EVALUATING FORMANT ESTIMATIONS AND DISCRETE COSINE TRANSFORM TO DIFFERENTIATE BETWEEN PHARYNGEAL FRICATIVES IN MEHWEB

Alexandre Arkhipov [a]*, Michael Daniel [a,b], Alexander Shiryaev [c], Ekaterina Shepel [d]

[a] Universität Hamburg, Germany; [b] Collegium de Lyon / Laboratoire Dynamique du Langage (Université de Lyon / CNRS), France; [c] Independent researcher, Singapore; [d] HSE University, Moscow
alexandre.arkhipov@uni-hamburg.de, misha.daniel@gmail.com, ashiryaev87@gmail.com, katerina11780@gmail.com

ABSTRACT

We investigate the laryngeal/pharyngeal part of the consonantal inventory of Mehweb Dargwa (East Caucasian), which has been shown to contrast plain laryngeals to pharyngeals, with additional contrasts due to pharyngealization feature.

We are using formant estimation on the fricative noise of laryngeal/pharyngeal fricatives to differentiate between contrasting segments and to approach their characterization in articulatory terms. Using formant analysis techniques is applicable here because the vocal tract configuration of pharyngeal and laryngeal segments is similar to that of vowels. While the formant structure of plain laryngeals accommodates to adjacent vowels, pharyngeals appear to have their own target formant values, which furthermore vary depending on the presence of pharyngealization.

We are comparing outcomes of our formant analysis to discrete cosine transform coefficients, which have been previously shown to differentiate well between sibilants.

Keywords: pharyngeal, pharyngealization, formants, DCT, East Caucasian.

1. MEHWEB LANGUAGE

1.1. Genealogical and sociolinguistic aspects

Mehweb (Glottocode: mege1234) is a one-village language of the Dargwa branch of the East Caucasian (Nakh-Daghestanian) family, sometimes considered to be a dialect of (Northern) Dargwa (Glottocode: darg1241, ISO 639-3: dar). It is spoken in the village of Mehweb (Russian: *Мегеб*) in central Daghestan at 1,800 m above sea level, geographically separated from other languages of the Dargwa branch and surrounded by speakers of Avar and Lak (belonging to other branches of East Caucasian).

As of 2019, the number of speakers is estimated to be between 800 and 900 [1], most of them living in Mehweb itself. “So far, there are no indications of language loss in Mehweb. All villagers speak Mehweb, and Mehweb is the first language acquired by children” [1: 2]. Mehweb has no written tradition; written languages used by Mehwebs are Russian and Avar. Most adults have a command of Russian, those born before the 1990s usually also speak Avar, and some of those born before the 1950s speak Lak.

Our starting point was the published description of Mehweb phonological system [2] based on fieldwork data from 2013–2016, briefly summarized in §1.2. A small sample of acoustic data from Mehweb, along with the data from several other East Caucasian languages, was used in [3], providing some phonetic observations presented in §1.3.

1.2. Mehweb phonology

The vowel system lists four vowel qualities: /i e a u/. All of them can be pharyngealized, whereby /uˤ/ is realized variably as [uˤ~oˤ].

Pharyngealization is claimed to be a ‘prosodic’ feature associated with a syllable (or root) rather than with a particular vowel or consonant; it may spread from lexically pharyngealized syllables (roots) to adjacent syllables. We follow [2] in marking the nucleus of the pharyngealized syllable with /ˤ/. Most pharyngealized syllables contain epiglottals [ʔ ʜ] or uvulars [q χ ʁ], but /aˤ/ is also attested in their absence (e.g. /laˤži/ ‘cheek’). Across lexicon, /aˤ/ is also the most frequent pharyngealized vowel, while front pharyngealized /iˤ eˤ/ are very rare.

The basic contrasting post-uvular segments are plain laryngeal (glottal) stop /ʔ/ and fricative /h/, and pharyngeal fricative /ħ/ (see examples in Table 1).

