Cognition 251 (2024) 105909

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Learning a language with vowelless words

Georgia Zellou [a,*], Santiago Barreda [a], Mohamed Lahrouchi [b], Rajka Smiljanić [c]

[a] Linguistics Department, University of California, Davis, USA
[b] CNRS & Université Paris 8, France
[c] Linguistics Department, University of Texas, Austin, USA

ARTICLE INFO

Keywords:
Speech perception
Word learning
Phonotactics
Phonetic variation
Vowelless words

ABSTRACT

Vowelless words are exceptionally typologically rare, though they are found in some languages, such as Tashlhiyt (e.g., *fkt* 'give it'). The current study tests whether lexicons containing tri-segmental (CCC) vowelless words are more difficult to acquire than lexicons not containing vowelless words by adult English speakers from brief auditory exposure. The role of acoustic-phonetic form on learning these typologically rare word forms is also explored: In Experiment 1, participants were trained on words produced in either only Clear speech or Casual speech productions of words; Experiment 2 trained participants on lexical items produced in both speech styles. Listeners were able to learn both vowelless and voweled lexicons equally well when speaking style was consistent for participants, but learning was lower for vowelless lexicons when training consisted of variable acoustic-phonetic forms. In both experiments, responses to a post-training wordlikeness ratings task containing novel items revealed that exposure to a vowelless lexicon leads participants to accept new vowelless words as acceptable lexical forms. These results demonstrate that one of the typologically rarest types of lexical forms - words without vowels - can be rapidly acquired by naive adult listeners. Yet, acoustic-phonetic variation modulates learning.

1. Introduction

Learning new words is a fundamental process in language acquisition. This task seems effortless when babies acquire language (Fenson et al., 1994) but becomes more challenging as we age. Adult learners of a new language have to master all levels of linguistic structure, including production and perception of the sound categories, syntactic structure, prosody, and its vocabulary. Furthermore, when learning new vocabulary, the speakers may contend with words that are significantly different in their structure from the patterns of their first language, or are even rare across languages. Indeed, some phonological patterns occur cross-linguistically more frequently than others. For instance, there is an overwhelming preference for words to contain vowels across languages of the world. Vowelless words are extremely typologically rare, though they are found in some languages, such as Tashlhiyt. Why are vowelless words so rare? Many common phonological alternations and patterns are argued to result from biases in learning (Blevins, 2004; Hayes & White, 2013). Are lexicons containing vowelless words harder to acquire than ones containing only words with vowels? The present study addresses this question.

A critical hypothesis about human language learning mechanisms is that they exhibit constraints that reflect the structures of natural languages (Chater & Christiansen, 2010). Learning studies are one approach to understand asymmetries in typological phonology patterns and also the attentional mechanisms that make some word forms harder to learn than others. Major findings in this line of work reveal that learners of real and artificial languages in lab settings show phonetically-grounded biases toward learning some phonological alternations over others. For example, babies exposed to languages that contain words with phonological and phonotactic patterns that align with natural attested patterns (e.g., C[+voice]VC[−voice] structures) infer and generalize and later prefer novel words with the same abstract structure, but not when the patterns are unattested in natural languages (i.e., word structures follow patterns that group a non-natural class of sounds - /p/, /d/, and /k/ - together) (Saffran & Thiessen, 2003). Adult participants also show biases in short learning experiments that reflect typological tendencies. For instance, vowel harmony (as well as vowel disharmony) can be learned by adults in an artificial language learning paradigm, but not a random vowel alternation pattern (Pycha, Nowak, Shin, & Shosted, 2003). Also, Wilson (2006) trained adults on different

---

* Corresponding author at: Linguistics Department, UC Davis, 1 Shields Avenue, Davis, CA 95615, USA.
E-mail address: gzellou@ucdavis.edu (G. Zellou).

https://doi.org/10.1016/j.cognition.2024.105909
Received 31 October 2023; Received in revised form 24 July 2024; Accepted 25 July 2024

palatalization processes (either /ki/ - > [tʃi] or /ke/ - > [tʃe]) and found that while the learning of velar palatalization in high and mid vowel contexts is similar, generalization of learning was found from [e] to [i], but not from [i] to [e]. It is argued that these biases in learning occur because there are articulatory, acoustic, and perceptual constraints that make the former patterns more typologically common than the latter patterns (e.g., Wilson, 2006; Smolek & Kapatsinski, 2018). Moreover, Creel, Aslin, and Tanenhaus (2006) investigated the role of segment similarity in adult learning of an artificial lexicon and found that consonant competition interferes with word learning more robustly than vowel competition. They also found that onset consonant competition leads to the most confusion during learning, suggesting a privileged status of consonants over vowels in the learning of new lexicons.

Thus, findings from adult learning studies indicate that looking at learning of different types of word structures provides insight into why some phonological patterns are preferred and others are dispreferred cross-linguistically. The current study applies this approach to the learning of vowelless words from Tashlhiyt, a language that contains many words without vowels. Why might vowelless words be harder to learn than words that contain vowels? Vowels are louder and provide a wider range of frequency information than most consonants (Peterson & Lehiste, 1960), which might make them more salient and perceptible for learners. Vowels also provide robust transitional cues about the identity of surrounding consonants (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967), which can be used by learners to perceive the internal phonological structure of words. These acoustic-phonetic properties of vowels might lead to a bias that makes voweled words easier to learn than vowelless words. Alternatively, the null hypothesis is that vowelless words will not be harder to learn due to the importance of consonants in word learning (Creel et al., 2006; Nazzi & Cutler, 2019). The current study tests whether adult English-speaking listeners show differences in learning a lexicon of Tashlhiyt items that contain mostly vowelless words, compared to a lexicon containing all voweled words, from brief listening experience. If the typological rarity of vowelless words stems in part from biases in learning, we predict that the lexicon containing vowelless words will be harder to learn.

Moreover, since acoustic-phonetic biases are argued to be one source of asymmetries in phonological patterns found cross-linguistically, we also examine the role of speech variation in learning. For one, it is well known that infant-directed speech (IDS) has distinct acoustic-phonetic variation from other speech registers (e.g., Cristia, 2013; Gutova, 2015) and, further, that infants show enhanced learning of words presented to them in IDS (Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Thiessen & Saffran, 2009). Whether clear speech, a speaking style with the aim of enhancing intelligibility for adult interlocutors, aids learning is less well understood. In particular, we test whether learning of vowelless words is affected by exposure to hyper- vs. hypo-articulated productions.

Since learning is a domain-general process, the present study can address longstanding issues about the nature of language, attention, and cognition. As we outline in the following sections, looking at learning (Section 1.1) and generalization (Section 1.2) of typologically unusual structures in an understudied language can be informative about the fundamental mechanisms underlying speech processing. We also explore the effect of acoustic-phonetic variation on learning of phonological patterns (Section 1.3) in order to more comprehensively explore the relationship between intelligibility-oriented variation and learning.

1.1. Vowelless words in Tashlhiyt

The present study explores whether vowelless words are harder to learn by teaching native English-speaking adults tri-segmental words in Tashlhiyt, a language that contains many vowelless words. Tashlhiyt (Afroasiatic; iso: [shi]) is an Amazigh language of southern Morocco with approximately 5 million speakers. Tashlhiyt is described as a “consonantal language”, since its phoneme inventory contains a high consonant-to-vowel ratio (35 consonants and 3 vowels[1]) (Ridouane, 2014). Tashlhiyt is known for its extremely permissive phonotactics: common lexical forms consist of word-initial sequences containing plateauing *kti* or falling *rku* sonority profiles, in addition to vowelless words containing only consonants in which lexical vowels are not present, e.g., *tftktstt* ‘you sprained it (FM)’ (Boukous, 1987, 2009; Dell & Elmedlaoui, 2012; Jebbour, 1996, 1999; Lahrouchi, 2010, 2018; Ridouane, 2008; Ridouane, 2014).

The study of language learning mechanisms for a language that permits vowelless words is vastly underrepresented in the literature. However, there is much work that has considered how the rarity of vowelless words in languages of the world speaks to fundamental issues in phonological typology. The robust typological generalization that languages prefer to have words with vowels is consistent with segmental sequencing accounts that there is a preference for syllables to contain rises in sonority from periphery to center (Clements, 1990; Zec, 1995). While some languages do permit syllabic consonants in restricted environments (e.g., sonorant consonants in unstressed syllables in German and English, such as bottle, button), allowance of consonant nuclei in stressed positions (e.g., monosyllabic words) is less typologically common. Within the small set of languages that allow vowelless words, most only prefer sonorant consonants to be syllabic, such as Slovak (syllabic liquids) and Yoruba (syllabic nasals) (Bell, 1970). Thus, the preference for voweled words, and within the subset of languages that allow vowelless words that there is greater allowance of syllabic sonorants than syllabic obstruents, has been argued to support a preference for highly sonorant units to be present in word forms.

The most cross-linguistically rare phenomenon is that of syllabic obstruents, the least sonorous segment category. Tashlhiyt allows any segment in its inventory to occupy the center of a word; there are few restrictions on consonant sequences, and vowelless words of all shapes are quite common within the language (Ridouane, 2014). Ridouane (2008: 328) reports that in a collection of texts of Tashlhiyt, close to a fourth of the lexical items are vowelless words. Lahrouchi (2010) compiled a list of over 200 bi- and tri-segmental Tashlhiyt verbs and categorized them into classes based on the types of consonant sequences they contained: only about a third of the verbs contained vowels (e.g., *knu* ‘lean’), the remaining were vowelless. Within vowelless items, there are roughly equal proportions of words containing initial sonorant-obstruent clusters (i.e., falling sonority; e.g., *rgl* ‘knock’), initial cluster of obstruents (i.e., plateauing sonority onsets; e.g., *bdr* ‘mention’), and initial obstruent-sonorant clusters (i.e., rising sonority, e.g., *krz* ‘plow’). Thus, vowelless words are commonly occuring word forms in Tashlhiyt and, within tri-segmental vowelless verbs, rising, falling, and plateauing sonority profiles are frequent forms.

There is a large amount of work investigating the articulatory and phonetic properties of vowelless words in Tashlhiyt and other languages were they are found, some with a focus on understanding what makes them phonologically stable within languages in which they are found (Fougeron & Ridouane, 2008; Pouplier & Beňuš, 2011; Ridouane, 2008; Ridouane & Fougeron, 2011). In particular, it appears that the coordination of articulatory gestures for adjacent consonants can be highly non-overlapping in Tashlhiyt (Hermes, Mücke, & Auris, 2017; Hermes, Ridouane, Mücke, & Grice, 2011; Ridouane, 2014). This means that sequences of consonants within Tashlhiyt words are minimally coarticulating, an articulatory setting that is also found in other languages that allow vowelless words (Pouplier & Beňuš, 2011). Wide gestural timing of consonant sequences within words can often result in the presence of transitional elements between segments, which can range in their realization from release bursts to voiceless vocoids to schwas (Fougeron & Ridouane, 2008). These elements are not considered epenthetic vowels, but rather a phonetic product of non-coarticulating consonant sequences

---

[1] /b, m, f, t, tˤ, d, dˤ, n, r, rˤ, s, sˤ, z, zˤ, l, lˤ, ʃ, (ʃˤ), ʒ, ʒˤ, j, k, kʷ, g, gʷ, w, q, qʷ, χ, χʷ, ʁ, ʁʷ, ħ, ʕ, h/ vs. /a, i, u/

(Dell & Elmedlaoui, 2012; Ridouane, 2008). Some recent work has explored the perception of vowelless words in Tashlhiyt. Zellou, Lahrouchi, and Bensoukas (2024) had both native and non-native Tashlhiyt listeners discriminate between minimal pairs varying in the middle segment that were either vowelless words (e.g., *ʁbr* vs. *ʁdr*) or voweled words (e.g., *fan* vs. *fin*). They found that for both types of listeners, discrimination performance was overall similar for vowelless and voweled minimal pairs. Within vowelless words, clearly spoken hyperarticulated words were easier to discriminate between than reduced forms of words suggesting that the acoustic-phonetic properties of vowelless words is a factor in their perceptibility.

In the current study, we test whether Tashlhiyt-naive listeners can learn a lexicon consisting of mainly vowelless words. There are several reasons why an examination of learning of vowelless words is relevant for understanding fundamental cognitive mechanisms of perception. First, vowelless words are the most rare phonological word forms. Are they rare because they are harder to learn? Addressing this question can shed light on major issues in phonological typology. Second, looking at learning of typologically unusual structures in an understudied language can also be informative about the fundamental mechanisms underlying speech processing. As mentioned earlier, there is work showing that consonants are more informative to early word acquisition than vowels (Creel et al., 2006; Nazzi & Cutler, 2019), therefore, it could be the case that vowelless words are not harder to learn than voweled words. This would further suggest that other underlying cognitive factors give rise to the cross-language rarity of these word forms.

