SOUND CHANGE IN THE LAB: EXPOSURE-INDUCED VOCALIC CHAIN SHIFTS

Václav Jonáš Podlipský[1] and Paul Olejarczuk[2]

[1]Palacký University Olomouc, Czechia, [2]California State University, Chico, USA
vaclav.j.podlipsky@upol.cz, polejarczuk@csuchico.edu

ABSTRACT

Previous research on perceptual adaptation in vowels has attested generalization of retuning across vowel-height counterparts. Our study tested for a lab-induced chain shift. In a lexical decision task, we exposed U.S. English participants to words with either raised or lowered tokens of /i/, along with unmanipulated back vowels but no other front vowels. In subsequent categorization of the members of a [i]~[ɪ]~[ɛ] continuum, we found a difference between the raised- and lowered-/i/ participants in the location of the /i/-/ɪ/ boundary (replicating perceptual adaptation in vowels) as well as the /ɪ/-/ɛ/ boundary. Thus, the exposure-induced shift of /i/-/ɪ/ triggered a similar shift for the unexposed /ɪ/-/ɛ/ contrast, pointing to adaptation at the level of the vocalic system rather than the level of perceptual cues or low-level auditory compensation.

Keywords: perceptual recalibration, generalization of adaptation, chain shifts, English vowels

1. INTRODUCTION

Languages tend to converge on vocalic and consonantal systems with optimal (articulatory and/or perceptual) distance between categories. Theoretical accounts of this phenomenon include the Dispersion Theory [1, 2], Dispersion-Focalization Theory [3], Self-Organizing Systems Theory [4], or evolutionary phonological accounts [5, 6]. One diachronic process purportedly motivated by the pressure for distance optimization is the *chain shift*, where an imbalance in perceptual/articulatory space created by an initial shift of one sound category is rectified by subsequent, cascading adjustments of other sounds throughout the system. Amongst the well-known examples of historical chain shifts, where the contrasts between categories may have remained while the members of these contrasts changed, are the Great Vowel Shift affecting Middle English, the Cockney Vowel Shift, Grimm's Law, and the Romance intervocalic lenition. Many diachronic chain shifts are also now underway, including Californian /ɪ/, /ɛ/, /æ/ lowering [7, 8].

Explanations like Dispersion Theory, which appeal to the notion of perceptual space, in essence constitute hypotheses about individuals (since perception is an aspect of human cognition). It therefore follows that, if perceptual space prefers optimal distance between elements, we should be able to observe this preference in a controlled laboratory setting. Our study tests this prediction using the paradigm of immediate perceptual adaptation to non-canonical speech-sound realizations. Listeners routinely face the challenge of adapting to high phonetic variability in natural speech input. Past research has demonstrated that listeners retune their speech sound categories when exposed to shifted sound tokens in the lab. In a seminal paper, Norris et al. [9] showed that a sound halfway between [f] and [s] was perceived as /f/ or as /s/ depending on lexical context in Dutch, and that this category adaptation persisted in subsequent perception. Apart from lexical cues, adaptation has been shown to also be guided by visual cues [10], or even simply by the distribution of phonetic properties in the input [11, 12]. Most previous research focused on the recalibration of Cs (see e.g. the review in [13]) and studies on Vs have, until recently, been scarce [11, 12, 14, 15, 16].

Interestingly, perceptual adaptation can generalize to other categories within the system. For instance, Kraljic & Samuel [17] revealed that exposure to shifted VOT in English [d]-[t] led to shifted categorization of /d/-/t/ and also /b/-/p/. Chládková et al. [11] found that exposure to shifted Greek [i]-[e] resulted in adjusted categorization of /i/-/e/ and also /u/-/o/ (Greek has a symmetrical 5-vowel system). This indicates that category recalibration is distinct from the lower-level auditory compensation attested e.g. by Ladefoged & Broadbent's study [18] showing that the same [bVt] word was more likely perceived as /bɛt/ than /bɪt/ when it followed a synthetic carrier phrase with a lowered F1 making the F1 of [bVt] seem relatively higher, i.e. more appropriate for /ɛ/ than /ɪ/. This was replicated with natural speech [19], but also reversed speech [20, 21] and spectrally rotated speech analogs [22] as precursors.

