



# Clear speech in Tashlhiyt Berber: The perception of typologically uncommon word-initial contrasts by native and naive listeners

Georgia Zellou, Mohamed Lahrouchi, Karim Bensoukas

## ► To cite this version:

Georgia Zellou, Mohamed Lahrouchi, Karim Bensoukas. Clear speech in Tashlhiyt Berber: The perception of typologically uncommon word-initial contrasts by native and naive listeners. *Journal of the Acoustical Society of America*, 2022, 152 (6), pp.3429-3443. 10.1121/10.0016579 . halshs-03903440

**HAL Id: halshs-03903440**

**<https://shs.hal.science/halshs-03903440>**

Submitted on 16 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Clear speech in Tashlhiyt Berber: The perception of typologically uncommon word-initial contrasts by native and naive listeners

Running Title: Clear speech in Tashlhiyt Berber

Georgia Zellou, Mohamed Lahrouchi, Karim Bensoukas

*Date of original submission: June 30, 2022*

Email address and affiliations of all authors:

Georgia Zellou  
University of California, Davis  
USA  
[gzellou@ucdavis.edu](mailto:gzellou@ucdavis.edu)

Mohamed Lahrouchi  
CNRS & Université Paris 8  
Paris, France  
[Mohamed.LAHROUCHI@cnrs.fr](mailto:Mohamed.LAHROUCHI@cnrs.fr)

Karim Bensoukas  
Mohammed V University  
Rabat, Morocco  
[k.bensoukas@gmail.com](mailto:k.bensoukas@gmail.com)

Mailing address of corresponding author:  
Georgia Zellou,  
Department of Linguistics, University of California, Davis  
469 Kerr Hall, One Shields Avenue,  
Davis, CA 95616

**Abstract.** Tashlhiyt Berber is known for having typologically unusual word-initial phonological contrasts, specifically word-initial singleton-geminate minimal pairs (e.g., *sin* vs. *ssin*) and sequences of consonants that violate the sonority sequencing principle (e.g., non-rising sonority sequences: *fsin*). The current study investigates the role of a listener-oriented speaking style on the perceptual enhancement of these rarer phonological contrasts. It examines the perception of word-initial singleton, geminate, and complex onsets in Tashlhiyt Berber across clear and casual speaking styles by native and naive listeners. While clear speech boosts the discriminability of pairs containing singleton-initial words for both listener groups, only native listeners performed better in discriminating between initial singleton-geminate contrasts in clear speech. Clear speech did not improve perception for lexical contrasts containing a non-rising-sonority consonant cluster for either listener group. These results are discussed in terms of how clear speech can inform phonological typology and the role of phonetic enhancement in language-universal vs. language-specific speech perception.

**Keywords.** clear speech, Tashlhiyt Berber, consonants, speech perception, phonological typology

## I. Introduction

Clear speech is a type of listener-oriented, intelligibility-enhancing mode of speaking. It has been shown to enhance the perceptibility of many different types of phonological contrasts, cross-linguistically (e.g., Croatian and English vowel contrasts in Smiljanić & Bradlow, 2005; Korean stop contrasts in Kang & Guion, 2008; Mandarin lexical tones in Tupper et al., 2021). An open question is whether all phonological contrasts are enhanced to an equivalent extent in clear speech. In the current study, we ask whether rarer phonological patterns receive less of a clear speech intelligibility boost, relative to more common phoneme contrasts. One theoretical approach in phonology is to search for phonetic explanations for why some sound contrasts are less common than others (e.g., Hayes & Steriade, 2004). Many researchers propose articulatory or perceptual motivations for typological patterns (e.g., Henke, Kaisse, & Wright, 2012; Wright, 2004). We explore the effect of speaking style on the perceptual discriminability of phonological patterns in Tashlhiyt Berber that are known for being typologically less common, which can shed light on questions in phonological typology.

Tashlhiyt Berber (iso: [shi]) is an Afroasiatic language spoken in Morocco. Tashlhiyt has been well studied for having typologically uncommon phonotactic properties: in particular, word-initial singleton-geminate contrasts (e.g., *sin* vs. *ssin*) and sequences of consonants that violate the sonority sequencing principle (e.g., sonority-plateauing sequences: *fsin*). This paper examines the effect of clear speech on the discrimination of these word-initial consonant contrasts in Tashlhiyt Berber. We predict that the more typologically uncommon contrasts (e.g., word pairs containing complex and geminate initial onsets) will have a smaller increase in perceptibility from casual to clear speech than more common contrasts (e.g., singleton contrasts). Furthermore, native and naive listeners' (here, American English speakers) discrimination of these contrasts across speech styles is compared, as well. Cross-language perception of clear speech provides a window to understanding the phonetic bases for cross-linguistic typological patterns (e.g., Peperkamp & Dupoux, 2007). This is a relatively understudied approach to examining these particular sound patterns and can inform theoretical perspectives on the perceptual factors that contribute to phonological contrasts being more or less common.

### A. Clear speech

Speakers naturally vary their speaking style based on their communicative environment (Lindblom, 1990). Speakers produce 'clear speech' when there is a reason to believe their listener will have trouble comprehending the signal. Clear speech is characterized by a variety of acoustic modifications relative to casual or conversational speech, such as slowing their speaking rate and producing more extreme segmental articulations (Picheny et al., 1986; Krause & Braidá, 2002; Uchanski, 2005; Smiljanić & Bradlow, 2009; Dilley et al., 2014; Cohn & Zellou, 2021;

Cohn et al., 2022). Speaking clearly has repeatedly been shown to benefit listeners by increasing intelligibility (e.g. Picheny et al., 1985; Smiljanić & Bradlow, 2011). And, clear speech is often tuned to particular types of listeners and communicative contexts; notably, speech produced with explicit consideration of the perceptual needs of interlocutors has been shown to be more intelligible to those particular interlocutors (Bradlow et al., 2003; Bradlow & Bent, 2002; Hargus Ferguson, 2004; Scarborough & Zellou, 2013; Zellou & Scarborough, 2015).

One of the reasons clear speech increases intelligibility is that it often involves enhancement of phonetic-acoustic distinctions between phonological categories. Duration is a common enhancement feature for sound contrasts. For instance, in English, speakers enhance initial consonant voicing contrast (“pit” vs. “bit”) by increasing the duration of voice onset time (VOT) on the voiceless stop, enhancing the property that makes those sounds distinctive (Schertz, 2013; Buz et al., 2016; Seyfarth et al., 2016). Another example is that the tense-lax vowel duration difference in English is even larger in clear speech (Leung et al., 2016).

However, since there is a large amount of variation in the phoneme inventory and the distribution of sounds in words across languages, it is important to examine clear speech effects across different languages. Many studies have focused on clear speech in English language contexts (e.g., the majority of studies reviewed in Uchanski (2005) examine clear speech in American or Canadian English). How speakers hyperarticulate different types of phonological contrasts and phonotactic patterns across a wide range of languages is important in order to understand what is universal and what is language-specific about clear speech. Smiljanić and Bradlow (2005, 2008) compared speech style modifications across English and Croatian and found similar spectral (vowel space expansion) and temporal (slower speech rate) enhancements despite language-specific differences in the realization of phonological contrasts in each language (10+ vs. 5 vowels respectively). Work such as this provides evidence that there is a universal listener-oriented, intelligibility-enhancement mode of speech that is used by talkers in similar ways despite differences in phonological inventories.

While the comparison of Croatian and English vowel enhancement is important for understanding how clear speech functions cross-linguistically, their vowel inventories are not particularly rare. How do speakers enhance typologically unusual phonological patterns? One example comes from a study of clear speech in Korean, which has a typologically less common three-way VOT contrast for initial voiceless stops (long (aspirated), short (fortis) and intermediate (lenis)). In a study comparing clear and conversational speech, Kang and Guion (2008) found that older adult Korean speakers produced both longer VOT durations for aspirated stops and shorter VOTs for lenis stops in clear speech, enhancing the distance between the three phoneme types. However, this contrast is in the process of undergoing a sound change: younger speakers show less use of VOT to convey the lenis vs. aspirated contrast, relying more on a lower  $f_0$  to signal lenis stops. Despite that, Kang and Guion (2008) observed that younger adults do produce a distinct three-way VOT pattern in clear speech. In perception, even though younger listeners showed greater use of  $f_0$  in phoneme identification, the clear speech forms containing enhanced VOT were still better perceived. Thus, even though the three-way VOT contrast is cross-linguistically uncommon and being lost in Korean, it is still enhanced in clear speech.

## **B. Phonetic bases of phonological typology**

Why are some phonological patterns less common than others? One of the earliest threads of research in phonology has been concerned with making generalizations about and explaining

cross-linguistic patterns in sound systems (e.g., Trubetzkoy, 1939; Jakobson, 1962). Empirically, some phonotactic patterns are more common in the world's languages while others are rarer. For instance, CV syllable structure is considered the most common permitted syllable form, cross-linguistically, and syllables with complex onsets are disallowed in many languages (e.g., Hawaiian) (Clements & Keyser, 1983; Rice, 2007). For languages that do allow complex onsets, there are commonalities in restrictions on the sequence of different consonant types. For instance, it has been proposed that sequences of segments within syllables generally follow the sonority sequencing principle (e.g., Selkirk, 1984; Clements, 1990; Blevins, 1995), which prefers sonority rises from the start of the onset to the nucleus and sonority decreases from the nucleus to the coda. Indeed, the types of onsets allowed in well-formed words across many languages can be explained by the sonority sequencing principle. For instance, in English, forms like [\*lbik] and [\*bnik] are disallowed because they do not contain maximally distinct sonority sequences: the former contains a fall in sonority before the nucleus, while the cluster in the latter does not contain sufficiently distinct sonority within the onset, viz. minimal sonority distance (cf. Steriade, 1982; Selkirk, 1984).

While typological distributions of sound patterns are one way to understand what makes sounds more or less likely to occur, another approach examines the phonetic bases of phonological tendencies (e.g., Hayes & Steriade, 2004). One hypothesis is that sound patterns that require more effort - either in terms of articulatory complexity and/or are more difficult to discriminate from other sounds - will be rarer because they are less robust in carrying lexical contrasts (Flemming, 2005). For instance, it has been argued that cross-linguistically preferred phoneme inventories tend to contain vowels that are positioned to be maximally acoustically distinct in order to be least confusable to listeners (Liljencrantz & Lindblom, 1972). We extend the argument that sound patterns in the world's languages are shaped by physical constraints on the production and perception to speech style variation. Since clear speech is considered a listener-oriented, intelligibility enhancing mode of speaking, investigating the perceptual consequences of clear speech on rarer vs. more common sound patterns is another way to understand how speech style variation impacts cross-linguistic sound patterns. The current study asks whether clear speech enhances the discriminability of typologically rare phonotactic patterns, such as those found in Tashlhiyt Berber, by native and naive listeners.