1.2. Generalization of learning

We also explore the question of how phonological patterns influence generalization of learning. Phonotactic patterns within the lexicon, specifically, is something that learners show abstract knowledge about, based on the structure of words in their lexicon, and apply to novel words. 9 month old infants, for instance, prefer to listen to nonce words that contain phonological patterns that are highly frequent in their ambient language (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994; Mattys & Jusczyk, 2001).

Prior work has shown that biases in learning can emerge in the ability to generalize patterns from learned words to novel words in an experimental setting (e.g., Wilson, 2006). In the current study, we test generalization of learning by having participants rate the well-formedness of novel words for the experimental language after learning. Wordlikeness tasks are a method used to gauge higher-level phonological processing since listeners compare the phonetic and phonological properties of stimuli to the characteristics of words in their memory. Prior work has shown that language-specific phonotactic patterns affect wordlikeness judgments (Hay, Pierrehumbert, & Beckman, 2004; Treiman, Kessler, Knewasser, Tincoff, & Bowman, 2000). And, nonwords with greater lexical support are rated as more wordlike than nonwords with less (Frisch, Large, Zawaydeh, & Pisoni, 2001; Munson, 2001).

Here, we use performance of novel vowelless words on a wordlikeness task as evidence of whether learning has generalized or not following exposure to the vowelless language by some of the participants. All participants rated both voweled and vowelless novel words, regardless of whether they had been trained on vowelless words or not in the learning phase. We predict that participants who had learned the lexicon containing vowelless words will be more likely to accept novel vowelless words as acceptable lexical forms than those who only learned voweled words. This would be evidence that even after short exposure, participants can learn typologically rare phonotactic constraints and apply them to novel words.

1.3. The role of within-talker acoustic-phonetic variation on learning

We were also interested in examining how exposure to Clear/hyperarticulated vs. Casual/reduced speech forms of words influence learning. Such naturally occurring speaking style variation has been shown to be better perceived by listeners, to enhance recognition memory and recall, and to improve speech segmentation (Guo & Smiljanic, 2023; Scarborough & Zellou, 2013; Smiljanić, 2021; Smiljanić & Bradlow, 2011). There is some evidence that exposure to hyperarticulated speech forms supports learning: Escudero, Benders, and Wanrooij (2011) showed that vowel categorization was improved for L1 Spanish learners of Dutch after the exposure to the vowels with extreme formant values compared to the exposure including only average values. Similarly, exposure to the more extreme third formant frequency (F3) and duration values led to better learning of the English /r/ - /l/ contrast by Japanese learners (Iverson, Hazan, & Bannister, 2005). One study to date, examined how clear speech affected artificial language (AL) learning (Guo & Smiljanić, 2021). In that study, English-speaking listeners were exposed to the speech streams containing uninterrupted repetitions of the new language's 'words.' After the exposure, they were asked to recognize the AL words in a two-alternative forced-choice test. Results showed that, compared to conversational speech, clear speech facilitated segmentation of the AL words by statistical learning, that is by tracking the probabilities of which syllables co-occurred more frequently within the speech streams. These results suggest that listeners use hyperarticulated/clear speech acoustic-phonetic enhancements to guide learning.

In Experiment 1, we extend this line of work to examine whether exposure to casual/reduced or hyperarticulated clear speech enhances learning for vowelless and voweled words. We exposed listeners to *either* Clear speech or Casual (Reduced) speech forms of words during learning. We predict that hyperarticulated forms would enhance learning across both language conditions. As mentioned above, vowelless words might be harder to learn because they have fewer acoustic cues to their internal phonological structure (Ohala & Kawasaki-Fukumori, 1997). However, in hyperarticulated speech, speakers often enhance the phonetic features to phonological contrasts making words more intelligible and easier to perceive across languages (Kang & Guion, 2008; Smiljanić & Bradlow, 2005; Zellou, Lahrouchi, & Bensoukas, 2022) which should lead to better learning compared to the casual forms. This would suggest that vowelless words are harder to learn, but only when they are produced in reduced speech conditions. Alternatively, if vowelless words are simply phonologically dispreferred, regardless of their phonetic realization, they should be harder to learn even when they are hyperarticulated.

In Experiment 2, we expose listeners to *both* Clear and Casual forms of words in training and testing. In real-life second language exposure contexts, learners hear variable forms of words. What is the effect of increased acoustic variation on learning? The evidence from prior work is mixed: in some cases, greater acoustic-phonetic variation during exposure leads to better learning; for instance, hearing minimal pairs across a range of voices facilitates learning (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Rost & McMurray, 2009), as does hearing a word produced in variable pitch and length (Galle, Apfelbaum, & McMurray, 2015). In those cases, it is argued that exposure to variable forms of words allows listeners to create more generalizable representations for items more quickly, supporting later recognition (Apfelbaum & McMurray, 2011). Other studies show that exposure to acoustic-phonetic variation can act like "noise" and result in weaker learning (e.g., Quam, Knight, & Gerken, 2017). We are specifically interested in whether exposure to variable acoustic-phonetic forms of words arising from the within-talker speaking style variation results in asymmetrical patterns of learning a vowelless vs. a voweled language. It is possible that increasing meaningful acoustic-phonetic variation will lead to more robust learning and generalization beyond the training stimuli, similar to the high-variability phonetic training and L2 lexical acquisition (e.g., Barcroft & Sommers, 2005). The greater variation could benefit learners in both voweled and vowelless conditions equally. It is also possible that the presence of words in both conversational and clear speech during exposure may draw learners' attention to the salient stylistic variation, parallel to hearing words produced by multiple talkers, thus incurring processing cost and hindering storing of information in memory. In that case, we would expect that any processing benefit found when the speaking style was held constant during exposure (Experiment 1) will disappear when speaking style is variable in exposure (Experiment 2). The effect of learning disruption would be equivalent for both types of words.

1.4. Current study

The current study tests whether lexicons containing tri-segmental (i.e., CCC) vowelless words are acquired at lower rates than lexicons not containing vowelless words by adult English speakers from brief listening experience. To that end, we ran two experiments. Across both experiments, one group of participants learned vowelless words, another group learned just words with vowels. Prior work has shown that phonotactic regularities not present in English could be learned by adult English speakers from brief listening experience (e.g., Onishi, Chambers, & Fisher, 2002). Tashlhiyt is a language spoken by millions of people and it has many vowelless words. A fundamental cognitive principle is that all natural languages are learnable. Therefore, we predict that English listeners can learn vowelless words from short-term auditory exposure. Yet, we ask whether there are asymmetries in learning of a lexicon including vowelless words that provide clues to why these lexical forms are so cross-linguistically rare. On the one hand, if typological patterns reflect biases in learning, then we predict that a lexicon containing vowelless words will be harder to learn than a lexicon containing only words with vowels. On the other hand, as outlined above, Tashlhiyt has unique phonetic properties that make cues to the internal structure of words robust. Therefore, another possibility is that a lexicon containing vowelless words in Tashlhiyt are just as learnable as a lexicon containing only voweled words. If there is no difference across these groups, this would indicate that the rarity of vowelless words cross-linguistically does not stem from a bias in auditory learning.

Since we use speech as a window to understand the question of learning vowelless words we must consider the role of acoustic-phonetic variation in learning. Specifically, speech is highly variable. Listeners encounter extreme, hyperarticulated forms of words, including speech directed to infants, children and adults, and also reduced, casual forms of words. What is the relationship between variation in the input and learning success? Therefore, the role of acoustic-phonetic form on learning these typologically rare word forms is also explored: In Experiment 1, participants were trained on words produced in either only Clear speech or Reduced speech productions of words; Experiment 2 trained participants on lexical items produced in both speech styles. Critically, we predict that the type and distribution of speech style learners hear at training will modulate learning of voweled and vowelless lexicons differently. This allows us to explore interactions between training language type and the training speech style type, as well as between these factors and the speaking style of the items at testing (specific predictions outlined in section 1.3).

Finally, we also explore if exposure to a lexicon with vowelless words leads to more generalized knowledge. In both studies, responses to a post-training wordlikeness ratings task containing novel items reveals if exposure to a vowelless lexicon leads participants to accept new vowelless words as acceptable lexical forms. Again, we expect that the type of lexicon and speech style variation listeners hear in training will affect their acceptance of novel vowelless words in the wordlikeness task. Specifically, we expect that exposure to the vowelless lexicon will lead to increased acceptance of new vowelless words as acceptable lexical forms. This effect may be modulated by speaking style such that exposure to clear speech forms may lead listeners to consider the new vowelless words forms to be better examples of new word forms but only in the matching speaking style. By examining what effect within-talker acoustic-phonetic variation has on learning, as well as patterns of generalization to new word form, our aim is a comprehensive examination of the factors that might shape how vowelless words are learned by adult listeners.

2. General methods

The main question addressed in the present study is whether English-speaking listeners learn a lexicon containing vowelless words to the same extent that they do a lexicon containing only voweled words. Our lexicons contained 16 real Tashlhiyt words each and participants are taught word-image correspondences for the lexicon of which they are assigned. Across the vowelless and voweled language conditions, the type of words within each lexicon varied and all other aspects of the experiment were held constant. Our vowelless lexicons were designed to approximate a language like Tashlhiyt, which contains many vowelless words varying in sonority profiles, as well as voweled words (note that, to our knowledge, there is no language in the world that is made up of exclusively vowelless words, so our goal in constructing the vowelless lexicons was to approximate one that does exist. The 3:1 ratio of vowelless to voweled words in our "vowelless" lexicon is similar to the rate of vowelless words for some parts of the lexicon with Tashlhiyt (Lahrouchi, 2010)). The voweled lexicons in the current study were also constructed of real Tashlhiyt words. In this condition, participants hear only words with vowels. The voweled language condition can be considered a "control" condition - since words with vowels are the most common word form across languages of the world, comparing how listeners acquire a lexicon containing only Tashlhiyt words with vowels serves as a baseline for word learning performance in this particular study.

2.1. Language conditions

All participants learned 16 tri-segmental Tashlhiyt word-image pairings. Participants were randomly assigned to one of the two Language conditions, based on the phonological structure of the training items: either the Vowelless Language condition or the Voweled Language condition.

In the *Vowelless Language condition*, participants learned a list of words that contained 12 vowelless (CCC) words, as well as 4 words with a vowel nucleus (CVC). Within the vowelless words, we selected an even amount of words with the three different sonority profiles: 4 contained rising sonority, 4 contained plateauing sonority, 4 contained falling sonority. We constructed two Vowelless training condition words - half of the participants within this condition were trained on one set of words, the other half were trained on the other lexical set.

In the *Voweled Language condition*, participants learned 16 voweled (CVC) words. Two sets of voweled words were constructed. Within this condition, half of participants were trained on one of the lexical sets, while the other half were trained on the second set of words.

For each language condition, we made two different lexicons to provide more variations of the languages (version "A" and "B") and participants were randomly assigned to one of the conditions and one of the lexicons within that condition. The lexical items used in this study (and image assignment across experimental lists) are provided in Appendix A.

2.2. Speech styles and stimulus preparation

All selected words were produced in a randomized order by a native speaker of Tashlhiyt (one of the authors, ML) in two speaking styles. The recording took place in a sound-attenuated booth using an AT 8010 Audio-technica microphone and USB audio mixer (M-Audio Fast Track), digitized at a 44.1 kHz sampling rate. To elicit Clear Speech, the speaker was given instructions to speak "clearly to someone who is having a hard time understanding you", similar to those used to elicit clear speech in prior work (e.g., Bradlow, 2002; Zellou et al., 2022). The speaker produced each word in two different frame sentences: *ini _ jat tklit* ‘say _ once’, *inna _ baħra* ‘he said _ a lot’. Following the Clear Speech style elicitation, the speaker produced the words in a fast, casual speaking style with the following instructions to speak “as if to a friend or family member who will have no trouble understanding you”, also modeled after those used in prior work (e.g., Bradlow, 2002). The speaker produced the words casually in each of the two frame sentences, as well.