Aiming at category recalibration rather than auditory compensation, this study (using a lexical decision task) exposed U.S. English listeners to words with the full high-to-low range of back Vs, but no front Vs except for /i/, whose tokens were lowered for one group of listeners and raised for another. Then, a

V categorization task tested for a potential difference between the groups in the location of the boundary between /i/ and /ɪ/ as well as the boundary between /ɪ/ and /ɛ/ (both latter absent from exposure). We asked whether a retuning of /i/~/ɪ/ would trigger a chain-shift reaction, namely adapted perception of /ɪ/~/ɛ/.

2. METHODOLOGY

2.1. Stimuli

All stimuli were based on sound-booth recordings of a male U.S. English speaker from Portland, OR. The initial phase, a lexical decision (LD) task, consisted of 208 exposure trials. These included 3× repetitions of 30 target items (English /i/-words, /i/ between obstruents, no other front Vs, none forming minimal pairs with /ɪ/-words, e.g. *cheese*, *fever*, *seafood*), 1×30 English word fillers (no front Vs, e.g. *assault*, *carbon*, *snuggle*), and 2×44 phonotactically plausible non-word fillers (no front Vs, dissimilar to words with front Vs, e.g. [ˈkupəɹə], [dəˈtɑɹ], [pʌlˈpun]).

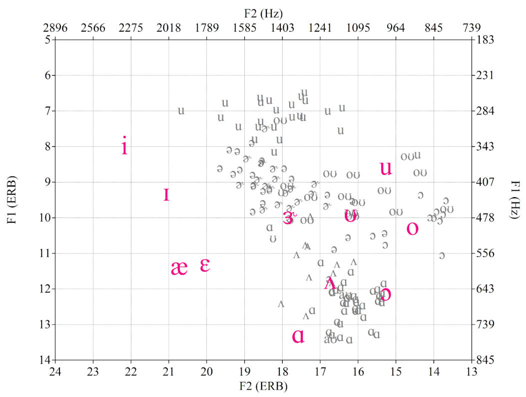

Figure 1: An F1 by F2 plot of all the Vs in the exposure words apart from /i/. U.S. English male means from [23] given for reference (in larger font).

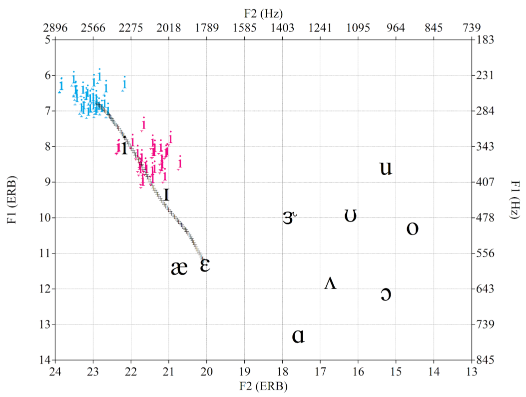

Figure 2: An F1 by F2 plot of all the /i/s in the exposure words, either lowered [i̞] (magenta) or raised [i̝] (cyan) artificially. The line of '+' symbols represents the 85-member [i]~[ɪ]~[ɛ] continuum used in the categorization task. U.S. English male means from [23] given for reference (in larger font).

Fig. 1 displays the F1 and F2 values of all the Vs contained in the exposure stimuli except for /i/, covering the full high-to-low range. Two resynthesized copies of each /i/-word were prepared, one with lowered [i̞] and the other with raised [i̝]. The resynthesis involved shifting the F1, F2, and F3 in Praat [24], using the method described in [11]. Table 1 gives the mean formant values of the original [i]s (not used in exposure) and the magnitude of each shift. The lowering was such that the resulting [i̞]s were between our speaker's prototypical /i/ and /ɪ/ but still more peripheral than /ɪ/ (see Fig. 2), and the raising ensured that the more extreme [i̝]s were without artifacts and still sounded natural. The shifted [i] tokens are shown in Fig. 2.

values in ERB	F1	F2	F3
Original [i] means	7.06	22.42	24.08
Lowering shift	+1.15	-1.11	-0.91
Raising shift	-0.575	+0.557	+0.455

Table 1: Mean F1, F2 and F3 frequencies of the [i] tokens before manipulation and the shift size for each direction and formant.

