TESTING THE PERCEPTUAL BASIS OF LARYNGEAL METATHESIS AND RARITY OF PREASPIRATED STOPS

Madeline Gilbert

Laboratoire de Phonétique et Phonologie (UMR 7018, CNRS/Sorbonne Nouvelle, Paris)
madeline-besse.gilbert@sorbonne-nouvelle.fr

ABSTRACT

Pre-stop coda [h] is often argued to be perceptually weaker than post-stop [h]. This perceptual property of [h] is often used to explain why laryngeal metathesis moves [h] from before a stop to after a stop (HC → CH), and why preaspirated stops are rarer than (post)aspirated stops. This study tests Arabic, English, and French-speaking listeners' perception of [h] in h-stop (HC) and stop-h (CH) sequences using an ABX perception task. The status and phonotactics of /h/ and aspirated stops differ in these languages. Results provide no evidence that [h] is easier to perceive after a stop than before a stop. For all groups, [h] was either equally perceptible in HC and CH sequences (Arabic, English), or more perceptible in HC sequences (French). Perception is largely driven by native language experience.

Keywords: cross-linguistic perception, metathesis, laryngeals, preaspiration

1. INTRODUCTION

The perceptibility of [h] is often argued to differ based on phonotactic position. According to Bladon [1], preconsonantal [h] is difficult to perceive because it has weak acoustic cues (low-intensity, dispersed energy), and because it has formants in the same regions as the preceding vowel, allowing the vowel to 'mask' it (see also [2]). In addition to weak acoustic cues, Silverman [2] and Kingston [3] argue that [h] is not salient before a stop because it occurs at the vowel offset, where air pressure is low. They argue that [h] is easier to perceive *following* a stop because stop releases are highly salient (high-intensity, high-volume, noisy), and because the auditory system is more sensitive to the onset of a stimulus (at a stop release, following silence) than to its offset (at the end of a vowel). The asymmetry in perceptibility between pre- and post-stop position is used to explain both laryngeal metathesis and the cross-linguistic rarity of preaspirated stops.

Metathesis is relatively uncommon, but laryngeals are frequent participants [4]. From one perspective, metathesis occurs to optimize the perception of one or both sounds [5, 6, 7].[1] Steriade [6] argues that regular, systematic metathesis improves perceptibility. This condition seems to hold for laryngeal metathesis, which typically operates in the direction [hC] → [Ch]. Most accounts assume that this directionality is because [h] is easier to perceive in the resulting [Ch] sequences than in the original [hC] sequences [8, 9, 4]. Cho [9] points out that the domain of laryngeal metathesis is often limited in ways that support this proposal. Metathesis in Cherokee and Korean is limited to contexts in which [h] would dock on a voiceless stop. In contrast, when [h] would dock on a sonorant, [h] fails to metathesize in Cherokee and deletes in Korean [9]. He argues that metathesis might be limited to occurring around voiceless stops because it is only these contexts where metathesizing improves perceptibility; docking [h] on a sonorant would not.

The supposed perceptual difficulty of preconsonantal [h] has also been used to explain the cross-linguistic rarity of preaspirated stops relative to postaspirated stops [2, 10]. Preaspirated stops are allophonically conditioned in ways that might optimize perception. For example, Icelandic aspirated stops are preaspirated word-medially and word-finally, but postaspirated word-initially [11, 12], where there would be no vowel preceding HC to provide the transitional cues necessary to perceive [h] [13]. Furthermore, the aspiration portion of preaspirated stops tends either to be lost, or to be strengthened by adding an oral constriction ([hp, ht, hk] → [fp, θt, xk]; [2, 10]), which might also suggest that preconsonantal glottal [h] is difficult to perceive on its own.

Thus, accounts of laryngeal metathesis and preaspiration rely on the hypothesis that [h] is less perceptible before a stop than on its release. Clayton [10] is one of the few studies that tests this hypothesis. In an AX task, he found that word-medial preaspiration ([VhCV]) was *not* universally more difficult than word-initial postaspiration ([#Ch]). However, the duration of

[h] was not controlled, and high accuracy on HC sequences could be due to longer [h] in these words.

I test the perceptual optimization account of laryngeal metathesis around voiceless stops, whereby [h] is supposedly easier to perceive in the CH sequences resulting from metathesis than in its original position (HC). An alternative hypothesis is that [h] perception is driven by native language experience. Listeners have been found to perceive sounds that exist in their native language better than sounds that do not (e.g., [14, 15, 16]), and Mielke [17] finds similar effects for [h]. To test language-specific perception, the current study was run on Arabic, English, and French-speaking listeners, whose native languages differ in the presence and status of HC and CH sequences.