Clear speech is characterized by a range of acoustic modifications, relative to casual speech (Smiljanić & Bradlow, 2009). Prior work has shown that word forms are consistently longer in clear speech than in casual speech (Picheny, Durlach, & Braida, 1986; Krause & Braida, 2002). Therefore, we measured word duration in the target words to verify a difference in clear and casual speech across our stimuli. The average word duration for clear speech productions was 430 milliseconds while the mean word duration for casual speech was 226 milliseconds. We modeled word duration (logged) with a linear mixed effects model using the *lme4* R package (Bates, 2015). The model included fixed effects of Speech Style (Clear vs. Casual, sum-coded) and Word Form (CCC vs. CVC, sum-coded) and a random intercept for Word. The model revealed an effect of Speech style (est. $= 0.14$, $t = 7.3$, $p < 0.001$). No other effects or interactions were significant. Thus, clear speech contained longer word durations than casual speech and the effect of speaking style was consistent across word types.

All items were segmented and excised from their frame sentences and amplitude normalized to 65 dB.

2.3. *General procedure: Training, testing, novel word rating*

The study began with a pre-test of their audio: participants heard one sentence presented auditorily (“She asked about the host”) and were asked to identify the sentence from three multiple choice options, each containing a phonologically close target word (host, toast, coast). All participants passed this audio check.

Participants then completed the experiment which consisted of three phases: training phase, testing phase, and novel word acceptability rating phase.

The training phase consisted of a paradigm adapted from Wong and Perrachione (2007). In this phase, participants were trained to identify word meanings as depicted by drawings of concrete nouns. Each participant was trained on a vocabulary of 16 items. The format of the training session is illustrated in Fig. 1. To facilitate learning, participants were trained on a group of 3 words at a time (i.e., a “trial group” of 3 items). Each of the 16 items was presented a total of 3 times, each item randomly assigned to one of 16 trial groups. In a learning session, participants heard a word production and saw an image which they were told corresponded to the meaning of the word (Fig. 1.A). After three trials, participants were then given a mini-test on the three words they just learned. One of the three words was played and participants selected the correct corresponding image from among three images (Fig. 1.B). After making a selection, feedback was provided to facilitate recognition of the items and to correct if a mistake was made. Participants heard 48 items (16 words * 3 repetitions across different trials groups).

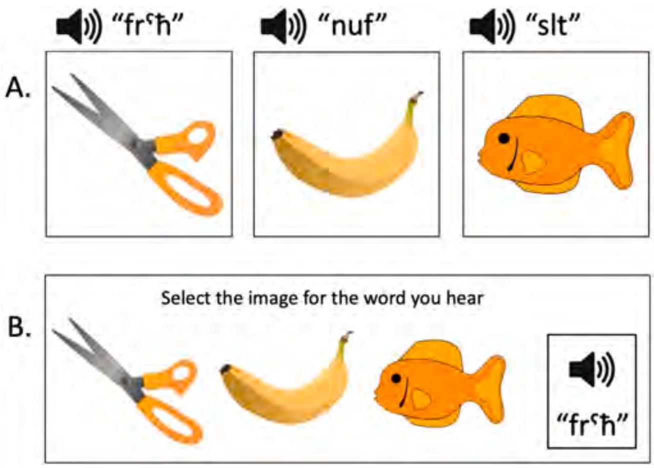

Fig. 1. Example of a training session trial group. Participants first heard each word within a trial group with its corresponding image (A.). Then, participants were quizzed on the words they had just learned in a trial group (B.). In this training phase, feedback was provided for whether they correctly or incorrectly selected the right image for a word. Participants learned 16 word-image associations in the training phase.

In the testing phase, subjects were presented with each of the 16 images. On each trial, they heard one of the 16 learned words and were asked to select the correct image. Words were presented randomly, twice each. No feedback was provided. Participants completed 32 total testing trials (16 words * 2 repetitions).

Following the testing phase, participants completed the novel word acceptability rating phase. In the word acceptability rating phase, participants heard 16 novel Tashlhiyt words (i.e., words not heard in training). The task was a nonword acceptability judgment task (Daland et al., 2011). Each trial consisted of the auditory presentation of a novel word. Listeners were instructed to rate how likely the word they heard could become a word in the language that they had just learned. Participants marked their rating on a sliding scale from 0 (“not at all likely”) to 100 (“very likely”), in increments of 5; the default position of the marker was reset to the midpoint (50) at the start of each trial.

Every participant heard a set of novel words in the rating phase that contained vowelless words and varied in phonological structure: 4 CVC, 4 CCC rising sonority, 4 CCC plateauing sonority, 4 CCC falling sonority. The ratings word set consisted of the items from one of the Vowelless training condition lists. If participants were assigned to a specific language training condition, during the ratings task they heard items from the *other* training set (e.g., if they heard list A in the learning phase, they heard list B in the ratings phase), counterbalanced across participants.

Participants heard each of the 16 novel words in both a clear speech and a casual speech production. Each item was presented to listeners twice in the ratings task. Participants completed 64 total ratings trials (16 words * 2 speech styles * 2 repetitions).

2.4. *Speech style conditions*

Table 1 summarizes the study design of the current experiments. In order to investigate the role of acoustic-phonetic variation on learning of vowelless and voweled words, we varied the nature of exposure to speaking style across two experiments. In Experiment 1, participants were exposed to only one speech style in the word learning phases of the experiment (either Clear or Casual items only). In Experiment 2, participants heard both Clear and Casual items during training and testing. In both experiments and across all conditions, the novel word acceptability phase contained vowelless and voweled words, as well as clear and casual productions of the items.

Table 1
Summary of study design across Experiments 1 and 2.

	Word learning (training and testing phases)	*Novel word acceptability phase*
Experiment 1	Vowelless or Voweled language between-subjects; Speech style between-subjects	Word type and Speech style within-subjects
Experiment 2	Vowelless or Voweled language between-subjects; Speech style within-subjects	Word type and Speech style within-subjects

3. Experiment 1: Speech style as between-subjects during learning

3.1. Participants

Eighty-five native English speakers (46 female, 4 non-binary, 1 genderqueer, 34 male; mean age = 21.3 years old) completed the experiment online via a Qualtrics survey. Participants were recruited via Academic Prolific and paid for their participation. Participants were instructed to complete the experiment in a quiet room without distractions or noise, to silence their phones, and to wear headphones. All participants completed informed consent before participating. None of the listeners reported having a hearing or language impairment.

All of the participants reported being native speakers of American English. Thirteen participants reported that they speak a language other than English in the home (Cantonese, $n = 2$; Japanese, $n = 1$; German, n = 1; Igbo, n = 1; Korean n = 1; Laos, n = 1; Marathi & Hindi, n = 1; Spanish, n = 2; Tagalog, n = 2; Vietnamese, n = 1), none of which are languages that allow vowelless words. We asked the English-speaking participants if they spoke or had studied Tashlhiyt or any of the languages of North Africa; none reported that this was the case.

Participants were randomly assigned to a Language and Speech Style condition: 40 were assigned to a Vowelless Language (20 Clear Speech Version, 20 Casual Speech Version); 42 were assigned to a Voweled Language (20 Clear, 22 Casual). The study began with a headphone check. They heard two sentences presented ("Bill heard we asked about the host", "I'm talking about the bench") and were asked to select the correct sentence from a set of options containing phonological competitors of the final word (e.g., "Bill heard we asked about the coast", "Bill heard we asked about the toast"). If they did not select the correct sentence, they were asked to complete the headphone again. Once participants passed the headphone check procedure, they instructed not to change the volume until the experiment ended.

3.2. Results

3.2.1. Word learning performance

Responses to the test portion of the learning phase were coded binomially for whether the participant accurately identified the correct image associated with the word in training (Correct = 1) or not (=0). Four participants whose accuracy was near zero (<10%) in the test phase (9, 40, 43, 74) were removed from the analyses. We fit a Bayesian multilevel logistic regression model with Stan (Team, 2024) using the *brms* package (Bürkner, 2017) in R (R Core Team, 2024). Fixed effects included Exposure Language Condition (Vowelless vs. Voweled) and Exposure Speech Style Condition (Clear vs. Casual) as well as the interaction between these predictors. Factors were sum-coded. The random effects structure consisted of random intercepts for participant and word, with by-word random slopes for Speech Style Condition. Bayesian inference relies on inspecting the posterior distribution of parameter estimates given the data and model structure. Results will be summarized using the mean μ, the standard deviation (s.d.) σ, and upper and lower bounds (α,β) of the 95% credible interval (C.I.) in the format: (mean = μ, s.d. = σ, 95% C.I. = $[\alpha,\beta]$).

$$\begin{gathered} Correct \sim \mathrm{Bernoulli}(p) \\ p = \mathrm{logit}^{-1}(z) \\ z = exp.\mathrm{clear}*exp.\mathrm{ccc} + (1|id) + (exp.\mathrm{clear}|word) \end{gathered} \tag{1}$$

Fig. 2 presents participants' learning performance across Exposure Language and Exposure Speech Style conditions. Information regarding the model fixed effects is provided in Table B1 in Appendix B and presented visually in Fig. 3.

The mean estimated word learning performance is 0.81 (s.d. = 0.21, 95% C.I. = [0.41, 1.23]), or a percentage of 69% accuracy. Chance level is 1/16 since participants had to pick one out of 16 choices in each test trial. This indicates that overall, participants successfully learned target names in the experiment, higher than the chance level performance of $p = 1/16 = 0.0625$.

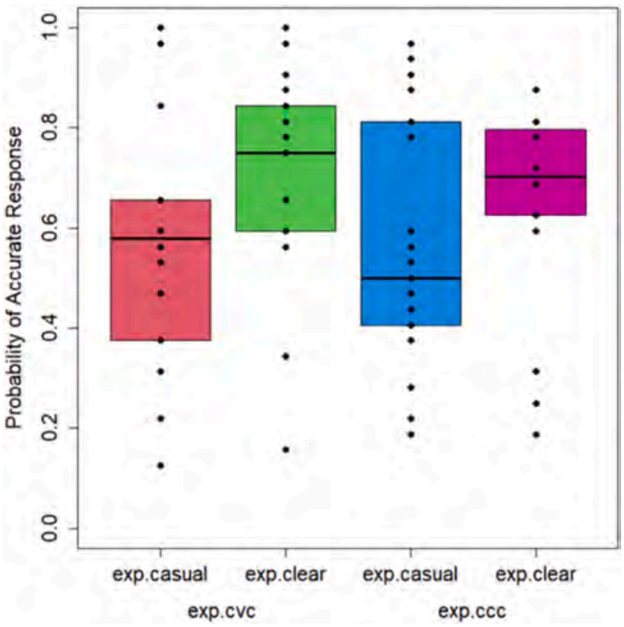

Fig. 2. Points represent accurate response rate by subject in each condition in Experiment 1. Boxes span interquartile ranges and lines indicate group medians.

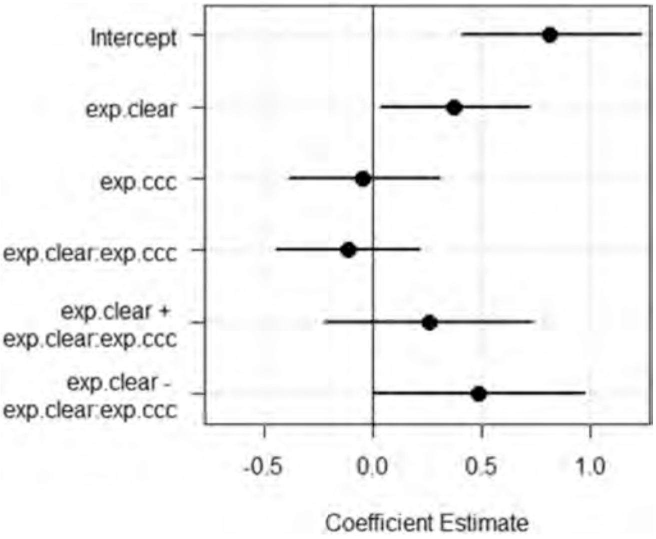

Fig. 3. Experiment 1 Learning performance model coefficients. Points indicate posterior means, lines indicate the 95% credible intervals for parameters.

There was not an effect of Exposure Language condition: word learning performance was the same across participants exposed to a vowelless language and a voweled language. Yet, there was an effect of Exposure Speech Style condition indicating that participants exposed to productions in Clear speech learned words better than those exposed to Casual speech productions (mean = 0.37, s.d. = 0.18, 95% C.I. = [0.03, 0.7]).

We inspected the magnitude of the simple effects for Exposure Speech Style across levels of Exposure Language Condition, also presented in Fig. 3. These indicate that the positive effect of clear speech is nearly two times greater in the voweled language condition (mean = 0.49, s.d. = 0.24, 95% C.I. = [0, 0.97]) relative to the vowelless language condition (mean = 0.26, s.d. = 0.27, 95% C.I. = [−0.23, 0.73]).

