel speakers and contexts, drawing on their experience from similar situations in the past. Adaptation is not limited to comprehension, however—communication is a two-way street, with speakers trading off producing and comprehending speech. Speakers not only modify their predictions of what they expect to hear; they also modify their own speech, imitating and aligning themselves with the phonetic characteristics of their interlocutor in spontaneous speech (e.g., Pardo, 2006) or while shadowing a model talker (e.g., Goldinger, 1998).

Phonetic imitation is likely motivated by a complex mixture of social and communicative factors (Pardo, Urmanche, Wilman, & Wiener, 2017), but it is only made possible by the systematicity of inter-speaker variation. Imitating an interlocutor—either to bring oneself in closer social alignment with an interlocutor in a position of power (e.g., Giles, 1973) or to simplify language processing for both speakers in a dialogue (e.g., Pickering and Garrod, 2004)—only has utility because individual speakers vary systematically, showing consistency from one utterance to the next. If speakers varied freely, such imitation would bring you no closer to your interlocutor's speech.

In addition to speakers adapting to the phonetic features of interlocutors and model talkers, in experimental settings speakers also adapt their speech production systems to

reflect novel phonotactic constraints embedded in laboratory speech or text (e.g., Dell et al., 2000; Onishi, Chambers, & Fisher, 2002). These experimental constraints are often arbitrary (e.g., [n] is constrained to onset position; [f] is constrained to coda position) and are, by definition, not characteristic of a speaker's native language. In tongue twister paradigms, for example, participants' speech errors reflect the phonotactic constraints of the laboratory text they are exposed to (e.g., errors resulting in [n] rarely appear in coda position).

In this study, we explore how the differences in the structure of variation for phonetics and phonotactics may result in differences in adaptation. While adaptation to the phonetic properties of an interlocutor is motivated by speakers' prior experience with the systematic nature of inter-speaker variation, phonotactic constraints vary little at the individual level. Instead, phonotactics vary extensively at the language- and dialect-wide level, with different language varieties exhibiting a broad range of possible syllable structures and sound sequences not found within language varieties.

We expect this difference in the structure of variation for phonotactics versus that for phonetics to result in differences between phonetic and phonotactic adaptation, as we saw in speech perception in Study 1. We argue that speakers make causal inferences about the source of phonotactic variation when adapting (e.g., Liu and Jaeger, 2018), guided by their prior experience with variation: that talkers within a language variety share a single phonotactic grammar, while talkers who differ in their language varieties do not. As such, we predict that when speakers encounter two talkers with differing phonotactics, they will only adapt to each talker's phonotactic grammar if they believe those talkers do not share a

language background. This hypothesis is consistent with Onishi et al. (2002), who found that in a speeded repetition task, participants were able to learn second-order consonant constraints conditioned on surrounding vowels, but not constraints conditioned on talker identity (both talkers were native English speakers).

We examine this question using a modified tongue twister paradigm. On each trial, participants will first shadow a model speaker producing a string of four syllables, and then repeat their productions again with orthographic support (previous tongue twister experiments have presented the twister in orthographic form only, with the exception of Smalle, Muylle, Szmalec, & Duyck, 2017). The strings of syllables associated with each model talker will reflect different phonotactic constraints. The language background of the model talkers was modulated in different conditions, with model talkers either native speakers of German or English. Three of the four conditions mirror the conditions in Study 1, with two talkers that either share or differ in their language backgrounds: *Native Shared* (two monolingual English model talkers), *Non-Native Shared* (two native German model talkers), and *Different* (one German, one English talker). In addition, a *Vowel* condition was included as a task control, ensuring that results from previous work are replicated (e.g., Gaskell, Warker, Lindsay, Frost, Guest, Snowden, & Stackhouse, 2014; Smalle et al., 2017; Warker, 2013; Warker & Dell, 2006) despite changes to the task. In the *Vowel* condition, participants are exposed to a single model talker, and learn a second-order constraint in which consonant restrictions are conditioned on the neighboring vowel (e.g., [ɛ] can be followed by [m] and preceded by [n]; the reverse is true for [i]). Adaptation is measured by analyzing the errors of participants: if errors involving an experimentally constrained

segment result in legal syllables (i.e., syllables that follow the phonotactic pattern embedded in the input) more than illegal syllables, it suggests participants are adapting.

If the same principles for adaptation apply in production as they do in perception, we expect a similar result to what was found in previous perception experiments: a higher degree of adaptation in the *Different* condition than in the shared conditions. At a more granular level, we may expect moderate adaptation in the *Non-Native Shared* condition, and a low degree of adaptation in the *Native Shared* condition, mirroring the results from perception.

3.2. Background

3.2.1. *Phonotactic learning in speech production*

Over the past twenty years, researchers have explored phonotactic adaptation using the tongue twister paradigm (Anderson, Holmes, Dell, & Middleton, 2019; Dell et al., 2000; Gaskell et al., 2014; Goldrick, 2004; Goldrick & Larson, 2008; Kittredge & Dell, 2016; Smalle et al., 2017; Taylor & Houghton, 2005; Warker, 2013; Warker & Dell, 2006; Warker et al., 2008; Warker, Xu, Dell, & Fisher, 2009). In the tongue twister paradigm, errors are elicited by having participants quickly repeat a string of nonsense syllables. These syllables follow an artificial, experimental phonotactic constraint, such as “syllables begin, but do not end, in [n]; vice versa for [f]”. In addition, twisters also include segments that are constrained in the participant’s native language, such as [ŋ], which is constrained to coda position in English. Speech errors virtually never violate such language-wide categorical phonotactic

constraints—this is sometimes referred to as the phonotactic regularity effect (Fromkin, 1971). In English, for example, producing *fip* when the target is *sip* is a possible error, but *ngip* for *sip* is an extremely unlikely error. Other consonants in the experiment are unconstrained, appearing in both onset and coda position. Such unconstrained errors tend to maintain their syllable position (Nooteboom, 1969): for example, *fik fip*, in which [f] maintains its position in the onset of the syllable, is a more likely error for the target sequence *fik sip* than *fik sif*, in which [f] switches to the coda position. In tongue twister experiments, this syllable position effect—which holds in roughly three-quarters of errors—serves as a baseline for experimental learning effects. Learning is measured as the degree to which errors involving experimentally constrained consonants maintain their syllable position above and beyond the syllable position effect for unconstrained consonants. In previous experiments, adaptation to the novel constraint causes error patterns for experimentally constrained consonants to resemble error patterns of language constrained syllables, often maintaining their syllable position 95% of the time or more (e.g., Dell et al., 2000).

Phonotactic adaptation experiments have shown that speakers are also sensitive to first- versus second-order constraints, with each type of constraint showing distinct learning patterns. First-order constraints are dependent only on syllable position (e.g., “[n] is constrained to onset position; vice versa for [f]”); second-order constraints are dependent on two factors (e.g., “if the vowel is [ɪ], [n] is constrained to onset and [f] is constrained to coda; if the vowel is [ɛ], vice versa”). Such second-order constraints naturally occur within the world’s languages. In English, for example, [b] appears in onset

position preceding the vowel [i] (as in *beet*), but is unattested in coda position following [i] (e.g., *teeb*; Kessler and Treiman, 1997). Previous work has shown that second-order constraints are not learned as thoroughly as first-order constraints, with errors less likely to maintain their syllable position. In addition, learning effects for second-order constraints do not appear during the first experimental session (see Warker, 2013). More specifically, sleep consolidation appears to play a critical role in the acquisition of such constraints, as participants require sleep in intervening periods between sessions to show effects of learning (Gaskell et al., 2014; Warker, 2013). The current study exposes participants to second-order constraints, with consonant position either dependent on vowel or model talker language background, depending on the condition. As such, we include three experimental sessions, each on different days, to ensure participants are able to consolidate the constraints.

3.2.2. *Causal inference in phonotactic adaptation in production*

We hypothesize that adaptation is spurred in contexts in which speakers are interacting with talkers that have different language backgrounds. If this is the case, why have participants in previous tongue twister experiments adapted at all, when they're simply reading twisters aloud and not interacting or shadowing other speakers? As noted in Study 1, Warker (2013) found that participants maintain the experimentally learned phonotactic constraints as long as one week after initial exposure, despite the massive intervening experience with English contrary to such constraints. Participants also learn

the new constraints rapidly, despite a lifetime of experience suggesting such constraints are not a part of their native language. Warker (2013) proposes a mechanistic account for such learning, in which participants create a copy of their general phonotactic grammar to use in experimental settings. In the experiment, the copy is updated to reflect only those features of the experimental context that are not true of the general phonotactic grammar.

A remaining question is why speakers create the copy in the first place—why do speakers have the adaptive ability to quickly modify their phonotactics in this way? Of course, speakers do not regularly find themselves in psycholinguistic experiments, so this does not seem like a possible motivation for this adaptive ability; moreover, speakers' native language phonotactics are likely relatively stable (Pierrehumbert, 2001), suggesting phonotactic adaptation is not driven by differences between speakers within one's native language. We argue that in tongue twister experiments, as in the perceptual adaptation explored in Study 1, participants are recruiting their second-language (L2) learning faculties to adapt to non-native phonotactic constraints. In the highly artificial laboratory setting, in which participants are exposed to nonsense syllables that are not presented in a semantically meaningful native language setting, participants may detect that they are in a non-native language context. Detecting such changes in context is a critical aspect of learning in a multi-context environment (Qian, Jaeger, & Aslin, 2012). Moreover, we have strong evidence that learners are able to detect different phonotactic environments and construct multiple phonotactic grammars in the form of multilingual speakers. Multilingual learners are exposed to conflicting phonotactic patterns in acquisition, and must separate evidence for different constraints based on the language context in which they appear

(although bilinguals' phonotactic grammars are not entirely separate; see Carlson, Blasingame, Goldrick, & Fink, 2016). Learners likely do this by leveraging a number of different cues, such as differing lexical items, phonetics, prosody, talker voices, and other factors (e.g., Bosch & Sebastián-Gallés, 2001; Weiss, Gerfen, & Mitchell, 2009). In this study, we predict participants will infer multiple linguistic contexts based on the phonetic differences between the native and non-native model talkers.

3.2.3. *Current Study*

Recent evidence in tongue twister studies suggests that speakers are relatively insensitive to high-level inferences about the sources of variation, however. Dell et al. (2000) informed some participants of the experimental constraints they would be exposed to, and found learning was unaffected by prior knowledge of the constraint (see also Smalle et al., 2017; Warker & Dell, 2006). Anderson et al. (2019), in a tongue twister paradigm, exposed learners to a first-order constraint and then reversed the phonotactic constraint partway through the experiment, before reversing it back again in the final block. If participants were sensitive to the change in context (i.e., the reversal of the phonotactic constraint) and the source of the variation, they should have been able to rapidly reverse the constraint, and learn the reversed constraint faster than the original constraint, given that the reversed phonotactic involved the same target consonants, simply in flipped syllabic position. Instead, participants learned the reverse constraint more slowly than the original, suggesting that phonotactic learning is incremental, and resistant to causal

inferences about changes in context. Anderson and colleagues argue that phonotactic adaptation is internal to the aspect of the production system that constructs syllables, and “not very cognitively penetrable” from outside systems. Moreover, the learning system may be domain general: participants show remarkably similar patterns for “phonotactic” learning in button-pushing tasks, in which different fingers are assigned different “consonants” (Anderson et al., 2019; Rebei, Anderson, & Dell, 2019). This further suggests that learners may be insensitive to language-specific inferences, such as those based on accent. Finally, adaptation in the production system may simply be more cognitively costly than adaptation in perception, and therefore slower and more constrained (Samuel, 2011). An account in which speakers are insensitive to top-down inferences about the source of variation, motivated by these previous tongue twister experiments, would predict learners are entirely unable to divide input between model talkers and form multiple phonotactic grammars. This account therefore predicts adaptation in the baseline *Vowel* condition in the current study. Adaptation in the *Different* condition, however, would suggest that inferences based on language context, specifically, may have a privileged status over other less relevant types of information.

The measure of interest in tongue twister experiments is the rate at which errors maintain syllable position. If participants are adapting to the experimental constraint, they should be more likely to make errors that maintain syllable position for experimentally restricted consonants (i.e., errors that follow the phonotactic constraint) than for unrestricted consonants. Adaptation will therefore be indicated by a higher proportion of errors maintaining their syllable position for constrained errors than unconstrained errors.

For second-order constraints, adaptation should specifically appear in the second and third experimental sessions, following previous results (e.g., Warker and Dell, 2006).

We predict a high degree of adaptation in the baseline *Vowel* condition, following results from previous experiments. If we do not replicate previous studies in the *Vowel* condition, it may indicate issues with the current design (especially having twistlers presented auditorily as well as orthographically). We also predict a high degree of adaptation in the *Different* condition, in which model talkers differ in their native language backgrounds, reflecting speakers' use of their prior experience with phonotactic variation to adapt to novel constraints. Both *Shared* conditions, on the other hand, should show a low to moderate degree of adaptation, given that both model talkers share a language background. More specifically, we may find low adaptation in the Shared Native condition, and moderate adaptation in the Shared Non-Native condition, as we did in perception in Study 1. Such intermediate adaptation may be due to speakers' asymmetric knowledge of native vs. non-native languages—speakers are likely quite certain when they are exposed to two native speakers of their native language that the two speakers share a language background, but likely much less certain when they are exposed to two non-native speakers.

