Hearing Hebrew Pharyngeals:
Experimental evidence for a covert phonemic distinction

Si Berrebi, Noa Bassel, and Roey J. Gafter*

1 Introduction

The pronunciation of pharyngeals is one of the key instances of phonological variation in Modern Hebrew. Whereas some speakers maintain the historical distinction between pharyngeal and non-pharyngeal consonants, the majority of speakers have merged these segments and produce only the non-pharyngeal variants. This has long been observed for production (Davis 1984; Lefkowitz 2004; Gafter 2016, *inter alia*), and may lead to the intuitive conclusion that the majority of Hebrew speakers also lack a representation of pharyngeals in their phonological system, as they do not produce these sounds. However, recent work that combines insights from sociolinguistics and psycholinguistics demonstrates that dialect perception does not always align with dialect production. For example, Sumner and Samuel (2009) demonstrate that rhotic speakers from the New York City area that grew up exposed to the non-rhotic New York City variety are primed by non-rhotic stimuli in a fashion similar to non-rhotic speakers, as opposed to rhotic speakers without such exposure. Similarly, exposure to pharyngeals may affect the perception of Hebrew speakers to the extent that speakers who do not produce them maintain some mental distinction between these categories.

We focus on one of the two pharyngeal consonants in Modern Hebrew, the voiceless pharyngeal fricative [ħ]. For speakers who produce it, it contrasts with a dorsal fricative, [χ], while most younger speakers only produce the latter variant.[1] Since the [ħ]-[χ] merger is an advanced change in progress in many Hebrew-speaking communities, it is quite common for parents who produce [ħ] to raise children who merge [ħ] and [χ] in production. Therefore, we distinguish not just between merged and non-merged speakers but between *three* groups of speakers (see Table 1): Pharyngeal Speakers, who produce both variants in their speech; Pharyngeal Listeners, who were exposed to pharyngeals in their parents' speech but do not produce them themselves; and Merged Listeners, who were exposed to and use only the merged variant [χ]. Pharyngeal Speakers differ from the other two groups at the level of *production*, while Merged and Pharyngeal Listeners produce the same merged realization and contrast each other at the level of the primary input received during childhood. In this study, we explore the lexical encoding of [ħ] as an independent or a merged category.

	Pharyngeal Speakers	Pharyngeal Listeners	Merged Listeners
Pharyngeal exposure	Yes	Yes	No
Pharyngeal production	Yes	No	No

Table 1: The three target groups of listeners in the study.

1.1 The Categories: Phonologically and Orthographically Distinct, One Variant Stereotyped

Although the [ħ]-[χ] merger is pervasive among speakers of Modern Hebrew, several cues in the language and writing system may facilitate speakers' maintenance of an underlying distinction in the representation of these phones. In Hebrew orthography, [ħ] and [χ] are represented by distinct

---

[1]The production of the merged category might vary between a uvular fricative and a uvular trill (Gafter 2020).

letters ([ח] and [כ] respectively). Hebrew speakers are taught about the etymological status of the pharyngeals at school, and their production is often praised as "the correct production" in language classes and by official language authorities such as the Hebrew Academy (Gafter 2021). From a phonological perspective, they trigger divergent alternations, where [χ] but not [ħ] alternates with [k] (as part of a spirantization process that also occurs with the stop-fricative pairs [b∼v] and [p∼f]).

Despite this support from prescriptive norms, pharyngeal production in Modern Hebrew suffers from a negative social evaluation. This is due to the so-called 'ethnic cleavage' between Jews of European descent, known as *Ashkenazi*, and Jews of Middle Eastern and North African descent who are referred to with the broad cover term *Mizrahi*. The persisting relevance of these social categories is reflected in the fact that Ashkenazis on average have higher earnings and education compared to Mizrahis (Lewin-Epstein and Cohen 2019; Dahan et al. 2003), and are considerably over-represented in the local cultural, judiciary and economic elite (Sasson-Levy 2013). Pharyngeal consonants are also present in the speech of Modern Hebrew as a second language by many Palestinian Israelis (Gafter and Horesh 2020), who are severely discriminated against in every aspect of life.

While its salient social status makes [ħ] easily recognizable to Merged Listeners (Gafter 2016), regardless of their input at acquisition, we argue that exposure to pharyngeals at a young age may lead to divergent patterns in word recognition.

1.2 Perception of Merged Categories

In order to understand what happens when merged speakers hear [ħ] – a consonant that exists in their native language but not in their own native variety – we can consider the difficulty caused to L2 speakers by encoding unfamiliar sounds in a new language. It has long been known from studies on L2 acquisition that the difficulty to produce L2 sounds that do not exist in one's L1 is rooted not only in articulatory factors, but also in perception. Cases of perceptually-merged L2 sounds occur even with highly proficient speakers. For example, Pallier et al. (2001) found that Spanish-dominant early bilinguals experienced repetition priming for Catalan /e/∼/ɛ/ minimal pairs, contrary to native Catalan speakers. This was interpreted as evidence that L1 Spanish speakers not only produce these categories as merged, but also perceive such pairs as homophones. Using the same paradigm, Dufour et al. (2007) found that comparing to French speakers from the Paris metro area, southern French speakers perceive /ɛ/∼/e/ minimal pairs to be homophones, showing that dialect-specific mergers can also induce significantly different lexical representations.

