GRZEGORZ KRYNICKI [1], GRZEGORZ MICHALSKI [2], JAROSŁAW WECKWERTH [3],
KAMIL KAŹMIERSKI [4], KATARZYNA DZIUBALSKA-KOŁACZYK [5]

Faculty of English, Adam Mickiewicz University, Poznań, Poland
[1] krynicki@amu.edu.pl
https://orcid.org/0000-0001-6015-1423
[2] grzegorz@amu.edu.pl
https://orcid.org/0000-0001-7268-7372
[3] wjarek@amu.edu.pl
https://orcid.org/0000-0001-9859-1571
[4] kamil.kazmierski@amu.edu.pl
https://orcid.org/0000-0002-5390-0964
[5] dkasia@amu.edu.pl
https://orcid.org/0000-0003-4884-3448

Grzegorz Krynicki – Grzegorz Michalski – Jarosław Weckwerth – Kamil Kaźmierski – Katarzyna Dziubalska-Kołaczyk: Are Polish affricates retroflex? A cross-linguistic electropalatographic study with a redesigned Reading-type palate
Alkalmazott Nyelvtudomány, XXII. évfolyam, 2022/1. szám, 223–242.
doi:http://dx.doi.org/10.18460/ANY.2022.1.014

Are Polish affricates retroflex? A cross-linguistic electropalatographic study with a redesigned Reading-type palate

Traditional descriptions of Polish consonants usually classify the phonemes /t͡ʃ d͡ʒ/ as laminal alveolar affricates, and /ʃ ʒ/ – as apical palato-alveolar but use symbols standardly used for post-alveolar consonants. However, there is a growing body of work treating them as retroflex based on their articulatory, phonetic and phonological characteristics. This view has been increasingly influential and is reflected in the increasingly widely accepted phonemic transcription forms /ʂ ʐ/ and /t͡ʂ d͡ʐ/. In this study, we consider the Polish apical postalveolar affricates /t͡ʂ d͡ʐ/ from the point of view of the articulatory properties of retroflex sounds cited in the literature. We contrast them with the articulatory properties of their English counterparts /tʃ dʒ/, which are considered apico-laminal palato-alveolars. The articulatory properties of the Polish and English affricates are analysed on the basis of a corpus collected from 3 Polish and 2 English native speakers using a standard 62-electrode EPG (electropalatographic) system with re-designed extra-thin Reading-type palates.

Keywords: Polish, English, affricates, retroflex, electropalatography

1. Introduction

Extant descriptive literature discusses a number of features of Polish and English consonant articulations that involve differences in the locations of tongue-to-palate contact and differences in the shape of the articulators. One of these features is the retroflex character of [t͡ʂ d͡ʐ ʂ ʐ].

In traditional descriptions, [tʃ dʒ ʃ ʒ] are treated as postalveolars with a flat posture of the central part of the tongue (Rubach 1984, 24, after Wierzchowska

1963), laminal alveolars (Jassem 2003, 104) or (post-)alveolars in general (Gussmann 2007, 7, 119).

Since the observation by Keating (1991, 36) that – based on linguagrams, palatograms and X-ray presented by Wierzchowska (1965, 1967, 1980) for the Polish fricatives and the corresponding affricates – the stop portions of the affricates are clearly apical, possibly partly sub-laminal, and therefore possibly retroflex, a number of articulatory, acoustic, perceptual and phonological arguments have been put forward in support of the claim that these phonemes are in fact retroflex (Lorenc, 2018, p. 132). The articulatory arguments are based on a range of properties common to many retroflex articulations (Hamann 2003, 32 ff.) and include their sublingual cavity size, apicality, retraction and the flapping out gesture that often accompanies them.

The **sublingual cavity size** was first observed in the articulation of Polish retroflex /ʂ ʐ/ and alveolo-palatal /ɕ ʑ/ fricatives by Halle & Stevens (1989, cit. after Keating 1991, 43), and measured to be on average approx. 13.5mm for /ʂ ʐ/ and 18.5mm for /t͡ʂ d͡ʐ/ (Lorenc 2018, 145).