The stimuli for the post-exposure categorization task were 85 unique, isolated vowels forming a finely sampled [i]~[ɪ]~[ɛ] continuum, produced by resynthesizing (the same method as /i/ shifting) a single token of our speaker's naturally-produced [ɛ] (265 ms in duration). The qualities ranged between the [ɛ] endpoint with F1 = 11.22, F2 = 20.08 and F3 = 23.65 ERB, and the [i] extreme with F1 = 6.79, F2 = 22.9, and F3 = 25.46 ERB, with near-perfect interpolation for the steps in between. The test continuum is shown in Fig. 2.

2.2. Procedure

The whole experiment was implemented as a Praat Demo window script [24] and run on a computer with a headset in a quiet environment. Participants first completed the LD task, i.e. the exposure phase. On each of the 208 trials, they heard one of the LD stimuli (in random order) and decided whether it was an existing English word by clicking on either a button labeled 'word' or one labeled 'not a word' (button position random). They were offered a short break after every 52 trials.

After the exposure task and a short silent break, they completed a 3-alternative forced-choice categorization test. On each of the 85 trials, they heard a stimulus (random order) and identified the vowel by clicking on one of 3 buttons marked '[i] *beef*, *see*', '[ɪ] *kiss*, *kid*', and '[ɛ] *fresh*, *egg*' (button order random). They took a short break halfway through.

2.3. Participants

The participants were 65 native U.S. English speakers (51 female, 14 male), aged 19 to 44 years (mean 23), all college students who participated in exchange for course credit. They were randomly assigned to either the lowered-/i/ (*n* = 33) or the raised-/i/ exposure condition (*n* = 32). About ⅔ in each condition were Californians, most others came from Nebraska, 5 from other (mid)western states, and 1 from Texas. None reported any hearing impairments.

3. RESULTS

All (anonymous) data and the analysis script in R [25-28] as well as the experiment Praat script and all stimuli are available at https://osf.io/czu5g/. We first inspected responses on the exposure task. Overall, the lexical decisions were correct 91% of the time. Words with lowered [i̞] were identified as English words 93% of the time, words with raised [i̝] 95% of the time, indicating that the manipulated Vs were mostly indeed perceived as /i/. No participants were thus excluded due to frequent misidentification of the manipulated words as non-words.

Inspection of the V categorization data revealed that the numbers of responses were not balanced across categories, with most listeners giving fewer /ɪ/ than /i/ or /ɛ/ responses. Four listeners with 5 or fewer /ɪ/ responses (out of 85 per listener) were excluded from subsequent analyses, resulting in the numbers of participants in the lowered and raised conditions of 29 and 32, respectively. Fig. 3 is a tile plot showing all responses for the included participants. Fig. 4 shows the proportion of each response category for each stimulus, pooling data across participants. The figures suggest a somewhat higher proportion of /i/ responses in the lowered than in the raised condition, and in turn a higher proportion of /ɛ/ in the raised than in the lowered condition.

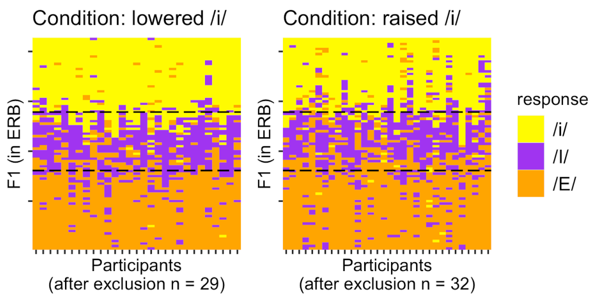

Figure 3: Tile plots displaying all categorization responses to the 85 stimuli (shown by their F1) in the two conditions. The dashed lines highlight the points of equal distance between category means, used for centering the F1 as predictor (see text). '/I/' = /ɪ/, '/E/' = /ɛ/.