These studies contribute the insight that highly proficient speakers – even native speakers, in the case of French – can vary significantly from each other in terms of their lexical encoding. Furthermore, there is a growing body of evidence suggesting that despite these seemingly symmetrical results (/ɛ/ words prime their minimal pair /e/ words and vice versa), lexical encoding is influenced by the phonetic resemblance of each category to a native L1 category of the listener. For example, Weber and Cutler (2004) found that L1 Dutch speakers who are highly proficient in English experience competition from /ɛ/ items upon hearing /æ/. That is, words that contain /æ/, such as *panda*, briefly activate words that contain /ɛ/, e.g. *pencil*. On the other hand, no competition effect was found in the opposite direction – upon hearing *pencil*, *panda* was not activated. This suggests that the category that is phonetically closer to an L1 category – in the Dutch-English case, /ɛ/, and in our case, /χ/ – is encoded more accurately in the lexicon, creating a more selective pattern of activation (in this case, reflected in the eye movements of participants in a Visual World Task). In other words, the category that is phonetically similar to the native category (hereafter the "familiar" category) is more lexically constraining than the "new" category, which is perceived and lexically encoded as a "bad exemplar" of the existing category.

Two additional studies support this view. L1 English speakers are known to have a difficulty producing the distinction between front and back round vowels that are found in German ("familiar" category = back round vowels), and between singleton and geminate consonants in Japanese ("familiar" category = singleton consonants). Darcy et al. (2013) found that while intermediate and advanced learners are almost on a par with native speakers in terms of their phonetic decoding (tested in an ABX task), their performance in a lexical decision task in which these sounds were switched is significantly worse. Again, performance with switched "familiar category" items was better: real

words were more likely to be accurately accepted if they included the familiar category, and non-words were more likely to be rejected if they included the new category. For example, the word *honig* 'honey' includes a back round vowel, while *könig* 'king' includes a front round vowel. L1 English learners of German more often correctly rejected a switch from the familiar to the new category (e.g.*hönig*) compared with the opposite direction of the switch (*konig*); they were also more likely to identify *honig* as a word, compared with *könig*. Melnik and Peperkamp (2019) showed similar findings with L1 French speakers learning English: performance on the /h/∼/Ø/ distinction was better with [Ø]-words (e.g. *officer*) compared with h-words (*husband*), and category-switched non-words with the added new category (*hofficer*) were easier to reject than category-switched non-words with the familiar category (*usband*).

Mergers between dialect-specific L1 categories, such as the Hebrew [ħ]-[χ] merger, have clear similarities to L2 learning, but also differences. While the sounds of another dialect may be new, the lexical items, of course, are not, as different dialects have a largely shared vocabulary. Therefore, speakers of a merged dialect receive direct evidence from their native lexicon during acquisition that there is *no difference* between the categories they might be facing when moving to a non-merged environment. For example, in the *pin-pen* merger in the US (Labov et al. 2008), a learner who had been exposed only or mostly to merging speakers during childhood is likely to perceive *pin* and *pen* as homophones (Conrey et al. 2005).

At the same time, the 'new' phonetic category (in the sense of Darcy et al. 2013) faced by merged native speakers is not necessarily entirely new: in many cases, speakers have some early exposure to the phonetic categories of non-merged varieties of their own language, which is often not the case with L2 learners. Phonological variation of this sort often bears social meaning, which is certainly the case with the highly stereotyped Hebrew pharyngeal fricative [ħ] (Gafter 2019). In this case, the 'new' phonetic category can be salient as a feature of a social group, and yet, as we show below, not constitute part of the lexical inventory for Merged Listeners.

2 Methods

We performed a lexical decision task where the target categories had been switched with each other in some of the real word items, predicting different outcomes in terms of grammatically based on the linguistic background of the listener. While the direction of ħ→χ is compatible with the common merger, the opposite direction, χ→ħ, is not compatible with any word in Hebrew (see section 2.2 for stimuli design). We used stimuli recorded by two speakers, introduced in separate blocks: one speaker merged the categories in his speech while the other maintained the pharyngeal pronunciation, alongside other markers of Mizrahi speech.