Table 3.1. Summary of conditions, along with model talker language background, vowels, gender, and degree of adaptation.

	<i>Native Shared</i>	<i>Non-Native Shared</i>	<i>Different</i>	<i>Vowel</i>
Model talkers	2 English	2 German	1 English, 1 German	1 English or 1 German
Vowels	[ɛ]	[ɛ]	[ɛ]	[ɛ] and [ɪ]
Predicted Degree of Adaptation	Low	Low-to-moderate	High	High

3.3. Methods

The design and analysis of the experiment—including predictions, number of participants, stimulus design, and model structure—were defined before data collection, unless otherwise noted, in a pre-registration on the Open Science Foundation platform (osf.io/uryc5/).

3.3.1. Participants

Sixteen participants were recruited, all native American English speakers from the Northwestern University community with no speech or hearing impairments. The experiment consisted of 3 one-hour sessions, each on a different day. Each session took place no further than a week apart. Participants were paid \$40: \$10 for each of the first two sessions, and \$20 for the final session, to incentivize participants to attend all three sessions. Participants were randomly assigned to different conditions.

Before the experiment began, participants completed a self-reported language

background questionnaire, reporting any second or third languages, as well as their age of acquisition and length of speaking for each language.

3.3.2. *Materials*

Participants were exposed to second-order constraints in all conditions: in the *Native Shared*, *Non-Native Shared*, and *Different* conditions, constraints were condition on model talker identity (e.g., if Talker A is the model talker, [m] is constrained to coda position and [n] is constrained to onset position; the converse is true for Talker B). In the *Vowel* condition, participants were exposed to a second-order constraint similar to past experiments, in which the possible positions of experimentally constrained consonants are conditioned on the vowel. Four model talkers were included: one male and one female German speaker, and one male and one female English speaker. In the *Vowel* condition, participants were exposed to a single model talker; a different model talker was assigned to each participant. In the *Native Shared* condition, participants were exposed to both English model talkers; in the *Non-Native Shared* condition, half of the participants were exposed to both German model talkers. In the *Different* condition, participants were exposed to one German model talker and one English model talker, who did not share a gender (e.g., male German talker and male English talker).

Following previous studies, stimuli consisted of CVC nonsense syllables made up from two vowels ([ɪ] and [ɛ]) and eight consonants ([f], [p], [k], [t], [m], [n], [ŋ], [h]). While most previous experiments used [g] and [s], [p] and [t] were used instead to avoid

phonotactic constraints in German: word-finally, [g] in German is produced as [k] due to word-final devoicing, and [s] is produced as [z] in word-initial position. [æ] is also usually used, but was replaced with [ɛ] because [æ] is not a vowel of German. [m] and [n] served as the experimentally restricted consonants following previous experiments (Dell et al., 2000; Warker, 2013; Warker and Dell, 2006). Another set of consonants ([k], [p], [t], [f]) was unrestricted, appearing freely in any position. [ŋ] and [h] served as the language-wide restricted consonants, with [ŋ] illegal in onset and [h] illegal in coda.

Each model talker recorded a set of 96 unique tongue twisters to be used in the *Different* and *Shared* conditions (the set was kept constant across model talkers). These twisters were randomly constructed, with the exception of phonotactic constraints: they obeyed language-wide constraints, and half of these twisters had [m] in onset and [n] in coda, while the other half had the reverse pattern. The phonotactic pattern assigned to each model talker was counter-balanced across participant. In the *Different* and *Shared* conditions, participants were exposed to alternating model talkers (and therefore alternating phonotactic constraints) on every trial. The [ɛ] vowel was used in each of these twisters. For example, a participant in one of the *Shared* or *Different* conditions might be exposed to the following two trials:

Trial 1 (*Shared/Different* conditions)

Talker A: *feng met hep ken*

Trial 2 (*Shared/Different* conditions):

Talker B: *neng fek hem pet*

Note that the position of [m] and [n] is dependent on the talker, which alternates each trial, while the vowel is consistent between trials.

With the exception of the alternation, the twister order was randomized. In each session, participants were exposed to the set of 96 twisters—half from each talker/phonotactic constraint—then exposed to the same set again in a different random order, for a total of 192 trials. There was no indication for participants that they were exposed to a smaller set twice, rather than one large set. In the *Vowel* condition, participants heard a single model talker, with each trial alternating between the [ɪ] and [ɛ] vowels, with the phonotactic constraint alternating based on the vowel. For example, a participant in the *Vowel* condition might be exposed to the following two trials:

Trial 1 (*Vowel* condition)

Talker A: *tik min pif hing*

Trial 2 (*Vowel* condition):

Talker A: *kep net feng hem*

Note that the position of [m] and [n] are dependent on the identity of the vowel, which alternates each trial, while the talker is consistent throughout the experiment.

Each model talker recorded an additional 48 twisters with the [ɪ] vowel for the *Vowel* condition, with the other 48 twisters re-used from the other conditions.

3.3.3. Procedure

Participants were recorded in a sound-proof booth using a head-mounted microphone, where they read the twisters from a computer monitor and heard twisters from model talkers over speakers. In addition, participants heard metronome beats from an earbud headphone placed in the left ear.

Before the experimental trials began each session, participants completed 4 practice trials, none of which contained the experimentally constrained consonants. Practice trials included the model talkers that would appear in the experimental trials. In each experimental trial, a sequence of four syllables was presented to participants (e.g., *fem heng ket nep*). In the slow repetition phase of each trial, participants were exposed to the sequence being read by the model talker at a slow pace (1 syllable/second). After the model talker audio stopped playing, the orthographic representation of the twister immediately appeared on the screen, and participants repeated the twister in time with the model talker. Eight seconds after initiating the trial (4 seconds for the model talker's productions, and 4 seconds for the participants repetition), participants were free to advance the experiment to the fast repetition phase. In the fast repetition phase, participants heard sixteen beats of a metronome at the speed of 2.5 syllables/second. Participants were instructed to wait during the first 4 beats, then repeat the four-syllable sequence three times over the final 12 beats. After all metronome beats were finished playing, participants were free to advance the experiment to the next trial.

After the final session, participants were asked a series of post-experiment

questions by the experimenter: first, did you notice anything noteworthy about the speakers? If participants did not bring up the non-native status of speakers on their own (in the conditions in which they were exposed to non-native speakers), they were asked a second question: did you think the speakers you heard were native speakers of English or non-native speakers? Finally, if participants responded that they heard one or more non-native speakers, they were asked to guess the speakers' native language backgrounds.

3.3.4. Analysis

Adaptation was statistically verified using logistic mixed-effects regressions. The dependent measure in all models was maintenance of syllable position. The first type of analysis fitted individual models to the data from each condition. Fixed effects included a contrast-coded fixed effect of consonant type (i.e., experimentally constrained vs. unrestricted) and a contrast-coded fixed effect of experimental session, in which session 1 was contrasted with sessions 2 and 3. An interaction term between session and consonant type was also included. This interaction term would indicate adaptation had taken place, with an increased maintenance of syllable position for constrained consonants in sessions 2 and 3 over session 1. Random effects will include random intercepts for target syllable; with only 4 participants per condition, there was not a sufficient number to include random effects for participant.

In a comparison across conditions, a model was fitted to the data from the *Different* and *Shared* conditions (excluding the *Vowel* condition). This model included contrast-coded

fixed effects consonant type, session, and their interaction, and random intercepts for target syllable. In addition, it included a contrast-coded fixed effect of condition, comparing the *Shared* conditions with the *Different* condition. If speakers make top-down inferences about language background during production, we would expect a significant interaction such that the effect of consonant type is stronger in the *Different* condition than in the *Shared* conditions.

A second contrast-coded term, comparing the *Native Shared* condition to the *Non-Native* condition and the *Different* condition (i.e., conditions with only native model talks vs. those with some or all non-native model talkers), was planned but ultimately not included.

Finally, alternative versions of the within-condition models and between-condition models were fitted to the data just from sessions 2 and 3. The structure of these models were identical to previous models, with the exception of session, which was not included as a fixed effect. Models without session 1 were unplanned analyses; their purpose was to simplify model structure and focus on sessions 2 and 3, where we had the strongest prior belief that adaptation would occur.

3.4. Results

Recordings were transcribed for errors. Errors made up of consonants not included in the experiment ($N = 64$) were noted but excluded from the analysis; individual vowel errors were generally not noted, but if participants made systematic vowel errors this was noted. Errors were coded for maintaining syllable position. For example, if the target twister was *fek tep hen meng* and the participant produced *tek tep hen meng*, the [t] error in

tek was coded as maintaining its position; if the participant produced *fet tep hen meng*, the [t] error in *fet* was coded as changing its position. Every twister was treated as if it had exactly 4 target syllables. If participants produced more or less than 4 syllables, participant productions were aligned with the 4 twister targets in the alignment that resulted in the fewest errors. Errors were also coded for rater confidence as either high confidence (i.e., very likely or certain error) or low confidence (i.e., possible/moderately likely error).

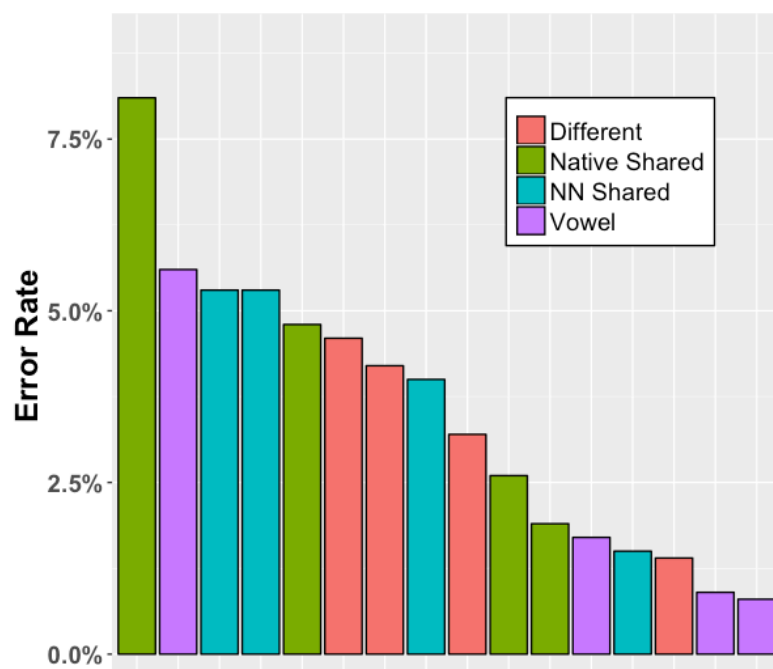
Across all conditions, there were 48 total sessions (16 participants, 3 sessions each), resulting in a total of 9,216 trials (2304 per condition) or 110,592 total syllables (27,648 per condition). Errors were coded by 2 coders; of the 48 total, one coder completed 43 sessions, while a second coder completed 6 sessions. One session (2,304 syllables) was coded for reliability by both coders. There was an overall agreement rate on the presence/absence of errors for 99.3% of all syllables. Looking at agreement on only those errors identified by the principal coder ($N = 27$), there was 74.1% conditionalized agreement, a number in line with previous experiments.

A total of 3864 errors were discovered, or 3.5% of all syllables, an error rate within the range of previous experiments (although on the low end; see Appendix F). Of these, 82 (2.1%) were coded as low confidence by the transcribers. A relatively wide range of error rates were found for different conditions (see Table 3.2), but that may simply reflect variation across participants (see Figure 3.1).

Table 3.2. Number of errors and error rates by condition.

Condition	<i>n</i> Errors	Error Rate
Different	928	3.4%
Native Shared	1201	4.3%
Non-Native Shared	1106	4.0%
Vowel	629	2.3%

Figure 3.1. Overall error rates by participant. Colors reflect experimental condition.



There were 807 errors on consonants subject to language-wide phonotactic constraints (i.e., errors resulting in [ŋ] and [h]). Of these, 99.8% followed English phonotactic constraints and maintained their syllable positions. There were 2,124 errors for unconstrained consonants ([p], [t], [k], and [f]), which maintained their syllable position 74.5% of time, in line with previous experiments. Of the 933 errors for experimentally constrained consonants ([m] and [n]), however, only 50.2% of errors maintained their

syllable position. This was a surprising result—even if no learning occurred at all, experimentally constrained consonants should have, all else being equal, maintained their syllable position at a similar rate to unconstrained consonants. Closer inspection of illegal constrained errors revealed that a large proportion consisted of swaps between [m] and [n]: for 40.5% of all [n] errors, the target was [m], and for 32.7% of all [m] errors, the target was [n] (see Table 3.3 for target-error matrix). Overall, swaps between nasal consonants occurred at a disproportionately high rate: tabulating the proportion of errors that were intended for a specific target consonant for all 64 possible error/target combinations revealed that 4 of the top 5 error/target combinations were nasal swaps. The locus for such swaps may be that nasal pairs were more phonetically similar than other consonant pairs; previous evidence from tongue twister experiments suggests phonetic similarity (Wilshire, 1999) and overlapping phonological features (Goldrick, 2004) affect speech errors. It's unclear, however, why [m] and [n] were more affected by phonetic similarity than other segments, as a number of previous experiments have included these segments (see following section for further discussion).