3.2.2. Novel word acceptability ratings

Acceptability ratings of the 16 novel items in the wordlikeness task could take on values from 0 to 100, which we scale to range from 0 to 1. Because these response values are bounded, and because 12% of responses were either 1 or 0, a zero-one inflated beta model is the most appropriate. These models predict responses using a mixture model comprising of three processes: 1) a value distributed according to a beta distribution with mean and variance parameters equal to μ and ϕ, which generates values between 0 and 1; 2) the probability that the variable will *not* be beta distributed, but will instead be a one or zero, α; and 3) given that the response is 1 or 0 and not beta distributed, γ is the probability that the response will be a 1 as opposed to a zero.

The model included parallel fixed effects structure for prediction of average wordlikeness (μ), probability of switching to a one or zero (α), and probability of observing a 1 (γ). Fixed effects were Exposure Language Condition (Vowelless vs. Voweled), Exposure Language Speech Style Condition (Clear vs. Casual), Novel Ratings Word Type (CVC vs. CCC), and Novel Ratings Word Speech Style (Clear vs. Casual). All factors were sum-coded. The three-way interaction between Exposure Language Condition, Exposure Speech Style, and Novel Ratings Word Type and the three-way interaction between Exposure Language Condition, Exposure Speech Style, and Novel Ratings Word Speech Style were included, as well as all possible two-way interactions subsumed within each three-way interaction. Variation in the variance (ϕ) was not modeled as no hypotheses relate to variation in this parameter. We also included random intercepts for participant and word, as well as by-participant random slopes for Novel Ratings Word Type, Novel Ratings Word Speech Style and their interaction for the prediction of average rating (μ). A model with this structure, presented in Eq. 2, was fit using *brms* in R.

$$\text{Rating} \sim \text{ZOIB}(\alpha,\gamma,\mu,\phi) = \begin{cases} \alpha(1-\gamma) & \text{Rating} = 0 \\ \alpha\gamma & \text{Rating} = 1 \\ (1-\alpha)\text{beta}(\mu\phi) & 0 < \text{Rating} < 1 \end{cases}$$

$$\begin{aligned} logit(\alpha) &= (exp.ccc + stim.clear)^{*}(exp.clear^{*}stim.ccc) + (1|id) + (1|word) \\ logit(\gamma) &= (exp.ccc + stim.clear)^{*}(exp.clear^{*}stim.ccc) + (1|id) + (1|word) \\ \text{logit}(\mu) &= (exp.\text{ccc} + \text{stim.clear})^{*}(exp.\text{clear}^{*}\text{stim.ccc}) + (1|id) + (exp.\text{clear}|word) \\ log(\phi) &= \text{Intercept} \end{aligned} \tag{2}$$

Fig. 4 displays participants' wordlikeness ratings for novel lexical items across exposure language, exposure speech style, and stimulus conditions. The model fixed effects are provided in Table B2 in Appendix B and fixed effects related to average wordlikeness rating (μ) are presented in Fig. 5 (fixed effects related to α and γ are plotted in Fig. B1 in Appendix B).

There were two interaction effects that had positive effects and small confidence intervals. First, as seen in the top panel of Fig. 4, there was an interaction between Exposure Language and Novel Ratings Word Type (mean = 0.14, s.d. = 0.02, 95% C.I. = [0.09, 0.18]): for instance, participants who were exposed to a vowelless lexicon in training (containing 12 vowelless words and 4 voweled words) gave higher wordlikeness ratings to novel vowelless words than to novel voweled words; the reverse pattern was observed for listeners exposed to the voweled language who gave higher ratings to novel CVC items.

Secondly, the interaction between Exposure Language Speech Style and Novel Ratings Word Speech Style was also positive (mean = 0.11, s.d. = 0.02, 95% C.I. = [0.06, 0.16]). As illustrated in the middle panel of Fig. 4, exposure to a Clear Speech Language leads to higher wordlikeness ratings of novel words produced in clear speech; but this reverses when participants learn words in a Casual Speech Language, in which they are subsequently more likely to rate novel casual speech words as likely words in the language than clear speech forms.

Additionally, three effects were associated with negative coefficients, but the intervals were quite large, or overlapping with zero to a small extent, suggesting a lot of noise in the data. First, there was an effect for Exposure Speech Style condition: wordlikeness ratings were overall lower when participants had been exposed to only Clear speech forms of words in training (mean = −0.10, s.d. = 0.05, 95% C.I. = [−0.2, −0.002]). Second, there was also a negative effect for vowelless word stimuli (mean = −0.05, s.d. = 0.04, 95% C.I. = [−0.14, 0.03]), indicating that novel vowelless words received overall lower wordlikeness ratings than novel voweled words. There was also a negative effect for the interaction between Exposure Speech Style and Novel Ratings Word Form: as seen in the bottom panel of Fig. 4, participants who were exposed to Clear speech in training gave lower wordlikeness judgments to novel vowelless words (mean = −0.04, s.d. = 0.02, 95% C.I. = [−0.08, 0]).

4. Experiment 2: Speech style as within-subjects during learning

4.1. Participants

Forty native English speakers (25 female, 2 non-binary, 1 genderqueer, 12 male; mean age = 21.5 years old) completed the experiment online via a Qualtrics survey. Participants were recruited via Academic Prolific and paid for their participation. As in Experiment 1, participants were instructed to complete the experiment in a quiet room without distractions or noise, to silence their phones, and to wear headphones and all completed informed consent before participating. None reported having a hearing or language impairment. They completed a headphone check before participating. All of the participants reported being native speakers of American English. Three participants reported that they speak a language other than English in the home (Hmong, *n* = 1; Urdu, n = 1; Patois, n = 1), none of which are languages that allow vowelless words. None of the participants reported that they spoke or had studied Tashlhiyt or any of the languages of North Africa.

Participants were randomly assigned to a Language condition: 20 were assigned to a Vowelless Language and 20 were assigned to a Voweled Language.

4.2. Results

4.2.1. Word learning performance

Correct (1) and incorrect (0) responses were modeled using a Bayesian mixed effects logistic regression using *brms* in R. The model (Eq. 3) included fixed effects for Language Condition (Vowelless vs. Voweled), Speech Style of the Item (Clear vs. Casual), and their interactions. Factors were sum-coded. Random effects structure consisted of random slopes for item speech style by participant and word.

$$\begin{aligned} Correct &\sim \text{Bernoulli}(p) \\ p &= \text{logit}^{-1}(z) \\ z = (exp.\text{clear}*\text{stim.clear})*\text{stim.clear} &+ (exp.\text{clear}|id) + (exp.\text{clear}|word) \end{aligned} \tag{3}$$

The full model output is provided in Table B3 in Appendix B, visualized in Fig. 7, and the summarized data are plotted in Fig. 6.

The mean estimated word learning performance is 0.29 (s.d. = 0.23, 95% C.I, = [−0.15, 0.74]) logits representing an accuracy of 57%. Chance level is 1/16 since participants had to pick one out of 16 choices in each test trial. This indicates that overall, participants successfully learned target names in the experiment, higher than the chance level performance of $p = 1/16 = 0.0625$. However, overall performance was lower than that in Experiment 1 (66%). Thus, learning performance is lower when participants are exposed to more variable acoustic-phonetic forms of words (Experiment 2), compared to where training consisted of only one type of speech style (Experiment 1).

There was an effect of Exposure Language condition: word learning performance was lower for participants exposed to a vowelless language than those who were exposed to a voweled language (mean = −0.48, s.d. = 0.19, 95% C.I. = [−0.86, −0.09]). There was also a negative coefficient associated with the interaction between Exposure Language and Stimulus Speech Style: participants exposed to the vowelless lexicon showed lower word learning performance for clear speech stimuli in test than for casual speech stimuli (mean = −0.12, s.d. = 0.07, 95% C.I. = [−0.26, 0.01]).

4.2.2. Novel word acceptability ratings

Novel word acceptability ratings (0–100; scaled to random from 0 to 1) were modeled using a zero-one inflated beta regression model fit using *brms* in R (for information of these models see section 3.2.2). As in Experiment 1, the model featured parallel fixed effects structure for three of the four model parameters (variation in ϕ was not modeled). The model included fixed effects of Exposure Language Condition (Vowelless vs. Voweled), Novel Ratings Word Type (CVC vs. CCC), and Novel Ratings Word Speech Style (Clear vs. Casual), and the two-way interactions between Exposure Language Condition and Novel Ratings Word Type, Novel Ratings Word Style and Novel Ratings Word Type were included. All factors were sum coded. Random intercepts for participant and word were included, as well as by-participant random slopes for Novel Ratings Word Type and Novel Ratings Word Speech Style for average wordlikeness ratings.

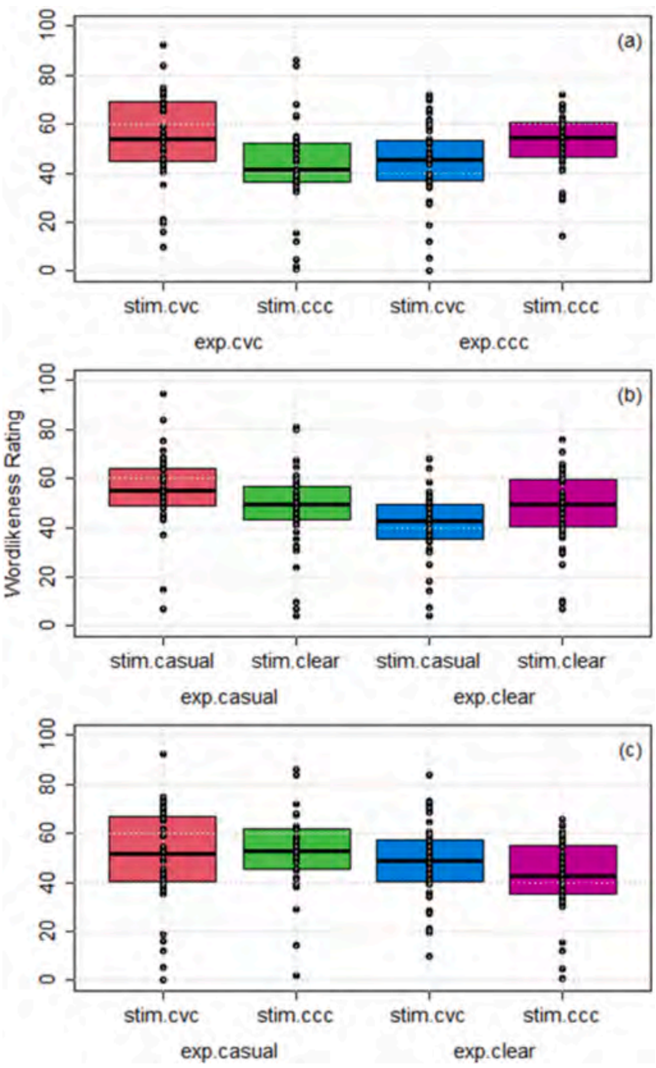

Fig. 4. Wordlikeness ratings for novel lexical items from Experiment 1. Plot A provides ratings by word type (stimulus CVC vs. stimulus CCC) and participant exposure language (Vowelless vs. Voweled). Plot B provides ratings by speech style of the novel item (clear stimulus vs. casual stimulus) and exposure language speech style (Clear Exposure vs. Casual Exposure). Plot C provides the interaction between novel word type and exposure speech style. Points represent average wordlikeness rating by subject in different conditions in Experiment 1. Boxes span interquartile ranges and lines indicate group medians.

$$\text{Rating} \sim \text{ZOIB}(\alpha, \gamma, \mu, \phi) = \begin{cases} \alpha(1-\gamma) & \text{Rating} = 0 \\ \alpha\gamma & \text{Rating} = 1 \\ (1-\alpha)\text{beta}(\mu\phi) & 0 < \text{Rating} < 1 \end{cases}$$

$$\begin{aligned} \text{logit}(\alpha) &= exp.\text{ccc} + (\text{stim.ccc}*\text{stim.clear}) + (1|id) + (1|word) \\ \text{logit}(\gamma) &= exp.\text{ccc} + (\text{stim.ccc}*\text{stim.clear}) + (1|id) + (1|word) \\ \text{logit}(\mu) &= exp.\text{ccc} + (\text{stim.ccc}*\text{stim.clear}) + (\text{stim.ccc}*\text{stim.clear}|id) + (exp.\text{clear}|word) \\ log(\phi) &= \text{Intercept} \end{aligned} \tag{4}$$

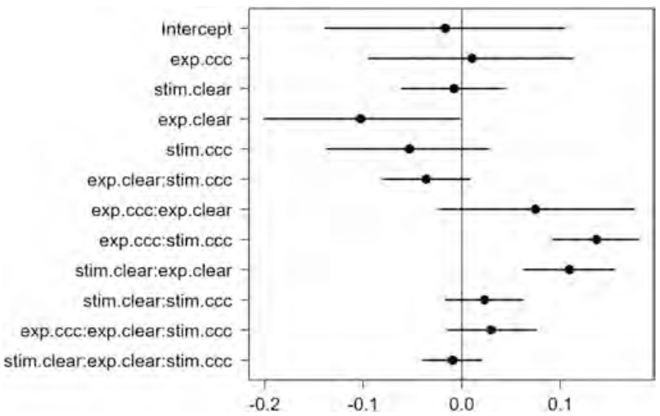

Fig. 5. Experiment 1 wordlikeness ratings model coefficients. Points indicate posterior means, lines indicate the 95% credible intervals for parameters.