Participants included Pharyngeal Speakers, Pharyngeal Listeners and Merged Listeners. We hypothesize that participants of the latter group have only the category /χ/ in their lexicon, and therefore expected that their reaction to items containing an etymologically correct ħ or a manipulated χ→ħ should be similar. In the lack of independent representation for [ħ] in the lexicon, the mechanism that maps any surface occurrence of [ħ] into the lexical category /χ/ should operate freely whether the input is natural or manipulated. In the ħ→χ condition, the result of the manipulation are compatible with the merger occurring in the participants' own production, and is therefore predicted to present no difficulty.

The status of switched-category productions is more complex for listeners from pharyngealizing backgrounds: since ħ is a separate category in their lexicon, and since the alternation χ→ħ does not occur naturally in either variety, such items are expected to be rejected. Items produced in the natural direction of the merger, i.e. the switched-category ħ→χ condition, are predicted to be lexically valid, but inconsistent with the identity of the pharyngealizing speaker, which may affect the processing of such items.

The perceived identity of a speaker was previously shown to affect various aspects of sense-making, including lexical expectations (Walker and Hay 2011), ambiguity resolution (Cai et al. 2017), acceptance of divergence in production (Weatherhead and White 2018) and compensating for known differences in phoneme boundaries (Strand 1999; D'Onofrio 2018; Lev-Ari et al. 2019).

For example, Walker and Hay (2011) found that words that are more likely to be used by older speakers according to a corpus were more quickly recognized when uttered by an old voice compared with a young voice, and vice versa. Cai et al. (2017) demonstrate that listeners tend to understand *bonnet* as a type of hat when uttered with an American accent, and as a car hood when uttered with a British accent. That is, listeners have different expectations regarding the words that a given speaker might use, and even regarding the particular meaning that the speaker intended in the case of ambiguous words. Given these findings, it can be expected that bi-dialectal speakers will be affected by the persona of the speaker, such that merged production items should raise more difficulty when produced by the pharyngealizing voice, compared with the merged voice.

Table 2 summarizes our predictions for the pattern of responses to switched-category items by each participant group. "V" indicates no expected difficulty in the processing of items under the given condition. "X" indicates expected rejection due to ungrammaticality – this is predicted only for items that do not occur in either variety, i.e. χ→ħ items. The mark "?" was used for items that are lexically valid, but inconsistent in terms of the identity of the speaker. While similar manipulations have been shown to affect various behavioral measures, as described above, it is unclear whether acceptance rates may be affected as well. Items produced in the merged speaker's voice were not manipulated and were predicted to be equally accepted by all groups.

Group	ħ→χ (Dialect-Inconsistent)	χ→ħ (Unattested)
Merged Listener	V	V
Pharyngeal Listener	?	X
Pharyngeal Speaker	?	X

Table 2: Prediction for pattern of acceptance of switched-category stimuli produced in the Pharyngealing speaker's voice, for the three groups of participants.

2.1 Participants

Seventy-eight native speakers of Modern Hebrew volunteered to participate in the experiment. Recruitment was done through the authors' friends and family. Participants who pronounce pharyngeals were listed as Pharyngeal Speakers; participants that do not pronounce pharyngeal themselves but had at least one primary caregiver who does were listed as Pharyngeal Listeners; Participants with no early exposure to pharyngeal dialects, including among extended family and friends, were listed as Merged Listeners (speakers in the two latter groups produce the same merged variant).[2]

The final number of participants included 24 Pharyngeal Speakers (mean age: 58, SD = 10), 25 Pharyngeal Listeners (mean age 31, SD = 10) and 29 Merged Listeners (mean age 34, SD = 16).[3]

Participants had no reported hearing problems. Six participants from the Pharyngeal Listeners group and one from the Pharyngeal Speakers group reported being diagnosed with ADHD.[4] Eight Merged Listeners and five Pharyngeal listeners reported having a second native language (Russian or English). Three Pharyngeal Speakers were native speakers of Yemeni Arabic, in addition to Modern Hebrew.

---

[2]Participants were not asked about their linguistic background before the task; if after the exit questionnaire it turned out that the participant did not fit neatly into one of the groups, their data was discarded (e.g. a Merged Listener was discarded because they had a pharyngealizing grandparent).

[3]Since Pharyngeal Speakers in Modern Hebrew speaking communities in Israel are generally older (due to the decline in usage of pharyngeals in younger generations) compared with the mean age of Pharyngeal Listeners (being the 2nd generation), we intentionally set out to test Merged Listeners of all ages, in order to have a wide variety of age representatives in this group that would match the Pharyngeal Speakers on the one hand and the Pharyngeal Listeners on the other. Hence the larger standard deviation in mean age in this group. Age data is missing for 5 participants due to a technical issue (3 Merged Listeners and 2 Pharyngeal Speakers).

[4]ADHD may induce lower performance in tasks that require attention on a single task. Since the participants who reported being diagnosed with ADHD were from groups that are expected to perform at ceiling, any effect it may have on the results would be in the opposite direction of our hypothesis.