Sound	Phar.	Example	Gloss
ʔ	no	/muʔ/	‘back’ (noun)
h	no	/haruš/	‘fermented drink’
ħ	no	/ħaq’ur/	‘burdock’
[ʡ] /ʔ/	yes	/ʡaˤt’a/	‘frog’
[ʜ] /ħ/	yes	/ħaˤbal/	‘three’
ʡ	no	/ʡat’/	‘flour’ (Avar loan)
ʡ	no	/ʡarʁal/	‘long’

Table 1: Examples with laryngeals and pharyngeal/epiglottal sounds. “Phar.” marks the presence/absence of pharyngealization.

Two segments appearing in pharyngealized syllables, designated as epiglottal stop [ʔˤ] and fricative [ʜ], are considered to be pharyngealized allophones respectively of /ʔˤ/ and /ħˤ/ [2: 32–35]. However, in some Avar loans as well as in a few native roots a similar epiglottal stop appears without pharyngealization [2: 19, fn. 3]. While its properties require further study, this /ʔˤ/ will be provisionally treated here as a separate phoneme.

1.3. Phonetic properties of pharyngeals/epiglottals

Phonetic properties of pharyngeals/epiglottals are not discussed in [2]. Pharyngealization effect on vowels is described as 'centering' [2: 32] and symbolized as /iˤ eˤ aˤ uˤ/ yielding [eˤ ɛˤ æˤ uˤ~oˤ].

The study in [3: 1552] considers a limited sample of Mehweb data and characterizes [ʔˤ] in the context of pharyngealization as a 'weak epiglottal stop or approximant,' and the non-pharyngealized [ʔ] as 'a stronger epiglottal stop.' The difference between [ħ] and its pharyngealized counterpart [ʜ] is attributed to an 'increase in laryngeal constriction, possibly with larynx raising' in the latter. It is reported to have 'a flat spectrum between 1–3 kHz and peak at 3–3.5 kHz', while the spectrum of [ħ] has 'a deep valley above 1 kHz and peak at 2.5 kHz' (numbers for adult male speakers). Additionally, both fricatives or only the non-pharyngealized one, depending on the speaker, may optionally be strengthened through aryepiglottic trilling.

2. OBJECTIVES, METHODS AND DATA

2.1. Objectives

We follow [3] in adopting the Laryngeal Articulator Model approach [4, 5] and understand East Caucasian 'pharyngeal' and 'epiglottal' as essentially the same place of articulation, with further possible distinctions based on a combination of such parameters as the degree of aryepiglottic laryngeal constriction, larynx height (larynx raising being an expected synergetic accompaniment to laryngeal constriction) and aryepiglottic trilling. We thus use 'pharyngeal' as a cover term for both 'pharyngeal' and 'epiglottal' employed in [2].

Our main objective in the present paper is to obtain a more systematic acoustic description of the laryngeal and pharyngeal fricatives, which would allow to make informed hypotheses on the articulatory features involved (a proper articulatory study currently not being possible for extralinguistic reasons). To achieve this, we perform spectral analysis of the fricatives across different vocalic contexts and in different positions within a word.

2.2. Methods

Acoustic metrics used in previous work to differentiate between fricative categories in various languages include spectral moments [6] and, more recently, discrete cosine transformation, or DCT coefficients [7]. The application of spectral moments is problematic for various reasons, highlighted in [8]. Therefore, we only recur to them for comparison with previous work and as to a baseline for other metrics.

DCT has been shown to be efficient in discriminating sibilants in rich sibilant systems such as Polish [9] and German [10]. However, while DCT analysis is good at capturing spectral differences between categories, the coefficient values may not always be directly interpretable in articulatory terms. This is why we supplement DCT with the formant analysis of the fricative spectra, more informative in this respect.