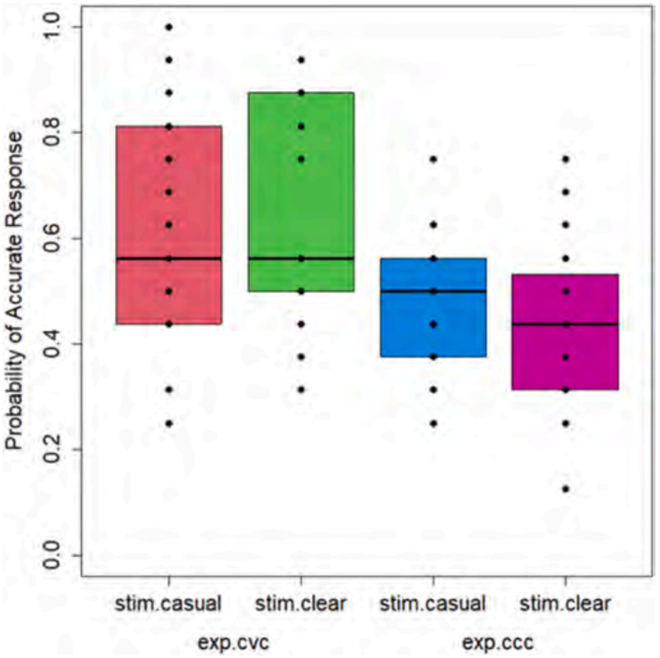

Fig. 6. Points represent accurate response rate by subject in each condition in Experiment 2. Boxes span interquartile ranges and lines indicate group medians.

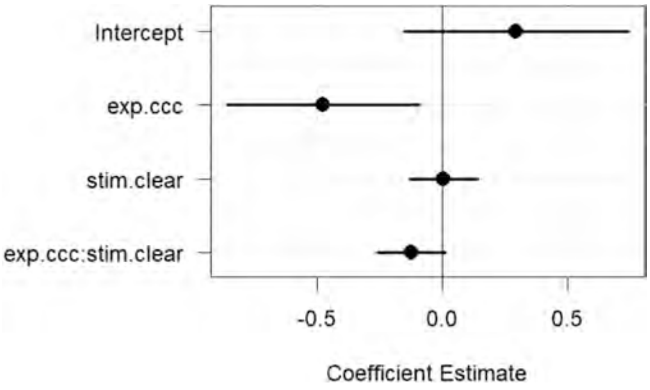

Fig. 7. Experiment 2 learning performance model coefficients. Points indicate posterior means, lines indicate the 95% credible intervals for parameters.

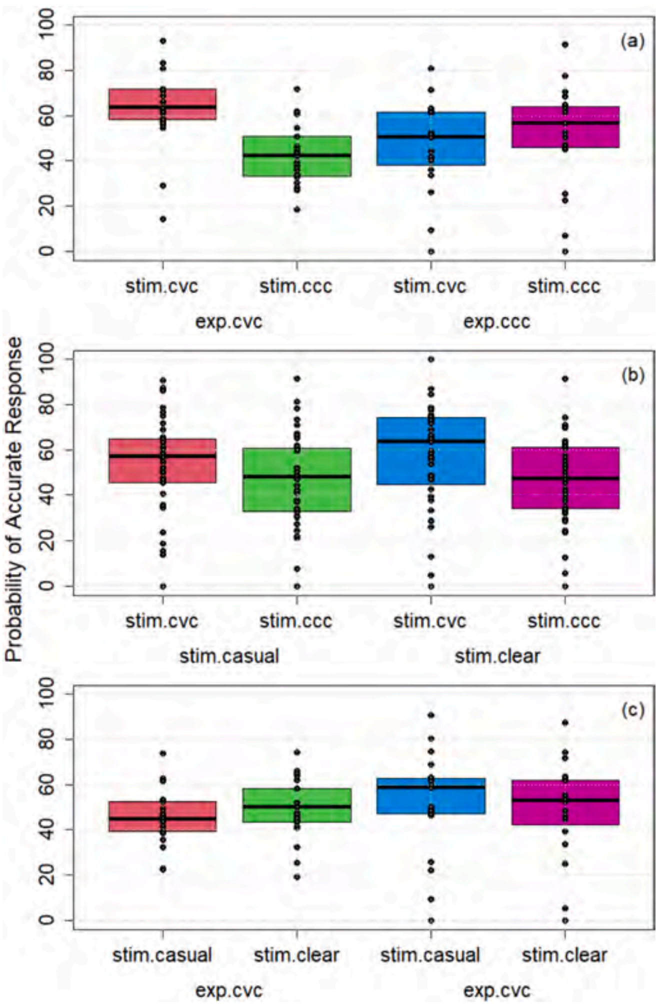

Fig. 8. Points represent average wordlikeness rating by subject in different conditions in Experiment 2. Boxes span interquartile ranges and lines indicate group medians.

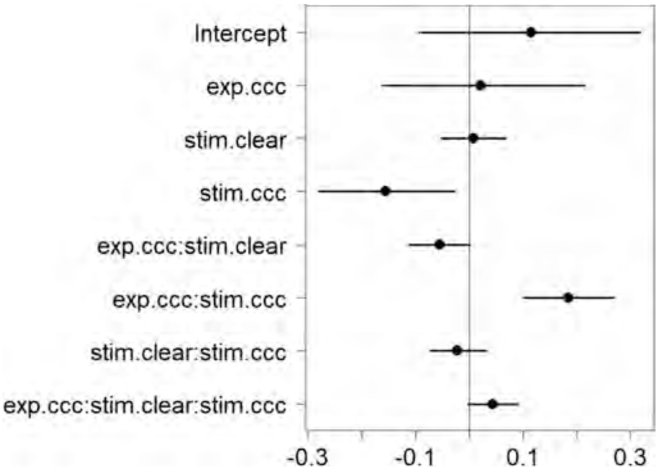

Fig. 9. Experiment 2 wordlikeness model coefficients. Points indicate posterior means, lines indicate the 95% credible intervals for parameters.

The summarized wordlikeness ratings are provided in Fig. 8. The full model output is provided in Table B4 in Appendix B and visualized in Fig. 9 (fixed effects related to α and γ are plotted in Fig. B2 in Appendix B). There was an effect of Novel Ratings Word Type: Overall, novel vowelless words received lower wordlikeness ratings than novel voweled words (mean = −0.16, s.d. = 0.06, 95% C.I. = [−0.28, −0.03]).

The interaction between Exposure Language and Novel Ratings Word Type was positive and is illustrated in the top panel of Fig. 8 (mean = 0.18, s.d. = 0.04, 95% C.I. = [0.1, 0.27]). Participants who were exposed to the voweled language rated novel CVC items higher than novel CCC items. Yet, participants exposed to a lexicon with vowelless words in the learning phase showed a smaller difference in ratings of novel CVC and CCC words.

The interaction between Exposure Language and Novel Ratings Word Speech Style had a negative coefficient (mean = −0.05, s.d. = 0.03, 95% C.I. = [−0.11, 0]), indicating that participants exposed to a vowelless lexicon in training provided lower wordlikeness ratings to novel words produced in clear speech.

Finally, there was also some evidence for an interaction between Novel Ratings Word Type and Novel Ratings Word Style (mean = −0.02, s.d. = 0.03, 95% C.I. = [−0.07, 0.03]), though the confidence interval was large. As seen in the bottom panel of Fig. 8, while wordlikeness ratings were higher for CVC items in Clear speech, there was no difference in wordlikeness ratings for CVC and CCC items in Casual speech.

5. General discussion

The current study explored whether learning a novel language is harder if that language contains vowelless words, a typologically rare lexical form. In two experiments, participants were auditorily trained on either a lexicon containing mostly vowelless word forms (tri-segmental CCC forms that are real words in Tashlhiyt) or a lexicon containing only words with vowels (also real Tashlhiyt words). Since recent work also has drawn attention to the role of acoustic-phonetic variation in the input on perceptual learning (e.g., Bradlow, Bassard, & Paller, 2023; Quam & Creel, 2021), speech style variation was manipulated in the training items across experiments: in Experiment 1 we compared the effect of exposure to only either Clear or Casual productions, while in Experiment 2 all participants were exposed to mixed speech style variation. We also investigated the effect of exposure on generalization to novel words - all listeners rated the wordlikeness of novel voweled and vowelless words produced in both clear and casual speech styles. We summarize and discuss all our findings below in turn.

5.1. *Learning vowelless words*

As outlined in the Introduction, vowelless words are extremely typologically rare. Are they less common across languages of the world because they are harder to *learn*? The findings from Experiment 1 suggest that this is not necessarily the case. When trained on a lexicon consisting of mostly vowelless words (12 vowelless and 4 voweled words), participants learned novel word forms equally well as participants exposed to a lexicon of only voweled words.

Why were vowelless words in Tashlhiyt not harder to learn than voweled words in this context? As outlined in the Introduction, there is a large body of work examining the articulatory and acoustic properties of these words finding that consonant sequences in Tashlhiyt are minimally overlapping in gestural timing. This creates salient acoustic cues present with vowelless words that make them perceptually robust. Our finding that naive learners can acquire vowelless words in Tashlhiyt supports articulatory, acoustic, and perceptual work on this language that speakers produce salient cues to the phonological structure of vowelless words (e.g., Ridouane & Fougeron, 2011; Zellou et al., 2024).

So, why are vowelless words so rare if they are not harder to learn following brief auditory exposure? Potentially, the learning bias lies in production. There is work suggesting that CV structure is preferred in articulatory terms. Cross-linguistically, children's early productions predominantly involve CV, and some CVC, syllables even if the ambient language contains vowelless words (Goldstein, Byrd, & Saltzman, 2006; Lahrouchi & Kern, 2018; MacNeilage & Davis, 1990, 1993). Prior work looking at the acquisition of non-native consonant sequences shows that they pose particular difficulty for adult naive learners (Davidson, 2011). This difficulty in L2 speech production is related to the phonological differences between the L1 and L2, learning of the novel coordination among articulatory gestures, and a host of language-independent phonetic characteristics (Davidson, 2005, 2006; Zsiga, 2003). This can be explored for vowelless words in Tashlhiyt in future work examining whether naive listeners can also acquire vowelless words in *production* after auditory exposure.

5.2. *Effects of speech style variation*

We found that overall learning performance was higher when participants were exposed to the Clear speech productions compared to the Casual speech productions, but only in Experiment 1. When speaking style was held constant, naive listeners benefited from the clear speech acoustic-articulatory modifications equally in learning vowelless and voweled words. The learning benefit for clear speech observed here is in line with other well-documented processing benefits during speech comprehension (Smiljanić, 2021), including increasing word intelligibility in adverse listening conditions (Grynpas, Baker, & Hazan, 2011; Picheny, Durlach, & Braida, 1985), boosting recognition memory and recall (Keerstock & Smiljanić, 2018, 2019; Van Engen, Chandrasekaran, & Smiljanic, 2012), and improving word segmentation (Guo & Smiljanic, 2023). Thus, the current study confirms another listener-based benefit that Clear speech provides: higher word learning performance (cf. Escudero et al., 2011). Clear speech word forms contain more distinctive and longer acoustic information and these properties also appear to support adult word learning (cf. phonetic training with All Enhanced cue manipulation in Iverson et al., 2005). It remains to be determined which specific conversational-to-clear speech modifications aid in the learning process and via what mechanism.