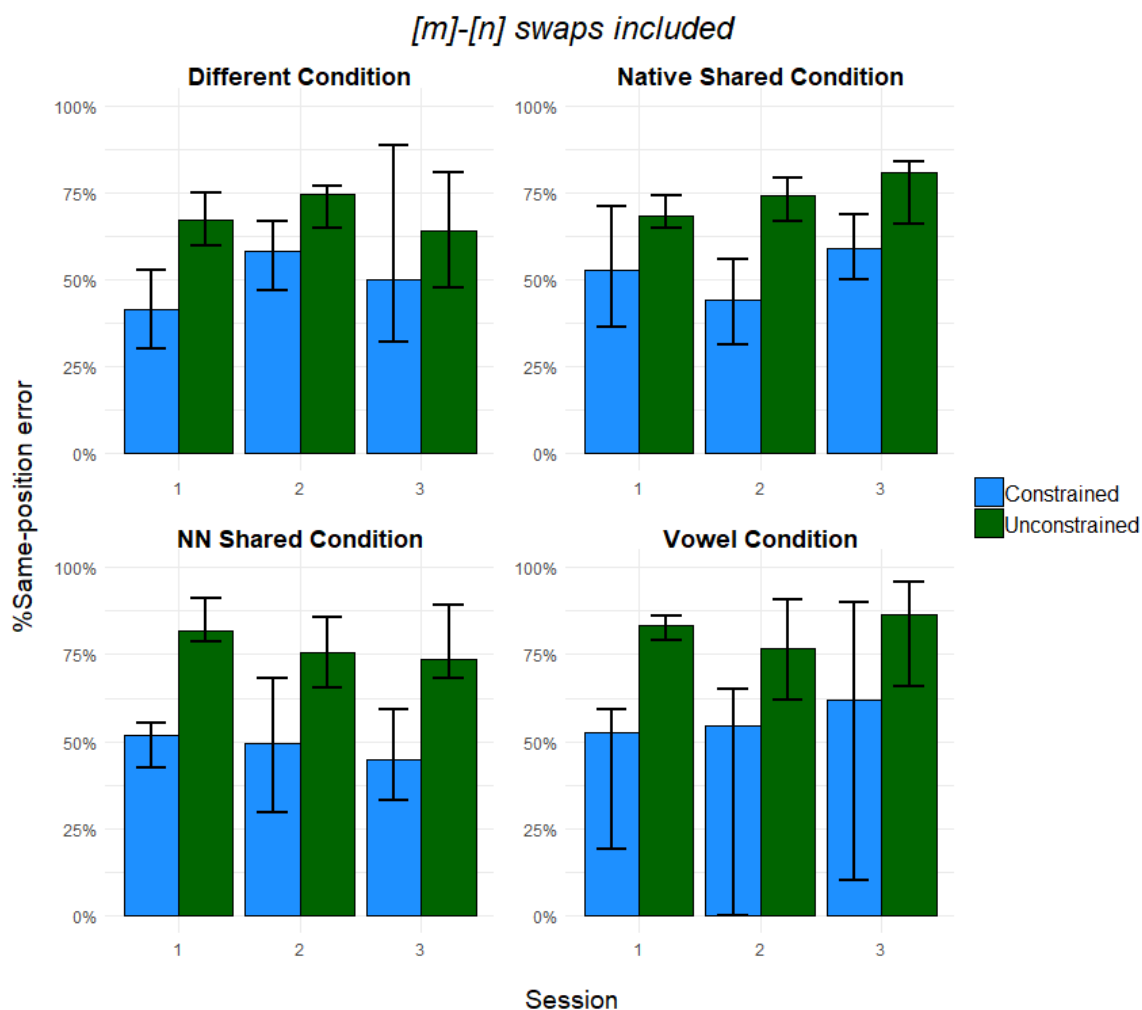
Table 3.3. Target/error matrix for all consonants and conditions. Targets are columns and errors are rows. The No Target column refers to errors that were made on extra syllables (i.e., when a participant produced more than 4 syllables in a single twister). Gradient color-coding reflects the number of errors for a given target/error combination.

	<i>f</i> targets	<i>h</i> targets	<i>k</i> targets	<i>m</i> targets	<i>n</i> targets	<i>ŋ</i> targets	<i>p</i> targets	<i>t</i> targets	No Target
<i>f</i> errors	-	69	85	44	19	12	133	33	29
<i>h</i> errors	55	-	129	32	54	0	52	51	32
<i>k</i> errors	103	158	-	20	11	35	150	126	23
<i>m</i> errors	24	18	22	-	119	99	50	12	20
<i>n</i> errors	16	39	27	231	-	188	13	34	21
<i>ŋ</i> errors	10	0	22	106	231	-	8	12	13
<i>p</i> errors	105	75	174	62	17	5	-	96	31
<i>t</i> errors	44	57	167	9	35	3	163	-	31

Such [m]-[n] swaps had a large negative effect on the maintenance of syllable position because [m] and [n] targets never occurred in the same syllable position in a given twister (i.e., [m]-[n] swaps were always illegal). As such, in addition to analyses conducted with the full data set, post-hoc analyses were also conducted with [m]-[n] swaps excluded from the data set to control for this phonetic similarity effect (Goldrick, 2004, followed a similar procedure). The goal of these analyses was to compare effects between conditions, once phonetic similarity was controlled. Note that while other nasal swaps involving [ŋ] also occurred at a disproportionately high rate, such swaps would not have the same effect on the syllable position effect because [ŋ] targets always occurred in coda position. Unlike [m]-[n] swaps, nasal swaps involving [ŋ] and another nasal did not, by definition, change syllable positions. As such, swaps including [ŋ] were at no point excluded from the data set. In addition, given the issues with perceptual similarity, low-confidence errors were excluded.

The percentage of errors that maintained syllable position, split by experimental session and condition, is shown in Figure 3.2. In within-condition models that included all sessions, the critical interaction was between the constraint term and the session term. No such interactions were significant for any condition except the *Native Shared* condition, which went in the opposite direction of adaptation (i.e., the syllable-maintenance effected decreased for experimentally constrained consonants in sessions 2 and 3; see Appendix G for full model results). Contra to our predictions, adaptation did not increase in later sessions within either the *Vowel* or *Different* conditions (see below for discussion of null result for the baseline *Vowel* condition).

Figure 3.2. Percentage of errors that maintain their syllable position for constrained vs. unconstrained consonants, broken down by session and condition. All data visualized, including [m]-[n] swaps. Error bars reflect 95% confidence interval over participant; however, note that participants contributed different numbers of errors to each bar.



In the between-condition model that included all sessions and compared adaptation in the *Different* condition to the *Shared* conditions, the critical interaction between

constraint, session, and condition was not significant (see Table 3.4)⁸. This suggests that, against our prediction, adaptation to the experimental constraint did not increase more for the *Different* condition in later sessions than for the *Shared* conditions. In the between-condition model that only included sessions 2 and 3, however, the critical interaction between constraint and condition was significant, suggesting that for the later sessions only, there was a larger adaptation effect in the *Different* condition than in the *Shared* conditions (see Table 3.5). This provides evidence for a stronger effect of adaptation in the later sessions of the *Different* condition than in the *Shared* conditions, although it should be noted this was an unplanned analysis.

Table 3.4. Between condition comparison; all sessions; [m]-[n] switches included.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	p
(Intercept)	1.24	0.11		
Constraint	0.3*	0.14	4.72	0.03
Session	0.03	0.13	0.05	0.83
Condition	-0.18	0.19	0.85	0.36
Constraint:Session	-0.12	0.26	0.23	0.63
Constraint:Condition	0.35	0.38	0.87	0.35
Session:Condition	0.46	0.38	1.44	0.23
Constraint:Session:Condition	0.9	0.77	1.36	0.24
Random Effects				Variance
Target Syllable				0.31

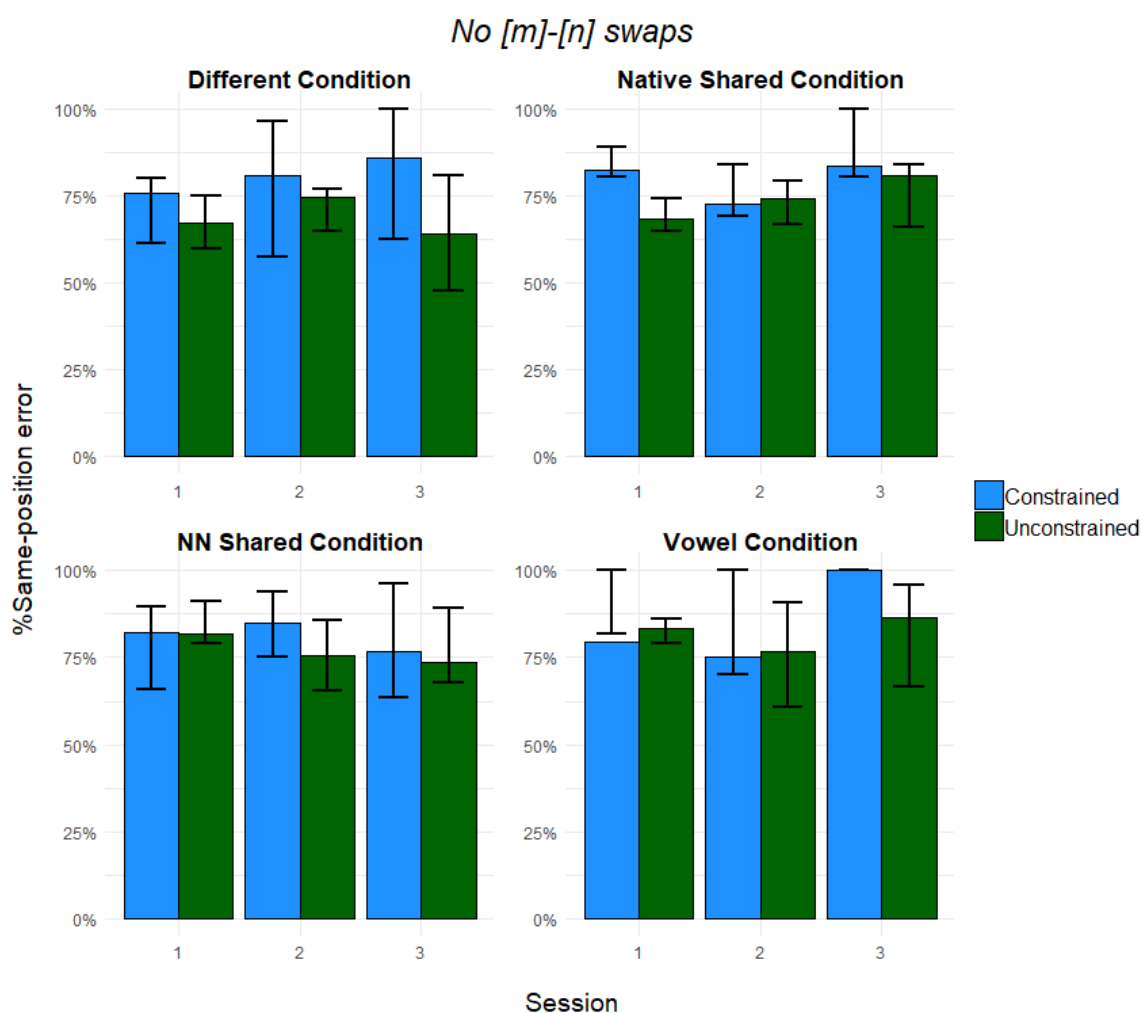
⁸ It should be noted that this critical interaction was marginally significant ($p = 0.07$) when low-confidence errors were excluded, illustrating the sensitivity of the effect to various analysis choices—see below for further discussion.

Table 3.5. Between condition comparison; session 1 excluded; [m]-[n] switches included

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	0.5	0.12		
Constraint	-1.12***	0.15	58.39	<0.001
Condition	0.05	0.19	0.06	0.8
Constraint:Condition	0.81*	0.39	4.32	0.04
Random Effects				Variance
Target Syllable				0.42

In the post-hoc analysis, all [m]-[n] switches were removed from the data set (see Figure 3.3). This has a large positive effect on the syllable maintenance effect for experimentally constrained consonants (raising it from 50.2% to 80.3%), bringing it much closer to the syllable maintenance effect for unconstrained consonants (74.5%). In the within-condition models that included all sessions, but excluded [m]-[n] swaps, the critical constraint:session interaction was not significant within any condition. This suggests that even when removing [m]-[n] swaps, there is no evidence that syllable maintenance increases after session 1 more for constrained consonants than for unconstrained consonants (see Appendix G for model results).

Figure 3.3. Percentage of errors that maintain their syllable position for constrained vs. unconstrained consonants, broken down by session and condition. [m]-[n] swaps were excluded. Error bars reflect 95% confidence interval over participant; however, note that participants contributed different numbers of errors to each bar. In the Vowel condition, one participant made the majority of the errors but had a much lower mean than other participants, resulting in a CI that does not overlap with the mean for Session 1, unconstrained errors.



In a separate within-condition analysis that excludes session 1, the critical main effect of constraint was significant in the *Different* condition (see Table 3.6), but not in any other condition (see Appendix G). This suggests that once [m]-[n] swaps were removed, syllable position was maintained at a higher rate for constrained consonants vs.

unconstrained consonants with sessions 2 and 3. Importantly, this effect does not emerge for any other condition, providing some limited evidence for greater adaptation in the *Different* condition.