Formants are characteristics of resonances of the vocal tract which determine the acoustic structure of vowels. Fricatives are produced with a noise source, hence their spectra are aperiodic. However, as shown in [11], spectral peaks in sibilants may show seamless transition into vowel formants. Spectral peaks in fricatives in the range of vowel formants F2–F5 are discussed in [12]. The length of the oral cavity in front of the main constriction has a strong influence on the spectral shape of fricatives. Spectra of anterior non-sibilants [f θ] are generally weak and have no sharp peaks since the front cavity is negligible and there is no significant obstacle downstream. Sibilants do present sharp peaks, but the lower frequency region corresponding to F1 has typically very little energy [11]. For more posterior consonants, the length of the resonance cavity increases; "[t]he more back fricatives, x, χ, ħ, have a spectral peak that decreases in frequency as the place of articulation approaches the glottis, and additional peaks in the higher part of the spectrum" [13]. Finally, it is (almost) the entire supraglottal tract which contributes to the production of laryngeals and pharyngeals. Consequently, their spectrum is most similar in structure to that of vowels.

Due to the aperiodicity of the spectrum, direct formant measurements in fricative noise are less reliable than in vowels. We use FastTrack [14], a plugin for Praat [15], to perform a set of automated measurements with different settings and select the optimal analysis for each individual sound based on the smoothness of the formant trajectories. Beyond calculations of spectra and formants, most other analyses and visualizations were done in R [16].

2.3. Data

The recordings were made in summer 2022 by the second author from four male and four female

speakers born between 1955 and 1971 and one male speaker born in 1994. All the older speakers formed married couples, and more generally were part of a tight social network within the speech community.

A list of target words was compiled to include laryngeal and pharyngeal stops and fricatives in word-initial, word-medial and word-final position in different vocalic contexts. Here, we limit our discussion to fricatives. A total of 134 stimuli were recorded, with the list of stimuli slightly different across speakers. The interviewer usually only uttered the Russian translation equivalent. Stimuli were presented in random order for each speaker, who were asked to produce the stimulus four times in isolation and once in a carrier phrase; the actual number of tokens varied across speakers and stimuli. Only isolated productions are considered in this paper.

The data were recorded with a Sennheiser HSP4-EW headset and Olympus DM-901 voice recorder as Linear PCM at 48 kHz/16 bit, manually segmented and labelled in Praat. Their annotation and analysis are currently in progress. In this paper, data from one female (*Mn*, 63 y.o.) and five male speakers (*Ab* (51), *An* (63 y.o.), *Kz* (67), *Mh* (28), *Mz* (59)) are used.

3. ACOUSTIC ANALYSIS

3.1. Preview

Generally, laryngeal [h] is markedly different from the two pharyngeals. It is shorter and less intensive; its mean duration in word-medial position was 44% to 73% of the mean duration of [ħ ʜ] across speakers. Therefore, the main focus was on distinguishing between the two pharyngeals.

Ensemble-averaged pre-emphasized spectra for the three target categories, plain laryngeal [h], pharyngeal [ħ] (labelled "ħ–") and its pharyngealized counterpart [ʜ] ("ħ+") in contexts before [a/aˤ] for speakers *An*, *Kz*, *Mn* are given in Fig. 2. Sharper peaks and valleys in [ħ] can be noticed compared to [ʜ].

3.2. Discrete cosine transformation

For DCT analysis, fricatives were resampled at 24 kHz and pre-emphasized by 6 dB/octave from 50 Hz. Series of 3 spectra at 10 ms step were extracted with 21.3 ms Hamming window around the temporal midpoint of the fricative, averaged and transformed to mel scale. Frequencies below 500 mel (414 Hz) were discarded to avoid possible effects of voicing. DCT coefficients (DCT0–DCT4) were computed with *emuR* package [17] for each averaged spectrum.

DCT coefficients behaved variably across speakers and contexts. DCT4 was the most stable and efficient in distinguishing between pharyngeals; it tended to be positive for [ħ] and negative for the pharyngealized [ʜ]. DCT3 tended to be positive for the laryngeal [h] and negative for both pharyngeals, so DCT4 and DCT3 combined performed well in distinguishing between the three fricative categories (see Fig. 3).