The clear speech learning benefit, however, was not observed in Experiment 2. In Experiment 2, listeners were exposed to more acoustic-phonetic variation with both clear and casual speech forms of words. In this context, not only was the clear speech benefit absent, learning performance for the vowelless lexicon was overall lower than for the voweled lexicon. Increasing acoustic-phonetic variation in the input, in the form of the speaking style changes, seems to create "noise" that makes learning more difficult, especially if learning involves vowelless words. Thus, learning of vowelless words can be similar to learning of voweled words, but only when there is no within-talker acoustic-phonetic variation. In more naturalistic language learning contexts, though, listeners are exposed to variation in the acoustic form of words, including in the form of listener-oriented speaking styles. Our results suggest that learning of a lexicon containing mainly phonologically-unusual structures is facilitated by a consistent acoustic-phonetic form in training and that increased variation disrupts learning.

The observation that increased acoustic-phonetic variation in input results in worse learning outcomes is also found in child language studies (Quam et al., 2017; van Heugten & Johnson, 2017). Some researchers have suggested that when acoustic-phonetic variation is "not relevant" to phonological learning it can have deleterious effects on acquisition, such as when one form of a word is produced by one talker and another form is produced by a different talker (Quam & Creel, 2021). In the present study, the variation was within-talker, but it was not presented in a meaningful or systematic way. Talkers often shift from clear to casual speaking modes in deliberate ways based on context and interlocutor. Random variation between these forms of words was not systematic, which could explain why this type of variation did not improve learning.

The finding that clear speech word forms when mixed with casual speech do not facilitate learning, and in fact, can hinder it, is somewhat unexpected. However, a variability-related processing cost encountered here may be similar to processing of other types of acoustic variability. Listener's speech processing is affected across multiple tasks when hearing acoustic-phonetic variation arising from multiple talkers (Heald & Nusbaum, 2014; Lim, Carter, Njoroge, Shinn-Cunningham, & Perrachione, 2021; Mullennix, Pisoni, & Martin, 1989; Stilp & Theodore, 2020). Processing cost is also associated with within-talker variation in speaking rate and speaking styles (Sommers & Barcroft, 2006; Sommers, Nygaard, & Pisoni, 1994). Even when there are not multiple talkers or multiple speech streams to contend with, listener-oriented speaking style variation can increase processing cost. This may be related to the increased uncertainty about whether the mapping of the variable incoming speech signal is appropriate, which is resource-demanding, or the variable input can disrupt auditory selective attention (see Luthra, 2023 for a review of these different accounts). This latter account is in line with the recent work showing that hearing clear speech sentences mixed with an energetic masker increased response times to the visual task compared to conversational speech (Meemann & Smiljanic, 2023). The results were taken to indicate that listeners directed their attentional resources to the more salient hyperarticulated clear speech. While current data do not allow us to tease apart these effects, both the resource-demanding mapping uncertainty and selective attention recruitment would make it harder to process incoming speech and harder to encode it in memory resulting in lower learning performance in the mixed-speaking-style conditions. The current results thus suggest that the within-talker variability related to intelligibility-enhancing clear speech modifications is substantial enough to disrupt processing.

It is important to note that the effect of speaking style variation was asymmetrical on learning of voweled and vowelless words. This observation could potentially shed light on why vowelless words are typologically rare. As discussed in the Introduction, vowels are louder and contain greater acoustic information than consonants, and they provide robust cues to their own identity as well as for those of surrounding segments. Yet, that alone cannot explain the patterns in the present study. Rather, our findings suggest that vowels provide listeners with auditory cues that are robust to stylistic *variation* such that learning of lexical items across clear and reduced forms is more likely. Words containing sequences of consonants only are not harder to learn per se, but rather, the variation across different stylistic productions of these words makes them harder to learn. In other words, these patterns indicate that phonological/phonetic learning biases cannot be fully understood without considering how robust they are to exposure to variable forms of words.

Together, our results indicate that Clear speech - here, produced with the goal of enhancing comprehension for an adult listener - can support adult language learning of novel words, including typologically rare ones, parallel to findings that IDS helps young children learn novel words (Ma et al., 2011; Singh, Nestor, Parikh, & Yull, 2009). High variation, particularly between clear and reduced forms, leads to less robust learning outcomes for adults, also as it does for children.

5.3. Generalization: Effect of exposure on wordlikeness ratings of novel words

Our study also tested what effect these phonological and phonetic patterns in the exposure language have on performance in a subsequent wordlikeness ratings task that contained vowelless and voweled words in both clear and casual speaking styles. We find that participants did generalize learning to novel word forms based on the lexical and acoustic-phonetic properties of the exposure language. Across both experiments, participants exposed to the vowelless language subsequently rated novel vowelless words as more likely to be lexical forms in that language than CVC words. This observation replicates related work showing that adults can learn non-native phonotactic patterns after brief auditory exposure (Onishi et al., 2002) and extends it to words without vowels, a highly rare phonotactic structure. It also provides evidence that participants exposed to the vowelless lexicon did acquire abstract knowledge that reflected the unique phonological properties of that language.

In Experiment 1, where speech style was between-subjects in training, we also observed an effect of acoustic-phonetic form on generalization: participants exposed to a language in either Clear or Casual speech style provided higher word-likeness ratings for novel items produced in the matching speech style. In other words, participants showed generalization of the *acoustic* form of the exposure lexicon. In Experiment 2, where speech style was within-subjects in training, the only effect of novel word style was that it interacted with word form: CVC clear speech items received the highest wordlikeness ratings overall, regardless of listener exposure language. This is consistent with prior work reporting that exposure to variations in acoustic-phonetic form leads to robust generalization effects (Quam & Creel, 2021).

We also observe that, across both experiments and not varying by listeners' exposure, novel CVC items receive higher wordlikeness rating when produced in Clear speech while no such boost is observed for novel CCC items. Why is this the case? One possibility is that spectral enhancement that comes with clear speech-related hyperarticulation leads voweled words to sound more like English words. Meanwhile, the Clear speech enhancement of vowelless words does not make the words sound more wordlike for English listeners. One possibility is that the effect of L1 phonotactic preferences emerges in clear speech since listeners can better compare novel words to those stored in memory. While questions such as this present avenues for future work, our finding that phonetic variation affects observed patterns of generalization highlight the importance of looking at different acoustic forms of words when investigating generalization from learning.

Another possible explanation for our observation that adults can generalize from a vowelless lexicon to the same extent as from a voweled lexicon could stem from how they approach the wordlikeness judgment task: perhaps our participants were learning that the test language simply contains word forms that are similar to their language or not. In that case, for the voweled lexicon, they can easily discriminate between novel CVC and CCC words by identifying the forms that are most similar to English; for the vowelless lexicon, they can discriminate between novel words based on being less similar to English or not (thus, rating vowelless words higher). The speech style result is similarly explainable: when the exposure language is presented in a consistent speech style, that style becomes a cue for a novel word in the wordlikeness task as being from the language. Prior word learning studies that have similar results also suggest this interpretation. For instance, Storkel, Armbrüster, and Hogan (2006) and Johnston and Kapatsinski (2011) found that phonotactically illegal words were easier to learn by English-speaking participants than legal ones in a word-picture matching task (like the one used in the present study in training). Johnston and Kapatsinski (2011) also found that the novel words that were phonotactically illegal in English were more likely to be judged as acceptable in the artificial language than phonotactically legal words, supporting the possibility that generalization of learning might be easier if there is a strong about the properties of that language (e.g., that it contains words that are very different from learners L1 or that it is produced in a specific speaking style).[2]

5.4. *Limitations and future directions*

There were also several limitations of the present study that present ripe directions for future work. For one, the current study presented listeners with small 16-word lexicons. In reality, language learning involves larger and more varied lexicons. Moreover, the current study also presented the items in only one voice. Exploring whether learning vowelless words is harder when exposed to different types of lexicons and across multiple sources of acoustic-phonetic variation is a direction for future work that can further shed light on the role of variation on acquisition of typologically rare structures.

Furthermore, the current study only focused on perceptual acquisition of vowelless words. The results from the present study lead to questions of whether vowelless words are harder to learn how to *produce* than vowelless words by L2 learners. There is much work examining the phonetic patterns of vowelless words in Tashlhiyt showing that they have unique articulatory implementation. We predict that learning how to produce vowelless words is difficult for L2 learners in ways that can reveal why they are so typologically rare (cf. Davidson, 2011).

6. Conclusion

Word learning is challenging for adult second language learners and also harder when the target language contains phonological structures that are not found in the first language. In the current study, we examined the learning of an exceptionally rare pattern across languages of the world: words without vowels. Following brief auditory exposure, native-English adult learners can learn Tashlhiyt lexicons containing mainly vowelless words. This suggests that vowelless words can be acquired by naive learners even in a constrained experimental setting. However, the acoustic-phonetic form of words modulates learning - speaking style variation during exposure reduces listeners' ability to learn vowelless words. These results show there is a relationship between acoustic-phonetic variation, word learning, and phonology typology and contribute to our understanding of the fundamental cognitive mechanisms underlying language acquisition.

Ethics and consent

Research was performed in accordance with the UC Davis IRB (IRB number 1328085–2).

CRediT authorship contribution statement

Georgia Zellou: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Santiago Barreda:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis. **Mohamed Lahrouchi:** Writing – review & editing, Writing – original draft, Methodology, Investigation. **Rajka Smiljanić:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors have no competing interests to declare.

Data availability

All data and code for this paper are available at: https://osf.io/539uc/?view_only=f2bf0190872c4aaa93068e03baa91037

Appendix A. Target words (with word structure and original Tashlhiyt gloss) and corresponding images used in the training phase

Vowelless Language Condition A	Vowelless Language Condition B	Voweled Language Condition A	Voweled Language Condition B	Corresponding Image
fan (CVC ‘they gave’)	ʁar (CVC ‘only’)	fan (CVC ‘they gave’)	ʁar (CVC ‘only’)	book
man (CVC ‘which’)	ruħ (CVC ‘go home’)	man (CVC ‘which’)	ruħ (CVC ‘go home’)	dog
nuf (CVC ‘we are better’)	sut (CVC ‘drink it’)	nuf (CVC ‘we are better’)	sut (CVC ‘drink it’)	banana
lʒdud (CVC ‘ancestors’)	lfal (CVC ‘omen’)	lʒdud (CVC ‘ancestors’)	lfal (CVC ‘omen’)	chair
zˤbr (CCC ‘prune’)	ʒbd (CCC ‘pull’)	ʕif (CVC ‘get tired of’)	tuf (CVC ‘she’s better’)	bike
bʕd (CCC ‘them, emphatic’)	ʒɦd (CCC ‘be strong’)	fat (CVC ‘give 2MS.PL’)	ʕum (CVC ‘swim’)	couch
fkt (CCC ‘give it’)	ftħ (CCC ‘operate’)	daʁ (CVC ‘again’)	mun (CVC ‘accompany’)	pen
bdr (CCC ‘mention’)	bzg (CCC ‘swell’)	mit (CVC ‘what’)	sir (CVC ‘go!’)	bird
nsˤħ (CCC ‘advise’)	nʒħ (CCC ‘pass a test’)	zˤurˤ (CVC ‘visit’)	zud (CVC ‘like, as’)	boat
rdˤl (CCC ‘borrow/lend’)	rgl (CCC ‘lock’)	luħ (CVC ‘throw’)	ran (CVC ‘they want’)	table
nʃf (CCC ‘scrape’)	ngr (CCC ‘between’)	sul (CVC ‘stay alive’)	sak (CVC ‘pass through’)	cow
rbħ (CCC ‘win’)	rˤħl (CCC ‘leave the city’)	liʁ (CVC ‘I married’)	fuħ (CVC ‘revel in’)	bus
frˤħ (CCC ‘be happy’)	ħrm (CCC ‘deprive’)	tid (CVC ‘these FM’)	tut (CVC ‘she hit’)	scissors
slt (CCC ‘leave on the sly’)	tlf (CCC ‘get mixed up’)	sin (CVC ‘two’)	dar (CVC ‘at’)	fish
ʒld (CCC ‘leather’)	zlm (CCC ‘glance’)	lan (CVC ‘they have’)	gan (CVC ‘they are’)	plane
krf (CCC ‘tie’)	ʕlf (CCC ‘feed’)	ʁir (CVC ‘only’)	riʁ (CVC ‘I want’)	hat

---

[2] Thanks to an anonymous reviewer for bringing this possibility to our attention.

Appendix B. Statistical model outputs

Table B1

Experiment 1 Learning Model Fixed Effect means, standard deviations (s.d.), and 2.5% and 97.5% credible Intervals (CI).