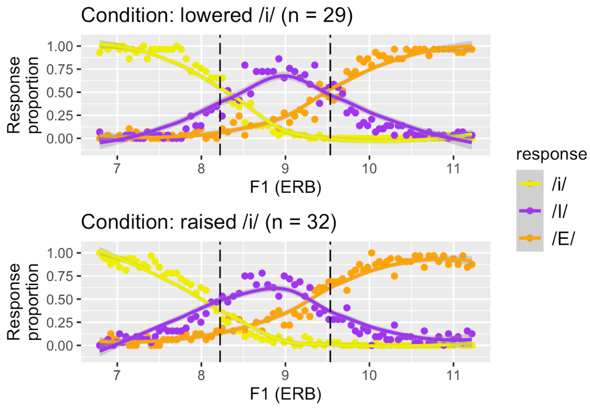

Figure 4: Scatterplots showing the proportion of each response category for each continuum stimulus (ranked by F1) across participants, with loess smoothing, split by condition. The dashed lines: same as Fig. 3.

The data were fitted to two mixed-effects logistic regression models, separately for the /i/-/ɪ/ contrast (excluding /ɛ/ responses from the data) and the /ɪ/-/ɛ/ contrast (excluding /i/ responses). Each model had Vowel Spectrum (centered) and Condition (sum-coded: lowered /i/ 1, raised /i/ -1) as the fixed effects and Participant as the random effect, with by-participant varying intercepts and slopes for Vowel Spectrum. Like in [11], F1 (in ERB) was used as a proxy of stimulus vowel spectrum (F2 and F3 steps being almost perfectly correlated with F1 steps across the continuum). Since the numbers of responses across categories were unbalanced (see above), the spectrum predictor was centered, differently for each contrast, to the points of equal distance between category means along the F1: first we computed mean F1 (in ERB) for responses elicited for each category and then the means of adjacent categories' means, shown as the dashed lines in Fig. 3 and 4, which served as the respective predictor centers.

Tables 2 and 3 give the models' coefficient estimates and Fig. 5 plots the fitted values for each model. The predicted probability of an /i/ response against /ɪ/ was 0.56 for the raised and 0.81 for the lowered condition (logit difference: +0.492, *SE* = 0.23, z = 2.14, p = 0.033), whereas the predicted probability of an /ɛ/ response against /ɪ/ was 0.74 for the raised and 0.67 for the lowered condition (logit difference: -0.206, *SE* = 0.093, z = -2.22, p = 0.026). The top panels of Fig. 5 also show the predicted category boundaries, i.e. points along the F1 continuum where *P(response)* = 0.5 for each Condition. Expressed as category boundary shifts, the effect of Condition was 0.18 ERB for the the /i/-/ɪ/ contrast and 0.26 ERB for the /ɪ/-/ɛ/ contrast. To visualize individual variability, the bottom panels of Fig. 5 plot by-participant logistic functions derived from the random effect coefficients of each model.