Table 3.6. Different condition; session 1 excluded; [m]-[n] switches excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.23	0.2		
Constraint	0.73*	0.35	4.76	0.03
Random Effects			Variance	
Target Syllable			0.81	

Finally, a between-condition analysis was performed on the data excluding [m]-[n] swaps. In a model that included all sessions, the critical three-way interaction between condition, constraint, and session was not significant (see Table 3.7). In a model that only included sessions 2 and 3, the critical two-way interaction was also not significant (see Table 3.8), differing from the effect found when [m] and [n] were included in the analysis.

Table 3.7. Between condition comparison; all sessions; [m]-[n] switches excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.23	0.11		
Constraint	0.32*	0.14	5.48	0.02
Session	0.01	0.13	0.01	0.92
Condition	-0.19	0.19	0.99	0.32
Constraint:Session	-0.08	0.26	0.09	0.76
Constraint:Condition	0.37	0.38	0.95	0.33
Session:Condition	0.49	0.38	1.61	0.2
Constraint:Session:Condition	0.81	0.76	1.12	0.29
Random Effects				Variance
Target Syllable				0.31

Table 3.8. Between condition comparison; session 1 excluded; [m]-[n] switches excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.26	0.13		
Constraint	0.23	0.19	1.55	0.21
Condition	0.05	0.26	0.04	0.85
Constraint:Condition	0.88	0.52	2.93	0.09
Random Effects				Variance
Target Syllable				0.29

The post-test questions revealed that participants generally had little explicit knowledge about the design of the experiment. No participants indicated they were aware of the phonotactic constraints based on talker/vowel. Only 2 participants, of the 10 who were exposed to at least one non-native talker, were able to detect that they had heard a non-native talker. Of these, one heard the non-native talker as being a native speaker of an East Asian language, while the other heard the non-native talker as being a native speaker of Vietnamese. It's unclear whether explicit detection of model talker's non-native language

background is necessary to successfully adapt in the *Different* condition. (Note that previous tongue twister experiments have shown that explicit knowledge of the experimental design have made little difference in participant behavior; e.g., Dell, et al. 2000.)

To summarize, we found inconsistent evidence for our initial predictions. Two significant results, both following from post-hoc analyses, supporting the hypothesis were found: first, that when [m]-[n] swaps were included, participants showed a higher syllable maintenance effect in later sessions of the *Different* condition than in the later sessions of the *Shared* conditions. And second, that when [m]-[n] swaps were not included, constrained consonants maintained their position more frequently than unconstrained consonants within the later sessions of the *Different* condition (and not in any other condition). Overall, however, these effects were quite brittle, and changed depending on which sessions were included in the analysis and whether or not [m]-[n] swaps were included. Moreover, we surprisingly did not find significant result of adaptation in the baseline *Vowel* condition, which makes it difficult to contextualize the results in other conditions.

3.4.1. Discussion

In the current study, participants were exposed to novel, non-native phonotactic constraints contingent on model talker in a tongue twister paradigm, with the language background of model talkers modulated across conditions. The results were ultimately inconclusive, with some weak evidence pointing towards participants acquiring the

constraints when model talkers differed in their language background, but not when they shared a language background. In addition, we unexpectedly did not replicate the well-established adaptation effects from previous tongue twister studies in the *Vowel* condition, which makes contextualizing results from other conditions difficult. Finally, the overall syllable maintenance effect for constrained consonants was surprisingly low due to a large number of [m]-[n] swaps, which further obscured any comparisons between conditions.

What is behind the high number of [m]-[n] swaps? Given that multiple previous studies have used these consonants without such complications, it's likely that the difference in the methodology of the current study—in which participants listen to the tongue twister before producing it—contributed to the abnormally high number of swaps. One possibility, given that this task involved perception, is the high perceptual confusability of [m] and [n]: listeners frequently misidentify [m] as [n], and vice-versa, in speech in noise (e.g., Miller and Nicely, 1955; Phatak and Allen, 2007). It's possible that the similarity of [m] and [n] in perception resulted in down-stream errors in production. While other consonants included in the experiment are also perceptually confusable (e.g., [p] and [k]; see Table 3.3), swaps between unconstrained consonants were not, by definition, always illegal, and thus would not have affected syllable maintenance rates in the same way.

Another contributing factor may have been the coder's perception of participants' [m] and [n] productions. If the coder had an atypical [m]-[n] category boundary, in which one perceptual category was wide and the other was fairly narrow, it could have resulted in frequent miscategorizations. Specifically, we would expect the coder to consistently choose

one nasal over the other. Indeed, there were 231 total [n] errors for [m] targets, and only 119 [m] errors for [n] targets (Miller & Nicely, 1955, find a similar asymmetry in perception). This was fairly consistent across coders, however, suggesting it's likely not related to one coder's idiosyncratic perceptual category boundary.

Smalle et al., (2017)—the only other study to use a similar paradigm—did not find depressed syllable maintenance effects for constrained syllables. Instead, the authors found a surprisingly high syllable-position effect for unconstrained syllables (87.4%, roughly 12% higher than previous studies), which was not replicated in this study. There were a number of important differences between the current study and Smalle et al., (2017) that might have resulted in the divergent findings (different set of consonants and vowels; Dutch participants/twisters, rather than English; 4 sessions; etc.)

Another perplexing finding was the lack of an effect in the *Vowel* condition. When [m]-[n] swaps are removed, the syllable position effect does reach 100% for constrained consonants in the third session, which is in line with previous results. This is over only 13 errors, however, which is far too small a sample to draw definitive conclusions from. This small sample size may have resulted in the lack of an effect—even with [m]-[n] swaps included, the 629 errors would be the second smallest number reported for any previous tongue twister experiment using 2nd-order constraints conditioned on vowels (see Appendix F). This was in large part due to the very low error rate in the *Vowel* condition (2.3%)—lower than any previous experiment with 2nd-order constraints except Smalle et al., (2017). Given that the error rate for the experiment as a whole was in line with previous work, the low error rate in the *Vowel* condition was most likely caused by the

small number of participants, and large variance between participant error rate: the two participants with the lowest error rates, both less than 1%, were in the *Vowel* condition. A related problem was that another participant in the *Vowel* condition had the second *highest* error rate. While this increased the size of the sample, 62% of the errors for the *Vowel* condition came from a single participant, which is clearly not a representative sample.

A second possible issue with the *Vowel* condition was the phonetic similarity of the vowels—[ɪ] and [ɛ] are much more similar than [ɪ] and [æ], the vowels usually chosen in previous tongue twister paradigms. In the post-experiment questionnaire, one participant mentioned they could not consistently tell the model talker's vowel productions apart, while two others sometimes heard the model talkers' [ɛ] vowel as [ɪ]. Given that the phonological features of the consonants involved in a 1st-order constraint affects errors in tongue twisters (Goldrick, 2004), it may also be the case that the similarity of the conditioning vowels in a 2nd-order constraint can weaken the syllable-position effect.

In future follow-up experiments using a similar design, a number of steps can be taken to avoid some of the issues highlighted above. Most obviously, a larger number of participants is required for a more representative sample. That said, there are some logistical obstacles to increasing the number of participants: because the recordings need to be coded for errors by hand, it is a highly time and money-intensive paradigm. One strategy to mitigate this cost is to shorten the length of the experiment. As in previous tongue twister experiments, the number of errors decreases over the course of the experiment, both within each session and across sessions—roughly half of all errors occurred in the first session alone—resulting in diminishing returns. Shortening the length

of the experiment to 96 trials per session and reducing the number of sessions to 2, while increasing the number of participants, will result in more efficient speech error coding, due to the increased density of errors. Second, a follow-up study should use less perceptually confusable experimentally constrained consonants. While most previous studies have used phonetically similar consonants (e.g., [k]-[g], [m]-[n], [f]-[s]), there is no a priori reason for this. Moving away from such pairs should mitigate the possible interaction between the auditory presentation of the twisters and the phonetic similarity of constrained segments, resulting in fewer illegal switches between experimentally constrained consonants.

It's possible a different design may be better suited to investigate the question at hand. One alternative is a modification of the constraint-switching design employed in Anderson et al. (2019). In the first experimental block, participants were exposed to a 1st-order constraint in a tongue twister experiment. This constraint was then reversed in the second block (e.g., if [f] was constrained to onset in Block 1, it was constrained to coda in Block 2), then reversed again in the third block. Anderson and colleagues found slower adaptation for the reversal of the constraint, suggesting an incremental phonotactic learning mechanism. Under the framework presented in this study, however, participants were learning and then unlearning the same laboratory “mini-grammar” in all blocks. In a modified version of the study, twisters could be presented auditorily, with the model talkers switching their language backgrounds between blocks (see Weiss et al., 2009, for a similar design in a statistical learning paradigm). This may serve as a cue to learners that they are in fact being exposed to different languages in each block, spurring faster adaptation, encouraging them to separate the evidence for the conflicting constraints into

different “mini-grammars”. In addition, this paradigm only requires a single session, allowing for the recruitment of a greater number of participants.

3.4.2. Production-perception dynamics and phonotactic adaptation

Recent research exploring the relationship between speech production and speech perception has placed a large emphasis on the role of prediction (e.g., Dell and Chang, 2014; Pickering and Garrod, 2007; Pickering and Garrod, 2013). Pickering and Garrod (2007) propose that when the perceptual system makes predictions—a ubiquitous process in comprehension that occurs at multiple levels of representation—it recruits the production system to construct a forward model that anticipates upcoming linguistic input. The discrepancies, or error, between the forward model’s predictions and the observed input drive future adaptation. Under this framework, phonetic imitation is a result of covert imitation during perception, during which speakers recruit the production system to predict their interlocutor’s speech. This in turn brings the speaker’s production system more in line with their interlocutor’s, resulting in phonetic imitation.

Kittredge and Dell (2016) investigated transfer of phonotactic constraints learned in perception to those learned in production. Participants alternated producing tongue twisters and hearing strings of syllables that conflicted in their phonotactic constraints. Critically, constraints participants were exposed to in perception only interfered with constraints in production when the perception task involved imitation, either by way of silent production or error monitoring without orthographic support (see also Warker et al.,

2009).

In the current study, participants shadowed the model talker's speech. Under Pickering and Garrod's framework, if this shadowing involves covert imitation, and therefore prediction, it should have recruited the production system. While the results of the current experiment are somewhat inconclusive, it's possible this link between perception and production is a critical pathway to pass information about causal inferences—in this case the relationship between phonotactic variation and the model speakers' backgrounds—from the perception system to the production system. It's possible that while the production system in isolation is relatively insensitive to the causal structure underlying phonotactic variation (e.g., Anderson et al., 2019), when it is co-active with the perception system, it gains access to such causal inferences. The inconclusive current results do not provide evidence for this account, however, and follow-up studies are critical to further explore this possibility.

3.5. Conclusion

In the current study, we investigated the role of causal inference in phonotactic adaptation in a tongue twister paradigm. We posited that speakers would use their prior experience with phonotactic variation—that it varies to a large degree between languages, and very little between speakers of the same language variety—to guide their adaptation to novel constraints. We exposed participants to 2nd-order phonotactic constraints conditioned on model talker, while modulating the language backgrounds of the model

talkers. We predicted that participants would adapt to a greater degree when talkers differed in their language backgrounds, as they would construct a separate phonotactic grammar for each talker (i.e., they would detect they were learning two separate “laboratory languages”). The results were ultimately inconclusive, although some evidence was found in the predicted direction. To the extent these effects are reliable, it suggests that participants may use the language background of the model talkers as a cue to a change in phonotactic context between model talkers. If these effects were confirmed in future studies, it would suggest that speech production is more sensitive to such high-level causal inferences than recent evidence has suggested, especially inferences involving the language backgrounds of talkers.

4. Conclusion

In this dissertation, we explored the ways in which the structure of phonotactic variation that speakers experience can induce adaptation to novel constraints in some contexts, and dampen it in others. We start from the assumption that adaptation, across domains, is motivated by prior experience. This is a computational level account of phonotactic adaptation (Marr, 1982): while previous work has focused on the mechanisms involved in phonotactic adaptation (e.g., Warker & Dell, 2006), this dissertation examines why we adapt in the first place, and under what circumstances. This line of inquiry has previously been investigated in the domain of phonetics (Kraljic, Samuel, & Brennan, 2008; Liu & Jaeger, 2018), in which listeners adapt to variation when they have evidence it is systematic and relevant for a given task (i.e., differences between individual talkers are relevant for recognizing speech sounds) but do not adapt when they have evidence the variation is incidental for a given task (i.e., disruptions from a pen in the mouth of a talker are not relevant for recognizing speech sounds in other contexts). This pattern of adaptation suggests that listeners must properly attribute variation to its underlying source for the given task when adapting. In the case of phonetics, listeners make causal inferences about the sources of variation based on the high degree of variation endemic to the phonetics of individual talkers, a contributing factor to the classic “lack of invariance” problem.

The structure of phonotactic variation is markedly different from that of phonetic variation: individual talkers are likely to have similar phonotactic grammars, while talkers of different languages are likely to have distinct grammars. As such, we argued that phonotactic adaptation would behave differently from phonetic adaptation. We predicted speakers would adapt to distinct phonotactic grammars for different talkers to a greater degree when those talkers differed in their language background than when they shared a language background. In each experiment reported in the dissertation, this prediction was tested by exposing participants to two talkers exhibiting different phonotactic constraints (i.e., 2nd-order constraints conditioned on talker identity), while modulating the language background of each talker. We explored adaptation in both perception (Study 1) and production (Study 2); here we summarize the results of each study, consider their implications for theories of adaptation, and posit future directions for this line of research.