Another common property of retroflex articulations, **apicality**, occurs when the active articulator is either the tip of the tongue (apical articulation) or its underside (subapical articulation). Polish /t͡ʂ d͡ʐ ʂ ʐ/ were considered as apical in the works of Wierzchowska (1980, 64). In more recent literature, it has been suggested that /t͡ʂ/ is apical because it has a long closure followed by short frication (Żygis, 2006, p. 105); only the tip was reported to be used in all 170 realisations of [ʂ] by 20 Polish natives as measured by EMA (electromagnetic articulography) in Lorenc (2018, p. 143).

Retracted articulation means that the tip is moved upwards and into the “posterior” region, and the tongue pulls backward, thus lowering its centre and retracting its back to the extent that there may be some velarisation or pharyngealization. Hamann (2003, 36) considers the shape of the tongue centre to be one of the main differences between retroflex and palato-alveolar or alveolo-palatal articulations: in the former, it is concave, and in the latter – convex (domed). Lorenc (2020, p. 173) found that the height of the sublingual cavity in Polish /t͡ʂ/ was greater than in Polish /t/ by about 6mm, even though this may be simply a reflection of the difference in POA, (post-)alveolar for /t͡ʂ/ vs. post-dental for /t/.

Retroflex realizations are also typically articulated at a **posterior** (Hamann 2003, 33: post-alveolar or palatal; Lorenc 2018, 135: from alveolar to palatal) place of articulation; based on Wierzchowska’s x-rays, Hamann (2003, 41) observes that /ʂ/ is articulated at the alveolar rather than any posterior place of articulation and this is also how Wierzchowska classifies this phoneme; a parallel observation is made by Żygis for /t͡ʂ/ (2006, 75). Hamann (2003, 41) acknowledges that although Polish [ʂ] does not meet the posteriority criterion as one of the four parameters of retroflexion, it is still retracted in the sense of being lowered and backed, it is apical, and it has a sublingual cavity, and that therefore it can still be classified as retroflex. As the recent EMA studies show, the fricative phase of /t͡ʂ/ occurs usually at the alveolar place of articulation (93.8%, Lorenc 2018, 143) whereas the closure phase occurs usually at the alveolar (52.2%) or postalveolar POA (41.6%). The fact that 41.6% of the /t͡ʂ/ realizations display the forward movement from the posterior to the anterior place of articulation, may in itself provide support for the retroflex character of these realizations as they may be a manifestations of the "flapping out" characteristic to retroflex.

Finally, the **flapping out** gesture, i.e., some type of movement of the tongue tip from a posterior to a more anterior position during the articulation is often though not universally associated with retroflex realizations (Hamann 2003, 32-34). The gesture is claimed to be more extensive in the context of an open vowel like /a(ː)/ than a close vowel like /i(ː)/ (Simonsen et al., 2008, p. 387). Flapping out has also been observed to look different for different manners of articulation, e.g., in fricatives it is found at the onset of the following segment, while in plosives it occurs during the closure phase. Because of this irregularity, Hamann (2003, 34) does not consider this feature a separate property of retroflexion. In this study, we do treat it as a separate criterion.

The methodology of this study allowed us to investigate only the last three of the five defining characteristics of prototypical retroflex consonants: retraction, posteriority and flapping out.

We considered the Polish /t͡ʂ d͡ʐ/ from the point of view of common articulatory properties of retroflexion listed above against the articulatory properties of their non-retroflex counterparts in Polish /t t͡ɕ ts d d͡ʑ dz/ and against their apico-laminal palato-alveolar English counterparts /t tʃ d dʒ/ (Cruttenden, 2014, p. 185).

2. Methodology

The articulatory properties of these sounds were analysed with the use of electropalatography (EPG). A standard EPG system was combined with extra-thin (~0.91 mm) Reading-type palate of new design that aims at minimizing the perceptible disturbance of speech, increasing the wearer's comfort, reducing adaptation time, while preserving the relatively high comparability of the EPG patterns obtained from different speakers due to the anatomical layout of electrodes. The EPG data were recorded through the *Linguagraph* EPG multiplexer by *Medical Solutions Ltd* (Kelly et al., 2000) using the normalized anatomical Reading-type palates. They were tested and proved to work with the multiplexer like the original manufacturer's palate. The sampling frequency was 100 Hz.