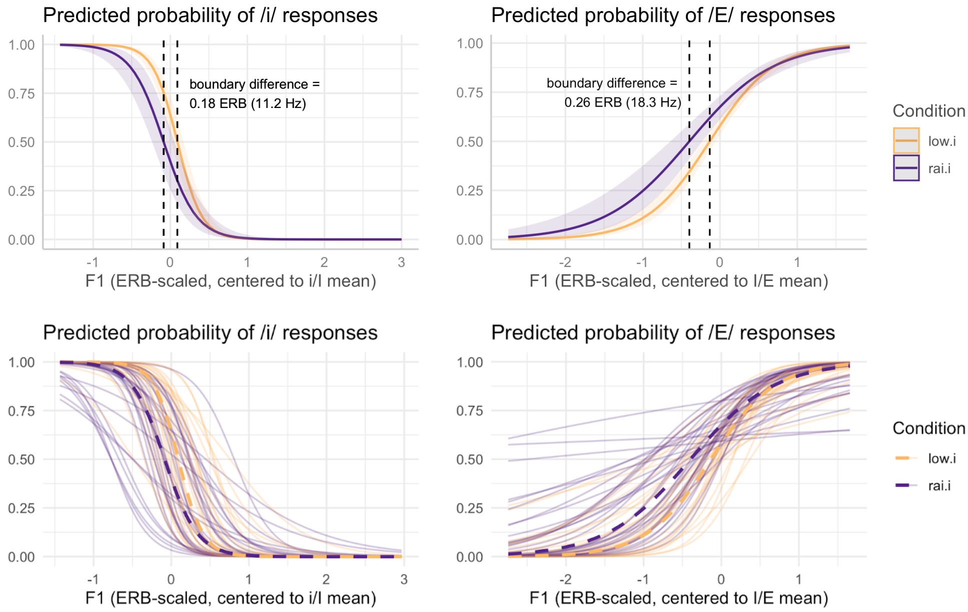

Figure 5: Values fitted by the logistic regression models for the /i/-/ɪ/ (left) and the /ɪ/-/ɛ/ contrast (right) split by Condition. Top panels show the overall logistic functions (repeated as thick dashed curves in the bottom panels) with dashed, vertical lines highlighting the corresponding category boundaries where *P(response)* = 0.5. Bottom panels show by-participant functions derived from the random effect coefficients.

	Estimate	SE	z	Pr(>|z|)
(Intercept)	0.088	0.235	0.375	0.707
F1_iɪ	-5.505	0.415	-13.260	**< 0.001**
Condition1	0.492	0.230	2.138	**0.033**
F1_iɪ : Condition1	-0.793	0.351	-2.255	**0.024**

Table 2: Coefficients estimated by the model of the probability of /i/ against /ɪ/ responses.

	Estimate	SE	z	Pr(>|z|)
(Intercept)	0.525	0.094	5.597	**< 0.001**
F1_ɪɛ	2.130	0.171	12.431	**< 0.001**
Condition1	-0.206	0.093	-2.221	**0.026**
F1_ɪɛ : Condition1	0.285	0.166	1.720	0.085

Table 3: Coefficients estimated by the model of the probability of /ɛ/ against /ɪ/ responses.

4. DISCUSSION

We exposed U.S. English listeners to words with either raised or lowered tokens of /i/, along with unmanipulated back vowels but no other front vowels. In the categorization test that followed, we found a difference between the raised- and lowered-/i/ listeners in the probability of the /i/ vs. /ɪ/ response and in the location of the /i/-/ɪ/ boundary in the expected direction, thus replicating the effect of perceptual recalibration for vowels documented previously [11, 12, 16].

At the same time, categorization differed in the expected direction for the /ɪ/-/ɛ/ contrast as well, although no tokens of these vowels had been presented in the exposure phase (unlike in [14, 15]), showing a generalization of the perceptual adjustment from /i/-/ɪ/ to an adjacent contrast. Previous studies found generalization of perceptual recalibration for vowels across vowel-height counterparts in Greek [11] and, using the shorter-term selective adaptation paradigm, also across vowel-frontness counterparts in German [29]. Our findings are indicative of chain-shift category retuning which has not been attested conclusively so far (cf. [14]).

Importantly, our results cannot be ascribed to speaker-specific formant cue calibration or auditory compensation as a general sensory mechanism of adjusting to the acoustic shifts of the stimuli (see [30] for these accounts) since all participants heard the same back Vs spanning the full high-to-low range, ensuring the two exposure conditions differed only in the realization of /i/ and not in the overall F1 span of the exposure stimuli.

Therefore, we interpret our findings as evidence of adaptation at the level of the vocalic system, induced in a controlled laboratory setting and potentially driven by a preference to maintain optimal distance between categories, whereby shifting one category triggered the shift of a neighboring category (though see [5, 31] for alternative explanations of the mechanisms underlying chain shifts). Assuming a link between perceptual processing and phonological categories, accumulated perceptual recalibration may thus serve as a mechanism of longitudinal chain shifts occurring in individual language users (e.g. [32]) and, by extension, in entire language communities.

4. ACKNOWLEDGEMENTS

This research was supported the Czech Science Foundation grant number 21-09797S.