Tashlhiyt Berber is well-known for allowing typologically rarer syllable structures. For instance, Tashlhiyt onset clusters can violate the sonority sequencing principle by containing sequences of consonants that 'plateau' /kti/ or decrease /rku/ in sonority (Dell & Elmedlaoui, 2002; Ridouane et al., 2014; Ridouane, 2016; Boukous, 1987, 2009). An alternative analysis of these consonant sequences is that they are not actually complex onsets; rather, they are heterosyllabic consonants reflecting Tashlhiyt's typologically unusual allowance of any consonant to be the nucleus of a syllable (e.g., /kti/ is syllabified as [k.ti]) (Jebbour 1996, 1999; Bensoukas, 2001). This hypothesis was explored by Ridouane, Hermes, and Hallé (2014) using both articulatory and perceptual evidence from native Tashlhiyt speakers. They found that as the number of consonants at the beginning of a word increased (e.g., /fik/ vs. /kfik/ vs. /tkfik/), the duration between the coda and the vowel-adjacent consonant remained stable. This is in contrast with true "complex onset" languages, which have been shown to compress the duration of vowel-adjacent consonants when the onsets contain more than one segment (e.g., in American English, the /k/ in "scab" and the /l/ in "plug" is shorter than the initial consonants in "cab" and "lug", respectively, Marin & Pouplier, 2010). Perceptual evidence reported by Ridouane et al. (2014) was consistent with this analysis: Tashlhiyt participants completed a game where they

heard a lexical item and repeated back only the “first part”. The majority of responses to words containing complex onsets were syllables containing singleton onsets, indicating that Tashlhiyt speakers parse initial consonant sequences into multiple independent syllables. Still, the allowance of these word-initial sequences of consonants in Tashlhiyt is typologically unusual, because they violate cross-linguistic preferences to either obey the sonority sequencing principle or to avoid non-sonorous segments to be syllable nuclei (Prince & Smolensky, 2004). In the current study, we do not attempt to disambiguate between these analyses; rather, in both accounts, certain types of word-initial consonant sequences that are allowed in Tashlhiyt words are not commonly found in other languages. We explore whether clear speech enhances the intelligibility of such consonant sequences in similar ways as for more typologically common contrasts (e.g., singleton onsets).

Tashlhiyt also has word-initial geminates that contrast with singleton onsets. While singleton-geminate contrasts are common cross-linguistically, word-initial geminates are typologically rare and found in only around 30 languages (Muller, 2001). Ridouane and Hallé (2017) examined the production and perception of word-initial singleton-geminate stops in Tashlhiyt which were produced in different phrasal contexts: under contrastive focus in phrase-initial, phrase-medial, and phrase-final position, as well as in isolation. Acoustically, geminates were longer than the singletons; the geminate voiceless stops also contained larger relative intensity differences between the burst and the following vowel compared to singleton counterparts; and this was even larger in focus position. They also evaluated Tashlhiyt listeners’ sensitivity to the initial singleton vs. geminate contrast using a perceptual discrimination task. Overall, listeners performed greater than chance at identifying the singleton-geminate contrasts, though the lowest perceptual discrimination was for the voiceless stops. Within voiceless stops, however, performance was greater for stimuli that were originally produced in focused position, compared to when they were produced in the other contexts. The authors argue that the geminate vs. singleton contrast is enhanced by speakers in focus position, though it is not clear whether this is due to articulatory strengthening or intelligibility-oriented motivation by speakers. Thus, we ask whether clear speech leads to an enhancement of the singleton vs. geminate onset contrast in Tashlhiyt.

Thus, there is considerable scholarly attention on the production and perception of “complex onsets” (i.e., multiple consonant sequences at the beginning of words that appear to syllabify into separate syllables) and word-initial geminates in Tashlhiyt, which can be used to inform theoretical models of speech production and cross-linguistic phonological typology. Consonant clusters are cross-linguistically more difficult to identify than single consonants because the presence of overlapping consonantal articulations attenuates the acoustic features that can be used by listeners to recover the segment identities (Henderson & Repp, 1982) and it has been argued that clusters containing acoustic cues that are more difficult to perceive will be less resistant to speech reduction effects than those that are easier to identify (Wright, 1996). Moreover, the typological rarity of obstruents serving as syllable nuclei has been argued to stem from the perceptual salience of high sonority peaks to serve as syllable-boundary marking units (Clements, 1990). Furthermore, initial singleton-geminate contrasts are uncommon because in phrase-initial position there are fewer cues to consonant duration (Muller, 2001). Thus, there are phonetic-perceptual explanations for why the word-initial geminate and consonant clusters, as those found in Tashlhiyt, are typologically uncommon. However, no prior work, to our knowledge, has directly compared the effect of different speaking styles (with words produced in the same sentence positions) on the intelligibility of these contrasts. Our approach is to compare

clear and casual speech productions of the perception of the three-way word-initial contrast in Tashlhiyt: singleton (e.g., *sin*), geminate (e.g., *ssin*), and multi-consonant (e.g., *fsin*, we refer to these as “consonant clusters” in the rest of the paper, though they appear to be treated as heterosyllabic) onsets.

### **C. Cross-language perception of clear speech**

How do naive listeners perceive the sounds of a foreign language? Examining cross-language speech perception is important to understand how the human mind copes with multiple languages (Bohn, 2017). In general, it is harder to discriminate between sounds and phonotactic patterns that are not present in a listeners’ L1 (e.g., Davidson, 2011; Best et al., 2001). However, there is work showing that clear speech improves intelligibility of non-native sound contrasts (Kabak & Maniwa, 2007). Examining how the perception of novel sound contrasts in a foreign language also has applications to second language acquisition as L2 adult learners usually begin as naive listeners (Bohn, 2017).

Cross-language perception has also been used to understand how language experience can be leveraged from an L1 to support the perception of typologically rarer sound contrasts in a non-native language. Relevantly, one study examined discrimination of the singleton-geminate contrast across different word positions (initial vs. final) in Tashlhiyt by naive listeners (whose L1 was French) (Hallé, Ridouane, & Best, 2016). They found that French listeners were better at discriminating between singleton and geminate segments when they were in word-final position, relative to word-initial position. Their interpretation of this finding is that French listeners were able to leverage differences in preceding vowel duration that covaries with final segment duration, a cue that is also present in their L1. This provides some support for our hypothesis that naive listeners in our present study will be able to leverage clear speech enhancements that are present on non-native Tashlhiyt words.

We ask whether clear speech, a cross-linguistically universal type of speech enhancement, improves the intelligibility of typologically rarer, non-native contrasts by naive listeners. Prior work has found that, while clear speech improves intelligibility for non-native listeners, the boost is substantially smaller than that for native listeners (Bradlow & Bent, 2002). This has been used to argue that clear speech is fundamentally geared toward native listeners who have extensive experience with the target language, though more recent work has suggested that this depends on multiple factors such as the type of contrast and background of the listener (Baese-Berk et al., 2020). However, again, prior work on non-native perception of clear speech has been limited to a small number of languages. Comparing native and naive listeners’ perceptual patterns in Tashlhiyt is a further way to address whether a listener-oriented mode of speaking translates to “native listener-oriented” enhancement of distinct sound patterns, or can benefit naive listeners, as well. Moreover, whether the clear speech benefit applies in parallel for rarer and more common sounds, or whether the asymmetries in how clear speech enhances sounds is even larger for naive listeners can inform phonological typology. We use the perception of clear speech in Tashlhiyt for naive listeners as a way to understand the nature of phonological enhancement.

### **D. Current study and predictions**

In the current study, we ask whether clear speech, an intelligibility-enhancing mode of speaking, can further inform the discussion about what makes some sound contrasts less common across languages of the world than others. We focus on a typologically-unusual three-way word-initial sound contrast in Tashlhiyt Berber: singleton (e.g., *sin*) vs. geminate (e.g., *ssin*) vs. complex sequences that are heterosyllabic (e.g., *fsin*). We elicit words containing these onset types in two types of speech styles: a casual and a clear speech mode. Then, we play pairs of words to native Tashlhiyt listeners and naive listeners who perform a discrimination task.

Our *first research question* is whether the more typologically rarer phonotactic contrasts receive less of a clear speech perceptual boost than typologically more common phonotactic contrasts?

On the one hand, clear speech is a mode of speaking with explicit motivation to make the underlying phonological structure recoverable to listeners. Also, cross-linguistically, even “more rare” phonological patterns are enhanced by speakers when produced in clear speech (cf. Korean 3-way VOT contrast: Kang & Guion, 2008). Therefore, one prediction is that all Tashlhiyt word-initial contrasts will be more intelligible in clear speech than in casual speech.

On the other hand, accounts that seek phonetic bases to explain typological patterns argue that some phonological contrasts are rarer than others because they are inherently more difficult, e.g. harder to perceive (Hayes & Steriade, 2004). We extend this hypothesis to a comparison across speaking styles: we hypothesize that rarer sound patterns receive less of a clear speech boost than more typologically common sound patterns. Following this hypothesis, a specific prediction for the current study is that discrimination of word pairs containing rarer contrasts (e.g., initial geminate or initial complex onsets) is increased to a lesser degree in clear speech than more common onset contrasts (e.g., words contrasting in singleton onsets only).

By examining the perception of typologically rarer phonotactic patterns in a clear speech mode, compared to a casual mode, we can better understand the phonetic bases for cross-linguistic sound patterns. Indeed, if certain sound contrasts are enhanced to a lesser extent in clear speech, relative to other contrasts, that could explain why they are less likely to be maintained across languages of the world (e.g., Blevins, 2004).

Our *second research question* is whether the clear speech benefit for the different types of Tashlhiyt onset types is the same for native vs. naïve listeners.