2.1. Language material: The FEAMU corpus

Over 6 hours of read speech recorded from 3 Polish and 2 English native speakers was time-aligned with articulatory EPG data and collected into a searchable articulatory corpus which we named FEAMU[1]. To collect the corpus, we recorded 2 male (M1, M2) and 1 female (F1) speakers of Standard Polish. M1, one of the authors of this paper, was one of the contributors (c.f. e.g., Moen et al. 2011, 1399 where 2 out of 4 contributors were the authors of the paper). The recordings were made before the formulation of any of the hypotheses put forward in this paper.

Recordings were conducted in quiet office conditions. For the audio, two cardioid microphones were used: one dynamic (Shure SV200) and one condenser (MXL 770 Mogami). The data were recorded in 32-bit samples at a rate of 44.1 kHz, later down-sampled to 22kHz for forced alignment.

2.2. Corpus texts and labelling

The FEAMU corpus contains read speech only. The Polish scripts were based on several sources. First, there were phonetically balanced lists from two earlier projects, PELT (the *Polish-English Literacy Tutor* by Dziubalska-Kołaczyk, Bogacka, Wypych, Pietrala, & Krynicki 2006) and HINT (the *Hearing in Noise Test* by Śliwińska-Kowalska, Kotyło, & Sol 2013). Second, we used warm-up texts from a *Course of English phonetics for Poles* developed for the purposes of a project on whose partial results the present study reports. The central novel component of the course was the presentation of visual feedback to the participants (Polish learners of English) of real-time articulatory feedback in the form of a visualization of EPG data.

The English script contained 460 phonetically balanced sentences used in MOCHA-TIMIT; 14 sentences, 110 English words and phrases as well as 120 nonce-words used in ESPRIT-ACCOR; and 1311 words from the English part of the above-mentioned *Course of English phonetics for Poles*.

FEAMU was aligned at the phoneme level using the *Montreal Forced Aligner* (MFA, McAuliffe et al. 2017). For Polish, MFA was run with the phonetic dictionary generated by *phonetizer* (Wypych 2011, 114; Szczyszek 2013). No manual corrections were introduced to the output of the automatic aligner except for files used for evaluation purposes: for Polish and English, one TextGrid each was compared before and after manual correction by a skilled annotator to calculate the mean displacements of segment boundaries. The results of the evaluation are given in Table 1.

[1] Acronym from the Faculty of English at Adam Mickiewicz University, where the authors are affiliated. Corpus samples available at https://fon.wa.amu.edu.pl/epg.

Table 1. Accuracy metrics of MFA forced alignment: mean displacement in milliseconds between phoneme boundaries placed by the MFA aligner and a human annotator

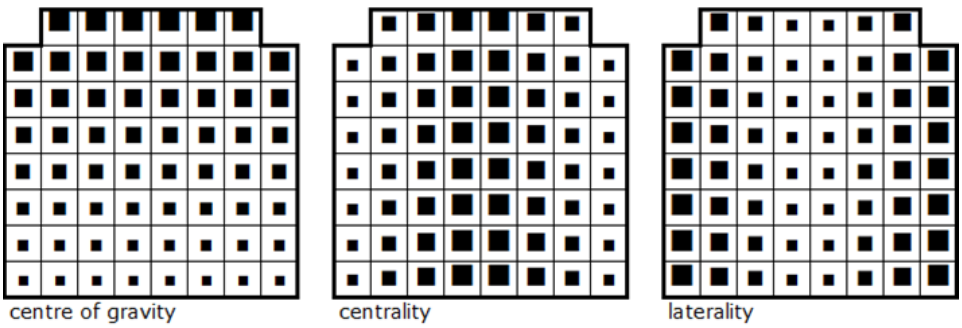
	English	Polish
Number of measurements compared	62	138
Mean displacement in ms	0.0166	0.0270
Displacement SD	0.0162	0.1443
Auto boundary within 10ms of human boundary	48.39%	78.26%
Auto boundary within 20ms of human boundary	83.87%	91.30%
Auto boundary within 30ms of human boundary	93.55%	94.20%

Although the methodology of the alignment quality test is limited (MacKenzie & Turton, 2020), it was deemed sufficient as a practical indicator that the alignment quality in both language parts of the corpus is consistent and acceptable.