2.2 Stimuli

Fifty-six Hebrew nouns served as critical items, of which 28 included [χ] and 28 [ħ] (none of the items contained both categories). Critical items were selected such that the status of [ħ]/[χ] could not be determined on the basis of phonological alternations (see section 1.1). 56 nouns which do not include either of the categories were used as distractors (Baseline Condition in Table 3), as well as 112 pseudo-nouns which were compatible with Hebrew nouns phonotactically and prosodically, and included [χ] and [ħ] to the same extent as the real words in the experiment: 28 with [χ], 28 with [ħ] (the same non-words used in Experiment 1) and 56 with neither [χ] nor [ħ].

We controlled for the position of the target sound (coda/intervocalic), concreteness (physical/non-physical item), number of syllables (two or three) and frequency; these features were evenly distributed between four mini-blocks. Each mini-block contained seven [χ] items, seven [ħ] items, 14 distractors and 28 non-words.

The stimuli were recorded by two cisgender men who are native speakers of Hebrew, aged 30 and 34. The former is a native merged speaker and his recordings included only the [χ] realization. The latter is a pharyngeal listener with parents and many others in the close environment who maintain the distinction. He was comfortable with producing pharyngeals and other markers of Mizrahi speech, including realization of the Hebrew rhotic as [r], as opposed to [ʁ]). Target items recorded in the Mizrahi voice had one version where [ħ] and [χ] were each used consistently in the appropriate lexical items, and a second version where the categories were switched with each other. These included "Inconsistent" items where [ħ] was replaced with [χ] and created merged items in a Mizrahi voice, and "Unattested" items where [χ] was replaced with [ħ], resulting in items which are not part of any native dialect.

The experiment proceeded in three blocks by recorded voice: a merged block, a pharyngeal block and a manipulated pharyngeal block composed of a mix of two mini-blocks, one of which consists of natural pharyngeal items and the other includes Unattested and Inconsistent trials. The order of the merged and the pharyngeal blocks was randomized, and the manipulated block was always presented last. This was done in order to make sure that the pharyngealizing speaker could initially be characterized as reliable (i.e. not as someone who makes production errors, from the point of view of a non-merged speaker's lexicon). The critical items were rotated between the conditions in a Latin square design, using the four mini-blocks such that participants heard an item exactly once. Sample items of each condition are provided in Table 3.

Voice	Baseline	/χ/	/ħ/	Inconsistent [χ] (ħ→χ)	Unattested [ħ] (χ→ħ)
Merged Voice	tʁufa	meχonit	aχot	N/A	N/A
Pharyngeal Voice	trufa	meχonit	aħot	ʔaχot	*meħonit
	'medicine'	'car'	'sister'		

Table 3: Sample stimuli in the merged and pharyngeal voices. Note the aforementioned difference in the production of /r/ as well, in accordance with the speaker's variety.

2.3 Procedure

The experiment was run on PsychoPy (Peirce and MacAskill 2018). Due to the outbreak of Covid-19, offline testing was replaced with an online version after roughly 2/3 of the data were collected. Participants were asked to wear headphones throughout the experiment; in the offline version, we used Sony MDRZX100 ZX Series Stereo Headphones, on a 13-inch MacBook Air computer with 2.2 GHz Intel Core i7. The Running mode (Online/Offline) was registered as a factor in both RT and Acceptance rates models. The offline data was analyzed again separately to make sure that the trends observed in the overall data are attested there as well. No differences in trends were observed.

3 Results

3.1 Data Analysis

Only real words were analyzed, including words produced in the Unattested Condition ($\chi \to \hbar$) and in the Inconsistent Condition ($\hbar \to \chi$ in the pharyngeal voice). One item from the Baseline Condition ([ʕiton] 'newspaper') was removed due to low acceptance rates (68%). All other real words were accepted in over 70% of all trials, excluding the Unattested Condition, where acceptance was predicted to vary across groups.

The accuracy criterion for participant inclusion was scoring above 70% success with real word stimuli produced in the Merged voice. This condition was chosen as the performance test since the majority of daily input from the media is produced in this variety, and rejecting such words probably stems from a general issue with the task. Three participants were excluded based on this criterion (2 Pharyngeal Speakers and 1 Pharyngeal Listener).

Analyses were conducted using the LmerTest package (Kuznetsova et al. 2017) in the R software environment (R Core Team 2017). A logistic model was used for accuracy/acceptance data, and a linear model for RT data. The models included random intercepts for participants and items, and the following fixed factors and their interactions:

- Voice was coded as a binary categorical variable, with the two levels, Merged and Pharyngeal Speaker;

- The baseline of Condition was items that included No [ħ]/[χ]; the conditions compared with the baseline were [ħ] items, [χ] items, Inconsistent and Unattested, the latter two being relevant only for the pharyngeal voice blocks;

- Group had two planned comparisons: Merged Listeners vs. Pharyngeal Speakers and Pharyngeal Listeners, and Pharyngeal Listeners vs. Pharyngeal Speakers (one comparison based on production, the other on received input).