4.1. Study 1

In Study 1, we tested adaptation in a recognition memory paradigm, exposing English-speaking listeners to nonsense syllables that reflected different talker-dependent phonotactic constraints (e.g., for Talker A, stops are constrained to onset and fricatives to coda; for Talker B, vice versa). Crucially, talkers either shared a language background (two French or two English talkers), or differed in their language backgrounds (one English, one French talker). In Experiment 1A, we also modulated the strength of the cue to the non-native speakers' language backgrounds, with more or less native English-like vowels ([u] vs. [y]).

Experiment 1A revealed a low degree of adaptation for the *Native Shared* condition, and a moderate degree of adaptation for every other condition. This result largely supported our hypothesis, with a greater degree of learning for the *Different* conditions than the *Native Shared* conditions. There were, however, aspects of the results that were unexpected: first, adaptation occurred in each condition, even those in which both talkers shared a language background. Our original prediction was that no adaptation should occur in such a context, as the shared native backgrounds of the two talkers should block adaptation. It appears that the strength of the bottom-up evidence of talker-specific phonotactic constraints, however, overcame listeners' top-down inferences about the source of the variation—but, critically, adaptation was still lower than in other conditions. A second unexpected result was that the strength of the cue to language background did not appear to affect adaptation. We argued that the weak cue was sufficient for listeners to determine the non-native status of the talker, making further cues redundant.

A third unexpected result, and the most troubling for the hypothesis, was the similar degree of adaptation for the *Different* conditions and the *Non-Native Shared* conditions. We suspected this surprisingly high degree of adaptation for the *Non-Native Shared* condition was a result of the two French talkers sounding dissimilar, due to one talker's somewhat aberrant productions. As such, in Experiment 1B we recorded a novel talker, and re-ran three of the four conditions of 1A (we did not include the *Native Shared* condition).

The results of this follow-up experiment strengthened the evidence for our hypothesis, as the *Different* conditions showed a high degree of adaptation, and the *Non-Native Shared* condition showed only a moderate degree of adaptation. But this result also

had two somewhat surprising aspects: first, why did adaptation increase in the *Different* conditions? It appears that participants specifically did not strongly adapt to aberrant talker's phonotactic constraints in Experiment 1A, regardless of condition. As such, replacing that talker resulted in an overall boost to adaptation in the *Different* conditions. A second surprise was that adaptation did not decrease in the *Non-Native Shared* condition. We argued that the moderate degree of adaptation in this condition was a result of listeners' asymmetric knowledge of native vs. non-native languages. When listeners are exposed to two talkers of a language they are highly familiar with, they are likely more confident these talkers share a language background than when exposed to talkers of a less familiar language. As such, they may be more likely to infer the non-native talkers don't share a language background, and adapt to a moderate degree.

As a whole, the results of Experiment 1 suggested that listeners adapted to a greater degree when talkers differed in their language background. Moreover, it suggested listeners have a structured model of phonotactics, assigning distinct phonotactic grammars to native vs. non-native languages. It remained unclear, however, whether listeners were capable of assigning distinct phonotactic grammars to different non-native languages. In Experiment 2, we addressed this question by exposing listeners to two non-native speakers of different languages (Hindi and Hungarian). We posited that if listeners made distinctions between non-native languages, they would treat any language differences as relevant variation, and therefore adapt to a high degree in this condition. If, on the other hand, listeners simply grouped all non-native talkers together, they would only adapt to a moderate degree. In addition, we included two conditions from Experiment 1: the *Mixed*

Different condition (i.e., one native and one non-native speaker) and the *Non-Native Shared* condition.

Results supported the within non-native distinctions hypothesis: listeners adapted to a high degree of adaptation when talkers differed in their language backgrounds, even when both talkers were non-native. These results suggest that the critical distinction listeners use when positing different phonotactic grammars for different speakers is language difference, rather than the more specific distinction of native vs. non-native. Additionally, the *Non-Native Shared* condition replicated Experiment 1B with a different stimulus design, languages, talkers, and number of items, showing a moderate degree of adaptation.

Study 1 sheds light on the motivation for phonotactic adaptation. Patterns of adaptation reflect the type of phonotactic variation listeners experience from various sources in their daily lives, such as accented English speech or exposure to non-native languages. These results also suggest that phonotactic adaptation is heightened in L2 contexts (see Warker, 2013), and may be a key part of the early stages of L2 acquisition. Learners are faced with profoundly difficult problems, such as discovering word boundaries, at the beginning of L2 acquisition. Rapid phonotactic adaptation in perception, particularly to subset phonotactics, may serve as a fast and efficient adjustment that allows learners to better segment words, as previous work suggests that phonotactic cues in the input are useful for segmentation (e.g., Brent & Cartwright, 1996) and that adults use these cues to guide segmentation (e.g., McQueen, 1998). Adapting to L2 subset phonotactics may allow learners to quickly narrow the set of possible word-forms they have to consider

during comprehension, aiding segmentation and lexical access. Such quick adjustments to L2 speech are likely not available for other phonological structures that are heavily influenced by low-level perceptual processes. L2 superset phonotactics (e.g., consonant clusters not present in the L1), for example, are very difficult to learn (e.g., Parlato, Christophe, Hirose, & Dupoux, 2010), which may reflect in part perceptual interference from listeners' L1s (e.g., Dupoux, et al., 1999). Phonotactic adaptation's role in L2 acquisition is a ripe avenue for future research (see future directions section below).

4.2. Study 2

In Study 2, we examined whether the principles of adaptation we tested in perception—that learners make inferences about the cause of variation based on their prior experience, and these inferences guide adaptation—also hold in speech production, using a modified tongue twister paradigm in which participants repeat after model talkers. Evidence from previous tongue twister experiments suggested that speakers built separate phonotactic grammars in laboratory contexts, and were able to separate evidence for constraints in the laboratory grammar from those used in their native language (Warker, 2013). As in perception in Study 1, we argued that speakers were recruiting their L2 acquisition faculties in adaptation. If adaptation is driven by L2 acquisition, then it follows that learners should be sensitive to cues regarding the language background of the model talkers they are repeating after. The language background of listeners may serve as cue for learners to detect a change in context, and separate the evidence for conflicting phonotactic constraints into separate grammars (e.g., Weiss, et al., 2009). However, a number of recent

studies have suggested that phonotactic adaptation in production is resistant to top-down inferences (e.g., Anderson et al., 2019), suggesting that learners will be unable to separate evidence for the conflicting phonotactic constraints based on talker language background.

Mirroring the design from Study 1, participants were exposed to model talkers who shared a native language background (2 English talkers), a non-native language background (2 German talkers), or differed in their language backgrounds (1 English, 1 German talker), with a 2nd-order phonotactic constraint conditioned on talker identity. A control condition was included in which the phonotactic constraint was conditioned on the identity of the vowel, following previous tongue twister studies (e.g., Warker & Dell, 2006). We predicted a high degree of learning in the *Different* and *Vowel* conditions, and a low-to-moderate degree of learning in the two *Shared* conditions.

The results were ultimately inconclusive. While there was some evidence of a higher degree of learning in the *Different* condition, the effect was brittle, and was not consistent among different analyses. Surprisingly, we found no effect of learning in the control condition, possibly due to a small number of errors, a majority of which came from a single participant. There was also an overwhelming effect of phonetic similarity for the target consonants ([m] and [n]) in all conditions, resulting in a high number of [m]-[n] swaps. Due to the design of the experiment, these swaps, by definition, violated the phonotactic constraint, resulting in surprisingly low overall maintenance of syllable position for the experimentally constrained consonants. This may have been due to the modified design, in which participants heard the twisters before repeating them: the perceptual similarity of nasal consonants may have interfered with the ensuing production.

4.3. Future Directions

The most obvious follow-up to both Studies 1 and 2 are replications. In Study 1, replications should include a greater number of talkers, as listeners appeared sensitive to fine-grained phonetic differences between talkers (see differences between Experiments 1A and 1B). In Study 2, replications should include a greater number of participants, with a design modified in a number of ways (less phonetically similar vowels in *Vowel* condition; less similar experimentally constrained consonants; etc.). An alternate design for Study 2 would expose listeners to a 1st-order constraint in an initial block, followed by a reversal in the following block, with model talker identity changing in each block (Anderson, et al., 2019). While previous studies have shown reduced learning for the reversed constraint, if speakers are sensitive to language background in adaptation in production they may detect a context change from one block to the other, and adapt to an equal degree after the reversal.

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. We predict that minutes of laboratory exposure in the learner's L1 will not

be enough to overcome these priors. One way to test this prediction would be to expose learners to real words in their native language, presented in semantically meaningful contexts. For example, a tongue twister made up of real words (e.g., *pot ham king fin*) could be initially presented to participants as pictures, before orthographic support appears (just as twisters were initially presented auditorily in Study 2). If rapid phonotactic adaptation is part of the process of L2 acquisition we would not expect it to be active in native speech contexts; subsequently, we would predict low rates of adaptation in such cases.

If phonotactic adaptation is indeed tied to L2 acquisition, we might also expect that participants' performance in these tasks may predict outcomes for longer term L2 learning. For example, if rapid phonotactic adaptation is one of the keys to unlocking word segmentation early in L2 acquisition, participants who adapt to a greater degree, and are therefore capable of quickly learning cues to novel word boundaries, may show greater ability to segment words in their L2 after exposure. A study along these lines could give participants a pre-test using the methods in Study 1 before an L2 immersion program, followed by a post-test word segmentation task (e.g., McQueen, 1998).

Another question that arises from this research is how presenting explicit information about talkers would affect participant behavior. 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 and 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* condition in Study 1, 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* vs. *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.

Another natural extension of this dissertation, given our findings that learners adapt to a greater degree when talkers differ in their language backgrounds, is whether learners would also show increased adaptation to talkers of different dialects of the same language (e.g., two regional dialects of the United States). The exact nature of listener's prior experience with dialects and phonotactic variation is not entirely clear. If dialects vary in their phonotactics, but to a lesser extent than languages vary, we would predict listeners would adapt to a moderate degree to talkers of different dialects. While dialects almost certainly have less pervasive variation in categorical constraints (e.g., there's no dialect of English that allows [ŋ] in onset position), they may show more variation for gradient phonotactic constraints (e.g., [s] appears more often in onset than [z]).

While quantifying phonotactic variation is a key challenge for future work, it is unclear how to operationalize and compare distance between phonotactic grammars. While we are confident that languages differ to a greater degree than individual talkers, there is no straightforward method to drawing either quantitative or qualitative distinctions between pairs of languages for a number of reasons. First, it may be the case that not all phonotactic differences are created equal. Differences that have a wider impact on the lexicon of a language, such as constraints on the shape of syllables (e.g., languages limited to only CV syllables vs. languages that allow more complex syllables) should be weighted more heavily than differences with regards to single sounds (e.g., languages that allow [ŋ] in onset vs. those that do not). Second, it is unclear how to compare language with

widely different sound inventories. Unlike, for example, Cantonese, English does not have phonotactic constraints on lexical tone, but of course lexical tone is not a feature of English at all. Finally, it is unclear if phonotactic distance should be quantified in an “objective” way, or if distance should be measured by how detectable phonotactic differences are between languages from the speaker’s point of view. This perceived distance may be the most important aspect for the purposes of 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.

4.4. Conclusions

In two studies, we have investigated the ways in which speakers' prior experience guides phonotactic adaptation. We hypothesized that speakers adapt to relevant or systematic variation, while ignoring irrelevant or incidental variation for the task at hand, based on their previous exposure to phonotactic variation. We found strong evidence for this effect in perception, but only weak/inconclusive evidence in production. This evidence extends theories of adaptation in which speakers make inferences about the causes of variation to a novel domain. In addition, we reframe the phenomenon of phonotactic learning as a part of the L2 acquisition faculty. As a whole, this dissertation explores how speakers contend with and adapt to endemic variation, shedding light on the mechanisms and motivations for adaptation.