For the English data, MFA was used with the British English MRC Machine Usable Dictionary v.2.00 (*MRC Psycholinguistic Database* 1997).

The corpus data was annotated for linguistically relevant articulatory features derived from EPG. A single EPG frame may be represented by a two-dimensional array in which binary values represent electrode activations. A time series of frames may be represented as a three-dimensional array. To make the arrays more manageable and relate them to linguistically meaningful articulatory features, a number of data reduction methods have been proposed in the literature (Hardcastle, Gibbon, & Nicolaidis 1991, 259; Carreira-Perpiñán and Renals 1998, 262; Gibbon & Nicolaidis 1999; Harrington 2010, 161).

Figure 1. Contact distribution indices: sums of rows or columns multiplied by different weights.



The reduction methods used in the current study included contact distribution indexes (whereby sums of specific rows or columns are multiplied by different weights before summation), as well as parameters derived from multiple EPG frames, including the direction of tongue movement during closure.

The contact distribution indexes were the **Centre of Gravity** and the **Centrality Index**. The Centre of Gravity (COG, Gibbon & Nicolaidis 1999, 238) is a measure of the frontness of the place of articulation (POA). The more advanced a given place of articulation (POA) is, the closer its COG value is to the maximum value of 7.6 (for which only row 1 of EPG frame is filled); the more retracted the POA, the closer its COG is to the minimum of 0.5 (for which only row 8 is filled). The **Centrality Index** (CI, Harrington 2010, 167) measures the extent of contact at the centre of the palate and varies between 0 and 1. The more the contacts are laterally distributed, the lower the value of CI.

2.3. A comparison of the Polish and English sections of FEAMU

The corpus contains a total of 06:07:48 hours of read speech, making it the second largest corpus containing EPG data available, after ESPRIT-ACCOR (07:38:55 hours; c.f. MOCHA-TIMIT which, as available today, contains 1:32:21hrs).

Tables 2 and 3 summarize the Polish and English parts of FEAMU side by side.

Table 2. Text, frame and phoneme statistics for the FEAMU corpus

	Polish	English
EPG frames	889032	809703
EPG frames with recognized phoneme	795506	467212
Phonemes	102774	42750
Tokens	11428	5136
Types (unique tokens)	4935	3013

Table 3. Speaker contributions to the FEAMU corpus.

Polish			English		
speaker	utterances	minutes	speaker	utterances	minutes
M1 (male)	905	1:25:57	M3 (male)	581	1:12:15
M2 (male)	894	1:40:32	M4 (male)	667	1:37:16
F1 (female)	115	0:11:48	total	1248	2:49:31
total	1914	3:18:17			

Within each language, the speakers read the same list of words and sentences. The female Polish speaker F1 read the first 10% of the script.

2.4. The FEAMU corpus vs. other corpora with EPG

Two major corpora containing articulatory EPG data are available for English: ESPRIT-ACCOR and MOCHA-TIMIT. They represent a wide range of speakers and are based on phonemically normalized scripts. Polish EPG data have been previously collected for the purpose of several studies (Pompino-Marschall &

Żygis 2003; Guzik & Harrington 2007; Harrington 2010; Bukmaier 2017), but, to the best of our knowledge, there are no publicly available EPG data for Polish except for the samples of Polish EPG articulations that accompany the EMU package (Winkelmann et al., 2017).

ESPRIT-ACCOR (Marchal & Hardcastle 1993; 'ACCOR-English' n.d.) is an acoustic and articulatory database recorded as part of the ESPRIT-ACCOR project investigating cross-language acoustic-articulatory correlations in coarticulatory processes. Southern British English, which is the focus of the current study, is represented by 5 speakers although data from one speaker's data (SN) is of limited use because of the blurred EPG image on most frames. ESPRIT-ACCOR covers 120 nonce words, 110 real words and 14 short sentences illustrating major connected speech processes. The original script was read 10 times by each speaker. The EPG used in ESPRIT-ACCOR was the Reading system (Jones & Hardcastle, 1995) with a sampling rate of 200 Hz. The English section of the corpus contains 7:38:55 hours of recordings.