Prior work has used naive listeners’ perceptual evaluation of non-native phonotactic structures in order to reveal what is “universal” about typological tendencies (Berent et al., 2009). For instance, English listeners are less likely to accurately perceive onset clusters with non-rising sonority (e.g., /bd/, /md/) than clusters with rising sonority (e.g., /bl/, /ml/) (Berent et al., 2007; Berent et al., 2009). Moreover, cross-linguistic studies of clear speech have found commonalities, despite cross-linguistic variation in phonological inventories and patterns, in how speakers enhance the acoustic signal to be more intelligible to listeners (Smiljanić & Bradlow, 2005). Thus, we might predict parallels in the ways that clear speech improves perceptibility of word-initial onset contrasts in Tashlhiyt for native and naive listeners.

On the one hand, clear speech has been found to enhance the distinction between language-specific phonological categories (Buz et al., 2016; Leung et al., 2016). Furthermore, the perception of non-native *consonant clusters* appears to be highly language specific (Davidson & Shaw, 2012; Peperkamp, 2007). From that perspective, an alternative prediction is that naive listeners, particularly American English listeners who do not have experience with the particular onset contrasts in Tashlhiyt, will not be able to leverage the acoustic enhancements of clear speech to the same extent as native listeners for these typologically less common contrasts.



## II. Experiment 1: Native Tashlhiyt listeners

### A. Methods

#### 1. Stimuli

Critical target words consisted of five sets of four monosyllabic Tashlhiyt lexical items that differ in the structure of the onset: one item contained a complex onset (e.g., *fsin* ‘they untied’), one item contained a geminate onset (e.g., *ssin* ‘with two/both two’), one item contained a single onset, where the consonant was the C2 of the complex onset (e.g., *sin* ‘two’), and one item contained a single onset where the consonant was the C1 of the complex onset (e.g., *fin* ‘they festered’). The full list of sets is provided in Table I. All sets were selected to contain minimal pairs across the critical Single C2, Complex and Geminate onset comparisons, except for 1 set (marked with an asterisk in Table I). Prior work examining initial singleton vs. geminate onset contrast in Tashlhiyt focused on initial stops (Ridouane & Hallé, 2017). However, in the current study, we chose minimal word sets that contain a wide range of segment types. Our aim is to generalize across a wide range of manners of articulation. Therefore, we selected word sets with different types of manners of articulation. Two of the sets included a voiceless obstruent geminate-singleton-cluster contrast. One set included a voiced obstruent contrast. The remaining two sets included a sonorant, /l/. Three out of five of the words with complex onsets contained sonority-plateau or sonority-decreasing consonant sequences. In addition to the critical target word sets, 14 CVC words were selected to serve as non-minimal pair items in the experiment, provided in Table II.

Table I. Complex, Geminate, Single C2, and Single C1 onset competitor items used in the current study.

Single C2 onset	Complex onset	Geminate onset	Single C1 onset
<i>sin</i> ‘two’	<i>fsin</i> ‘they untied’	<i>ssin</i> ‘with two/both two’	<i>fin</i> ‘they festered’
<i>lik</i> ‘I married’	<i>kwlik</i> ‘I went up’	<i>llik</i> ‘I exist/I am somewhere’	<i>kwik</i> ‘I caught’
<i>lan</i> ‘they have’	<i>flan</i> ‘someone’	<i>llan</i> ‘they exist’	<i>fan</i> ‘they gave’
<i>tid</i> ‘these.FM’	<i>ktid</i> ‘recall/remember’	<i>ttid</i> ‘washing powder (brand)’	<i>mit</i> ‘what’
<i>dar</i> ‘at’	<i>kdar</i> ‘at, (in+at)’	<i>ddan*</i> ‘they went’	<i>kar</i> ‘only’

Table II. Non-minimal pair words (different trials paired words in list 1 and list 2 or list 3 and list 2 to maximize phonological distinctiveness).

Non-minimal pair list 1	Non-minimal pair list 2	Non-minimal pair list 3
<i>r<sup>h</sup>uh</i> ‘go home’	<i>sir</i> ‘go!’	<i>mun</i> ‘accompany’
<i>kir</i> ‘only’	<i>sut</i> ‘drink it’	<i>rik</i> ‘I want’
<i>gan</i> ‘they are’	<i>zud</i> ‘like, as’	<i>z<sup>h</sup>ur</i> ‘visit’

<i>tuf</i> ‘she’s better’	<i>luḥ</i> ‘throw’	<i>ran</i> ‘they want’
<i>sul</i> ‘stay alive’	<i>sak</i> ‘pass through’	

In addition to target and non-minimal pair words, a list of 14 concrete, picture-able nouns were selected (*ajjis* ‘horse’, *tafukt* ‘sun’, *afullus* ‘rooster’, *az‘alim* ‘onion’, *ajdi* ‘dog’, *izi* ‘fly’, *ajjur* ‘moon’, *akr‘um* ‘bread’, *anz‘ar* ‘rain’, *idukan* ‘slippers’, *tafunast* ‘cow’, *ilm* ‘skin’, *tafruxt* ‘girl’, *afrux* ‘boy’). These words would be used in a pre-test language comprehension assessment for the native Tashlhiyt listeners (see section II.A.4.).

## 2. Speech styles

The target and non-minimal pair words were produced 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 Audio-technica ATM33a microphone and USB audio mixer (Sound Devices, USB Pre 2), digitized at a 44.1 kHz sampling rate. To elicit Clear Speech, the speaker produced the words to a real listener (one of the authors, GZ, who was not a speaker of Tashlhiyt) with the instructions “In this condition, speak the words clearly to someone who is having a hard time understanding you.” In addition, the speaker was told that the listener had an unordered list of the lexical items and their job was to put the words in the correct order as they were spoken. These two instructions were designed to provide the motivation for the speaker to produce maximally intelligible productions and is similar to that used to elicit clear speech in prior work (e.g., Bradlow, 2002 elicited clear speech by having speakers imagine they are talking to someone from a different language background). The speaker produced each word twice, in two different frame sentences: *ini \_\_\_ jat tklit* ‘say \_\_\_ once’, *inna \_\_\_ bahra* ‘he said \_\_\_ a lot’. Clear Speech was elicited first following prior paradigms (e.g., similar interlocutor-based clear vs. casual speech styles were elicited in this order by Scarborough & Zellou (2013)).

Following the Clear Speech style elicitation, the speaker produced the words in a casual, casual speaking style. To elicit Casual Speech, the speaker was given the following instructions: “now, speak the list as if you are talking to a friend or family member you have known for a long time who has no trouble understanding you, and speak quickly”. These instructions are similar to those taken from prior work; for instance, Bradlow (2002) elicited a casual speech style by instructing speakers to talk as if to someone familiar with their voice (similar instructions were provided by Kang and Guion (2008)). In this case, the speaker was also notified that the listener was following the productions with a list that matched the exact ordering of the speaker’s list. As for the Clear speech, the speaker produced the words in Casual speech twice, in the two frame sentences. The listener did not provide feedback or respond during word elicitation at any point.

## 3. Stimuli preparation

All items were segmented and excised from their frame sentences to be presented in isolation. These recordings were amplitude normalized to 60 dB. Typically, signal detection paradigms present speech embedded in noise in order to increase the difficulty of the task. Therefore, all items were mixed with white noise (following, e.g., Bradlow & Bent, 2002) at a signal-to-noise

ratio (SNR) of +3 dB (similar to prior work which used an SNR of +2 dB (e.g., Bradlow & Alexander, 2006)).

#### 4. Participants and procedure

Participants consisted of 28 native speakers of Tashlhiyt Berber (mean age: 42 years old,  $sd = 10$  years; 7 female, 21 male), with no reported hearing impairments. Participants were recruited through email flyers. All participants reported that Tashlhiyt was their first language and that both parents speak Tashlhiyt. Participants also spoke French (all instructions were provided in French). Participants reported growing up in cities or towns in Southern Morocco, including Agadir ( $n = 16$ ), Marrakech ( $n = 2$ ), Essaouira ( $n = 1$ ), or smaller villages ( $n = 9$ ). None of the listeners reported having a hearing or language impairment.

The experiment was conducted online using Qualtrics. As most computer-literate Tashlhiyt speakers are fluent in French, all instructions in the experiment were presented in French. Participants completed informed consent before participating. Then, before beginning the study, participants were instructed to complete the experiment in a quiet room without distractions or noise, to silence their phones, and to wear headphones.

The experiment began with pre-test language comprehension assessment in order to verify that the listeners were native speakers of Tashlhiyt. The paradigm was a word identification task. On each trial, listeners heard a production of one of 14 lexical comprehension nouns, presented once only, and were presented with two black and white images on the screen. Participants were instructed to select the image corresponding to the word they heard. Participants completed 14 of these trials, hearing each of the nouns once. All of the participants correctly identified every noun.

Following the word identification task, participants completed an “AX” discrimination task. On a given trial, listeners heard two different stimulus items, presented with a 400 ms inter-stimulus interval, and they were instructed to indicate whether the two items are the same word (repeated twice) or two different words. The stimuli were presented once only, with no possibility to repeat the trial. Before starting the task, participants were given instructions for this paradigm and presented two example trials, one containing the same word repeated twice and one containing different words, and feedback on the correct response for those examples.

During each discrimination trial, participants responded whether the two productions were the same word (“le même mot”) or two different words (“deux mots différents”) via a mouse click. For the critical target word sets, there were four types of different-word discrimination trial types: 1) Single C2 onset vs. Another single onset (e.g., *sin* v. *fin*); 2) Single C2 onset vs. Complex onset (e.g., *sin* v. *fsin*); 3) Single C2 onset vs. Geminate onset (e.g., *sin* v. *ssin*); 4) Complex onset vs. Geminate onset (e.g., *fsin* v. *ssin*). In addition, ten non-minimal pair word pairings were created using the non-minimal pair items. The non-minimal pair word pairings were selected to contain non-minimal pair words, with multiple different phonemes (e.g., *mun* vs. *sir*).

There were an equal number of “same” and “different” trials: half of the trials consisted of different lexical items; the other half of trials consisted of two different productions of the same word. Each trial type was presented in the Clear Speech style and in the Casual Speech style (i.e., both items within all trials were taken from the same speech style elicitation; same-word discrimination trials contained different productions of the same word produced in the same speaking style). There were a total of 120 discrimination trials: 20 critical word trials + 10

non-minimal pair trials \* 2 speech styles (Clear, Casual) \* 2 word pairing types (same, different). The trials were randomly presented. Four lists were generated, consisting of the two different orderings of the words within each pairing and two different repetitions of each word across the lists. Participants were randomly assigned to a list. The experiment took approximately 20 minutes to complete.

