On the perceptual reliability of articulation without acoustics.

Rachid Ridouane, Pierre Hallé

ON THE PERCEPTUAL RELIABILITY OF ARTICULATION WITHOUT ACOUSTICS

Rachid Ridouane & Pierre Hallé

Laboratoire de Phonétique et Phonologie (CNRS/Sorbonne-Nouvelle), France
rachid.ridouane@univ-paris3.fr; pierre.halle@univ-paris5.fr

ABSTRACT

This study explores how articulation recovery might be accomplished in the absence of clear acoustic output consequences. Based on perception data from Tashlhiyt Berber utterance-initial voiceless singleton and geminate stops (e.g. tut vs. ttut), we show that auditory information alone is not sufficient for native listeners to elicit the standard perception performance expected from native listeners on a native contrast. Implications of the results on the general issue of the nature of speech targets are briefly discussed.

Keywords: initial voiceless geminate stops: production, perception, representation

1. INTRODUCTION

Quantity contrasts with consonants are common in the languages of the world, but occur mainly in word-medial position. Word-initial geminates are typologically rare [3]. Even less frequent is the occurrence in the languages of the world of word-initial voiceless singleton/geminate stops, such as /t/-/tt/. To our knowledge this has been phonetically documented in four languages: Pattani Malay [1, 2], Cypriot Greek [7, 10], Thurgovian Swiss German [5, 6], and Tashlhiyt Berber [8]. Cross-linguistically, the main correlate of geminated stops is a longer closure duration [9]. When voiceless, stop closure translates acoustically into a silent gap whose duration cannot be perceived in utterance-initial position, since nothing is heard until the release. In this context, voiceless geminates may thus only slightly acoustically differ from their singleton counterparts.

The /t/-/tt/ contrast in word-initial position raises a puzzling issue in both production and perception: do speakers produce the length contrast between these segments, even in the absence of acoustic information? Are there any additional acoustic attributes enhancing the distinction between singletons and geminates in this position? Are native listeners sensitive to these attributes, if any?

Contradictory prior results have been reported in literature. In Pattani Malay [1, 2], significant acoustic differences were found between initial singletons and geminates phrase-medially in terms of closure duration. In the absence of this information (i.e. in utterance-initial position), listeners were still capable of accurately recovering the lexical contrast. Their correct identification was based on combined secondary cues including relative amplitude, the fundamental frequency of the following vowel, and the relative weights of the first and second syllables. In Cypriot Greek, closure duration as well as VOT duration were found to be consistent acoustic cues distinguishing the two series, with geminates displaying longer closure and VOT [10]. In utterance-initial position, Cypriot listeners also reliably recover the contrast, their judgements being based mainly on VOT differences [7]. For Thurgovian, however, a preliminary perceptual study failed to find identification performance above chance level for the contrast in utterance-initial position [5]. However, the distinction, as estimated by tongue-palate contact, is very clear in terms of articulatory gestures, contact duration being more than twice longer in geminates than in singleton stops [6].

These conflicting results might be explained: in Pattani Malay, the distinction seems to entail a difference in accentuation. Abramson [2] speculates it will undergo transphonemisation, switching from a segmental to an accentual pattern distinction. In Cypriot Greek, the contrast between singletons and geminates is also a laryngeal contrast between unaspirated and aspirated stops, respectively [10]. In Thurgovian, minimal pairs with this distinction are very infrequent and may be treated as homophones. The situation with Tashlhiyt, the language investigated in this study, is different: the distinction does not correlate with accentual or laryngeal acoustic differences and is highly productive.

1.1. Initial voiceless geminate stops in Tashlhiyt

Each consonant in Tashlhiyt has a geminate counterpart at the lexical level. For voiceless stops, this distinction is attested in initial and final positions in addition to the typologically more common medial position:

(1) [tut] "she hit" [ttut] "forget him"
[juti] "he exceeds" [jutti] "he hit him"
[fit] "give it, masc." [fitt] "give it, fem."