MOCHA-TIMIT (Wrench & Hardcastle 2000; 'MOCHA-TIMIT' n.d.) was acquired in a sentence-reading task by 3 speakers at the Centre for Speech Technology Research, Edinburgh University. As in ESPRIT-ACCOR, the Reading EPG system was used at a sampling frequency of 200Hz. The speakers read a set of 460 British TIMIT-style sentences designed to provide phonetically diverse material and capture English connected speech processes with good coverage (Garofolo et al., 1993). The corpus as it is available today contains 1:32:21 hours of speech.

3. Hypotheses

Based on the literature presented in the Introduction, we hypothesized that, if the Polish /t͡ʂ d͡ʐ/ are to be understood as retroflex, then they will display less side activations than the alveolo-palatal /t͡ɕ d͡ʑ/ reflecting their concave rather than convex tongue shape (retraction) and apicality. They will also be articulated with backer contact locations than /t d/ reflecting their postalveolar rather than post-dental place of articulation (posteriority). Finally, they will be produced with a flapping out motion from the posterior to anterior position during the closure phase.

4. Results

We considered our results from the perspective of three properties common to many retroflex articulations, namely the degree of retraction, posteriority and flapping out.

4.1. Retraction

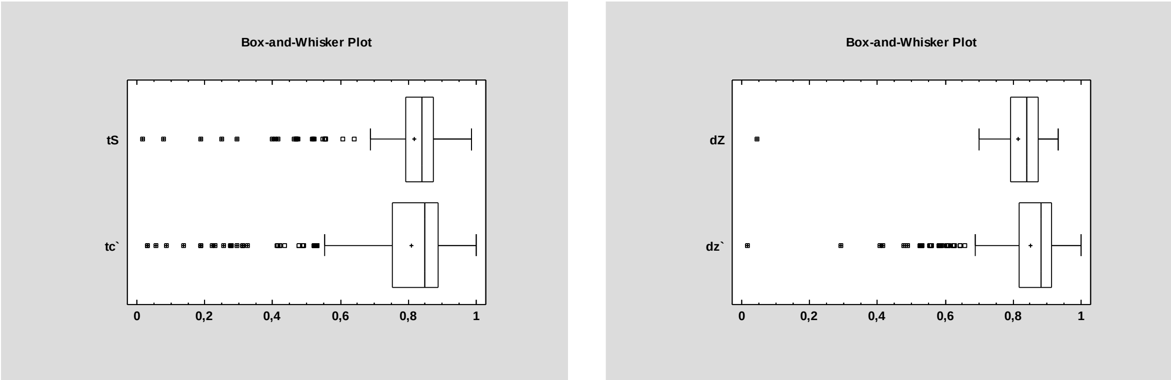
We followed Hamann (2003, 36) in the assumption that the concave vs. convex tongue centre shape distinction is an important predictor of the distinction between the retroflex vs. non-retroflex articulations. To test the possibility of capturing the concave/convex tongue shape difference, as a data reduction index, we utilised the Centrality Index (CI). The choice of the index was based on the premise that bunching of the tongue centre (convex tongue shape) will generate roughly parallel lines of activations in the outermost columns of the EPG frame, thus increasing the number of side activations, whereas concaving of the tongue centre for retroflex articulation will decrease the number of side activations. The *Centrality Index* was used by Harrington (2010, 170) to distinguish between consonants that have a wide vs. narrow central groove, in particular, to distinguish Polish fricative /ʃ/ (our /ʂ/) (low CI, wide groove) from /s ɕ/ (mid or high CI, narrow groove). Based on a set of 10 palatograms, he observed that /ʃ/ showed a lower number of contacts in the columns 4–5 and therefore a lower CI than /s ɕ/.

Table 4. Centrality Index as a possible indicator of convex vs. concave shape of the tongue centre

	Mean maximum CI in all segment frames				CI in middle frame of segment			
	t͡ʂ	t͡ɕ	d͡ʐ	d͡ʑ	t͡ʂ	t͡ɕ	d͡ʐ	d͡ʑ
No. of segments	1174	1582	76	658	1173	1581	76	657
CI	0.82	0.80	0.82	0.85	0.63	0.65	0.61	0.73
SD	0.08	0.13	0.10	0.10	0.22	0.17	0.21	0.17
W-test	W = 1.06, p <.001		W = 36339, p <.001		W = 969170, p = .05		W = 34562, p <.001	