## 5. Acoustic Analysis

Clear speech is characterized by a range of acoustic modifications, relative to casual speech. For instance, prior work has shown that clear speech is consistently produced with longer consonant and vowel durations, as well as higher vowel intensity (Picheny et al., 1986; Krause & Braida, 2002). Therefore, we analyze these acoustic features in the target words to verify a difference in clear and casual speech across our stimuli.

Consonant duration, in particular, has been shown to vary across the onset types examined here. Prior work has identified that word-initial geminates are generally produced with longer closure duration, cross-linguistically (Muller, 2001), and in Tashlhiyt (Ridouane, 2007; Ridouane & Hallé, 2017; Ridouane, 2022)<sup>1</sup>. Furthermore, prior work has examined the temporal properties of consonants in onset clusters of Tashlhiyt in order to make claims about their phonological status (Ridouane, Hermes, & Hallé, 2014). There is also work looking at temporal variation in CC onset clusters shows that changes in speech rate affect the duration of C2 more than C1.

The vowel-adjacent consonant duration (e.g., the duration of the /s/ in *sin*, *ssin*, and *fsin*) and vowel duration in each target word was hand-segmented by a phonetically trained research assistant. Vowel and consonant durations were measured, in addition to vowel intensity.

The consonant duration and vowel duration data were logged. Each of the three dependent variables were centered and scaled. Three separate mixed effects linear regression models were run on each acoustic variable, using the *lme4* (Bates et al., 2014) and p-values were calculated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) in R (version 4.1.2, R Core Team, 2021). The models each contained identical fixed and random effects structures: fixed effects of Onset Type (Single [reference level], Complex, Geminate; treatment coded), Speech Style (Clear, Casual [reference level]; treatment coded), and their interaction. By-word random intercepts were also included.

## 6. Perception Analysis

In order to quantify a standardized measure of sensitivity to the word contrasts, we converted listeners' discrimination responses to d-prime. D-prime is a commonly used measure of sensitivity used in speech discrimination and signal detection studies (Macmillan & Creelman, 1991). We computed a d-prime score for each participant for each contrast type and for each speech style based on their hit rate (i.e., proportion of correctly selecting "different" for trials where the productions were indeed two different words) and their false alarm rate (i.e., proportion of incorrectly selecting "different" for trials where the same word was presented

---

<sup>1</sup> Acoustic analyses have revealed secondary acoustic characteristics associated with geminate production; for instance, in Tashlhiyt, initial voiceless stop geminates are produced with longer closure duration and greater VOT than singletons (Ridouane & Hallé, 2017). However, since our target items contained a range of manners of articulation, we focus just on consonant duration.

twice) within each condition. D-prime is calculated as the difference between the z-scored hit rate and the z-scored false alarm rate. A higher relative d-prime score indicates that listeners are better able to discriminate between two sound contrasts while correcting for each individual's rate of misses and/or false alarms. This resulted in ten d-prime scores for each listener, a value for each of the five Contrast Types (Single Onsets [e.g., *sin-fin*], Single v. Complex [e.g., *sin-fsin*], Single v. Geminate [e.g., *sin-ssin*], Geminate v. Complex [e.g., *ssin-fsin*], Non-minimal pair [e.g., *sir-mun*]) across the two Speech Styles.

D-prime scores were analyzed using linear mixed effects models, using the *lme4* (Bates et al., 2014) and p-values were calculated using the *lmerTest* package (Kuznetsova et al., 2017) in R (version 4.1.2, R Core Team, 2021). The model included fixed effects of Speech Style (Clear, Casual [reference level]; treatment coded) and Contrast Type (Single Onsets [e.g., *sin-fin*], Single v. Complex [e.g., *sin-fsin*], Single v. Geminate [e.g., *sin-ssin*], Geminate v. Complex [e.g., *ssin-fsin*], Non-minimal pair [e.g., *sir-mun*] [reference level]; treatment coded). The interaction between style and contrast type was also included. The random effects included by-participant random intercepts and by-participant random slopes for contrast type and style.

Prior work has used a technique for interpreting lmer models that contain a factor with more than two levels. In this procedure, a model is run multiple times on the same data set with different reference levels for the factor of comparison (Clopper, 2013; Tremblay et al., 2017). We follow this procedure: the model was run with the reference level for Contrast Type re-leveled in order to produce estimates for multiple pairwise contrasts. (Model 1 reference level = Non-minimal pair, Model 2 reference level = Single v. Complex; Model 3 reference level = Single v. Geminate). Post-hoc tests of pairwise comparisons of style within each contrast type were done using a *Tukey* correction for multiple comparisons using the *emmeans()* function in the *emmeans* R package (V. 1.5.3; Lenth et al., 2021).

## B. Results

### 1. Acoustic properties of the stimuli

Acoustic measurements (by onset type and speech style) are shown in Table III. As seen, for all the measurements and all the onset types, values are higher for Clear speech than for Casual speech, indicating hyperarticulation in the clear speaking mode.

*Table III.* Means (and standard errors) of acoustic values for each onset type by speech style.

	Single		Complex		Geminate	
	Clear	Casual	Clear	Casual	Clear	Casual
Vowel-adjacent Consonant duration (ms)	106.4 (13.3)	71.0 (10.0)	93.5 (7.8)	65.6 (6.8)	191.9 (21.1)	141.2 (13.0)
Vowel duration (ms)	129.6 (1.9)	52 (3.8)	115.6 (5.7)	57.8 (3.4)	120.6 (6.6)	65.6 (4.9)
Vowel intensity (dB)	77.7 (1.1)	73.7 (0.6)	76.6 (1.4)	73.6 (0.5)	76.6 (1.1)	73.3 (0.5)

The consonant duration model revealed a significant main effect of Speech Style ( $coef. = 0.72, t = 2.8, p < 0.01$ ) confirming that, across the board, consonant durations are longer in the clear speaking mode. The model also computed a main effect of Geminate Onset type, indicating that initial geminates are longer than initial singletons ( $coef. = 1.29, t = 2.8, p < 0.05$ ). Yet, there was no difference in consonant duration between singleton and complex onsets ( $p = 0.86$ ). The observation that singletons and vowel-adjacent consonants in complex onsets are not different in duration is consistent with the claim that pre-vocalic consonant clusters in Tashlhiyt are heterosyllabic (e.g., Ridouane, Hermes, & Hallé, 2014). In other words, the similar duration profiles for the [s] in /sin/ and /fsin/ is consistent with prior articulatory and phonological evidence supporting the claim that the latter is syllabified as [f.sin]. There were no significant interactions between Onset Type and Speech Style in the consonant duration model, indicating that the effect of speaking mode is consistent across word types.

The vowel duration model also computed a significant effect of Speech Style, wherein vowels in clear speech are longer than in casual speech ( $coef. = 2.22, t = 13.4, p < 0.001$ ). Vowels following geminate onsets were longer than vowels following singletons ( $coef. = 0.55, t = 2.5, p < 0.05$ ), but vowels following complex onsets were not different in length than vowels following singletons ( $p = 2.5$ ). Significant interactions between Style and Onset type revealed that the difference in vowel duration between Clear and Casual Styles was smaller for words with initial geminates than for words with initial singletons ( $coef. = -0.76, t = -3.3, p < 0.01$ ). Similarly, the difference in vowel durations between Clear and Casual for complex onset words was also smaller than that for singleton words ( $coef. = -0.56, t = -2.4, p < 0.05$ ).

For the vowel intensity model, a significant effect of Speech Style also confirmed that vowels in Clear speech are produced with greater intensity than vowels in Casual speech ( $coef. = 1.11, t = 4.0, p < 0.001$ ). No other main effects or interactions were significant in the intensity model.

## 2. Native Tashlhiyt listeners' perception

Figure 1 displays d-prime scores for contrast comparisons across clear and casual speaking styles. A Table with summary statistics from the model is provided in Appendix A (with outputs from when the Contrast type reference level was non-minimal pair and when it was relevelled to Single v. Geminate and then Single v. Complex; in the text, effects are reported for the non-minimal pair reference level output, except where relevelings are explicitly mentioned). There was a significant main effect of Speech Style ( $coef. = 1.71, t = 7.92, p < 0.001$ ), indicating that, overall, Tashlhiyt listeners are better able to discriminate contrasts in clear speech, compared to casual speech. There were also main effects of Contrast Type. As seen in Figure 1, perceptual discrimination is lower for single onset contrasts than non-minimal pairs ( $coef. = -2.05, t = -8.64, p < 0.001$ ). D-prime is even lower for the other contrasts: Single v. Complex ( $coef. = -2.70, t = -11.59, p < 0.001$ ); Single v. Geminate ( $coef. = -3.39, t = -15.95, p < 0.001$ ); and Geminate v. Complex ( $coef. = -3.92, t = -18.44, p < 0.001$ ). The model with Contrast Type relevelled to be Single v. Geminate revealed that listeners are significantly less sensitive to that contrast compared to the Single v. Complex contrast ( $coef. = 0.7, t = 2.9, p < 0.01$ ); that releveling also revealed that listeners show less perceptual sensitivity toward Complex v. Geminate contrasts than the Single v. Geminate contrasts ( $coef. = -0.53, t = -2.49, p < 0.05$ ). In other words, comparisons with the lowest d-prime scores contain an initial geminate. The ranking of the

perceptual sensitivity of the contrasts is as follows: Non-Minimal Pair > Single Onsets > Single v. Complex > Single v. Geminate > Geminate vs. Complex.

The model also computed interactions between Speech Style and Contrast Type. The difference in listener sensitivity to the lexical contrasts between clear and casual speech is larger for the Single Onset contrasts from the Non-Minimal Pair contrasts (*coef.* = 2.01, *t* = 7.31, *p* < 0.001). As seen in Figure 1, this is due to a greater reduction in perceptual sensitivity to the single onset contrasts from clear to casual speech; in other words, while in clear speech, single onset contrasts are just as discriminable as non-minimal pair items, casual speech leads to a larger decrease in the perceptual discriminability between single onset contrasts. On the other hand, the difference in sensitivity scores across speech types is larger for the non-minimal pair trials than for the other contrast types, indicated by lower coefficient values for the interaction term. Post hoc *Tukey* tests were performed on the model with *emmeans* to unpack the effect of Speech Style within each Contrast Type and the significance levels from these tests are provided in Figure 1. Clear speech leads to a small increase in discriminability for the single vs. geminate onset contrast (*p* < 0.001). However, there is no difference in discriminability across speech styles for the single vs. complex onset contrast (*p* = 0.1) or for the geminate vs. complex onset contrast in casual speech, compared to the clear speech condition (*p* = 0.1).