5. REFERENCES

[1] Lindblom, B. (1986). Phonetic universals in vowel systems. In: Ohala, J., Jaeger, J. (eds.) *Experimental Phonology*. Academic Press, 13–44.
[2] Becker-Kristal, R. (2010). *Acoustic typology of vowel inventories and Dispersion Theory: Insights from a large cross-linguistic corpus*. PhD diss., UCLA.
[3] Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *J. Phonetics*, 25(3), 255–286.
[4] de Boer, Bart. 2001. *The origins of vowel systems*. OUP.
[5] Boersma, P., & Hamann, S. (2008). The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology*, 25(2), 217–270.
[6] Vaux, B., & Samuels, B. (2015). Explaining vowel systems: dispersion theory vs natural selection. *Linguist. Rev.*, 32(3), 573–599.
[7] Kennedy, R., & Grama, J. (2012). Chain shifting and centralization in California vowels: An acoustic analysis. *Am. Speech*, 87(1), 39–56.
[8] D'Onofrio, A., Pratt, T., & Van Hofwegen, J. (2019). Compression in the California vowel shift: Tracking generational sound change in California's Central Valley. *Lang. Var. Change*, 31(2), 193–217.
[9] Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychol.*, 47(2), 204–238.
[10] Reinisch, E., Wozny, D. R., Mitterer, H., & Holt, L. L. (2014). Phonetic category recalibration: What are the categories? *J. Phonetics*, 45, 91–105.
[11] Chládková, K., Podlipský, V. J., & Chionidou, A. (2017). Perceptual adaptation of vowels generalizes across the phonology and does not require local context. *J. Exp. Psychol. Human*, 43(2), 414–427.
[12] Colby, S., Clayards, M., & Baum, S. (2018). The role of lexical status and individual differences for perceptual learning in younger and older adults. *J. Speech Lang. Hear. R.*, 61(8), 1855–1874.
[13] Nelson, S., & Durvasula, K. (2021). Lexically-guided perceptual learning does generalize to new phonetic contexts. *J. Phonetics*, 84, 101019.
[14] McQueen, J., & Mitterer, H. (2005). Lexically-driven perceptual adjustments of vowel categories. *Proc. of ISCA Workshop on Plasticity in Speech Perception* London, 233–236.
[15] Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive Sci.*, 32(3), 543–562.
[16] Kim, D., Clayards, M., & Kong, E. J. (2020). Individual differences in perceptual adaptation to unfamiliar phonetic categories. *J. Phonetics*, 81, 100984.
[17] Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychon. B. Rev.*, 13(2), 262–268.
[18] Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *J. Acoust. Soc. Am.*, 29(1), 98–104.
[19] Ladefoged, P. (1989). A note on "Information conveyed by vowels." *J. Acoust. Soc. Am.*, 85(5), 2223–2224.
[20] Watkins, A. J. (1991). Central, auditory mechanisms of perceptual compensation for spectral–envelope distortion. *J. Acoust. Soc. Am.*, 90(6), 2942–2955.
[21] Watkins, A. J., & Makin, S. J. (1994). Perceptual compensation for speaker differences and for spectral-envelope distortion. *J. Acoust. Soc. Am.*, 96(3), 1263–1282.
[22] Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Constraints on the processes responsible for the extrinsic normalization of vowels. *Atten. Percept. Psycho.*, 73(4), 1195–1215.
[23] Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *J. Acoust. Soc. Am.*, 97(5), 3099–3111.
[24] Boersma, Paul & Weenink, David (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.20, retrieved 24/09/2022 from http://www.praat.org/.
[25] R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL https://www.R-project.org/.
[26] Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* 67, 1–48.
[27] Singmann, H., et al. (2022). Afex: Analysis of Factorial Experiments. R package version 1.1-1. https://CRAN.R-project.org/package=afex.
[28] Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag.
[29] Llompart, M., & Reinisch, E. (2018). Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels. *J. Phonetics*, 67, 34–48.
[30] Chodroff, E., & Wilson, C. (2020). Acoustic–phonetic and auditory mechanisms of adaptation in the perception of sibilant fricatives. *Atten. Percept. Psycho.*, 82(4), 2027–2048.
[31] Ettlinger, M. (2007). An exemplar-based model of chain shifts. *Proc. 16th ICPhS* Saarbrücken, 685–688.
[32] Harrington, J., Palethorpe, S., & Watson, C. (2000). Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts. *J. Int. Phon. Assoc.*, 30(1-2), 63–78.