Two additional fixed factors for which we did not consider interaction were Running mode (Offline/Online) and age (continuous). The BOBYQA optimizer was used to allow the models to converge (Bates et al. 2007).

3.2 Acceptance Rates

The acceptance rates model did not converge when Age was included as a fixed factor, therefore it was removed, leaving the final model as follows (*N = 8325*):

Acceptance rates model = response $\sim$ Condition * Voice * Group + running mode + (1 | participant) + (1 | item)

The logistic regression analysis revealed a main effect of Condition, with the Unattested and Inconsistent items conditions yielding significantly lower acceptance rates (Unattested: β = -3.17, SE = 0.28, z = -11.44, p <0.0001; Inconsistent: β = -0.84, SE = 0.32, z = -2.66, p <0.008). Interactions were found between Participant Group and Condition, such that the Unattested Condition affected listeners from pharyngealizing backgrounds (Pharyngeal Speakers and Pharyngeal Listeners) more than it did the Merged Listeners group, and the Pharyngeal Speakers more than Pharyngeal Listeners (Merged vs. Pharyngeal Speakers + Pharyngeal Listeners: β = -1.04, SE = 0.13, z = -8.13, p <0.0001; Pharyngeal listeners vs. Pharyngeal Speakers: β = -1.38, SE = 0.21, z = -6.51, p <0.0001). The same was true with the Inconsistent Condition (Merged vs. Pharyngeal Speakers + Pharyngeal Listeners: β = -0.44, SE = 0.17 , z = -2.6, p <0.009; Pharyngeal Listeners vs. Pharyngeal Speakers: β = -0.61, SE = 0.27, z = -2.29, p <0.02). Finally, there was a triple interaction between Voice, Condition and Group, such that Merged Listeners made more errors with [χ] items in the Merged voice (β = 0.29, SE = 0.14, z = 2.03, p <0.04). Figure 1 is a visual summary of the model, using the Emmeans package of R (Lenth et al. 2018).
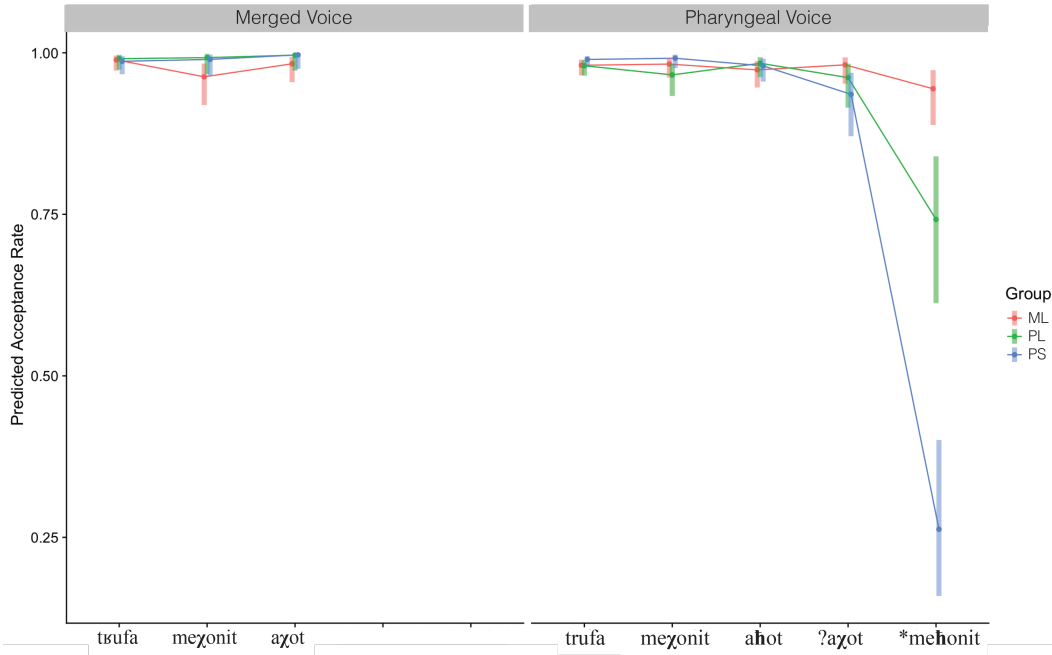
Figure 1: A summary of the logistic regression model of acceptance rates. Blue: Pharyngeal Speakers (PS), Green: Pharyngeal Listeners (PL), Red: Merged Listeners (ML). The merged voice condition is on the left panel; note that the Inconsistent and Unattested conditions are irrelevant for this voice, since his productions are merged. On the right – the pharyngeal voice. Conditions from left to right within each panel: Baseline, /χ/, /ħ/, Inconsistent [χ], Unattested [ħ].