3.3. Spectral moments

The four spectral moments (center of gravity, standard deviation, skewness and kurtosis) were obtained on the spectra (in Hertz scale). Although laryngeal [h] was clearly separated from pharyngeals in many speaker x moment combinations, the results for the two pharyngeals were inconsistent and often contradictory across speakers. Specifically in the case of post-velars it is not surprising since the spectral moments do not capture the complex spectral shapes.

Speaker	Ab	An	Kz	Mh	Mn	Mz	Total
h	44	50	24	39	40	38	235
ħ ('ħ–')	82	108	81	80	91	87	529
ʜ ('ħ+')	70	77	55	59	62	81	404
Total	196	235	160	178	193	206	1168

Table 2: Number of tokens per speaker for three fricative categories. The pharyngeals [ħ] and [ʜ] are labelled in the data resp. as 'ħ–' and 'ħ+'.

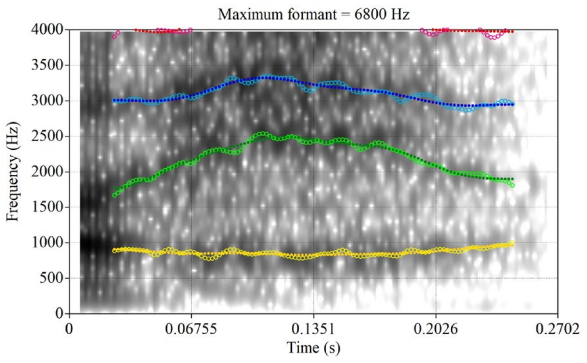

Figure 1: Sample image of formant analysis in noise by FastTrack. Speaker *Mn*, [ħ] in *tamaħ* 'consciousness'.

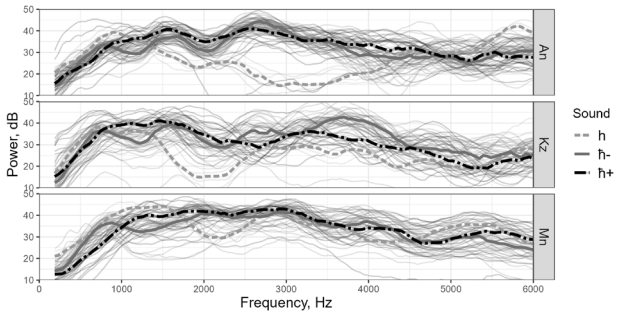

Figure 2: Averaged spectra for [h] (light grey), [ħ] (dark grey), [ʜ] (black) in contexts before [a/aˤ].

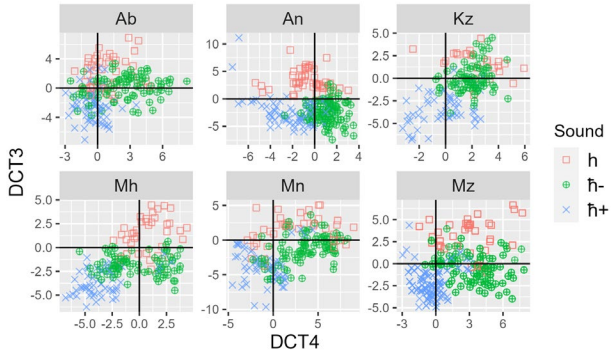

Figure 3: Fricatives in DCT3xDCT4 space.

3.4. Formant analysis

For this analysis, fricatives were submitted to the formant extraction algorithm of FastTrack. Most tokens were processed with the maximum frequency setting within 4000–6000 Hz for male speakers and 5200–6800 Hz for the female speaker. In some cases manual correction of the maximum frequency range was performed to force a better analysis.