### C. Interim discussion

In Experiment 1, native Tashlhiyt listeners show variation in their sensitivity to different types of word-initial contrasts when performing a discrimination task with the sounds embedded in noise. As expected, minimal pairs were harder to discriminate between than non-minimal pairs. Moreover, listeners are less sensitive to word-initial contrasts involving typologically rarer sounds: contrasts involving complex onsets have lower sensitivity scores and contrasts involving initial geminates have the lowest perceptual discriminability scores. Thus, even for listeners who use these contrasts natively, the more typologically rarer contrasts are the most difficult to discriminate.

Our first research question was whether differences in discriminability across clear and casual speech are smaller for typologically rarer word-initial contrasts in Tashlhiyt, relative to more common ones. We predicted that the perceptual increase seen in clear speech is smaller when listeners discriminate between word pairs that contain a typologically rarer pattern (e.g., an initial geminate *ssin* or initial cluster with falling sonority *fsin*) relative to discrimination of words contrasting in word-initial singletons (e.g., *sin* vs. *fin*).

Overall, we observed a clear speech effect for Tashlhiyt. When the words are produced in a listener-oriented, intelligibility-enhancing mode, listeners are better able to discriminate between two different items. This basic finding adds to growing work on the intelligibility benefit of clear speech, cross-linguistically (e.g., Croatian in Smiljanić & Bradlow, 2008; Korean in Kang & Guion, 2008; Mandarin in Li et al., 2011). With respect to our specific question, the results indicate that clear speech does not enhance intelligibility for all word-initial contrasts equally for native Tashlhiyt listeners. While the intelligibility-oriented speaking mode improves discriminability for single consonant onset contrasts and also for the initial singleton-geminate contrast, it does not improve performance for contrasts containing initial clusters (compared to either a single or geminate onset). Moreover, the size of the effect was smaller for contrasts containing initial geminates than contrasts containing initial singletons only, supporting our hypothesis.

[ PLACE FIGURES 1 AND 2 ROUGHLY HERE]

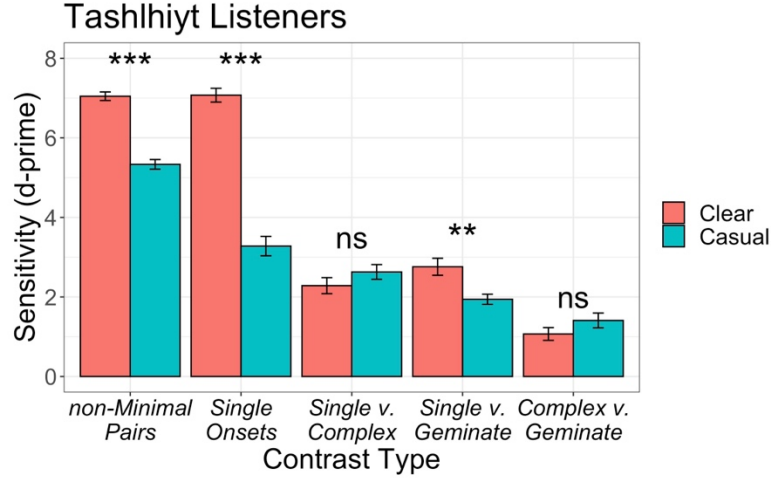


Figure 1. (Color Online). Native Tashlhiyt listeners' mean sensitivity (d-prime scores) for Tashlhiyt contrasts by speaking style (error bars depict standard error of the mean). Significance levels reflect pairwise comparisons of speaking style within each contrast type as computed using post hoc *Tukey* tests performed on the model with *emmeans*.

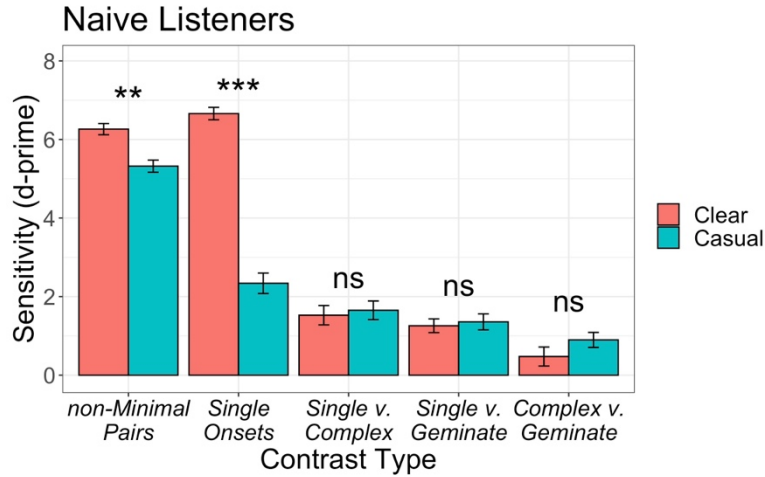


Figure 2. (Color Online). Naive listeners' mean sensitivity (d-prime scores) for Tashlhiyt contrasts by speaking style (error bars depict standard error of the mean). Significance levels reflect pairwise comparisons of speaking style within each contrast type as computed using post hoc *Tukey* tests performed on the model with *emmeans*.

### III. Experiment 2: Naive listeners

#### A. Methods



## 1. Participants and procedure

Twenty-eight non-Tashlhiyt-speaking participants (mean age = 19.5 years old,  $sd = 1.5$ ; 17 female, 2 non-binary, 9 male), who had no reported hearing impairment, were recruited from the UC Davis Psychology subjects' pool. All of the participants reported being native speakers of American English. Some participants reported being bilingual, speaking a language other than English in the home (Spanish,  $n = 6$ ; Urdu,  $n = 1$ ; Vietnamese,  $n = 3$ ; Chinese,  $n = 2$ ; Kannada,  $n = 1$ ; Russian,  $n = 1$ ; Ilocano,  $n = 1$ )<sup>2</sup>. None of the participants reported that they spoke or had studied Tashlhiyt or any of the languages of North Africa. None of the listeners reported having a hearing or language impairment.

## 2. Procedure and analysis

Experiment 2 consisted of a discrimination task. Stimuli, trial, and list structure for the discrimination task was identical to that in Experiment 1. The experiment was conducted online using Qualtrics. Participants were instructed to complete the experiment in a quiet room, where they would be undisturbed for the duration of the task, and to wear headphones. The experiment began with a sound calibration procedure: participants heard one sentence presented auditorily ("She asked about the host"), presented in silence at 60 dB, and were asked to identify the sentence from three multiple choice options, each containing a phonologically close target word (host, toast, coast).

Subjects were told that they would be hearing utterances produced in a language they have never heard before. They were told that it is fine that they don't understand the words. Participants were instructed to identify whether the two utterances in a trial contain the same or different sounds. Participants were also told that they could guess if they were not sure. Before starting the task, participants were given instructions about this paradigm and they heard two practice trials, a "same" example and a "different" example, and feedback on the correct response for those examples. Then, participants completed the discrimination task, indicating "same" or "different" on each trial with a mouse click.

D-prime scores were calculated and analyzed following the same procedure as that in Experiment 1.

## B. Results

The mean d-prime scores by contrast type and speaking style are plotted in Figure 2. A Table with the summary statistics from the model is provided in Appendix B (with outputs from when the Contrast type reference level was non-minimal pair and when it was relevelled to Single v. Geminate and then Single v. Complex; in the text, effects are reported for the non-minimal pair reference level output, unless explicitly mentioned). Speaking style was a main effect: naive listeners display higher sensitivity to contrasts overall in the clear speaking mode ( $coef. = 0.94$ ,  $t = 3.72$ ,  $p < 0.001$ ). There were also differences in naive listeners' perceptual sensitivity across contrast types. Perceptual discrimination is lower for single onset contrasts than for non-minimal

---

<sup>2</sup> Word-initial geminates are typologically uncommon and not found in the native language(s) of the naive listeners (Muller, 2001). To our knowledge, none of the native languages of the naïve listeners are reported to have heterosyllabic complex onsets.

pair contrasts ( $coef. = -2.98, t = -11.53, p < 0.001$ ). The model with contrast type relevelled to Single v. Complex showed that perceptual sensitivity is greater for single onset contrasts relative to single v. complex onset contrasts ( $coef. = 0.69, t = 2.64, p < 0.01$ ). Naive listeners are also better at discriminating between Single v. Complex than Complex v. Geminate contrasts ( $coef. = -0.75, t = -2.6, p < 0.05$ ). When the reference level was set to be Single v. Geminate, we find that naive listeners show no difference in perceptual sensitivity between the single v. geminate and the geminate v. complex onset contrast ( $p = 0.3$ ) or to the single v. geminate and the single v. complex onset contrast ( $p = 0.1$ ). Thus, naive listeners' sensitivity hierarchy is as follows: Non-Minimal Pair > Single Onsets > Single v. Complex > Single v. Geminate = Geminate vs. Complex; although, there is a difference within complex onset pairs: Single v. Complex > Geminate vs. Complex.

There were also interactions between Speaking Style and Contrast Type. The difference in perceptual sensitivity between clear and casual speech is larger for the single onset contrasts than for the non-minimal pairs ( $coef. = 3.38, t = 9.43, p < 0.001$ ). Figure 2 illustrates this effect: while the discriminability of single onset minimal pairs and non-minimal pairs are similar in clear speech, the ability of naive listeners to discriminate between words that contrast in simple onsets decreases greatly in casual speech, paralleling the pattern observed for Tashlhiyt listeners. Post hoc *Tukey* tests indicate that naive listeners show no difference in perceptual sensitivity across clear and casual speech productions for each of the contrasts involving complex and/or geminate onsets.

#### IV. Comparison of listener groups

In order to directly compare perceptual sensitivity patterns of Tashlhiyt onset contrasts in clear and casual speech across the native and naive groups, a linear mixed effects model was run on the combined d-prime data from Experiments 1 and 2. The model contained three fixed effects: Listener Group (Naive listeners [reference level], Tashlhiyt native listeners; treatment coded), Speaking Style (Casual [reference level], Clear; treatment coded), and Contrast Type (Non-minimal pair [reference level], Single Onsets, Single v. Complex, Single v. Geminate, Geminate v. Complex; treatment coded). All two- and three-way interactions were also included. The model included random intercepts for each participant and by-participant random slopes for Contrast Type (the model did not converge when by-subject random slopes for style were included).