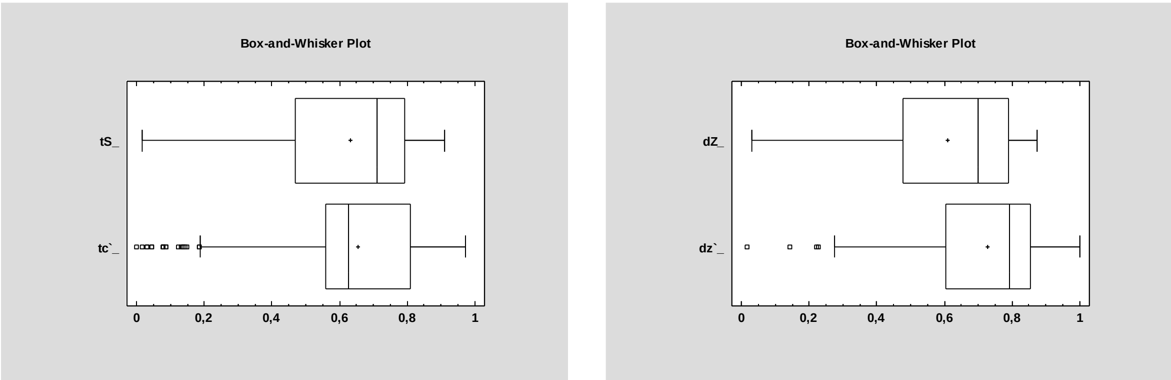
In our data, however, the difference between the centrality index of /t͡ʂ/ vs. /t͡ɕ/ is not obvious (Table 4). It is more visible in e.g., /d͡ʐ/ vs. /d͡ʑ/. In all samples, the standardized skewness and kurtosis values were outside the normal range, which would tend to invalidate the results of the parametric t-test. For this reason, the Mann-Whitney (Wilcoxon) W-test was also used to compare medians. W-test confirmed the hypothesis at the 0.001 level of significance. W-test also confirmed that /d͡ʑ/ is more central than /d͡ʐ/, which is consistent with our hypothesis.

Figure 2. Plots for mean maximum CI in all segment frames. tS = /ʈʂ/; tc` = tɕ; dZ = /ɖʐ/; dz` = /dʑ/



Results based on peak CI measured across a given affricate (Figure 2) show the CI during the stop phase, since it is the stop component of the affricate that activates more central electrodes than its fricative phase: M = 0.82, SD = 0.11 for all four affricates, compared to M = 0.66, SD = 0.20 for CI measured in the middle frame of the affricate, when it is likely that the fricative phase is taking place (Figure 3).

Figure 3. Plot for CI in middle frame of segment



Considering the large number of outliers visible in Figure 2 and Figure 3, and the small number of cases representing /ɖ͡ʐ (76, viz. Table 4)/, the above results should be considered as inconclusive.

The limited applicability of CI for distinguishing /ʈ͡ʂ ɖ͡ʐ/ from /t͡ɕ d͡ʑ/ using the number of side activations results from the fact that CI attributes high weights to two middle columns, within which the differences between these phonemes are subtle. A visual inspection of the greyscale images in Figures 4 and 5 reveals that the majority of the difference between /ʈ͡ʂ/ and /t͡ɕ/ resides in columns 2 and 7.

Figure 4. A greyscale image from 1174 frames of /t͡ʂ/ at the mid-point of segment duration

0	.14	.22	.14	.14	.23	.3	0
.62	.54	.54	.39	.2	.53	.51	.73
.95	.9	.38	.21	.21	.34	.75	.96
.99	.73	.1	.05	.03	.11	.7	.99
1	.18	.01	0	0	.01	.46	1
1	.17	.01	0	0	0	.18	1
1	.32	0	0	0	.01	.2	1
1	.76	.27	.34	.3	.09	.86	1

Figure 5. A greyscale image from 1582 frames of /t͡ɕ/ at the mid-point of segment duration

0	.41	.31	.2	.08	.17	.33	0
.96	.84	.73	.34	.12	.62	.7	.94
.99	.98	.79	.1	.17	.68	.91	.99
.99	.98	.44	.02	.03	.67	.97	1
1	.98	.45	.01	.02	.53	.97	1
1	.98	.34	.01	.01	.47	.98	1
1	.89	.03	.01	0	.4	.95	1
1	.87	.3	.38	.32	.18	.97	1