	Mean	s.d.	2.5% CI	97.5% CI
Intercept	0.813	0.212	0.405	1.233
exp.clear	0.372	0.178	0.029	0.722
exp.ccc	−0.047	0.175	−0.387	0.304
exp.clear:exp.ccc	−0.115	0.170	−0.452	0.214

Table B2

Experiment 1 Rating Model Fixed Effect means, standard deviations (s.d.), and 2.5% and 97.5% credible Intervals (CI). Coefficient prefixes indicate the modeled parameter each coefficient relates to.

	Mean	s.d.	2.5% CI	97.5% CI
mu_Intercept	−0.017	0.062	−0.139	0.104
phi_Intercept	1.578	0.020	1.538	1.617
zoi_Intercept	−3.597	0.366	−4.349	−2.902
coi_Intercept	−1.073	0.511	−2.128	−0.088
mu_exp.ccc	0.010	0.052	−0.094	0.113
mu_stim.clear	−0.008	0.026	−0.061	0.043
mu_exp.clear	−0.103	0.051	−0.201	−0.002
mu_stim.ccc	−0.053	0.042	−0.137	0.028
mu_exp.clear:stim.ccc	−0.036	0.022	−0.080	0.008
mu_exp.ccc:exp.clear	0.075	0.051	−0.023	0.175
mu_exp.ccc:stim.ccc	0.137	0.022	0.092	0.179
mu_stim.clear:exp.clear	0.109	0.024	0.063	0.155
mu_stim.clear:stim.ccc	0.023	0.020	−0.016	0.062
mu_exp.ccc:exp.clear:stim.ccc	0.030	0.023	−0.014	0.075
mu_stim.clear:exp.clear:stim.ccc	−0.009	0.015	−0.039	0.020
zoi_exp.ccc	−0.570	0.344	−1.279	0.081
zoi_stim.clear	0.194	0.067	0.059	0.328
zoi_exp.clear	−0.291	0.339	−0.954	0.382
zoi_stim.ccc	−0.227	0.125	−0.470	0.024
zoi_exp.clear:stim.ccc	0.016	0.074	−0.129	0.160
zoi_exp.ccc:exp.clear	−0.590	0.342	−1.253	0.072
zoi_exp.ccc:stim.ccc	−0.225	0.073	−0.369	−0.081
zoi_stim.clear:exp.clear	0.084	0.069	−0.048	0.220
zoi_stim.clear:stim.ccc	−0.071	0.067	−0.204	0.058
zoi_exp.ccc:exp.clear:stim.ccc	0.133	0.072	−0.008	0.275
zoi_stim.clear:exp.clear:stim.ccc	−0.111	0.068	−0.247	0.023
coi_exp.ccc	0.114	0.424	−0.705	0.964
coi_stim.clear	0.302	0.168	−0.015	0.640
coi_exp.clear	−0.245	0.417	−1.044	0.579
coi_stim.ccc	0.065	0.216	−0.361	0.498
coi_exp.clear:stim.ccc	0.005	0.190	−0.357	0.379
coi_exp.ccc:exp.clear	−0.019	0.423	−0.857	0.802
coi_exp.ccc:stim.ccc	1.230	0.198	0.857	1.631
coi_stim.clear:exp.clear	0.725	0.172	0.399	1.063
coi_stim.clear:stim.ccc	−0.060	0.165	−0.380	0.263
coi_exp.ccc:exp.clear:stim.ccc	0.055	0.193	−0.313	0.434
coi_stim.clear:exp.clear:stim.ccc	0.153	0.167	−0.176	0.472

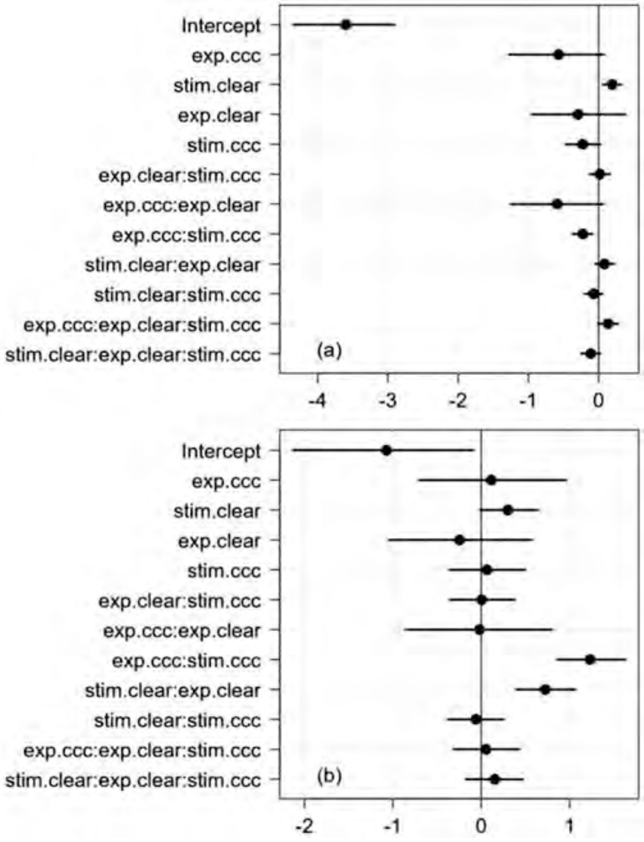

Fig. B1. Experiment 1 wordlikeness model coefficients for (a) α and (b) γ. Points indicate posterior means, lines indicate the 95% credible intervals for parameters.

Table B3
Experiment 2 Learning Model Fixed Effect means, standard deviations (s.d.), and 2.5% and 97.5% credible Intervals (CI).

	Mean	s.d.	2.5% CI	97.5% CI
Intercept	0.292	0.228	−0.154	0.743
exp.ccc	−0.476	0.193	−0.860	−0.089
stim.clear	0.003	0.070	−0.131	0.142
exp.ccc:stim.clear	−0.124	0.069	−0.260	0.011

Table B4
Experiment 2 Rating Model Fixed Effect means, standard deviations (s.d.), and 2.5% and 97.5% credible Intervals (CI). Coefficient prefixes indicate the modeled parameter each coefficient relates to.

	Mean	s.d.	2.5% CI	97.5% CI
mu_Intercept	0.115	0.106	−0.093	0.319
phi_Intercept	1.546	0.030	1.488	1.604
alpha_Intercept	−2.997	0.543	−4.096	−1.947
alpha_Intercept	1.171	0.905	−0.611	2.965
mu_exp.ccc	0.022	0.097	−0.162	0.216
mu_stim.clear	0.007	0.031	−0.052	0.069
mu_stim.ccc	−0.156	0.064	−0.280	−0.027
mu_exp.ccc:stim.clear	−0.055	0.029	−0.112	0.001
mu_exp.ccc:stim.ccc	0.185	0.043	0.100	0.270
mu_stim.clear:stim.ccc	−0.022	0.027	−0.074	0.033
mu_exp.ccc:stim.clear:stim.ccc	0.044	0.024	−0.002	0.091
gamma_exp.ccc	0.598	0.518	−0.405	1.630
gamma_stim.clear	0.247	0.091	0.072	0.424
gamma_stim.ccc	−0.219	0.147	−0.501	0.070
gamma_exp.ccc:stim.clear	0.109	0.091	−0.069	0.291
gamma_exp.ccc:stim.ccc	0.259	0.091	0.082	0.437
gamma_stim.clear:stim.ccc	−0.104	0.091	−0.284	0.072
gamma_exp.ccc:stim.clear:stim.ccc	−0.129	0.089	−0.307	0.048

(*continued on next page*)

Table B4 (*continued*)

	Mean	s.d.	2.5% CI	97.5% CI
alpha_exp.ccc	−0.674	0.745	−2.156	0.789
alpha_stim.clear	0.521	0.240	0.073	0.997
alpha_stim.ccc	−0.777	0.495	−1.788	0.200
alpha_exp.ccc:stim.clear	−0.568	0.239	−1.054	−0.119
alpha_exp.ccc:stim.ccc	1.513	0.289	0.972	2.105
alpha_stim.clear:stim.ccc	−0.109	0.237	−0.587	0.358
alpha_exp.ccc:stim.clear:stim.ccc	−0.067	0.231	−0.522	0.387

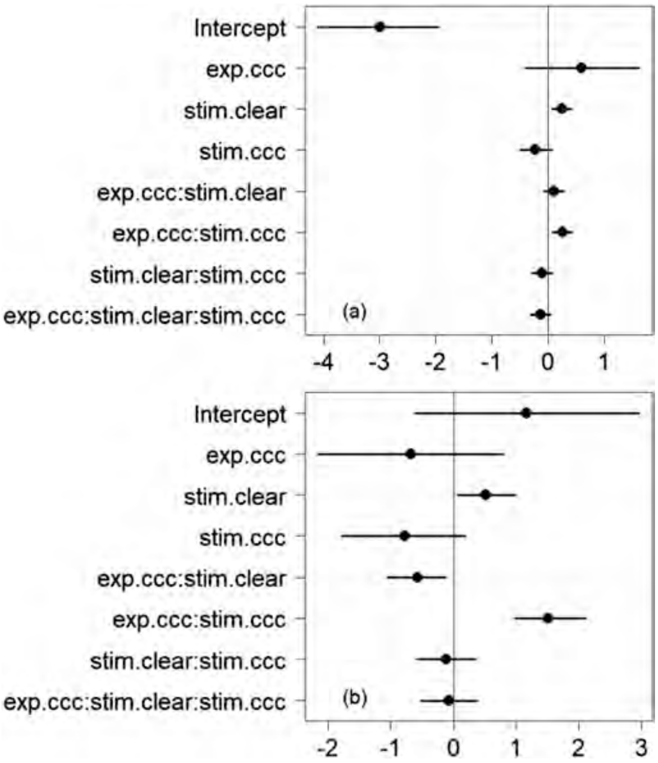

Fig. B2. Experiment 2 wordlikeness model coefficients for (a) α and (b) γ. Points indicate posterior means, lines indicate the 95% credible intervals for parameters.

References

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science, 35*(6), 1105–1138. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27*(3), 387–414. https://doi.org/10.1017/S0272263105050175

Bates, D. M. (2015). lme4: mixed-effects modeling with R *https.cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf*.

Bell, A. (1970). *Syllabic consonants. Working papers on language universals. 4* pp. B1–B49). Stanford University.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge University Press.

Boukous, A. (1987). *Phonotactique et Domaines Prosodiques en Berbère*. Doctorat d'Etat dissertation. Saint-Denis: Université Paris VIII-Vincennes.

Boukous, A. (2009). *Phonologie de l'Amazighe*. Rabat: IRCAM publications.

Bradlow, A. R. (2002). Confluent talker-and listener-oriented forces in clear speech production. *Laboratory Phonology, 7*, 241–273. https://doi.org/10.1515/9783110197105

Bradlow, A. R., Bassard, A. M., & Paller, K. A. (2023). Generalized perceptual adaptation to second-language speech: Variability, similarity, and intelligibility. *The Journal of the Acoustical Society of America, 154*(3), 1601–1613. https://doi.org/10.1121/10.0020914

Bürkner, P. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Chater, N., & Christiansen, M. H. (2010). Language acquisition meets language evolution. *Cognitive Science, 34*(7), 1131–1157. https://doi.org/10.1111/j.1551-6709.2009.01049.x

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and physics of speech* (pp. 283–333). Cambridge: CUP.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2006). Acquiring an artificial lexicon: Segment type and order information in early lexical entries. *Journal of Memory and Language, 54*(1), 1–19. https://doi.org/10.1016/j.jml.2005.09.003

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology, 28*(2), 197–234. https://doi.org/10.1017/S0952675711000145

Davidson, L. (2005). Addressing phonological questions with ultrasound. *Clinical Linguistics & Phonetics, 19*(6–7), 619–633. https://doi.org/10.1080/02699200500114077

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics, 34*(1), 104–137. https://doi.org/10.1016/j.wocn.2005.03.004

Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass, 5*(3), 126–139. https://doi.org/10.1111/j.1749-818X.2010.00266.x

Dell, F., & Elmedlaoui, M. (2012). *Syllables in Tashlhiyt Berber and in Moroccan Arabic* (Vol. 2). Springer Science & Business Media.

Escudero, P., Benders, T., & Wanrooij, K. (2011). Enhanced bimodal distributions facilitate the learning of second language vowels. *The Journal of the Acoustical Society of America, 130*(4). https://doi.org/10.1121/1.3629144. EL206-EL212.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), 1–185. https://doi.org/10.2307/1166093

Fougeron, C., & Ridouane, R. (2008). On the phonetic implementation of syllabic consonants and vowel-less syllables in Tashlhiyt. *Estudios de Fonética Experimental,* 140–175. https://revistes.ub.edu/index.php/experimentalphonetics/article/view/44293.