While there was not an overall effect of Listener Group ( $p = 0.9$ ), there were interactions between Group and the other predictors. The interaction between Group and Speaking Style ( $coef. = 0.77, t = 2.3, p < 0.05$ ) revealed that the difference in overall perceptual sensitivity for Tashlhiyt speakers across clear (mean d-prime = 4.1) and casual (mean = 2.9) is larger than for naive listeners (clear mean = 3.2, casual mean = 2.3).

There were interactions between Listener Group and Contrast Type. Notably, while Tashlhiyt listeners show higher perceptual sensitivity than naive listeners to the contrast types with the highest overall d-prime scores (Group\*Single Onsets ( $coef. = 0.97, t = 2.7, p < 0.01$ ; Tashlhiyt mean = 5.3, Naive mean = 4.5) and Group\*Single v. Complex ( $coef. = 0.90, t = 2.6, p < 0.05$ ; Tashlhiyt mean = 2.5, Naive mean = 1.6)), there are no differences across groups in discrimination performance of the contrasts containing initial geminates (Group\*Single v. Geminate  $p = 0.2$ ; Group\*Geminate v. Complex  $p = 0.1$ ).

There was a three-way interaction between Group, Speaking Style, and Contrast Type, which was explored using *emmeans* to perform *Tukey*-adjusted pairwise comparisons of estimated marginal means for d-prime for the Naive vs. Native means at each level of comparison. The output of these post hoc comparisons, provided in Table IV, reveals that, for the rarer contrasts, the native Tashlhiyt listeners show higher perceptual sensitivity than the naive listeners across both clear and casual speech. However, for the easier contrasts, language experience played a role in whether naive listeners perform similarly to the native listeners. For non-minimal pair words in casual speech, naive listeners perform equivalently to native listeners; yet, native listeners perform reliably better than naive listeners for these words in clear speech, indicating that language-specific experience is helpful to leverage clear speech for the easier comparison type. Meanwhile, for the single-onset minimal pair words, clear speech boosts performance for naive and native listeners equally.

Table IV. Listener comparisons with *Tukey*-adjusted p-values.

	<i>Est.</i>	<i>SE</i>	<i>df</i>	<i>Lower CI</i>	<i>Upper CI</i>	<i>t</i>	<i>p (Tukey)</i>
Naive vs. Native (Casual, <i>sir-mun</i> )	-0.01	0.25	112.95	-0.52	0.49	-0.05	0.96
Naive vs. Native (Clear, <i>sir-mun</i> )	-0.78	0.25	114.20	-1.28	-0.28	-3.09	<0.001
Naive vs. Native (Casual, <i>sin-fin</i> )	-0.94	0.33	83.43	-1.59	-0.29	-2.89	<0.001
Naive vs. Native (Clear, <i>sin-fin</i> )	-0.41	0.30	90.46	-1.01	0.18	-1.38	0.17
Naive vs. Native (Casual, <i>sin-fsin</i> )	-0.98	0.27	101.55	-1.52	-0.43	-3.57	<0.001
Naive vs. Native (Clear, <i>sin-fsin</i> )	-0.76	0.29	94.93	-1.33	-0.19	-2.63	<0.05
Naive vs. Native (Casual, <i>sin-ssin</i> )	-0.58	0.23	132.05	-1.05	-0.12	-2.50	<0.05
Naive vs. Native (Clear, <i>sin-ssin</i> )	-1.50	0.24	121.44	-1.98	-1.02	-6.14	<0.001
Naive vs. Native (Casual, <i>ssin-fsin</i> )	-0.51	0.25	113.63	-1.01	-0.01	-2.01	<0.05
Naive vs. Native (Clear, <i>ssin-fsin</i> )	-0.59	0.27	104.14	-1.13	-0.06	-2.21	<0.05

### A. Comparison of onset clusters based on sonority profile

Our main research question is whether there are differences in perceptual sensitivity across clear and casual speech for onset types in Tashlhiyt. In the main analysis above, we grouped together all of the complex clusters as one onset type based on the position that all onset clusters are heterosyllabic in Tashlhiyt (Ridouane et al., 2014). However, as described in the methods (section II.A.1), the word sets selected for this study contained onset clusters that varied in having an increase in sonority or not: Three out of five of the words with complex onsets contained sonority-plateau or sonority-decreasing consonant sequences (/kt/, /fs/, /ɤd/), while the other two contained increasing sonority (/fl/, /ɤ<sup>w</sup>l/). Since the former type of sequence is actually more typologically unusual, we compare these two types of clusters in order to test the hypothesis that examining clear speech perception can inform theories of phonological typology.

Thus, we performed a post-hoc analysis on just the trials containing complex onsets to compare perceptual sensitivity to clusters containing non-rising sonority to those with rising

sonority across speaking styles. We computed a d-prime score for rising and non-rising sonority clusters for each participant and for each speech style. We ran an lmer with these d-prime scores as the dependent variable. The model included fixed effects of Sonority Type (Rising, Non-rising [reference level]), Speech Style (Clear, Casual [reference level]) and Listener Group (Tashlhiyt native listeners, Naive listeners [reference level]). The model included random intercepts for each participant and by-participant random slopes for Sonority Type and Speech Style.

The model computed a main effect of Listener group: Tashlhiyt listeners have higher perceptual sensitivity than non-native listeners ( $coef. = 1.4, t = 3.4, p < 0.01$ ). There was also a significant interaction between Sonority Type and Speech Style ( $coef. = 1.3, t = 2.7, p < 0.05$ ), which can be seen in Figure 3. In effect, while listeners' ability to discriminate word pairs that contain a rising-sonority complex onset improves in clear speech, discrimination of non-rising sonority complex onsets *decreases* in clear speech. Thus, there is a clear speech benefit for onset clusters in Tashlhiyt, but only for clusters that increase in sonority, i.e. the more typologically common onset cluster types. No other effects or interaction were significant.

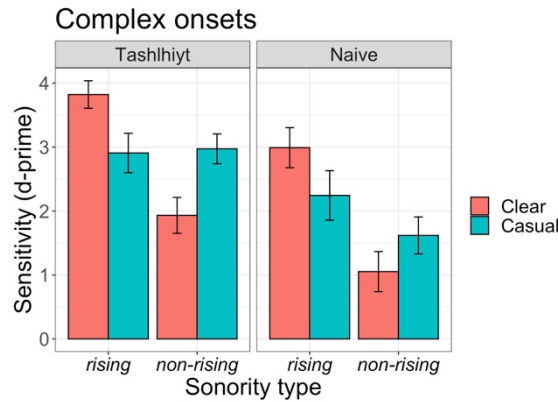


Figure 3. (Color Online). Native and Naive listeners' mean sensitivity (d-prime scores) for Tashlhiyt cluster contrasts by sonority type and speaking style (error bars depict standard error of the mean).

## V. General Discussion

The current study examined how clear speech influences the perception of word-initial phonological contrasts in Tashlhiyt Berber by native and naive listeners. We find that, in general, clear speech can improve listeners' ability to discriminate between Tashlhiyt Berber words, relative to casual speech. This observation adds to growing cross-linguistic empirical evidence about the intelligibility benefit of listener-oriented speech modes.

A main goal in our investigation of the perception of clear speech in Tashlhiyt Berber is an examination of the typologically rare phonotactic patterns in the language. Specifically, we explored the effect of clear speech on perceptual sensitivity of the contrast between word-initial singleton (e.g., *sin*), geminate (e.g., *ssin*), and complex onsets (e.g., particularly those that violate sonority sequencing by having non-rising sonority: *fsin* and *ktid*). Overall, across listener groups and speech styles, discrimination performance of pairs containing more typologically rare onset types is lower. For native Tashlhiyt listeners, their ability to discriminate between different onset types decreases with the typological rarity of those contrasts ( $sin-fin > sin-fsin > sin-ssin > ssin-fsin$ ). This is consistent with findings reported in prior work – for instance, that initial geminate vs. singleton contrast is difficult for native Tashlhiyt Berber listeners (Ridouane & Hallé, 2017) – and supports the hypothesis that the cross-linguistic frequency of phonological patterns can be at

least partially explained by phonetic perceptual biases. In this case, it could be argued that what makes initial geminates and complex onsets (i.e., like those found in Tashlhiyt with non-rising sonority) less common in languages of the world is that they are harder to discriminate, even for native listeners, and lead to less robust lexical contrast maintenance (e.g., Wright, 2004).

While there is an overall clear speech benefit, we find that not all the Tashlhiyt phonological contrasts are enhanced for listeners to an equal extent in clear speech. Clear speech does not enhance the perceptibility of the rarer sound patterns as much as for the more common ones. Clear speech does not improve the discriminability of falling-sonority onset clusters. Also, for the singleton-geminate contrast, where clear speech does lead to better discrimination by the native listeners only, the effect size of the speech style difference was smaller than that for the single onset and non-minimal pair contrasts. This observation is consistent with the claim that harder to perceive contrasts will be less common, cross-linguistically (e.g., Liljencrantz & Lindblom, 1972; Flemming, 2005; Hayes & Steriade, 2004). Specifically, the smaller, or non-existent, clear speech enhancement for geminate and complex onset contrasts might explain why they are rarer, cross-linguistically. In this case, harder-to-perceive contrasts appear to be harder-to-perceive even when speakers make explicit efforts to be intelligible.

In contrast, clear speech did improve discriminability of single onset contrasts, as well as onset clusters containing rising-sonority consonant sequences. That this was observed for both native and naive listeners is evidence that clear speech provides a universal, cross-language boost for these contrasts. This is consistent with the hypothesis that there is a cross-language preference for rising sonority clusters (cf. Berent et al., 2010: 419). Moreover, the observation that native speakers show a difference in the clear speech boost for rising vs. falling onset clusters could be further taken as evidence of a phonological difference between these categories.