In the 8 × 8 rectangle in Table 5, each cell contains the result of subtracting the value of a cell for /t͡ʂ/ from the corresponding cell for /t͡ɕ/; e.g., for the cell in the 1st row and 2nd column, 0.41 – 0.14 = 0.27. In the row below the 8 × 8 rectangle, the averages for the columns are given, while the averages for the rows are given in the rightmost column. The average difference for all 62 cells is 0.17. The bulk of the difference is focused in columns 2 and 7 and rows 5 and 6.

Table 5. The result of subtracting a cell for /t͡ʂ/ from the corresponding cell for /t͡ɕ/

	0.27	0.09	0.06	-0.06	-0.06	0.03		0.06
0.34	0.30	0.19	-0.05	-0.08	0.09	0.19	0.21	0.15
0.04	0.08	0.41	-0.11	-0.04	0.34	0.16	0.03	0.11
0.00	0.25	0.34	-0.03	0.00	0.56	0.27	0.01	0.18
0.00	0.80	0.44	0.01	0.02	0.52	0.51	0.00	0.29
0.00	0.81	0.33	0.01	0.01	0.47	0.80	0.00	0.30
0.00	0.57	0.03	0.01	0.00	0.39	0.75	0.00	0.22
0.00	0.11	0.03	0.04	0.02	0.09	0.11	0.00	0.05
0.05	0.40	0.23	-0.01	-0.02	0.30	0.35	0.04	0.17

The modification of the CI offered by Simonsen et al. (Simonsen et al., 2008, p. 391) or the use of most other standard contact indexes, e.g., the Laterality Index (LI, Carreira-Perpiñán and Renals 1998, 262) would also probably fail to capture the concave/convex distinction. All those indexes measure articulatory differences along a single dimension (e.g., CoG – front–back, LI – left–right in the transverse plane) whereas the concave 3-dimensional tongue shape, when mapped onto the 2-dimensional EPG frame, seems to vary in both dimensions, as suggested by the elliptical shape formed during the articulation of /t͡ʂ/ by the missing contacts at the crossing of rows 6–8 and columns 4–5 (Figure 4). Such a shape may result from a concave tongue centre and stands in contrast to the roughly parallel lines of activations on the four outermost columns, which may result from the bunching of the tongue centre.

A two-dimensional metric should therefore be used to capture these distinctions. Such a metric could be related to the number of activations in that elliptical area, which roughly overlaps with the region identified by Barry (1992, 394) as correlated with palatalization in Russian (delineated by the thick line in Table 5). Visual inspection of the articulations suggests the possibility of determining the concave tongue shape by the low number of activations in the palatal region and high number of activations in columns 1 and 8. However, even a two-dimensional metric on its own probably could not unambiguously determine the concave/convex shape of the tongue.

The application of any of such metrics would require that it is mapped onto the concave/convex categories and calibrated against a training set where this distinction was grounded through instrumentation other than EPG, following e.g., the approach presented by Simonsen et al. (2008, 391).

Therefore, the metric we used here does not allow us to unequivocally determine whether the shape of the tongue during the articulation of /t͡ʂ/ is concave and convex in /t͡ɕ/ but visual inspection of the greyscale images would suggest that it is.