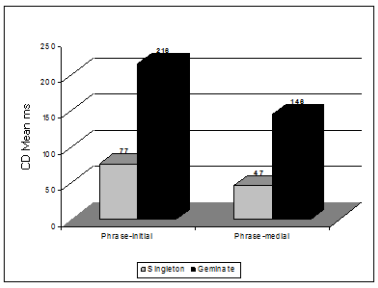
Also, as shown in (2), certain verbs form their imperfective by prefixing a geminated /tt/ to the basic stem, allowing for the contrast with corresponding perfective verbs, 3f. These words most frequently occur in spontaneous speech at the onset of a dialogue turn.

(2)

Stem	impf	perf, 3f	
[asi]	[ttasi]	[tasi]	"to take"
[ara]	[ttara]	[tara]	"to write"

Ridouane [8] described the Tashlhiyt geminate versus singleton differences in production for stops and fricatives in different prosodic positions, providing both articulatory and acoustic measurements. Initial singleton and geminate voiceless stops did not differ reliably on any classic acoustic measurement (VOT, RMS amplitude, F0 perturbations). They did differ substantially, however, for closure duration, as estimated from electropalatographic measurements (figure 1). In other words, the articulatory target is achieved (i.e. longer contact duration) although the corresponding acoustic consequence cannot be recovered from the signal.

Figure 1: Mean contact durations (CD) in *ms* for word-initial singleton and geminate voiceless stops in two prosodic contexts (EPG data from [8]).



Given that closure duration of voiceless stops leaves no audible cue phrase-initially, the question raises as to whether native listeners can still distinguish e.g. *tut* from *ttut* in utterance-initial position. We hypothesize that while they must be subtle indeed since they escape classic acoustic investigation, cues to underlying articulation are still present and native listeners are sensitive to them. As many recent studies have shown, listeners can exploit very subtle acoustic-phonetic cues to almost fully retrieve intended meaning in cases of potential ambiguity (see [10], for a review). Yet, native listeners may not be able to distinguish *tut* from *ttut* in the case of auditory only presentation. This would imply that auditory information is not sufficient to recover underlying articulation, and that additional information (e.g. visible information possibly associated with underlying articulation) must be available for the intended phonemic category to be heard.

While the first finding could be consistent with either articulatory or auditory accounts of experiential effects on speech perception, the second would support the idea that articulatory gestures rather than acoustic cues per se are the basis of phonological contrasts. A related issue concerns the mental (phonological) representation of geminates: prosodic structure (quantity: two X slots) or segmental feature specification (quality: [tense] feature)? A [+tense] representation would predict good perception performance, whereas a two X slot representation would predict that in the absence of temporal information the contrast will not be reliably recovered. The following perceptual experiments put these different possibilities at test.

2. PERCEPTUAL EXPERIMENTS

Two experiments have been conducted to see whether native listeners of Tashlhiyt can recover the contrast between voiceless singleton stops and geminates phrase-initially. The experiments consisted of a categorial AXB discrimination test and a forced-choice identification test.

2.1. Method

2.1.1. Participants

Twenty volunteers from the University Ibnou Zohr in Agadir (Morocco) took part in this experiment (aged 19 to 37, mean 26.1, SD 4.9, 6 females and 14 males). All were native speakers of Tashlhiyt and none reported any hearing deficit or any kind of language impairment.

2.1.2. Stimuli and design

A Tashlhiyt native speaker was recorded as he produced the eight minimal-pair words with initial singleton/geminate contrast shown in (3). Four

repetitions of each item were retained as experimental stimuli. Minimal pairs in Set 1 were recorded in three sentential contexts aimed at manipulating the perceptual salience of the singleton-geminate contrast: (1) embedded in a neutral carrier sentence (*inna --- jat twalt*, "he said --- once"). (2) in citation form (i.e., in isolation) where the word is equivalent to an entire phrase. (3) "contrasting focus", in which one minimal pair word is stressed against the other (e.g.: ***ttili*** *as nniʁ maʃi tili* "I said ***ttili*** and not *tili*"). The minimal pairs in set 2 and set 3 contrast initial singletons and geminates for voiced stops and fricatives. These control pairs allowed to compare the perceptual impact of clear acoustic closure-duration differences in voicing and frication against that of the minimal acoustic traces offered by initial voiceless stops. For these consonants, minimal pairs were recorded in the "isolated" context only. Word-stimuli were extracted from their context for presentation in the discrimination and identification experiments.