Why is the singleton-geminate contrast enhanced (to a small extent) in clear speech for Tashlhiyt listeners, but not the contrasts involving falling-sonority complex onsets? One possibility is that this is consistent with the hypothesis that complex onsets are actually heterosyllabic in Tashlhiyt, e.g., that /fsin/ is syllabified as [f.sin] (Ridouane et al., 2014). In the heterosyllabicity scenario, the [sin] portion of /fsin/ is phonologically identical to that of /sin/, where the [s] is a single onset. A heterosyllabicity analysis would mean that the vowel-adjacent consonants in *sin* and *fsin* are in structurally similar positions, which might lead to similar acoustic enhancements and, subsequently, greater confusability when the listening conditions (i.e., in noise) mask the presence and identity of the C1. Indeed, the acoustic analysis of our stimuli revealed that there is no difference in duration between the vowel-adjacent consonants across “complex” and single onset target words, consistent with this possibility. Actually, cross-linguistic work examining acoustic enhancements of clusters is somewhat mixed. For instance, while Wright (1996) found that stop+stop clusters in Tsou that could be categorized as being more perceptually difficult contained less auditory cue reduction in casual speech, this was not observed in Russian by Pouplier et al. (2017), who found no difference in cue reduction based on cluster difficulty. For the talker and signal-to-noise ratio used in the current study, we observe that more carefully articulated clusters are not easier to perceive in Tashlhiyt. Thus, all three possible scenarios have been observed across different languages: cue enhancement of more difficult clusters in Tsou, no difference based on cluster difficulty in Russian, and less cue enhancement in more difficult clusters in Tashlhiyt. Indeed, there is much work outlining how onset cluster production is highly cross-linguistically variable (e.g., Pouplier et al., 2022). Exploring how listener-oriented speaking styles influence onset cluster production and

perception across different languages is an avenue that future work can explore this issue more extensively.

Another goal of the present study was to examine cross-language perception of clear speech in Tashlhiyt. We found that clear speech does not improve the perceptual sensitivity of the more difficult contrasts containing initial geminates for naive (here, American English-speaking) listeners, in contrast to the clear speech benefit seen for native listeners (recall that there was no clear speech benefit for falling-sonority clusters for either listening group). The fact that clear speech does not improve perceptibility of these sounds for naive listeners is further support for the hypothesis that rarer sounds are those that are harder-to-perceive; here, this is borne out in naive listeners' inability to leverage the enhanced cues present on clearly produced initial geminates.

However, we observe that clear speech enhances the discriminability of some Tashlhiyt Berber words for naive listeners. This demonstrates that the perceptual benefit of intelligibility-oriented speech styles applies even for speech from a language that listeners have never heard before and adds to growing work examining non-native speech perception of Tashlhiyt (cf. Ridouane et al., 2016). However, the perceptual enhancement was only for the phonotactically simple word structures (CVC pairs of words contrasting in onset phonemes, or non-minimal pairs). Note that many of these words contained *phonemes* that the English-speaking listeners would have had no prior exposure to (e.g., /h/, /ɣ/, /zʰ/), yet, their discriminability improves in clear speech. We find that naive listeners perform just as well as native listeners in perceiving these words in clear speech, which does not provide additional support for the claim that clear speech is strictly “native listener-oriented” (Bradlow & Bent, 2002). Rather, non-native listeners can harness clear speech effects under certain conditions to achieve perceptual patterns similar to native listeners (Baese-Berk et al., 2020). One possible explanation for the high performance of naive listeners in these conditions is that our stimuli were elicited to a non-native Tashlhiyt listener. Indeed, prior work has shown that non-native-directed speech is better perceived by both native and non-native listeners (Kangatharan et al., 2022). While the present study looked only at non-native directed speech, an avenue for future work is to investigate how clear speech patterns vary across different types of listeners and with different types of instructions.

## A. Limitations and future directions

In addition to the limitations and future directions mentioned above, there are other limitations to the present study that open other avenues for future work in this line of research. For one, much work in phonetics has examined cross-linguistic differences in the phonetic implementation of phonological systems across languages of the world. Future work looking at the clear speech benefit, or lack thereof, in the perception of diverse phonological contrasts across different languages and dialects is a ripe area to continue to explore how listener-oriented speech can inform our understanding of what makes some phonological patterns less common than others.

It is also worth noting that just because the clear speech elicited in the current study did not enhance the perceptual sensitivity of contrasts involving the non-rising clusters, does not mean that a listener-oriented speech mode would not boost discriminability of these words in other communicative contexts. Perhaps a different type of clear speech would lead to different enhancement patterns (e.g., infant- or foreigner-directed speech). Moreover, prior work has shown that the type of instruction given to speakers affects their speech patterns (e.g., Lam et al., 2012). Hence, the way the stimuli were elicited in the current study could have affected our

pattern of results. Also, although the stimuli in the present study were presented in noise, it is a higher SNR relative to other speech-in-noise experiments which often present items at lower SNRs (e.g., -3 and -6 in Pycha et al., 2022; -4 and -8 in Bradlow & Bent, 2002). Furthermore, the use of white noise as the environmental masker could have made the perception of fricative clusters particularly challenging. Future work exploring how other types of listener-directed speech (e.g., child- or hard-of-hearing-directed speech), speaker instructions, different listening situations (interactive vs. non-interactive, with feedback, etc.), and other types and levels of noise (e.g., multi-talker babble, pink noise) influence the perceptibility of typologically less common phonotactic patterns in Tashlhiyt and other languages is a ripe area for future work.

Another limitation of the present study is that the two listener groups differed in other characteristics, such as their age and gender balance, as well as native-speaker status. Future work can explore our hypotheses across different types of listener groups. Finally, while our post-hoc analysis comparing rising vs. non-rising sonority onset clusters grouped together clusters that vary on other phonological dimensions (manner of articulation, falling vs. plateauing sonority clusters), the question of how clear speech affects the discriminability for specific types of non-rising sonority clusters can be explored in future work, as well.

## B. Conclusion

The present study examined the perception of clear speech in Tashlhiyt Berber, and we demonstrated that the perceptual benefit of clear speech varies across typologically more common and rarer phonotactic patterns. This study contributes to cross-linguistic investigations of clear speech, shedding light on the question of what is universal vs. language-specific about intelligibility-enhancing speaking modes. Our results also suggest that examining the perception of different speech styles can inform longstanding issues in phonological theory. Comparison of the production and perception of distinct speech styles across different languages, and across different listener types, has a rich tradition in linguistics (e.g., Lindblom, 1990; Bradlow & Bent, 2002; Smiljanić & Bradlow, 2009; Baese-Berk et al., 2020) and can be used to uncover the underlying phonetic complexity of different types of lexical contrasts (Kang & Guion, 2008; Zellou & Scarborough, 2019).

## References

- Baese-Berk, M. M., McLaughlin, D. J., & McGowan, K. B. (2020). "Perception of non-native speech," *Language and Linguistics Compass* **14**(7), e12375.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., and Grothendieck, G. (2014). "Package lme4: Linear mixed-effects models using Eigen and S4," *CRAN Repos.* **1**, 1–113.
- Bensoukas, K. (2001). *Stem Forms in the Nontemplatic Morphology of Berber*. Doctorat d'État dissertation, FLHS, Rabat.
- Berent, I., Balaban, E., Lennertz, T., & Vaknin-Nusbaum, V. (2010). "Phonological universals constrain the processing of nonspeech stimuli," *J. Exp. Psych.: Gen.* **139**(3), 418-435.
- Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). "Listeners' knowledge of phonological universals: Evidence from nasal clusters," *Phonology* **26**(1), 75–108.

- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). "What we know about what we have never heard: Evidence from perceptual illusions," *Cognition* **104**(3), 591–630.
- Best, C. T., McRoberts, G. W., & Goodell, E. (2001). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," *J. Acoust. Soc. Am.* **109**(2), 775-794.
- Blevins, J. (1995). "The syllable in phonological theory," in *The Handbook of Phonological Theory*, edited by John Goldsmith (Oxford, Blackwell), pp. 206-244.
- Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
- Bohn, O. S. (2017). "Cross-Language and Second Language Speech Perception," in *The Handbook of Psycholinguistics*, edited by Eva M. Fernández and Helen Smith Cairns (Wiley, Oxford), pp. 213–239.
- Boukous, A. (1987). *Phonotactique et Domaines Prosodiques en Berbère*. Doctorat d'Etat dissertation, Université Paris VIII- Vincennes, Saint-Denis.
- Boukous, A. (2009). *Phonologie de l'Amazighe*. Rabat: IRCAM publications.
- Bradlow, A. R. (2002). "Confluent talker-and listener-oriented forces in clear speech production," *Lab. Phon.* **7**, 241-273.
- Bradlow, A. R., & Bent, T. (2002). "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.* **112**(1), 272–284.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). "Speaking Clearly for Children with Learning Disabilities: Sentence Perception in Noise," *J. Speech, Lang., and Hear. Res.* **46**, 80–97.
- Bradlow, A. R., & Alexander, J. A. (2007). "Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners," *J. Acoust. Soc. Am.* **121**(4), 2339-2349.
- Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). "Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations," *J. Mem. and Lang.* **89**, 68–86.
- Clements, G. N. (1990). "The role of the sonority cycle in core syllabification," *Lab. Phon.* **1**, 283–333.
- Clements, G. N. & Keyser, S. J. (1983). *CV phonology*. Linguistic Inquiry Monographs, **9**. Cambridge: MIT Press.
- Clopper, C. G. (2013). "Modeling multi-level factors using linear mixed effects," *Proc. of Meet. on Acoust. ICA 2013* **19**(1), 1-6.
- Cohn, M., Ferenc Segedin, B., & Zellou, G. (2022). "Acoustic-phonetic properties of Siri-and human-directed speech," *J. Phon.* **90**, 101123.
- Cohn, M., & Zellou, G. (2021). "Prosodic differences in human-and Alexa-directed speech, but similar local intelligibility adjustments," *Frontiers in Comm.*
- Davidson, L. (2011). "Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts," *J. Exp. Psych.: Human Perc. and Perf.* **37**(1), 270–282.
- Davidson, L., & Shaw, J. A. (2012). "Sources of illusion in consonant cluster perception," *J. Phon.* **40**(2), 234–248.
- Dell, F., & Elmedlaoui, M. (2002). *Syllables in Tashlhiyt Berber and in Moroccan Arabic* (Vol. 2). Springer Science & Business Media.