3.3 Reaction Times

Only correct responses were considered for the reaction times analysis. RTs were measured from the onset of the stimulus and log-transformed before being entered into the model. Responses below 0 or above 5 seconds were removed (0.5% of the responses), and responses of above 2.5sd of the participant's mean were replaced by the mean+2.5sd (7 trials). The final analysis included 7877 trials, and the model was:

RT model = logRT ∼ Condition * Voice * Group + age + running mode + (1 | participant) + (1 | item)

The model revealed a main effect of Condition, such that /χ/, /ħ/ and Unattested items were responded to more slowly compared with the Baseline (/χ/: β = 0.03, SE = 0.01, df = 176.6, t = 2.06, p <0.05; /ħ/: β = 0.057, SE = 0.015 , df = 176.2, t = 3.71, p <0.001; Unattested: β = 0.14, SE = 0.022, df = 281.9, t = 6.215, p <0.0001). The main effect for /χ/ and /ħ/ is likely to be due to the overall high rate of items (words and non-words alike) that include [χ] or [ħ] – two thirds of the stimuli. The non-words in particular might have created a general carefulness around trials containing these sounds. The lack of a similar effect for the Inconsistent Condition, which also contains [χ], is due to interactions in opposite directions with Group, as discussed below.

Another main effect was that of recorded voice: the Merged voice was responded to faster than the pharyngealizing voice (β = -0.025, SE = 0.011, df = 147.7, t = -2.281, p <0.024).

Mimicking the acceptance rates, interactions were found between Group and Condition, such that the Unattested Condition affected listeners from pharyngealizing backgrounds more than it did the Merged Listeners group, and the Pharyngeal Speakers more than Pharyngeal Listeners (Merged vs. Pharyngeal Speakers + Pharyngeal Listeners: β = 0.033, SE = 0.008, df = 7702, t = 4.057, p <0.0001; Pharyngeal Listeners vs. Pharyngeal Speakers: β = 0.046, SE = 0.018, df = 7707, t =

2.583, p <0.01). The Inconsistent Condition affected speakers from a pharyngealizing background more than Merged Listeners (β = 0.018, SE = 0.007, df = 7645, t = 2.71, p <0.007); here, there was no significant difference between Pharyngeal Speakers and Pharyngeal Listeners.

Another interaction between Group and Condition was that Merged Listeners responded more slowly to the /ħ/ condition compared with listeners from pharyngealizing backgrounds (β = -0.011, SE = 0.004, df = 7633, t = -2.501, p <0.012); a triple interaction between Voice, /ħ/ and Group, suggests that this effect was mainly due to consistent productions of [ħ] (i.e. items produced in the Pharyngeal voice; β = 0.01, SE = 0.008, df = 7621, t = 2.283, p <0.03).[5] That is, while listeners from Pharyngealizing backgrounds responded to consistent /ħ/ productions faster compared with Inconsistent productions, the opposite was true with Merged Listeners.
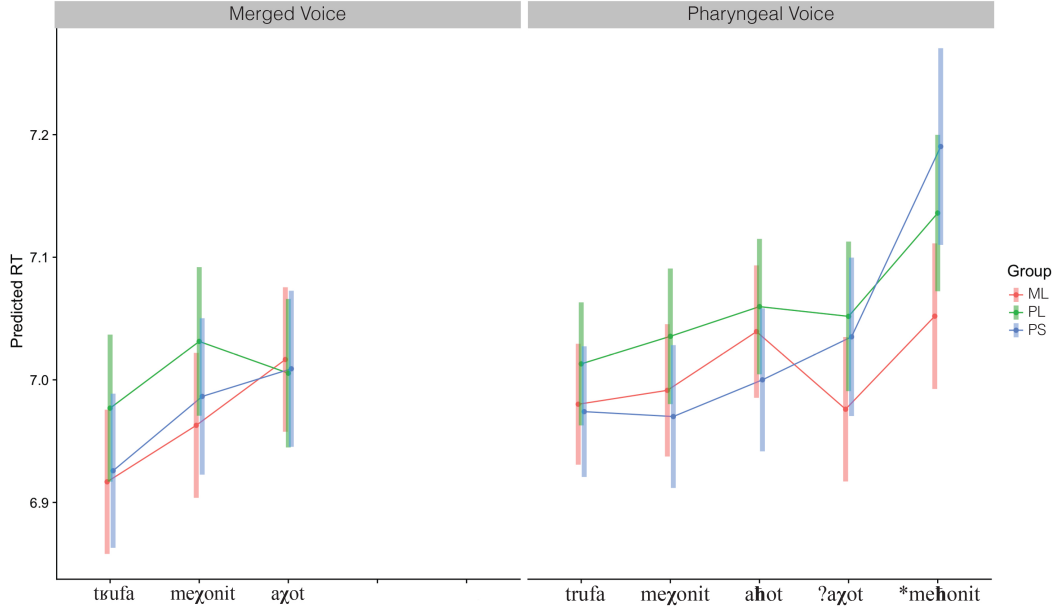

Figure 2: A summary of the linear regression model for RTs. Blue: Pharyngeal Speakers (PS), Green: Pharyngeal Listeners (PL), Red: Merged Listeners (ML). Left panel: Merged voice, right: Pharyngealizing voice. From left to right: Baseline, /χ/, /ħ/, Inconsistent [χ], Unattested [ħ].