(3) Stimuli used for the perceptual experiments:

Set1: words contrasting voiceless stops:

tut 'she hit'	vs.	ttut 'forget him'
tili 'ewe'	vs.	ttili 'have'
kijji 'you'	vs.	kkijji 'take a road'
ks 'feed on'	vs.	kks 'take off'

Set2: words contrasting voiced stops:

gar 'bad'	vs.	ggar 'be last'
diʁ 'with us'	vs.	ddiʁ 'I went'

Set3: words contrasting voiceless fricatives:

fit 'give it'	vs.	ffit 'pour it'
sir 'go'	vs.	ssir 'lace'

2.1.3. Procedure

Participants were tested individually in a quiet room and received the speech stimuli through professional quality covering headphones. On each AXB trial, participants were presented with three stimuli and had to indicate whether second item X matched better the first or the third stimulus, by pressing the response key labeled '1' or '3'. The inter-stimulus (offset to onset), inter-trial, and inter-block intervals were set to 1 s, 4 s, and 9 s, respectively. Response times were measured from the onset of the X stimulus. For the identification test, subjects were asked to identify the correct item produced by choosing one of the two written response alternatives on the left and right side of computer screen. The subjects had to choose by pressing the response key labeled either 'left' or 'right' on the keyboard. The identification and discrimination experiments were run using the DMDX software [7]. Each test was preceded by training trials on contrasts different from those used in the test trials (e.g. *kijji-gijji*, *jutid-juttid*).

2.2. Results

The results of the identification and categorization tests are displayed in figures 2, 3 and 4. Figure 2 shows the overall identification accuracy for the three types of initial consonants. Of all the contrasted types, participants encountered difficulty only with the utterance-initial voiceless stops. They perform poorly, just above chance level (61.8%), on identification of initial voiceless stops, but near ceiling for initial voiced stops 96.7% and fricatives 95.2%. The differences between voiceless stops on the one hand and voiced stops and fricatives on the other hand are significant at p <.0001. Regarding reaction time, listeners' performances also differed depending on the contrast type, with the longest RT's for voiceless stops (figure 3). The fact that reaction times are slower for T's provides additional evidence the perceptual distance between stimuli is smaller. Again the difference between T's on the one and D's and F's on the other hand are significant at p <.0001.

Figure 2: Correct identification rates of the singleton-geminate contrast for initial voiceless stops (T's), voiced stops (D's), and voiceless fricatives (F's). The rate for T's corresponds to the mean of the three sentential contexts (standard errors as positive and negative error bars).

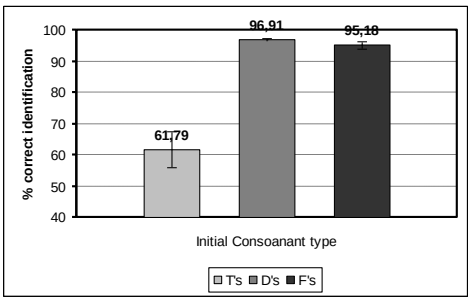


Figure 3: Response time data for the three types of initial consonants. Symbols as in figure 2.

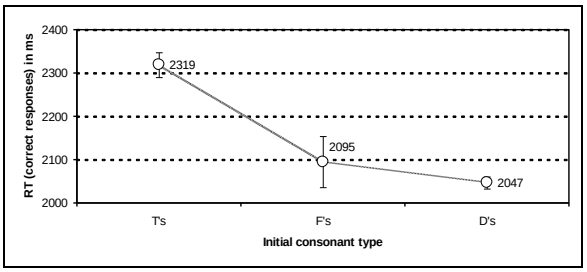
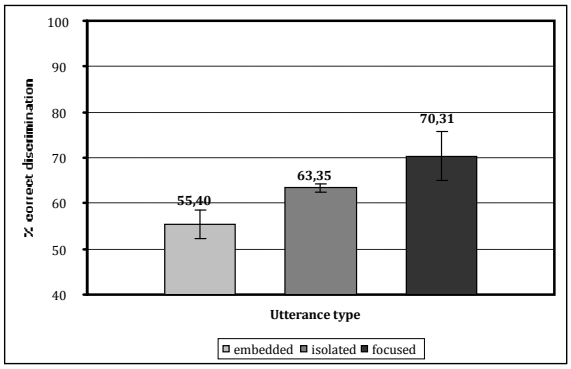