- Dilley, L. C., Millett, A. L., McAuley, J. D., & Bergeson, T. R. (2014). "Phonetic variation in consonants in infant-directed and adult-directed speech: the case of regressive place assimilation in word-final alveolar stops," *J. of Child Lang.* **41**(1), 155–175.
- Flemming, E. (2005). "Speech perception and phonological contrast," in *The Handbook of Speech Perception*, edited by David B. Pisoni and Robert E. Remez (Blackwell, Oxford), pp. 156–181.
- Fowler, C. A., Levy, E. T., & Brown, J. M. (1997). "Reductions of spoken words in certain discourse contexts," *J. Mem. Lang.* **37**, 24–40.
- Hallé, P. A., Ridouane, R., & Best, C. T. (2016). "Differential difficulties in perception of Tashlhiyt Berber consonant quantity contrasts by native Tashlhiyt listeners vs. Berber-naïve French listeners," *Frontiers in Psych.* **7**, 209.
- Hargus Ferguson, S. (2004). "Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners," *J. Acoust. Soc. Am.* **116**(4), 2365–2373.
- Hayes, B., & Steriade, D. (2004). "Introduction: the phonetic bases of phonological Markedness," in *Phonetically Based Phonology*, edited by Bruce Hayes, Robert M. Kirchner, and Donce Steriade (Cambridge, Cambridge University Press), pp. 1–33.
- Henderson, J. B., & Repp, B. H. (1982). "Is a stop consonant released when followed by another stop consonant?" *Phonetica* **39**(2-3), 71–82.
- Henke, E., Kaisse, E. M., & Wright, R. (2012). "Is the sonority sequencing principle an epiphenomenon," in *The Sonority Controversy*, edited by Steve Parker (Berlin, De Gruyter), pp. 65–100.
- Hermes, A., Mücke, D., Auris, B., & Ridouane, R. (2018). "Structural Effects on Properties of Consonantal Gestures in Tashlhiyt," In *Proceedings of Interspeech*, 197–201.
- Jakobson, Roman (1962). *Selected writings* (vol. I). The Hague: Mouton.
- Jebbour, A. (1996). *Morphologie et Contraintes Prosodiques en Berbère (Tachelhit de Tiznit)-Analyse Linguistique et Traitement Automatique*. Doctorat d'Etat thesis, Mohammed V University, Faculty of Letters and Human Sciences, Rabat.
- Jebbour, A. (1999). "Syllable Weight and Syllable Nuclei in Tachelhit Berber of Tiznit," *Cahiers de Grammaire 24: Phonologie: théorie et variation*, 95–116.
- Kabak, B., & Maniwa, K. (2007). "L2 perception of English fricatives in clear and conversational speech: The role of phonemic, phonetic, and acoustic factors," In *Proceedings of ICPhS* (Vol. 2007).
- Kang, K. H., & Guion, S. G. (2008). "Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies," *J. Acoust. Soc. Am.* **124**(6), 3909–3917.
- Kangatharan, J., Uther, M., & Gobet, F. (2022). "The effect of hyperarticulation on speech comprehension under adverse listening conditions," *Psych. Res.* **86**(5), 1535–1546.
- Krause, J. C., & Braida, L. D. (2004). "Acoustic properties of naturally produced clear speech at normal speaking rates," *J. Acoust. Soc. Am.* **115**(1), 362–378.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). "lmerTest package: tests in linear mixed effects models," *J. Stat. Software* **82**, 1–26.
- Lam, J., Tjaden, K., & Wilding, G. (2012). "Acoustics of clear speech: effect of instruction," *JSLHR* **55**(6), 1807–1821.
- Lenth, R., Herve, M., Love, J., Riebl, H., & Singman, H. (2021). "Package 'emmeans'," [Software Package] (<https://github.com/rvnlenth/emmeans>).
- Leung, K. K., Jongman, A., Wang, Y., & Sereno, J. A. (2016). "Acoustic characteristics of clearly spoken English tense and lax vowels," *J. Acoust. Soc. Am.* **140**(1), 45–58.

- Li, Y., Zhang, G., Kang, H. Y., Liu, S., Han, D., & Fu, Q. J. (2011). Effects of speaking style on speech intelligibility for Mandarin-speaking cochlear implant users. *J. Acoust. Soc. Am.* **129**(6), EL242–EL247.
- Liljencrants, J. & Lindblom, B. (1972). “Numerical simulation of vowel quality systems: the role of perceptual contrast,” *Language* 48, 839–862.
- Lindblom, B. (1990). “Explaining phonetic variation: A sketch of the H&H theory,” In *Speech production and speech modelling* (Springer, Dordrecht), pp. 403–439.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.
- Marin, S., & Pouplier, M. (2010). “Temporal organization of complex onsets and codas in American English: Testing the predictions of a gestural coupling model,” *Motor Control* **14**(3), 380–407.
- Muller, J. S. (2001). *The phonology and phonetics of word-initial geminates*. The Ohio State University.
- Peperkamp, S. (2007). “Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin,” *Cognition* **104**(3), 631–637.
- Peperkamp, S., & Dupoux, E. (2002). “A typological study of stress ‘deafness’,” *Lab. Phon.* **7**, 203–240.
- Picheny, M. A., Durlach, N. I., & Braidă, L. D. (1985). “Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech,” *J. Speech, Lang., and Hear. Res.* **28**(1), 96–103.
- Picheny, M. A., Durlach, N. I., & Braidă, L. D. (1986). “Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech,” *J. Speech, Lang., and Hear. Res.* **29**(4), 434–446.
- Pouplier, M., Marin, S., Hoole, P., & Kochetov, A. (2017). “Speech rate effects in Russian onset clusters are modulated by frequency, but not auditory cue robustness,” *J. Phon.* **64**, 108–126.
- Pouplier, M., Pastätter, M., Hoole, P., Marin, S., Chitoran, I., Lentz, T. O., & Kochetov, A. (2022). “Language and cluster-specific effects in the timing of onset consonant sequences in seven languages,” *J. Phon.* **93**, 101153.
- Prince, A., & Smolensky, P. (2004). *Optimality Theory: Constraint interaction in generative grammar*. John Wiley & Sons.
- Pycha, A., Cohn, M., & Zellou, G. (2002). “Face-masked speech intelligibility: the influence of speaking style, visual information, and background noise,” *Frontiers in Comm.*, 102.
- R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. 2012.
- Rice, K. (2007). “Markedness in phonology,” In *The Cambridge Handbook of Phonology*, edited by Paul de Lacy (Cambridge, Cambridge University Press), pp. 79–97.
- Ridouane, R. (2007). “Gemination in Tashlhiyt Berber: an acoustic and articulatory study,” *J. Inter. Phon. Ass.* **37**(2), 119–142.
- Ridouane, R. (2008). “Syllables without vowels: phonetic and phonological evidence from Tashlhiyt Berber,” *Phonology* **25**, 321–359.
- Ridouane, R. (2016). “Leading issues in Tashlhiyt phonology,” *Lang. and Ling. Compass* **10**, 644–660.
- Ridouane, R. (2022). *Phonetics and phonology of Tashlhiyt geminates: An overview*. <https://halshs.archives-ouvertes.fr/halshs-03511107>.

- Ridouane, R., Hermes, A., & Hallé, P. (2014). "Tashlhiyt's ban of complex syllable onsets: Phonetic and perceptual evidence," *Lang. Typ. and Univ.* **67**(1), 7–20.
- Ridouane, R., & Hallé, P. A. (2017). "Word-initial geminates," In *The Phonetics and Phonology of Geminate Consonants*, edited by Haruo Kubozono (Oxford, Oxford University Press), pp. 66–84.
- Scarborough, R., & Zellou, G. (2013). "Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception" *J. Acoust. Soc. Am.* **134**(5), 3793–3807.
- Schertz, J. (2013). "Exaggeration of featural contrasts in clarifications of misheard speech in English," *J. Phon.* **41**(3-4), 249–263.
- Selkirk, E. (1984). "On the major class features and syllable theory," in *Language sound structure*, edited by Mark Aronoff and Richard T. Oehrle (Cambridge, MIT Press), pp. 337–383.
- Seyfarth, S., Buz, E., & Jaeger, T. F. (2016). "Dynamic hyperarticulation of coda voicing contrasts," *J. Acoust. Soc. Am.* **139**(2), EL31–EL37.
- Smiljanić, R., & Bradlow, A. R. (2005). "Production and perception of clear speech in Croatian and English," *J. Acoust. Soc. Am.* **118**(3), 1677–1688.
- Smiljanic, R., & Bradlow, A. R. (2008). "Stability of temporal contrasts across speaking styles in English and Croatian," *J. Phon.* **36**(1), 91–113.
- Smiljanić, R., & Bradlow, A. R. (2009). "Speaking and hearing clearly: Talker and listener factors in speaking style changes," *Lang. and Ling. Comp.* **3**(1), 236–264.
- Smiljanić, R., & Bradlow, A. R. (2011). "Bidirectional clear speech perception benefit for native and high-proficiency non-native talkers and listeners: Intelligibility and accentedness," *J. Acoust. Soc. Am.* **130**(6), 4020–4031.
- Steriade, D. (1982). *Greek prosodies and the nature of syllabification*. Unpublished doctoral dissertation. MIT, Cambridge.
- Tremblay, A., Namjoshi, J., Spinelli, E., Broersma, M., Cho, T., Kim, S., Martínez-García, M.T. & Connell, K. (2017). "Experience with a second language affects the use of fundamental frequency in speech segmentation," *PloS one*, **12**(7), p.e0181709.
- Trubetzkoy, N. S. (1939). "Gedanken über das Indogermanenproblem," *Acta Linguistica* **1**(1), 81–89.
- Tupper, P., Leung, K. W., Wang, Y., Jongman, A., & Sereno, J. A. (2021). "The contrast between clear and plain speaking style for Mandarin tones," *J. Acoust. Soc. Am.* **150**(6), 4464–4473.
- Uchanski, R. M. (2005). "Clear Speech," in *The Handbook of Speech Perception*, edited by David B. Pisoni and Robert E. Remez (Blackwell, Oxford), pp. 207–235.
- Wright, R. A. (1996). *Consonant clusters and cue preservation in Tsou*. University of California, Los Angeles.
- Wright, R. (2004). "A review of perceptual cues and cue robustness," in *Phonetically Based Phonology*, edited by Bruce Hayes, Robert M. Kirchner, and Donce Steriade (Cambridge, Cambridge University Press), pp. 34–57.
- Zellou, G., & Scarborough, R. (2015). "Lexically conditioned phonetic variation in motherese: age-of-acquisition and other word-specific factors in infant-and adult-directed speech," *Lab. Phon.* **6**(3-4), 305–336.
- Zellou, G., & Scarborough, R. (2019). "Neighborhood-conditioned phonetic enhancement of an allophonic vowel split," *J. Acoust. Soc. Am.* **145**(6), 3675–3685.