2. METHOD

2.1. Materials and procedure

The stimuli were trisyllabic nonce words (penultimate stress) with medial consonants /p, t, k/ flanked by /a/. Word sets were built off three base items: *famapa*, *lamaka*, *lanata*. Each word set contained versions with no [h] (NoH; [VCV]), pre-posed [h] (HC; [VhCV]), and post-posed [h] (CH; [VChV]). These words were organized into HC/C and CH/C comparisons, which test listeners' accuracy in perceiving [h] in pre-stop and post-stop position, respectively (Table 1). Listeners heard all possible presentation orders (AAB, ABA, BAA, BAB), with an ISI of 500ms between words. These stimuli are part of a larger study with other medial consonants and an additional comparison type, which are not reported here.

Table 1: Sample trials for medial consonant /t/

Type	Order	Example
HC/C	HC-None-HC	lanahta-lanata-lanahta
CH/C	CH-None-CH	lanatha-lanata-lanatha

Stimuli were recorded by a male native speaker of Turkish (a trained linguist) in a sound booth, using a Shure KSM44 microphone with a pop filter. Turkish was used because it allows clusters with [h] and gives no group a native-listener advantage. Two productions of each word were used, so that X would not be the same physical token as A or B. Test nonce words were manipulated so that words in the same set (*lanata, lanahta, lanatha*) matched for duration of the preceding vowel, duration of [h], pitch contour, and intensity.

The ABX task was run on PCIbex. Listeners heard three words, and pushed a key indicating whether the final word (X) was the same as the first (A) or second (B). The full experiment consisted of 132 test trials and 72 control trials. Again, this paper reports only a subset of the results.

2.2. Participants

Listeners were 20 native speakers each of Arabic (Levantine), English (U.S.), and French (France), recruited on Prolific. Most English and French listeners resided in their respective countries, while most Arabic listeners resided in European countries. All listeners reported a single native language, except for two Arabic speakers who reported native multilingualism with English and Hebrew. Participants had intermediate proficiency or higher in other languages (English speakers: French, Japanese, Spanish; French speakers: English, German, and Spanish; Arabic speakers: English, Italian, French, German, Hebrew, Czech, Hungarian). I do not believe this knowledge substantially affected the results (Section 4). Four participants were excluded due to low accuracy on controls (1 Arabic, 2 English, 1 French).

2.3. Hypotheses

Under the perceptual optimization account, listeners should perceive [h] more accurately in CH than in HC sequences. In contrast, if perception is driven by native language experience, listeners should perceive [h] more accurately in contexts where it occurs in their native language (Table 2). Arabic listeners should accurately distinguish [h] in both HC and CH sequences, since Arabic has phonemic /h/ and is phonotactically permissive.[2] French lacks /h/ and aspirated stops; these listeners should have low accuracy on both HC and CH sequences. English has aspirated stops and phonemic /h/, but /h/ is not permitted as a coda. They should perceive [h] in CH better than in HC, and [h] in CH more accurately than French listeners.

Table 2: Properties of listeners' native languages

Language	HC	CH
Arabic	Yes - /h/	No
English	No	Yes - Asp. Stops
French	No	No

3. RESULTS

Linear mixed-effects models (*lme4* [18], in R) predicted Accuracy (Incorrect, Correct) from the

independent variables Language (Arabic, English, French), Condition (HC/C, CH/C), and their interaction. Models also had random intercepts of Participant and Word Set. There was a significant interaction between Language and Condition, so I report results from *emmeans* pairwise comparisons (p-values Tukey adjusted) [19].

The results suggest that [h] is easier to hear before a stop than after it. Figure 1 shows each group's accuracy by condition. Arabic and English listeners had higher accuracy in HC/C than in CH/C, but the difference between conditions was not statistically significant (Table 3). French listeners had significantly higher accuracy in HC/C than in CH/C. Figure 2 compares groups within each condition. In the CH/C condition, the groups all differ significantly in accuracy, in the following order: Arabic > English > French (Table 4a). In the HC/C condition, Arabic listeners' accuracy is higher than French listeners', but Arabic-English and English-French groups do not differ significantly (Table 4b).