The results confirm the high variability of the spectrum of [h]. Very often, formants are steady between [h] and the vowel, i.e. the formant structure of the laryngeal adapts completely to the vowel. In contrast, the two pharyngeals exhibit a far more stable formant configuration across vocalic contexts, its own for each sound and distinct from the adjacent vowels. The non-pharyngealized [ħ] exhibits has a high F1 and a relatively high F3 (up to ca. 3 kHz in male speakers). Its pharyngealized counterpart [ʜ] has much more tightly spaced formants, with an extremely high F1 and a low F3. It is worth noting that no formant configuration in pharyngeals is apparently continued into the vowel. Formant transitions thus deserve further study.

3.5. Comparison

We used interpretative machine learning (iML) to compare the efficiency of different types of metrics to distinguish between the two pharyngeal categories, [ħ] and [ʜ]. Three datasets were created, each including the 933 pharyngeal tokens, with Speaker and Gender as predictors. The competing dataset-specific predictors were spectral Moments (1 to 4), DCT coefficients (0 to 4), and Formants (F1 to F3, measured around the fricative midpoint). We also tested Formants datasets augmented with formant bandwidths (FormBW), formant frequencies at 0.3, 0.5 and 0.7 relative timepoints (FormDyn), and both (FormDynBW).

Three different ML algorithms were run on each dataset: logistic regression (LR), random forest (RF) and gradient boosting (GB). Each algorithm was trained 1,000 times on random 30% of the data and tested on the other 70%. F1-score was calculated each time as a quality metric (once for each of [ħ, ʜ]). The pipeline of the model was realized in Google Colabs using Scikit-learn package [18]. The median of the quality metric is reported in Table 3.

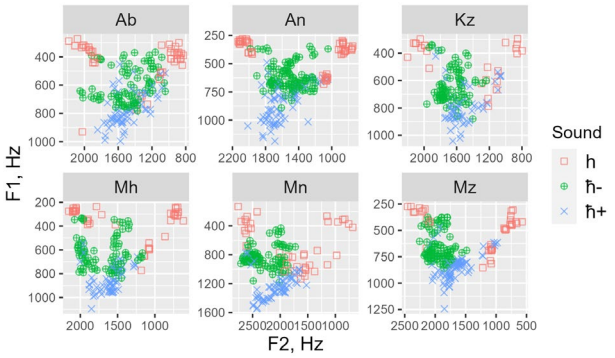

Figure 4: Fricatives in F2xF1 space.

	LR1	GB1	RF1	LR2	GB2	RF2
Moments	50.8	63.8	66.0	70.1	74.4	75.3
DCT	82.5	85.9	87.2	86.8	89.2	90.2
Formants	86.2	86.0	88.0	90.0	90.0	91.2
FormBW	87.6	89.0	90.3	91.0	92.1	93.0
FormDyn	89.8	91.6	92.8	92.7	93.8	94.6
FormDynBW	90.1	91.9	92.7	92.7	94.0	94.5

Table 3: Median F1-score (%) of models predicting the pharyngeal categories. LR1, GB1, RF1 is the score for [ʜ], LR2, GB2, RF2 the score for [ħ].

In all settings, RF achieved better results than GB and LR. The score for [ħ] was always better than for [ʜ], presumably because of the unbalanced token quantity. Results on Moments were by far the worst, starting at 50.8% for LR1. Formants yielded slightly better F1-score than DCT, while introducing formant dynamics gave a noticeable improvement (a stronger one than adding bandwidths).

4. DISCUSSION

The two contrasting pharyngeal categories have demonstrated consistent formant configurations distinct from each other and from that of vowels. Formant values allow for an articulatory interpretation; the high F1 values for [ħ ʜ] are consistent with a narrowing in the lower pharyngeal domain, and low F3 values have also been reported previously for pharyngealized vowels in East Caucasian [13; 19].

Directions for future work include the analysis of formant transitions, of voice quality in adjacent vowels, and of the pharyngeal and laryngeal stops.