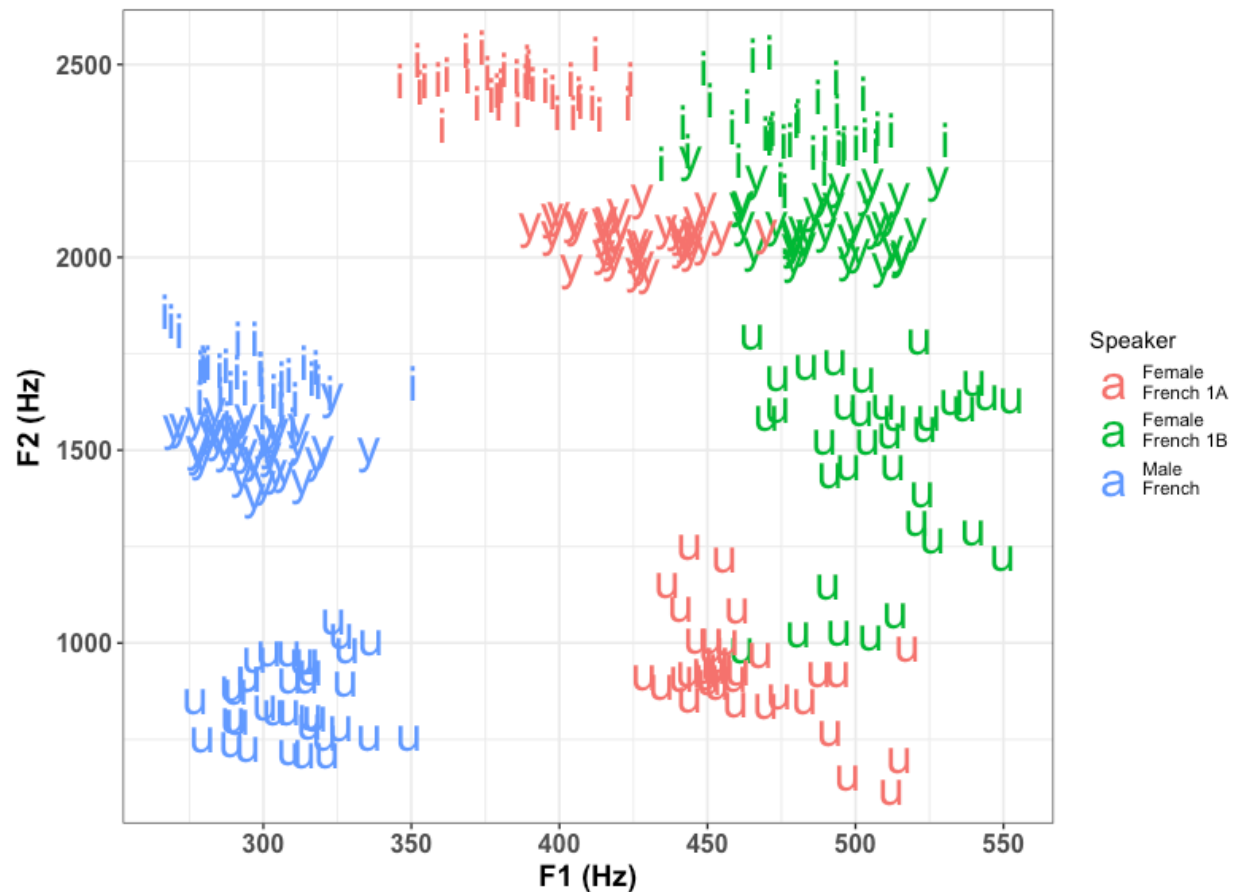
5. Appendix

5.1. Appendix A – Study 1 vowel acoustics analysis

Results from Experiments 1A and 1B found no difference between *Weak* and *Strong Different* conditions, which were differentiated by either a relatively weak phonetic cue to the French talkers' non-native language background (use of the French vowel [u]) or a stronger cue (use of the French vowel [y]). One explanation for this lack of effect is that the *Weak* condition included a sufficient number of acoustic cues such that listeners could confidently infer that the French talkers were non-native. An alternative, however, is that the French talkers' productions of [u] and [y] were not sufficiently acoustically distinct, and therefore the vowel manipulation made little difference to listeners. To investigate this possibility, a post-hoc acoustic analysis was completed of the French speakers' vowel productions in Experiments 1A and 1B.

Vowels intervals for each item were hand-marked in Praat. F1 and F2 values were then automatically measured at the mid-point of each vowel using a script. A total of 324 vowels were measured (108 for each of the French talkers). Note that different LPC (linear predictive coding) settings were used to accurately capture male and female formant values.

Figure 5.1. Scatterplot of French talkers' first and second vowel formants. Each point is a vowel, with vowel identity indicated by the appropriate IPA symbol. Color indicates talker differences.



As shown in Figure 5.1, [y] and [u] are acoustically distinct based on the differences in F1 and F2 for the female French speakers. This suggests the lack of an effect based on cue strength does not stem from a lack of acoustic differences across conditions. Further evidence that English listeners can distinguish these [y] and [u] stimuli comes from Steele, et al. (2015). Using the same stimuli as Experiment 1A, they found that native English listeners can acquire phonotactic constraints conditioned on [y] vs. [u] –suggesting the stimuli are perceptually distinct.

5.2. Appendix B – Study 1 pilot study

Prior to Experiment 1, a pilot study was run that included 3 of the 4 conditions included in Experiment 1A: *Native Shared* (2 English talkers), *Weak Different* (1 English talker and 1 French talker with the [u] vowel), and *Strong Different* (1 English talker and 1 French talker with eh [y] vowel). Other than the exclusion of the *Non-native Shared* condition, the stimuli, design, and procedure of the pilot were identical to those of Experiment 1A. The power analysis for Experiment 1 (see Appendix C) was based on the results from this pilot.

Participants

To reach the target of 48 participants (16 per condition) who passed the experimental criteria, a total of 85 native speakers of English were recruited on AMT (passing rate of 56.5%).

Data Analysis

Data analysis was identical to that in Experiment 1A with one exception: the accent term, comparing the two *Shared* conditions, was not included, as there was only one shared condition. The pilot data was analyzed using a logistic mixed-effects regression, with participant responses as the dependent measure. Fixed effects included **legality**, and two contrast-coded terms: **language difference** (i.e., *Shared* vs. *Different* conditions), and **strength** (i.e., *Weak* vs. *Strong Different* conditions). Interaction terms were included

between legality and both contrast-coded terms. Random effects included random intercepts and random slopes by legality for both participants and items. In addition, follow-up analyses were run on individual conditions, which included a fixed effect of legality, and random intercepts, as well as slopes by legality for items (the models did not converge with random slopes by participant).

Pilot Results

The analysis revealed a main effect of legality ($\beta = 0.49$, s.e. $\beta = 0.12$, $\chi^2(1) = 15.23$, $p < 0.001$), suggesting that participants were, overall, able to learn the constraint. In addition, there was a marginal interaction between the shared term and legality ($\beta = -0.63$, s.e. $\beta = 0.34$, $\chi^2(1) = 3.31$, $p = 0.07$), providing weak evidence that participants adapted to a greater degree in the *Different* conditions. This was consistent with a follow-up analysis showing that there was a significant difference between participant responses on legal and illegal syllables in both of the *Different* conditions (*Strong*: $\beta = 0.73$, s.e. $\beta = 0.19$, $\chi^2(1) = 13.1$, $p < 0.001$; *Weak*: $\beta = 0.46$, s.e. $\beta = 0.19$, $\chi^2(1) = 6.14$, $p < 0.05$), but no such difference in the *Shared Native* condition ($\beta = 0.19$, s.e. $\beta = 0.20$, $\chi^2(1) = 0.92$, $p = 0.34$). This suggested that listeners only adapted to talker-specific phonotactic constraints if speakers differed in their language background, as was found in Experiments 1 and 2. There was no significant interaction between strength and legality ($\beta = 0.32$, s.e. $\beta = 0.30$, $\chi^2(1) = 1.16$, $p = 0.28$), suggesting both conditions provided sufficient cues for listeners to identify a difference in language background, similar to Experiment 1.

5.3. Appendix C – Study 1 power analysis

The number of participants was set to yield sufficient statistical power ($\beta > .8$). Power was estimated by Monte Carlo simulations based on results from a pilot study (see Appendix B for details). Using the estimates for each fixed and random effect in the logistic mixed effects model fit to these pilot data, we generated 1000 simulated data sets. For each simulated data set, we randomly and independently sampled each fixed effect value from a normal distribution (with the mean set to the respective coefficient estimate and standard deviation set to the corresponding standard error estimate) and independently sampled each random effect based on the estimated random effect distributions (correlations between coefficients were not incorporated into our sampling procedure). These were then used to generate a set of recognition memory test response. We then fit the same regression model to these simulated responses. (If the model failed to converge, we generated a new simulated data set.) Statistical power β was estimated by the proportion of the 1000 models in which the crucial interaction term—between the fixed effects for legality and shared/different language background—was found to be significant. We increased the number of participants iteratively, generating novel simulated data sets and running new models with each iteration, until we reached the threshold of $\beta > .8$. This threshold was reached with 64 participants per condition (estimated $\beta = .804$).

5.4. Appendix D – Study 1 passing rates

Recall that experimental criteria in Experiments 1A and 1B were as follows: in the generalization phase, participants had to correctly accept at least 90% of previously heard items (i.e., a hit rate of over 90%) and correctly reject at least 10% of novel items. These criteria ensured that participants were able to recall items they had previously heard multiple times, and that they were able to differentiate between previously heard items and novel items. In Experiment 2, these criteria were loosened due to the phonological confusability of the stimulus set: the hit rate criterion was lowered from 90% to 85%. A third criterion was also added: listeners' hit rate could not exceed their false alarm rate, to ensure that listeners could differentiate novel and familiar items. For example, a participant with a hit rate of 85% must correctly reject at least 15% of novel items (i.e., a false alarm rate no higher than 85%).

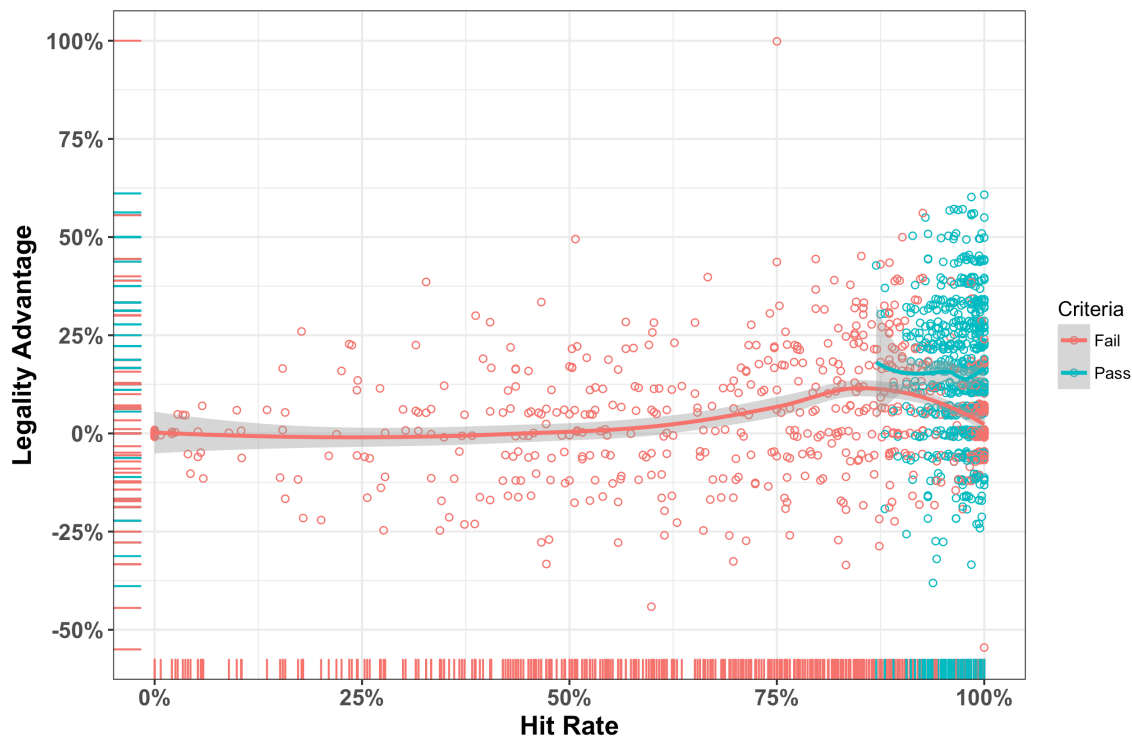
Table 5.1. Passing rates for each condition and experiment in Study 1.

Experiment	Condition	Total participants	Passing Participants	Passing Rate
1A	Native Shared	124	64	51.6%
1A	Non-Native Shared	101	66	65.3%
1A	Weak Different	118	66	55.9%
1A	Strong Different	112	64	57.1%
1B	Non-Native Shared	158	66	41.8%
1B	Weak Different	119	64	53.8%
1B	Strong Different	141	64	45.4%
2	Non-Native Shared	160	66	41.3%
2	Non-Native Different	157	67	42.7%
2	Mixed Different	124	69	55.6%

n.b. The same number of participants (64) were analyzed in each condition. The number of passing participants sometimes exceeded this due to technical limitations in our experimental pipeline.

As shown in Table 5.1, passing rates ranged from 41.3% to 65.3% between conditions. Overall, 655 participants passed the criteria out of 1314 participants (49.8%). This was in line with previous results using this paradigm (Denby, et al., 2018). In a post-hoc analysis, we investigate the relationship between the hit rate in the generalization phase and the legality effect. As Figure 5.2 shows, participants with a hit rate lower than roughly 75% show little to no legality effect. This is unsurprising: if, for example, a participant correctly accepts only half of familiar items, they are simply at chance, and therefore will not show any differences between legal and illegal generalization items, as they are likely guessing. Participants whose hit rate is much lower than 50% may have misinterpreted the experimental instructions, and simply answered “no” to any items they had not encountered prior to the experiment (rather than within the experiment).

Figure 5.2. Scatterplot of hit rate (%yes on familiar items) in generalization phase by legality advantage (false alarm rate for legal items minus false alarm rate on illegal items) for all experiments in Study 1. Each dot represents a single participant; colors represent whether participants passed or failed criteria. Lines represent Loess regression; shading represents 95% confidence interval



Note that the relationship between hit rate and legality advantage is non-linear—for participants who fail the criteria, the legality advantage peaks around a hit rate of 85%; as the hit rate increases to 100%, the legality advantage falls back down to almost 0%. This is also an expected result, as such participants are failing the criteria based on a high false alarm rate: they are responding “yes” to almost every item, regardless of whether it is familiar or novel. Detecting differences in response patterns between legal and illegal novel items is essentially impossible with such a high overall false alarm rate.