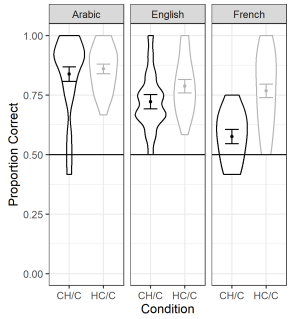

Figure 1: Group x Condition

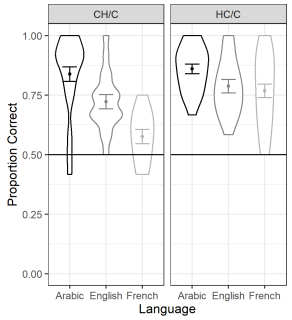

Figure 2: Condition x Group

Table 3: *emmeans* Comparisons within Group

		Est	SE	z.ratio	p
Ar.	HC/C-CH/C	0.17	0.26	0.66	0.51
Eng.	HC/C-CH/C	0.36	0.23	1.58	0.11
Fr.	HC/C-CH/C	0.95	0.21	4.56	0.00***

Table 4: *emmeans* Comparisons within Condition

		Est	SE	z.ratio	p
a. CH/C	Ar. - Eng.	0.70	0.25	2.76	0.02*
	Ar. - Fr.	1.41	0.24	5.78	0.00***
	Eng. - Fr.	0.71	0.22	3.16	0.00***
b. HC/C	Ar. - Eng.	0.52	0.27	1.92	0.13
	Ar. - Fr.	0.63	0.26	2.40	0.04*
	Eng. - Fr.	0.12	0.25	0.47	0.88

4. DISCUSSION

The results provide no evidence that [h] is easier to hear in CH than in HC sequences, and thus no evidence that HC → CH metathesis improves the perceptibility of [h] around voiceless stops. No listener group perceives [h] better in CH than in HC sequences. Instead, perception is driven by native language background. Arabic listeners have high accuracy in both HC/C and CH/C conditions, as expected since Arabic has /h/ and allows it as a coda. Although Arabic does not have CH sequences, /h/ occurs after other consonants, which may explain high accuracy in CH/C [17]. English listeners appear to benefit from experience with aspirated stops: their accuracy on CH/C is higher than French listeners'. However, this experience only puts accuracy in CH/C on par with HC/C. Given that English lacks coda [h], we might expect higher accuracy on CH/C. The fact that English has both /h/ and aspirated stops may have tempered the boost for CH/C. Beyond deciding if [h] was present, these listeners must decide if it is a realization of /h/ or aspiration on a stop. The CH sequences in my stimuli were also onsets to unstressed syllables, a position where English stops are not aspirated.

French listeners' accuracy was unexpectedly high, given that French lacks both HC and CH sequences. The Perceptual Assimilation Model [20] provides possible explanations. French listeners may have mapped (glottal) [h] to French /ʁ/, which occurs as a coda and in onset clusters. While /ʁ/ is typically uvular, there is extensive variation [21]. If [h] were mapped to /ʁ/, discrimination is expected to be good. Another possibility is that [h] was uncategorized, which could lead to good discrimination because listeners may attend to strictly acoustic properties [22].

Mielke [17] also tested the general and language-specific perception of [h]. Listeners heard nonce words recorded by a Turkish speaker and clicked where they heard <h> (if at all) on a word printed on the screen. When they clicked, <h> appeared in that location. He reports sensitivity to [h], not accuracy. His results differ from mine in ways that may be driven by task effects. In his experiment, Arabic listeners were more sensitive to [h] in HC sequences than English and French listeners; my Arabic listeners had higher accuracy than English and French listeners HC sequences, but Arabic and English listeners did not differ from each other. In Mielke's task, English-speaking listeners may not have clicked on orthographic <h> in HC stimuli,

since coda [h] does not exist in English. Another difference is that Mielke's English and French listeners had similar sensitivity in CH sequences; numerically, French listeners were actually *more* sensitive to [h] in CH sequences than English listeners. My study finds the opposite. Mielke suggests that English listeners may have categorized both the intervocalic stops and CH sequences in his stimuli as aspirated stops. In this case, they might not click on <h> since aspiration on stops is not represented orthographically in English. My ABX task removes the interference of orthography that may have affected English listeners' sensitivity to [h] in Mielke's study. A final difference is that Mielke's English and French listeners were more sensitive to [h] in CH than in HC sequences. In my study, English listeners' accuracy is similar in both conditions, and French listeners have lower accuracy CH than on HC sequences. This may be due to phonetic differences in the stimuli.