Frisch, S. A., Large, N. R., Zawaydeh, B., & Pisoni, D. B. (2001). Emergent phonotactic generalizations in English and Arabic. *Typological Studies in Language, 45*, 159–180.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development, 11*(1), 66–79. https://doi.org/10.1080/15475441.2014.895249

Goldstein, L., Byrd, D., & Saltzman, E. (2006). The role of vocal tract gestural action units in understanding the evolution of phonology. In M. A. Arbib (Ed.), *Action to language via the mirror neuron system* (pp. 215–249). Cambridge: Cambridge University Press.

Grynpas, J., Baker, R., & Hazan, V. (2011). Clear speech strategies and speech perception in adverse listening conditions. In *Proceedings of the International Phonetic Association* (pp. 779–782).

Guo, Z. C., & Smiljanić, R. (2021). Speaking clearly improves speech segmentation by statistical learning under optimal listening conditions. *Laboratory Phonology, 12*(1), 14. https://doi.org/10.5334/labphon.310

Guo, Z.-C., & Smiljanic, R. (2023). Clear speech facilitates word segmentation: Evidence from eye-tracking. In R. Skarnitzl, & J. Volín (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 177–181).

Gutova, E. (2015). Baby talk in Berber and Maghrebian Arabic. *Études et documents berbères, 34*(1), 95–123.

Hay, J., Pierrehumbert, J., & Beckman, M. (2004). Speech perception, well-formedness, and the statistics of the lexicon. *Papers in Laboratory Phonology VI*, 58–74.

Hayes, B., & White, J. (2013). Phonological naturalness and phonotactic learning. *Linguistic Inquiry, 44*(1), 45–75. https://doi.org/10.1162/LING_a_00119

Heald, S. L., & Nusbaum, H. C. (2014). Talker variability in audio-visual speech perception. *Frontiers in Psychology, 5*, 698. https://doi.org/10.3389/fpsyg.2014.00698

Hermes, A., Mücke, D., & Auris, B. (2017). The variability of syllable patterns in Tashlhiyt Berber and polish. *Journal of Phonetics, 64*, 127–144. https://doi.org/10.1016/j.wocn.2017.05.004

Hermes, A., Ridouane, R., Mücke, D., & Grice, M. (2011). Gestural coordination in Tashlhiyt syllables. In *ICPhS* (pp. 859–862).

van Heugten, M., & Johnson, E. K. (2017). Input matters: Multi-accent language exposure affects word form recognition in infancy. *The Journal of the Acoustical Society of America, 142*(2). https://doi.org/10.1121/1.4997604. EL196-EL200.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/−/l/to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267–3278. https://doi.org/10.1121/1.2062307

Jebbour, A. (1996). *Morphologie et Contraintes Prosodiques en Berbère (Tachelhit de Tiznit)-Analyse Linguistique et Traitement Automatique*. Doctorat d'Etat thesis. Rabat: Mohammed V University, Faculty of Letters and Human Sciences.

Jebbour, A. (1999). Syllable weight and syllable nuclei in Tachelhit Berber of Tiznit. In *Cahiers de Grammaire 24: Phonologie: théorie et variation* (pp. 95–116).

Johnston, L. H., & Kapatsinski, V. (2011). In the beginning there were the weird: A Phonotactic novelty preference in adult word learning. In *ICPhS* (pp. 978–981).

Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*(3), 402–420. https://doi.org/10.1006/jmla.1993.1022

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Kang, K. H., & Guion, S. G. (2008). Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America, 124*(6), 3909–3917. https://doi.org/10.1121/1.2988292

Keerstock, S., & Smiljanić, R. (2018). Effects of intelligibility on within-and cross-modal sentence recognition memory for native and non-native listeners. *The Journal of the Acoustical Society of America, 144*(5), 2871–2881. https://doi.org/10.1121/1.5078589

Keerstock, S., & Smiljanić, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America, 146*(6), 4604–4610. https://doi.org/10.1121/1.5141372

Krause, J. C., & Braida, L. D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *The Journal of the Acoustical Society of America, 112*(5), 2165–2172. https://doi.org/10.1121/1.1509432

Lahrouchi, M. (2010). On the internal structure of Tashlhiyt Berber triconsonantal roots. *Linguistic Inquiry, 41*(2), 255–285. https://doi.org/10.1162/ling.2010.41.2.255

Lahrouchi, M. (2018). Syllable structure and vowel-zero alternations in Moroccan Arabic and Berber. In A. Agwuele, & A. Bodomo (Eds.), *The Routledge handbook of African linguistics* (pp. 168–180). Oxon, UK: Routledge.

Lahrouchi, M., & Kern, S. (2018). From babbling to first words in Tashlhiyt language acquisition: Longitudinal two-case studies. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 63*(4), 493–526. https://doi.org/10.1017/cnj.2018.6

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

Lim, S. J., Carter, Y. D., Njoroge, J. M., Shinn-Cunningham, B. G., & Perrachione, T. K. (2021). Talker discontinuity disrupts attention to speech: Evidence from EEG and pupillometry. *Brain and Language, 221*, Article 104996. https://doi.org/10.1016/j.bandl.2021.104996

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242–1255. https://doi.org/10.1121/1.408177

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874–886. https://doi.org/10.1121/1.1894649

Luthra, S. (2023). Why are listeners hindered by talker variability? *Psychonomic Bulletin & Review, 31*, 104–121. https://doi.org/10.3758/s13423-023-02355-6

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development, 7*(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

MacNeilage, P. F., & Davis, B. L. (1990). Acquisition of speech production: Frames, then content. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 453–475). Hillsdale, N.J: Lawrence Erlbaum.

MacNeilage, P. F., & Davis, B. L. (1993). Motor explanations of babbling and early speech patterns. In B. de Boysson-Bardies, S. de Schönen, P. Jusczyk, P. MacNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 341–352). Dordrecht: Kluwer.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*(2), 91–121. https://doi.org/10.1016/S0010-0277(00)00109-8

Meemann, K., & Smiljanic, R. (2023). Clear speech improves word recognition but may increase listening effort. In R. Skarnitzl, & J. Volín (Eds.), *Proceedings of the 20th international congress of phonetic sciences* (pp. 246–250).

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America, 85*(1), 365–378. https://doi.org/10.1121/1.397688

Munson, B. (2001). Phonological pattern frequency and speech production in adults and children. *Journal of Speech, Language, and Hearing Research, 44*(1–4), 778–792. https://doi.org/10.1044/1092-4388(2001/061)

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics, 5*, 25–47. https://doi.org/10.1146/annurev-linguistics-011718-011919

Ohala, J., & Kawasaki-Fukumori, H. (1997). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. *Language and its ecology: Essays in memory of Einar Haugen, 100*, 343.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*(1), B13–B23. https://doi.org/10.1016/S0010-0277(01)00165-2

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693–703. https://doi.org/10.1121/1.1908183

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 28*(1), 96–103. https://doi.org/10.1044/jshr.2801.96

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech, Language, and Hearing Research, 29*(4), 434–446. https://doi.org/10.1044/jshr.2904.434

Pouplier, M., & Beňuš, Š. (2011). On the phonetic status of syllabic consonants: Evidence from Slovak. *Laboratory Phonology, 2*(2), 243–273. https://doi.org/10.1515/labphon.2011.009

Pycha, A., Nowak, P., Shin, E., & Shosted, R. (2003, March). Phonological rule-learning and its implications for a theory of vowel harmony. In *, Vol. 22. Proceedings of the 22nd west coast conference on formal linguistics* (pp. 101–114). Somerville, MA: Cascadilla Press.

Quam, C., & Creel, S. C. (2021). Impacts of acoustic-phonetic variability on perceptual development for spoken language: A review. *Wiley Interdisciplinary Reviews: Cognitive Science, 12*(5), Article e1558. https://doi.org/10.1002/wcs.1558

Quam, C., Knight, S., & Gerken, L. (2017). The distribution of talker variability impacts infants' word learning. *Laboratory. Phonology, 8*(1). https://doi.org/10.5334/labphon.25

R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ridouane, R. (2008). Syllables without vowels: Phonetic and phonological evidence from Tashlhiyt Berber. *Phonology, 25*(2), 321–359. https://doi.org/10.1017/S0952675708001498

Ridouane, R. (2014). Tashlhiyt Berber. *Journal of the International Phonetic Association, 44*(2), 207–221. https://doi.org/10.1017/S0025100313000388

Ridouane, R., & Fougeron, C. (2011). Schwa elements in Tashlhiyt word-initial clusters. *Laboratory Phonology, 2*(2), 275–300. https://doi.org/10.1515/labphon.2011.010

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science, 12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Saffran, J. R., & Thiessen, E. D. (2003). Pattern induction by infant language learners. *Developmental Psychology, 39*(3), 484–494. https://doi.org/10.1037/0012-1649.39.3.484

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America, 134*(5), 3793–3807. https://doi.org/10.1121/1.4824120

Singh, L., Nestor, S., Parikh, C., & Yull, A. (2009). Influences of infant-directed speech on early word recognition. *Infancy, 14*(6), 654–666. https://doi.org/10.1080/15250000903263973

Smiljanić, R. (2021). Clear speech perception: Linguistic and cognitive benefits. *The Handbook of Speech Perception*, 177–205.

Smiljanić, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *The Journal of the Acoustical Society of America, 118*(3), 1677–1688. https://doi.org/10.1121/1.2000788

Smiljanić, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 3*(1), 236–264. https://doi.org/10.1111/j.1749-818X.2008.00112.x

Smiljanić, R., & Bradlow, A. R. (2011). Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America, 130*(6), 4020–4031. https://doi.org/10.1121/1.3652882

Smolek, A., & Kapatsinski, V. (2018). What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology, 9*(1), 10. https://doi.org/10.5334/labphon.93

Sommers, M. S., & Barcroft, J. (2006). Stimulus variability and the phonetic relevance hypothesis: Effects of variability in speaking style, fundamental frequency, and speaking rate on spoken word identification. *The Journal of the Acoustical Society of America, 119*(4), 2406–2416. https://doi.org/10.1121/1.2171836

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *The Journal of the Acoustical Society of America, 96*(3), 1314–1324. https://doi.org/10.1121/1.411453

Stilp, C. E., & Theodore, R. M. (2020). Talker normalization is mediated by structured indexical information. *Attention, Perception, & Psychophysics, 82*, 2237–2243. https://doi.org/10.3758/s13414-020-01971-x

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research.*. https://doi.org/10.1044/1092-4388(2006/085)

Team, Stan Development. (2024). *Stan Modeling Language Users Guide and Reference Manual*, 2.3.4.1. https://mc-stan.org.

Thiessen, E. D., & Saffran, J. R. (2009). How the melody facilitates the message and vice versa in infant learning and memory. *Annals of the New York Academy of Sciences, 1169*(1), 225–233. https://doi.org/10.1111/j.1749-6632.2009.04547.x

Treiman, R., Kessler, B., Knewasser, S., Tincoff, R., & Bowman, M. (2000). English speakers' sensitivity to phonotactic patterns. *Papers in laboratory phonology V: Acquisition and the lexicon, 5*, 269.

Van Engen, K. J., Chandrasekaran, B., & Smiljanic, R. (2012). Effects of speech clarity on recognition memory for spoken sentences. *PLoS One, 7*(9), Article e43753. https://doi.org/10.1371/journal.pone.0043753

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science, 30*(5), 945–982. https://doi.org/10.1207/s15516709cog0000_89

Wong, P. C., & Perrachione, T. K. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied PsychoLinguistics, 28*(4), 565–585. https://doi.org/10.1017/S0142716407070312

Zec, D. (1995). Sonority constraints on syllable structure. *Phonology, 12*(1), 85–129. https://doi.org/10.1017/S0952675700002396

Zellou, G., Lahrouchi, M., & Bensoukas, K. (2022). Clear speech in Tashlhiyt Berber: The perception of typologically uncommon word-initial contrasts by native and naive listeners. *The Journal of the Acoustical Society of America, 152*(6), 3429–3443. https://doi.org/10.1121/10.0016579

Zellou, G., Lahrouchi, M., & Bensoukas, K. (2024). The perception of vowelless words in Tashlhiyt. *Glossa: a journal of general linguistics, 9*(1). https://doi.org/10.16995/glossa.10438

Zsiga, E. C. (2003). Articulatory timing in a second language: Evidence from Russian and English. *Studies in Second Language Acquisition, 25*(3), 399–432. https://doi.org/10.1017/S0272263103000160