These results suggest that the criteria were necessary to filter out participants who were biased towards always responding “yes” or always responding “no”, as well as those who answered randomly (i.e., a hit rate of ~50%). It also appears as though the criteria may have been slightly too restrictive, as participants whose hit rate was above roughly 75% appeared to be tracking the constraint, as shown by their increased legality advantage. In a second post-hoc analysis, we re-plotted the data while loosening the criteria to include participants with a hit rate as low as 75%. As in Experiment 2, we included a criterion that participants’ false alarm rate must be lower than their hit rate, to ensure they are able to differentiate familiar and novel items.

Loosening the criteria resulted in an additional 152 participants passing, increasing the overall passing rate from 49.8% to 61.4%. As can be seen in Figures 5.3 – 5.5, the results of the experiment do not qualitatively change with the addition of these participants. This suggests the criteria as originally set were somewhat overly restrictive, with an additional ~10% of participants unnecessarily excluded. Based on these results, we recommend that future experiments with similar designs should loosen the criteria to 75%.

Figure 5.3. Legality advantage for Experiment 1A, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.

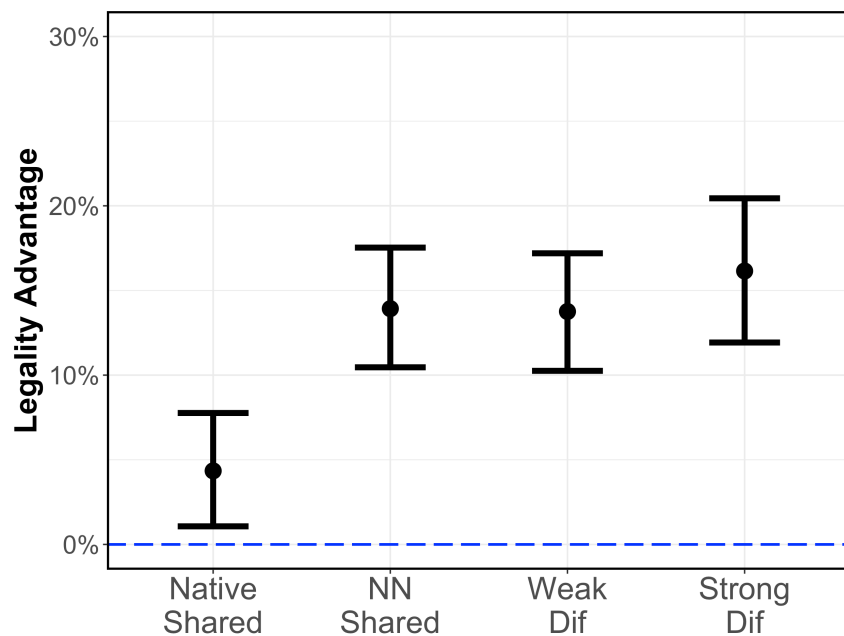


Figure 5.4. Legality advantage for Experiment 1B, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.

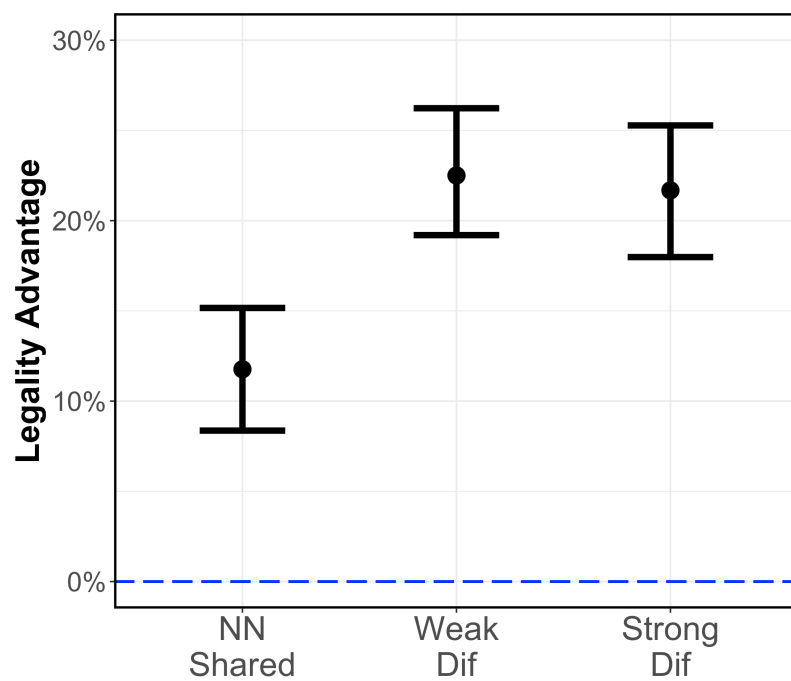
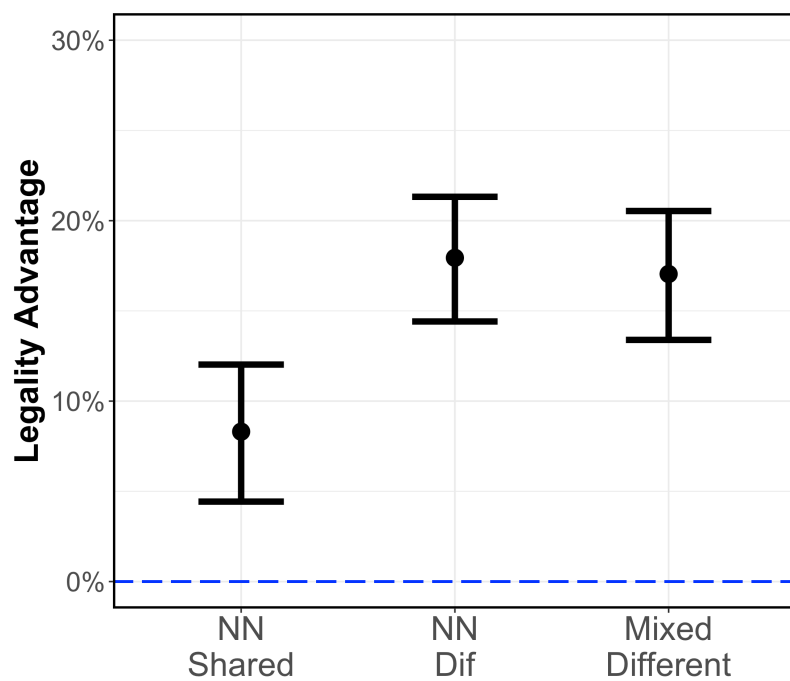


Figure 5.5. Legality advantage for Experiment 2, with hit rate criterion lowered to 75%. Error bars reflect bootstrapped 95% confidence interval.



5.5. Appendix E – Study 1 model results

Table 5.2. Experiment 1A Fixed Effects.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	P
(Intercept)	0.30**	0.09		
Legality	0.64***	0.06	73.63	<0.0001
Language Difference	0.00	0.26	0.00	0.99
Strength	-0.28	0.18	2.43	0.12
Accent	-0.34	0.18	3.41	0.06
Legality:Language Difference	0.72***	0.20	12.59	<0.001
Legality:Strength	-0.05	0.14	0.15	0.70
Legality:Accent	0.45**	0.14	9.78	<.01

Table 5.3. Experiment 1A random effects.

Random Effects	Variance
Intercept (item)	0.31
Slope (item by legality)	0.06
Intercept (participant)	0.91
Slope (participant by legality)	0.06

Table 5.4. Experiment 1B fixed effects.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	p
(Intercept)	0.52	0.09		
Legality	1.02***	0.07	127.64	<0.0001
Language Difference	-1.06***	0.19	28.82	<0.0001
Strength	0.16	0.16	0.99	0.32
Legality:Language Difference	0.57***	0.18	9.81	<0.01
Legality:Strength	-0.02	0.15	0.03	0.87

Table 5.5. Experiment 1B random effects.

Random Effects	Variance
Intercept (Item)	0.36
Slope (item by Legality)	0.06
Intercept (Participant)	0.67
Slope (Participant by Legality)	0.09

Table 5.6. Experiment 2 fixed effects.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	P
(Intercept)	0.97***	0.08		
Legality	0.76***	0.08	62.83	<0.0001
Language Difference	-0.20	0.19	1.13	0.29
Native	0.14	0.19	0.52	0.47
Legality:Language Difference	0.71***	0.21	11.08	<0.001
Legality:Native	-0.04	0.21	0.03	0.87

Table 5.7. Experiment 2 random effects.

Random Effects	Variance
Intercept (item)	0.15
Slope (item by Legality)	0.08
Intercept (participant)	0.45
Slope (participant by Legality)	0.09

5.6. Appendix F – Tongue twister sample size analysis

An analysis of 15 previous tongue twister experiments for which design details and results were accessible found a wide range of total number of syllables analyzed and error rates (see Table 5.8). The average number of items analyzed (words or syllables, depending on the design of the experiment) was 49,152 (range of 9,216-184,320). The current experiment exposes participants to second-order constraints; the average number of items for previous experiments with such constraints was 53,453 (range of 18,432-82,944).

Table 5.8 Number of items analyzed (words or syllables), errors, error rates, and constraint order for previous tongue twister experiments.

Paper	Experiment	Total Items	Total Errors	Error Rate	Order
Dell et al. (2000)	1	36,864	3065	8.3%	1 st
Dell et al. (2000)	2	36,864	3584	9.7%	1 st
Dell et al. (2000)	3	36,864	1769	4.8%	2 nd
Goldrick (2004)	1	184,320	6762	4.8%	1 st
Kittredge & Dell (2016)	1	64,512	5010	5.7%	1 st
Smalle et al., (2017)	1 - Adults	55,296	1240	2.2%	2 nd
Taylor & Houghton (2005)	1	36,864	1313	3.6%	1 st
Taylor & Houghton (2005)	2	9,216	745	8.1%	1 st
Taylor & Houghton (2005)	3	9,216	729	7.9%	1 st
Warker (2013)	1	82,944	3926	4.7%	2 nd
Warker (2013)	2	55,296	3355	6.1%	2 nd
Warker & Dell (2006)	1a	18,432	500	2.7%	2 nd
Warker & Dell (2006)	1b	18,432	1074	5.8%	2 nd
Warker & Dell (2015)	1	55,296	4043	7.3%	1 st
Warker et al. (2008)	1	73,728	5460	7.4%	2 nd
Warker et al. (2009)	1	18,432	391	2.1%	1 st

An initial pilot experiment was previously conducted with 9,216 total items per condition (36,864 total). 11 of the 32 participant sessions were analyzed for errors, for a total of 12,672 items. A total of 433 errors were found, for an error rate of 3.4%. This error rate was within the range of previously found error rates, although it was on the low end of the range. The low error rate, however, and the relatively small number of items, rendered the results uninterpretable, especially when split by condition. The current experiment increases the sample size to 110,592 (27,648 per condition; see above for details).

5.7. Appendix G – Study 2 model results

5.7.1. Within-condition models; all sessions; [m]-[n] swaps included

Table 5.9. Different condition; all sessions; [m]-[n] swaps included.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	P
(Intercept)	0.32	0.14		
Constraint	-0.97***	0.19	25.86	<0.001
Session	0.28	0.18	2.34	0.13
Constraint:Session	0.44	0.36	1.48	0.22
Random Effects				Variance
Target Syllable				0.46

Table 5.10. Native Shared condition; all sessions; [m]-[n] swaps included.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	P
(Intercept)	0.43	0.13		
Constraint	-0.92***	0.17	30.74	<0.001
Session	0.1	0.16	0.38	0.54
Constraint:Session	-0.61	0.31	3.91	0.05
Random Effects				Variance
Target Syllable				0.43

Table 5.11. Non-Native Shared condition; all sessions; [m]-[n] swaps included.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	P
(Intercept)	0.62	0.2		
Constraint	-1.64***	0.21	64.2	<0.001
Session	-0.27	0.18	2.25	0.13
Constraint:Session	0.42	0.36	1.36	0.24
Random Effects			Variance	
Target Syllable			1.2	

Table 5.12. Vowel condition; all sessions; [m]-[n] swaps included.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	0.94	0.18		
Constraint	-1.47***	0.28	26.9	<0.001
Session	0.11	0.27	0.16	0.69
Constraint:Session	0.66	0.54	1.51	0.22
Random Effects			Variance	
Target Syllable			0.86	

5.7.2. Within-condition models; [m]-[n] swaps excluded

Table 5.13. Different condition; all sessions; [m]-[n] swaps excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.12	0.16		
Constraint	0.49	0.26	3.8	0.05
Session	0.27	0.25	1.21	0.27
Constraint:Session	0.39	0.49	0.63	0.43
Random Effects			Variance	
Target Syllable			0.34	

Table 5.14. Native Shared condition; all sessions; [m]-[n] swaps excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	p
(Intercept)	1.17	0.14		
Constraint	0.38	0.21	3.33	0.07
Session	-0.06	0.21	0.09	0.76
Constraint:Session	-0.78	0.41	3.58	0.06
Random Effects				Variance
Target Syllable				0.26

Table 5.15. Non-Native Shared condition; all sessions; [m]-[n] swaps excluded

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	p
(Intercept)	1.52	0.17		
Constraint	0.01	0.25	0	0.97
Session	-0.3	0.23	1.68	0.2
Constraint:Session	0.38	0.45	0.69	0.41
Random Effects				Variance
Target Syllable				0.42

Table 5.16. Vowel condition; all sessions; [m]-[n] swaps excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	p
(Intercept)	1.81	0.25		
Constraint	0.13	0.38	0.11	0.74
Session	0.19	0.37	0.28	0.59
Constraint:Session	0.8	0.73	1.22	0.27
Random Effects				Variance
Target Syllable				0.81

5.7.3. Within-condition models; session 1 excluded; [m]-[n] swaps excluded

Table 5.17. Different condition; session 1 excluded; [m]-[n] swaps excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.23	0.2		
Constraint	0.73*	0.35	4.76	0.03
Random Effects				Variance
Target Syllable				0.81

Table 5.18. Native Shared; session 1 excluded; [m]-[n] swaps excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.23	0.2		
Constraint	-0.01	0.31	0	0.98
Random Effects				Variance
Target Syllable				0.54

Table 5.19. Non-Native Shared condition; session 1 excluded; [m]-[n] swaps excluded.