Recall that Clayton's [10] AX task also found that listeners' ability to discriminate [h] depended on their native language. Pre-stop [h] was not universally difficult. His Gaelic listeners, whose native language has pre- and postaspirated stops, had the highest accuracy on all sequences and perceived CH and HC sequences equally well. His Polish listeners, whose native language does not have coda [h] or aspirated stops, had higher accuracy on HC than on CH sequences, like my French listeners. Similar to my suggestion for French, Clayton proposes that they may have mapped coda [h] onto a native phonemic category. His English listeners were the only ones who had more difficulty with medial HC sequences than with initial CH sequences, which he argues is because only the CH sequences are native-like.

Why, then, are HC sequences rare as preaspirated stops, and 'resolved' by metathesis in the direction HC $\rightarrow$ CH? For preaspirated stops, Clayton argues that they are perceptible, but are difficult to innovate. For both preaspirated stops and metathesis, articulatory pressures, or nuanced perceptual factors not captured in either Clayton's or my study (e.g., listening conditions affecting [h] perception more in some positions, details of coarticulation), may also be at play.

Listeners' knowledge of other languages could have influenced the results, although participants mostly knew languages where the distribution of [h] is *less* permissive than in their native language (e.g., no French listeners knew Arabic or Hebrew, which have coda [h]). However, most Arabic and French-speaking listeners knew English, which could have provided them with experience with aspirated stops. This experience does not seem to have been overly helpful. Arabic listeners actually outperform English listeners on CH sequences, and French listeners have lower accuracy than English listeners. If knowledge of English helps French listeners, the effect is weak.

HC sequences of various types are accompanied by other cues in natural speech, including differences in the duration of the preceding vowel and stop closure (e.g., [23, 2, 24, 10]). This study isolated only the perceptibility of [h] in different positions. The other cues may enhance perceptual contrastiveness of the stops or [h], and future studies on preaspiration and laryngeal metathesis would benefit from studying the interaction between cues.

Another interesting avenue of research would be to analyze the differences in [h] perception in languages where the status of preaspiration differs. In some languages, preaspirated stops are usually considered to be a phonological category (e.g., Icelandic, see [25] for overview). In others, preaspiration is optional and variable (e.g., [23, 26, 27]). In these latter cases, [h] may not need to be as perceptible since it functions as a secondary cue or sociolinguistic variant.

5. CONCLUSION

This study tested the claim that [h] is more difficult to perceive in h-stop sequences than in stop-h sequences, an assumption often used to explain laryngeal metathesis and the rarity of preaspirated stops. A cross-linguistic ABX task provides no evidence for this assumption. Instead, [h] is either equally difficult to perceive in HC and CH sequences (Arabic and English speakers), or harder to perceive in CH sequences (French speakers). Instead of a general difficulty with preconsonantal [h], listeners' ability to perceive [h] depends on the distribution of [h] and aspirated stops in their native language.

6. ACKNOWLEDGMENTS

Thanks to Maria Gouskova, Lisa Davidson, Gillian Gallagher, Juliet Stanton, and Joseph Casillas for helpful feedback. Thanks also to Deniz Özyıldız for recording Turkish stimuli, and to the Department of Language and Literature at NTNU for use of their recording facilities. This work has also benefited from a government grant managed by the ANR under the 'Investissements d'Avenir' program (ANR-10-LABX-0083) (LABEX-EFL).

7. REFERENCES

[1] A. Bladon, "Phonetics for hearers," in *Language for hearers*, G. McGregor, Ed. Great Britain: Pergamon Press, 1986, pp. 1–24.

[2] D. Silverman, "On the rarity of pre-aspirated stops," *Journal of Linguistics*, vol. 39, no. 3, pp. 575–598, 2003.

[3] J. Kingston, "Articulatory binding," in *Papers in laboratory phonology 1: Between the grammar and physics of speech*, J. Kingston and M. E. Beckman, Eds. Cambridge: Cambridge University Press, 1990.

[4] T.-J. Yoon, "Laryngeal metathesis as perceptual optimization," *Studies in Phonetics, Phonology and Morphology*, vol. 18, no. 3, pp. 413–435, 2012.

[5] M. Grammont, *Traité phonétique*. Paris: Librarie Delagrave, 1933.

[6] D. Steriade, "Directional asymmetries in place assimilation: A perceptual account," in *Perception in Phonology*, E. Hume and K. Johnson, Eds. San Diego: Academic Press, 2001, pp. 291–250.