Age and Running Mode were both marginally significant, such that older participants were slower (β = 0.002, SE = 0.001, df = 74.9, t = 1.95, p = 0.054), as were online participants (β = 0.056, SE = 0.029, df = 75.03, t = 1.95, p <0.055). Figure 2 is a visual summary of the model.

4 Discussion

The results revealed a significant difference in the patterns of reactions to switched-category items between Pharyngeal Speakers, Pharyngeal Listeners, and Merged Listeners. Particularly, participants from the first two groups rejected Unattested items (χ→ħ), or responded to them significantly more slowly, indicating a difference in the status of this category in their lexicon. Pharyngeal Speakers and Pharyngeal Listeners additionally rejected and/or responded more slowly to Inconsistent items (ħ→χ in the pharyngealizing voice) at a higher rate than Merged Listeners, indicating that the perceived identity of the speaker affected their lexical decision as well.

Moreover, for some participants in the Pharyngeal Speakers and Pharyngeal Listeners groups, ħ→χ ('Dialect-Inconsistent') items were also unacceptable. Such items were responded to significantly more slowly by these groups, contrary to the Merged Listeners group.

---

[5]Recall that in the Merged speaker's voice, /ħ/ always surfaces as [χ].

Merged Listeners were slower than the other groups in lexical decision for consistent [ħ] items produced in the pharyngeal voice. This may be an 'accent effect', i.e. the price of processing a variant that is not often met by these speakers. Similarly, Pharyngeal Speakers were slowed by items produced with [χ] significantly more than Pharyngeal Listeners, suggesting a comparable price in the opposite direction.

Overall, the results confirm that some speakers who produce a merged [χ] category, namely the Pharyngeal Listeners group, still maintain a covert lexical distinction between [ħ] and [χ], despite the identical surface realization. With respect to the Merged Listeners group, our interpretation of the results is that the lexical representations of Merged Listeners does not include the category [ħ], which is therefore always mapped onto an underlying /χ/. As mentioned before, Merged Listeners can perceive [ħ] as a distinct sound, but there is no category in their lexicon that matches it – [ħ] is perceived as a surface variant of /χ/ that has social, but not lexical consequences. These results are in accordance with the prediction of the "fuzzy representations" hypothesis (Darcy et al. 2013): [ħ] is perceived as distinct phonetically, but is assimilated at the phonological encoding level to [χ]. A crucial difference between the current case and the L2 cases that Darcy et al. set out to explain is the distribution of the categories in the input listeners are exposed to. For L2 learners, the perceptual merger (e.g. of [o] and [ö]) is due to the absence of a category in their L1, while the input they receive in L2 is typically *not merged*, reflecting L1 production. In the context of a native merger, listeners are exposed to merged and non-merged speech at varying degrees.

In our study, Merged Listeners seem to attribute the use of [ħ] to a relevant social distinction, but do not employ it as a lexical variable, although they are aware of its lexical status through direct instruction and the orthographic system. Pharyngeal Speakers and Pharyngeal Listeners are equipped to make the lexical distinction, as demonstrated by their rejection of real words in which [χ] had been replaced with [ħ] (Unattested Condition). In that respect, this result is similar to that reported by Sumner and Samuel (2009), who showed that NYC speakers who produce a rhotic variety perform more like non-rhotic speakers, depending on their degree of exposure to the non-rhotic variety. As in the NYC case, our results are in the context of an ongoing change in progress, where socially marked variables become less common alongside (or due to) social stereotyping. It may be the case that a discrepancy between dialect production and dialect perception and representation is a transitory stage typical of such contexts.

References

Bates, Douglas, Deepayan Sarkar, Maintainer Douglas Bates, and L Matrix. 2007. The lme4 package. *R package version* 2:74.

Cai, Zhenguang G, Rebecca A Gilbert, Matthew H Davis, M Gareth Gaskell, Lauren Farrar, Sarah Adler, and Jennifer M Rodd. 2017. Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology* 98:73–101.

Conrey, Brianna, Geoffrey F Potts, and Nancy A Niedzielski. 2005. Effects of dialect on merger perception: Erp and behavioral correlates. *Brain and Language* 95:435–449.