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	1.39	0.2		
Constraint	0.14	0.33	0.18	0.67
Random Effects				Variance
Target Syllable				0.38

Table 5.20. Vowel condition; session 1 excluded; [m]-[n] swaps excluded

Fixed Effect	Coefficient	Standard Error	χ^2 (1)	<i>p</i>
(Intercept)	2.13	0.42		
Constraint	0.62	0.61	1.13	0.29
Random Effects				Variance
Target Syllable				1.57

6. References

- Anderson, N. D., Holmes, E. W., Dell, G. S., & Middleton, E. L. (2019). Reversal shift in phonotactic learning during language production: Evidence for incremental learning. *Journal of Memory and Language*, 106(March 2018), 135–149.
<https://doi.org/10.1016/j.jml.2019.03.002>
- Atagi, E., & Bent, T. (2013). Auditory free classification of nonnative speech. *Journal of Phonetics*, 41, 509–519. <https://doi.org/10.1016/j.wocn.2013.09.003>
- Baese-Berk, M. M., Bradlow, A. R., & Wright, B. a. (2013). Accent-independent adaptation to foreign accented speech. *The Journal of the Acoustical Society of America*, 133(3), EL174-80. <https://doi.org/10.1121/1.4789864>
- 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(0), 255–278. <https://doi.org/http://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(2015), 18–32.
<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(2017), 138–154.
<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.

<https://doi.org/10.1121/1.1332378>

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. (2014). Journal of Experimental Child The benefits of being bilingual : Working memory in bilingual Turkish – Dutch children. *Journal of Experimental Child Psychology*, 128, 105–119.

<https://doi.org/10.1016/j.jecp.2014.06.007>

Bosch, L., & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination Abilities in Infants from Bilingual Environments. *Infancy*, 2(1), 29–49.

https://doi.org/10.1207/S15327078IN0201_3

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2), 707–729. <https://doi.org/10.1016/j.cognition.2007.04.005>

Brent, M. R., & Cartwright, T. a. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61(1–2), 93–125.

[https://doi.org/10.1016/S0010-0277\(96\)00719-6](https://doi.org/10.1016/S0010-0277(96)00719-6)

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. <https://doi.org/10.1177/1745691610393980>

Caramazza, A., & Yeni-Komshian, G. H. (1974). Voice onset time in two French dialects. *Journal of phonetics*, 2(3), 239–245.

Carlson, M. T., Goldrick, M., Blasingame, M., & Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism*, 19(5), 939–954.

<https://doi.org/10.1017/S1366728915000334>

- Chomsky, N., & Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics*, 1(2), 97–138. <https://doi.org/10.1017/S0022226700001134>
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: the role of talker variation in lexical access. *Cognition*, 106(2), 633–664. <https://doi.org/10.1016/j.cognition.2007.03.013>
- Dart, S. N. (1998). Comparing French and English coronal consonant articulation. *Journal of Phonetics*, 26(1), 71–94. <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. *J Exp Psychol Learn Mem Cogn*, 26(6), 1355–1367. <https://doi.org/10.1037/0278-7393.26.6.1355>
- Dell, Gary S, & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 369(1634), 20120394. <https://doi.org/10.1098/rstb.2012.0394>
- Denby, T., Schechter, 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. <https://doi.org/10.1037/xlm0000465>
- Drager, K. (2010). Sociophonetic variation in speech perception. *Linguistics and Language Compass*, 4(7), 473–480. <https://doi.org/10.1111/j.1749-818X.2010.00210.x>
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in

- Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1568–1578. <https://doi.org/10.1037/0096-1523.25.6.1568>
- Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, 67(2), 224–238. <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. Retrieved from http://jimflege.com/files/Flege_new_similar_JP_1987.pdf
- Fromkin, V. A. (1971). The Non-Anomalous Nature of Anomalous Utterances. *Language*, 47(1), 27–52. <https://doi.org/10.2307/412187>
- Giles, H. (1973). Accent Mobility: A Model and Some Data. *Anthropological Linguistics*, 15(2), 87–105. <https://doi.org/10.1111/j.1460-2466.2008.00398.x>
- Goldinger, S. D. (1996). Words and voices: episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 22(5), 1166–1183. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8926483>
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9577239>
- Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, 51, 586–603. <https://doi.org/10.1016/j.jml.2004.07.004>

- Goldrick, M. (2011). Linking Speech Errors and Generative Phonological Theory. *Language and Linguistics Compass*, 5(6), 397–412. <https://doi.org/10.1111/j.1749-818X.2011.00282.x>
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48(4), 1689–1699. <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 Pt 1), 3099–3111. <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(1), 57–83. <https://doi.org/10.1016/j.cognition.2012.10.013>
- Kessler, B., & Treiman, R. (1997). Syllable Structure and the Distribution of Phonemes in English Syllables. *Journal of Memory and Language*, 37, 295–311.
- Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production. *Journal of Memory and Language*, 89(2016), 8–22. <https://doi.org/10.1016/j.jml.2015.08.001>
- Kleinschmidt, D. F. (2018). Structure in talker variability: How much is there and how much can it help? *Language, Cognition and Neuroscience*, 3798. <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.

<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. <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.

<https://doi.org/10.1016/j.cognition.2011.08.015>.Perceptual

Kraljic, T., Samuel, A. G., & Brennan, S. E. (2008). First Impressions and Last Resorts How Listeners Adjust to Speaker Variability, 19(4), 332–338.

Labov, W. (1969). Contraction, Deletion, and Inherent Variability of the English Copula. *Language*, 45(4), 715. <https://doi.org/10.2307/412333>

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), 2670–2682. <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.

<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. <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.

<https://doi.org/10.1037/0096-1523.33.2.391>

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman and Company. Retrieved from https://mechanism.ucsd.edu/teaching/f18/David_Marr_Vision_A_Computational_Investigation_into_the_Human_Representation_and_Processing_of_Visual_Information.chapter1.pdf

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception.

Perception & Psychophysics, 34(4), 338–348. <https://doi.org/10.3758/BF03203046>

Mattys, S. L., & Jusczyk, P. W. (2001). *Phonotactic cues for segmentation of fluent speech by infants. Cognition* (Vol. 78). [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. <https://doi.org/10.1006/jmla.1998.2568>

Miller, G. A., & Nicely, P. E. (1955). An Analysis of Perceptual Confusions Among Some English Consonants. *The Journal of the Acoustical Society of America*, 27(2), 338–352. <https://doi.org/10.1121/1.1907526>

Nooteboom, S. G. (1973). The tongue slips into patterns. *Speech errors as linguistic evidence*, 144-156.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception.

Perception & Psychophysics, 60(3), 355–376. <https://doi.org/10.3758/BF03206860>

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83(1), 13–23. [https://doi.org/10.1016/S0010-0277\(01\)00165-2](https://doi.org/10.1016/S0010-0277(01)00165-2)

Otake, T., Yoneyama, K., Cutler, A., & van der Lugt, A. (1996). The representation of

- Japanese moraic nasals. *The Journal of the Acoustical Society of America*, 100(6), 3831–3842. <https://doi.org/10.1121/1.417239>
- 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*, 1–45. <https://doi.org/10.1111/lang.12168>
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393. <https://doi.org/10.1121/1.2178720>
- Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1), 190–197. <https://doi.org/10.1016/j.wocn.2011.10.001>
- Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception, and Psychophysics*, 79(2), 637–659. <https://doi.org/10.3758/s13414-016-1226-0>
- Park, H. (2013). Detecting foreign accent in monosyllables: The role of L1 phonotactics. *Journal of Phonetics*, 41(2), 78–87. <https://doi.org/10.1016/j.wocn.2012.11.001>
- Parlato-Oliveira, E., Christophe, A., Hirose, Y., & Dupoux, E. (2010). Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *The Journal of the Acoustical Society of America*, 127(6), 3738–3748. <https://doi.org/10.1121/1.3327792>
- Phatak, S. A., & Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *The Journal of the Acoustical Society of America*, 121(4), 2312–2326. <https://doi.org/10.1121/1.2642397>

- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169–190; discussion 190-226. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15595235>
- Pickering, M. J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11(3), 105–110. <https://doi.org/10.1016/j.tics.2006.12.002>
- Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04), 329–347. <https://doi.org/10.1017/s0140525x12001495>
- Pierrehumbert, J. (2001). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, 16(5–6), 691–698. <https://doi.org/10.1080/01690960143000218>
- Pitt, M. a., & McQueen, J. M. (1998). Is Compensation for Coarticulation Mediated by the Lexicon? *Journal of Memory and Language*, 39, 347–370. <https://doi.org/10.1006/jmla.1998.2571>
- Qian, T., Jaeger, T. F., & Aslin, R. N. (2012). Learning to Represent a Multi-Context Environment: More than Detecting Changes. *Frontiers in Psychology*, 3(July), 1–9. <https://doi.org/10.3389/fpsyg.2012.00228>
- Rebei, A., Anderson, N. D., & Dell, G. S. (2019). Learning the Phonotactics of Button Pushing: Consolidation, Retention, and Syllable Structure. *Journal of Experimental Psychology: Learning Memory and Cognition*. <https://doi.org/10.1037/xlm0000686>
- 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), 539–555. <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).
<https://doi.org/10.1515/labphon.2011.005>
- Schumacher, R. A., Pierrehumbert, J. B., & LaShell, P. (2014). Reconciling Inconsistency in Encoded Morphological Distinctions in an Artificial Language. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*. Retrieved from
<https://mindmodeling.org/cogsci2014/papers/500/paper500.pdf>
- 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. <https://doi.org/10.1121/1.3101452>
- Smalle, E. H. M., Muylle, M., Szmalec, A., & Duyck, W. (2017). The Different Time Course of Phonotactic Constraint Learning in Children and Adults: Evidence From Speech Errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1821.
<https://doi.org/10.1037/xlm0000405>
- Staub Casassanto, L. (2008). Does Social Information Influence Sentence Processing ? *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, (July 2008), 799–804. Retrieved from
<http://csjarchive.cogsci.rpi.edu/Proceedings/2008/pdfs/p799.pdf>
- Steele, A., Denby, T., Chan, C., & Goldrick, M. (2015). Learning Non-Native Phonotactic Constraints Over the Web. In The Scottish Consortium for ICPhS 2015 (Eds.)

Proceedings of the XVIIIth 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.

[https://doi.org/10.1044/1092-4388\(2006/085\).Differentiating](https://doi.org/10.1044/1092-4388(2006/085).Differentiating)

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., & Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: Phonetic context effects. *The Journal of the Acoustical Society of America*, *122*(2), 1111–1129.

<https://doi.org/10.1121/1.2749716>

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.

<https://doi.org/10.1017/S1366728918000421>

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

<https://doi.org/10.1006/jmla.1998.2618>

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. <https://doi.org/10.1037/a0028648>

Warker, J. A., Xu, Y., Dell, G. S., & Fisher, C. (2009). Speech errors reflect the phonotactic constraints in recently spoken syllables, but not in recently heard syllables. *Cognition*, *112*(1), 81–96. <https://doi.org/10.1016/j.cognition.2009.03.009>

Warker, J. a, & Dell, G. S. (2006). Speech errors reflect newly learned phonotactic

- constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 387–398. <https://doi.org/10.1037/0278-7393.32.2.387>
- Warker, J., Dell, G. S., Whalen, C., & Gereg, S. (2008). Limits on Learning Phonotactic Constraints from Recent Production Experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(5), 1289–1295. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *The Journal of the Acoustical Society of America*, 119(1), 597–607. <https://doi.org/10.1121/1.2141003>
- Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, 428(6982), 557–561. Retrieved from <http://dx.doi.org/10.1038/nature02420>
- Weiss, D. J., Gerfen, C., & Mitchel, A. D. (2009). Speech Segmentation in a Simulated Bilingual Environment: A Challenge for Statistical Learning? *Language Learning and Development*, 5(1), 30–49. <https://doi.org/10.1080/15475440802340101>
- Wright, B. a, Baese-berk, M., Marrone, N., & Bradlow, A. R. (2016). Enhancing speech learning by combining task practice with periods of stimulus exposure without practice, 138(2), 1–26. <https://doi.org/10.1121/1.4927411>
- Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, 97, 30–46. <https://doi.org/10.1016/j.jml.2017.07.005>