5. ACKNOWLEDGEMENTS

*Contribution by this author was made in the context of the joint research funding of the German Federal Government and Federal States in the Academies' Programme, with funding from the Federal Ministry of Education and Research and the Free and Hanseatic City of Hamburg. The Academies' Programme is coordinated by the Union of the German Academies of Sciences and Humanities.

Segmentation of audio files was done by Anna Rozzorenova-Helle, Ekaterina Shepel and Georgiy Bulgakov. We are indebted to Yuliy Daniel for implementing the interpretative machine learning pipeline.

We are grateful to two anonymous reviewers for their valuable comments.

6. REFERENCES

[1] Dobrushina, N. 2019. The language and people of Mehweb. In: Daniel, M., Dobrushina, N. & Ganenkov, D. (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Language Science Press, 1–15. DOI:10.5281/zenodo.3402054

[2] Moroz, G. 2019. Phonology of Mehweb. In: Daniel, M., Dobrushina, N. & Ganenkov, D. (eds.), *The Mehweb language: Essays on phonology, morphology and syntax*. Language Science Press, 17–37. DOI:10.5281/zenodo.3402056

[3] Arkhipov, A., Daniel, M., Belyaev, O., Moroz, G., Esling, J. H. 2019. A reinterpretation of lower-vocal-tract articulations in Caucasian languages. *Proc. 19th ICPhS* Melbourne, 1550–1554.

[4] Esling, J. H. 2005. There are no back vowels: The laryngeal articulator model. *Cdn. J. Ling.* 50, 13–44.

[5] Esling, J. H., Moisik, S. R., Benner, A., Crevier-Buchman, L. 2019. *Voice Quality: The Laryngeal Articulator Model*. Cambridge: CUP.

[6] Forrest, K., Weismer, G., Milenkovic, P., Dougall, R. N. 1988. Statistical analysis of word-initial voiceless obstruents: Preliminary data. *J. Acoust. Soc. Am.* 84(1), 115–123.

[7] Harrington, J. 2010. Acoustic phonetics. In: Hardcastle, W. J., Laver, J., Gibbon, F. E. (eds.), *The Handbook of Phonetic Sciences*. Wiley-Blackwell, 81–129.

[8] Shadle, C. 2023. Alternatives to moments for characterizing fricatives: Reconsidering Forrest et al. (1988). *J. Acoust. Soc. Am.* 153, 1412–1426.

[9] Bukmaier, V., Harrington, J. 2016. The articulatory and acoustic characteristics of Polish sibilants and their consequences for diachronic change. *J. Int. Phonetic Assoc.* 46, 311–329.

[10] Jannedy, S., Weirich, M. 2017. Spectral moments vs discrete cosine transformation coefficients: Evaluation of acoustic measures distinguishing two merging German fricatives. *J. Acoust. Soc. Am.* 142(1), 395–405.

[11] Soli, S. D. 1981. Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *J. Acoust. Soc. Am.* 70(4), 976–984.

[12] Stevens, K. N. 1998. *Acoustic Phonetics*. MIT Press.

[13] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Blackwell.

[14] Barreda, S. 2021. Fast Track: fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard*, 7(1). https://doi.org/10.1515/lingvan-2020-0051.

[15] Boersma, P., Weenink, D. 2022. *Praat: Doing phonetics by computer* [Computer program]. Version 6.2.23. http://www.praat.org/ (accessed 08.10.2022).

[16] R Core Team. 2022. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. https://www.R-project.org/.

[17] Harrington, J. 2010. *The Phonetic Analysis of Speech Corpora*. Blackwell.

[18] Pedregosa, F., Varoquaux, G., Gramfort, A. et al. 2011. Scikit-learn: Machine Learning in Python. *J. Machine Learning Research* 12, 2825–2830.

[19] Arkhipov, A. 2015. The acoustic correlates of vowel pharyngealization in Archi (East Caucasian). *Proc. 18th ICPhS* Glasgow, paper 1014.