[7] E. Hume, "The indeterminacy/attestation model of metathesis," *Language*, vol. 80, no. 2, pp. 203–237, 2004.

[8] E. Flemming, "Laryngeal metathesis and vowel deletion in Cherokee," in *Cherokee papers from UCLA*, ser. UCLA Occasional Papers in Linguistics. Los Angeles: Department of Linguistics, UCLA, 1996, vol. 16, pp. 23–44.

[9] H. Cho, "Laryngeal feature mobility in Cherokee and Korean," *Language and Linguistics*, vol. 56, pp. 283–306, 2012.

[10] I. Clayton, "On the natural history of preaspirated stops," PhD Dissertation, University of North Carolina Chapel Hill, Chapel Hill, 2010.

[11] P. Ladefoged and I. Maddieson, *The sounds of the world's languages*. Cambridge, MA: Blackwell Publishers, 1996.

[12] C.-K. Suh, "Aspects of phonetics and phonology of Icelandic preaspiration," *The Phonology-Morphology Circle of Korea*, vol. 6, pp. 63–83, 2001.

[13] D. Steriade, "Phonetics in phonology: The case of laryngeal neutralization," 1997.

[14] J. F. Werker and R. C. Tees, "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," *Infant Behavior and Development*, vol. 7, no. 1, pp. 49–63, 1984.

[15] E. Dupoux, K. Kakehi, Y. Hirose, C. Pallier, and J. Mehler, "Epenthetic vowels in Japanese: A perceptual illusion?" *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, no. 6, pp. 1568–1578, 1999.

[16] L. Davidson and J. A. Shaw, "Sources of illusion in consonant cluster perception," *Journal of Phonetics*, vol. 40, no. 2, pp. 234–248, 2012.

[17] J. Mielke, "The interplay of speech perception and phonology: Experimental evidence from Turkish," *Phonetica*, vol. 60, pp. 208–229, 2003.

[18] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[19] R. Lenth, "emmeans: Estimated Marginal Means, aka Least-Squares Means," 2020. [Online]. Available: https://CRAN.R-project.org/package=emmeans

[20] C. T. Best and M. D. Tyler, "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language experience in second language teaching and learning: In honor of James Emil Flege*, ser. Language Learning & Language Teaching, O.-S. Bohn and M. J. Munro, Eds. Amsterdam/Philadelphia: John Benjamins Publishing Company, 2007, vol. 17, pp. 13–34.

[21] E. R. Webb, "Minimalism and French /R/: Phonological representations in phonetically based phonology," *Journal of French Language Studies*, vol. 19, no. 1, pp. 87–115, 2009.

[22] C. T. Best, G. W. McRoberts, and N. M. Sithole, "Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 14, no. 3, pp. 345–360, 1988.

[23] P. Helgason, "Preaspiration in the Nordic Languages," PhD Dissertation, Stockholm University, Stockholm, Sweden, 2002.

[24] P. O'Neill, "S-aspiration and occlusives in Andalusian Spanish: Phonetics or phonology?" *Oxford University Working Papers in Linguistics, Philology & Phonetics*, vol. 12, pp. 73–85, 2009.

[25] L. Heimisdóttir, "The phonology of aspiration in Icelandic: A gesture-based approach," PhD Dissertation, Cornell University, Ithaca, NY, 2015.

[26] M. Stevens and U. Reubold, "Pre-aspiration, quantity, and sound change," *Laboratory Phonology*, vol. 5, no. 4, pp. 455–488, 2014.

[27] M. Hejná, "Pre-aspiration in Welsh English: A case study of Aberystwyth," PhD Dissertation, University of Manchester, Manchester, 2015.

[28] J. Blevins and A. Garrett, "The evolution of metathesis," in *Phonetically based phonology*, B. Hayes, D. Steriade, and R. Kirchner, Eds. New York: Cambridge University Press, 2004, pp. 117–156.

[1] Alternatively, metathesis may be phonetically natural, but not functionally motivated [28].

[2] Arabic does not have /p/. I included words with /p/ for Arabic-speaking listeners because (a) all of my Arabic-speaking listeners had exposure to English, which does have /p/; (b) their accuracy on /p/ was parallel to English and French listeners'. All listener groups had lower accuracy on /p/ than on /t, k/, but Arabic listeners' accuracy was higher than English and French listeners on all three consonants. In a model including Consonant (/p, t, k/), no interaction terms were significant.