Dahan, Momi, Eyal Dvir, Natalia Mironychev, and Samuel Shye. 2003. Have the gaps in education narrowed? on factors determining eligibility for the israeli matriculation certificate. *Israel Economic Review* 1.

Darcy, Isabelle, Danielle Daidone, and Chisato Kojima. 2013. Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon* 8:372–420.

Davis, Lawrence M. 1984. The pharyngels in hebrew: Linguistic change in apparent time. *Folia Linguistica Historica* 18:25–32.

D'Onofrio, Annette. 2018. Personae and phonetic detail in sociolinguistic signs. *Language in Society* 47:513–539.

Dufour, Sophie, Noël Nguyen, and Ulrich Hans Frauenfelder. 2007. The perception of phonemic contrasts in a non-native dialect. *The Journal of the Acoustical Society of America* 121:EL131–EL136.

Gafter, Roey J. 2016. What's a stigmatized variant doing in the word list? authenticity in reading styles and hebrew pharyngeals. *Journal of Sociolinguistics* 20:31–58.

Gafter, Roey J. 2019. Using pharyngeals out of context: Linguistic stereotypes in parodic performances of mizrahi hebrew speakers. In *The Routledge Companion to the Work of John R. Rickford*, ed. Renée Blake and Isabelle Buchstaller, 232–241. New York: Routledge.

Gafter, Roey J. 2020. Trilling as a sociolinguistic variable: Ethnicity and variation in the hebrew dorsal frica-
    tives. *Sociolinguistic Studies* 14:483–503.

Gafter, Roey J. 2021. Reconciling seemingly conflicting social meanings. In *Social Meaning and Linguistic
    Variation: Theorizing the Third Wave*, ed. Lauren Hall-Lew, Emma Moore, and Robert J Podesva, 222–242.
    Cambridge University Press.

Gafter, Roey J, and Uri Horesh. 2020. Two languages, one variable? pharyngeal realizations among arabic-
    hebrew bilinguals. *Journal of Sociolinguistics* 24:369–387.

Kuznetsova, Alexandra, Per B Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmertest package: tests
    in linear mixed effects models. *Journal of Statistical Software* 82.

Labov, William, Sharon Ash, and Charles Boberg. 2008. *The atlas of North American English: Phonetics,
    phonology and sound change*. Berlin: Walter de Gruyter.

Lefkowitz, Daniel. 2004. *Words and stones: The politics of language and identity in Israel*, volume 26. Oxford
    University Press on Demand.

Lenth, Russell, Henrik Singmann, Jonathon Love, Paul Buerkner, and Maxime Herve. 2018. Emmeans: esti-
    mated marginal means, aka least-squares means. r package vers 1.2. 4.

Lev-Ari, Shiri, Robin Dodsworth, Jeff Mielke, and Sharon Peperkamp. 2019. The different roles of expectations
    in phonetic and lexical processing. In *Interspeech*, 2305–2309.

Lewin-Epstein, Noah, and Yinon Cohen. 2019. Ethnic origin and identity in the jewish population of israel.
    Journal of Ethnic and Migration Studies 45:2118–2137.

Melnik, Gerda Ana, and Sharon Peperkamp. 2019. Perceptual deletion and asymmetric lexical access in second
    language learners. *The Journal of the Acoustical Society of America* 145:EL13–EL18.

Pallier, Christophe, Angels Colomé, and Núria Sebastián-Gallés. 2001. The influence of native-language
    phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science* 12:445–
    449.

Peirce, Jonathan, and Michael MacAskill. 2018. *Building experiments in PsychoPy*. London: Sage publications
    Ltd.

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical
    Computing, Vienna, Austria.

Sasson-Levy, Orna. 2013. A different kind of whiteness: Marking and unmarking of social boundaries in the
    construction of hegemonic ethnicity. In *Sociological Forum*, volume 28, 27–50. Wiley Online Library.

Strand, Elizabeth A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of language
    and social psychology* 18:86–100.

Sumner, Meghan, and Arthur G Samuel. 2009. The effect of experience on the perception and representation
    of dialect variants. *Journal of Memory and Language* 60:487–501.

Walker, Abby, and Jen Hay. 2011. Congruence between 'word age' and 'voice age' facilitates lexical access.
    Laboratory Phonology 2:219–237.

Weatherhead, Drew, and Katherine S White. 2018. And then i saw her race: Race-based expectations affect
    infants' word processing. *Cognition* 177:87–97.

Weber, Andrea, and Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal
    of memory and language* 50:1–25.

Si Berrebi
Department of Linguistics
Tel Aviv University
siberreb@mail.tau.ac.il

Noa Bassel
Department of Linguistics
The Hebrew University in Jerusalem
noa.bassel@mail.huji.ac.il

Roey J. Gafter
Department of Hebrew Language
Ben-Gurion University of The Negev
roeyg@bgu.ac.il