Figure 4 shows that listeners' performance is modulated by the context in which minimal pairs are produced. The highest performance obtains for the "focused" and in a lesser degree for the 'isolated'' condition, suggesting that more reliable acoustic cues to underlying articulation are used in these conditions (the difference between 'isolated' and 'focused' is not significant, while the other pairwise comparisons are significant at $p<.05$). Whatever these additional cues are, however, they only help a little. The performance of native speakers is still poorer compared to the reliable identification and categorization of voiced stops and fricatives.

Figure 4: Correct categorization rates of the singleton-geminate contrast for initial voiceless stops in three different utterance types.



3. GENERAL DISCUSSION

This study's motivation was to explore how articulation recovery might be accomplished in the absence of clear acoustic output consequences. Based on the singleton/geminate contrast for voiceless stops in phrase-initial position, the results obtained show that available bottom-up information is not sufficient to elicit the standard perception performance expected from native listeners on a native contrast. Even when the acoustic cues are enhanced in specific prosodic contexts (e.g. under focus) these cues are not enough. Clearly, perception cannot be based solely on auditory-acoustic representations. Does only top-down information help recover intended gemination or non-gemination? Such a diagnostic would predict contrast neutralization in the near future. But the contrast is alive (exploited both by the lexicon and morphology) and is systematically maintained at the articulatory level. The contrast is not neutralized presumably because it is not limited to voiceless stops as it concerns other consonants with clearly audible acoustic closure-duration differences in voicing and frication. In addition, native listeners generally are not aware of increased difficulty with the /t/-/tt/ contrasts word-initially, suggesting they routinely recover underlying articulation rather than comparing an auditory-acoustic input with stored auditory representations. The fact that a phonological or a morphological contrast can be systematically encoded even in the absence of acoustic/auditory consequences implies that, at least in some cases, the targets of speech production can be articulatory. Related to this is the question of the phonological representation of geminates. Our data support a 2 X-slot representation (where X = timing unit). This structural representation is reflected in the observed articulatory differences in consonant duration. In the absence of this temporal information the contrast can no longer be appropriately recovered.

4. REFERENCES

[1] Abramson, A.S. 1986. The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association* 16, 8-16.

[2] Abramson, A.S. 1999. Fundamental frequency as a cue to word-initial consonant length: Pattani Malay. *Proc. of the 14th ICPhS* Berkeley, University of California, 591-594.

[3] Davis, S. 1999. On the representation of initial geminates. *Phonology* 16, 93-104.

[4] Forster, K., Forster, J. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35, 116-124.

[5] Kraehenmann, A. 2001. Swiss German stops: Geminates all over the word. *Phonology* 18(1), 109-145.

[6] Kraehenmann, A., Lahiri, A. 2008. Duration differences in the articulation and acoustics of Swiss German word-initial geminate and singleton stops. *J. Acoust. Soc. Am.* 123(6), 4446-4455.

[7] Muller, S.J. 2003. The production and perception of word initial geminates in Cypriot Greek. *Proceedings of the 15th International Congress of Phonetic Sciences*, *ICPhS 2003* Barcelona, 1867-1870.

[8] Ridouane, R. 2007. Gemination in Tashlhiyt Berber: an acoustic and articulatory study. *Journal of the International Phonetic Association* 37(2), 119-142.

[9] Ridouane, R. 2010. Geminates at the junction of phonetics and phonology. In Fougeron, C., Kuhnert, B., Delais-Roussarie, E. (eds.), *Papers in Laboratory Phonology X*, 61-90.

[10] Spinelli, E., Welby, P., Schaegis, A.-L. 2007. Fine-grained access to targets and competitors in phonemically identical spoken sequences: The case of French elision. *Language and Cognitive Processes* 22, 828-859.

[11] Tserdanelis, G., Arvaniti, A. 2001. The acoustic characteristics of geminate consonants in Cypriot Greek. *Proceedings of the 4th International Conference on Greek Linguistics*, 29-36.