4.2. Posteriority (closure fronting)

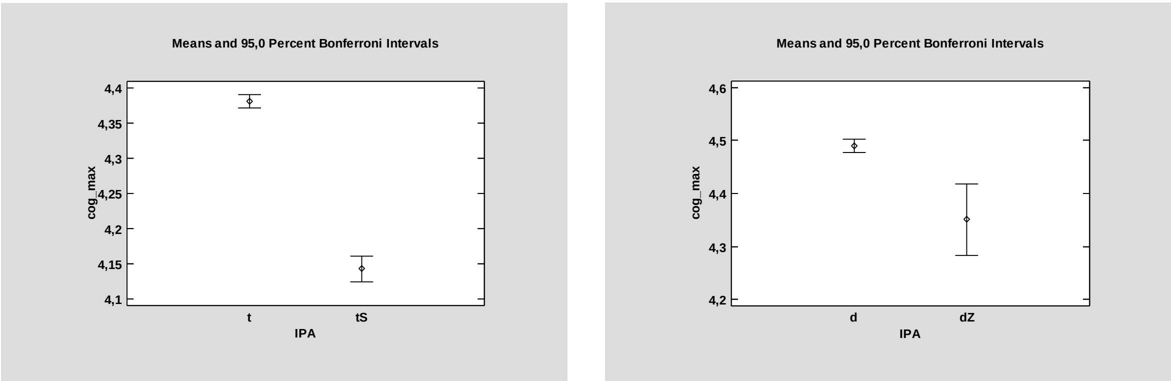
One of the contact indices that quantifies the distribution of the place of articulation between the front and back of the palate is the centre of Centre of Gravity (COG, Gibbon & Nicolaidis 1999, 238). One-Way ANOVA was conducted to compare the mean values of the maximum COG within a given segment. The procedure was conducted twice, first for /t/ vs. /t͡ʂ/ and then for /d/ vs. /d͡ʐ/. For both pairs, the F-test showed significant differences between their mean COG's at the 5% significance level.

Table 6. Centre of Gravity contrasted between Polish /t/ vs /t͡ʂ/ and /d/ vs. /d͡ʐ/

	t	t͡ʂ	d	d͡ʐ
number of segments	4467	1174	2251	76
mean max. COG in	4.38	4.15	4.49	4.35
SD	0.44	0.47	0.42	0.48
F-test	$F(1, 5585) = 259.40, p < 0.001$		$F(1, 2306) = 7.76, p < 0.01$	

As a post-hoc test, Bonferroni’s multiple comparison procedure was selected. The means plots below illustrate the mean COG for each pair and the standard error of each mean. They indicate that in each pair, the stop has a significantly higher COG and therefore is more advanced that its affricated counterpart.

Figure 6. Post-hoc comparison of Polish /t/ vs /t͡ʂ/ and /d/ vs. /d͡ʐ/ in terms of COG and therefore front vs. back place of articulation



This shows that Polish /t͡ʂ d͡ʐ/ display significantly backer contact locations than /t d/.

To identify the exact place of articulation for the segments in each pair, first the time span of each segment was identified, the corresponding EPG frames were extracted, and the frames showing full closure in rows 1–5 were identified. Segments without full closure in the relevant rows were disregarded, and the average row number where the mid-point of the full closure occurred was recorded. This way, we could pinpoint the closure to either the postalveolar or the post-dental region.

The span of rows selected above was restricted to avoid, in the final frames of a postalveolar phoneme, mistaking the velar closure that may be derived from following velar stop as inherent to the postalveolar phoneme, e.g., *miasteczko, wszystkich, odgórnego*. The selected rows included rows 1–5 (Figure 9). Row 1 was assumed to correspond to the postdental POA; row 2 – to the alveolar POA; and rows 3–4 – to the postalveolar POA. The status of row 5 varies in the literature. It is postalveolar in e.g. Wrench (Wrench, 2007, p. 5) and palatal in Harrington (2010, 153). However, an inspection of the occurrences of palatal /k/ and /g/ in the FEAMU corpus with the closure occurring on M = 6.68 row, with a relatively narrow SD = 0.42, indicated that including row 5 should be still considered postalveolar and useful for correct classification of closures in postalveolar phonemes in palatalized contexts.

Figure 7. The phonetic zoning scheme adapted after (Gibbon & Nicolaidis, 1999, p. 234; Wrench, 2007, p. 2) except that in rows 1–2 we make the distinction into (post)dental and alveolar articulations instead of grouping them into one alveolar zone

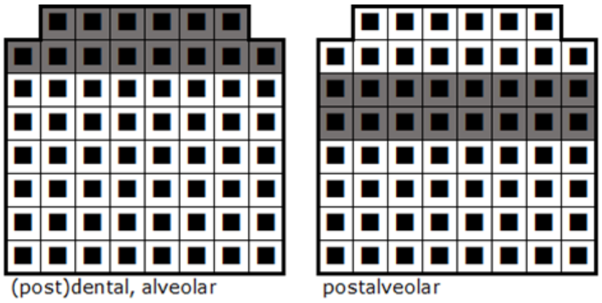
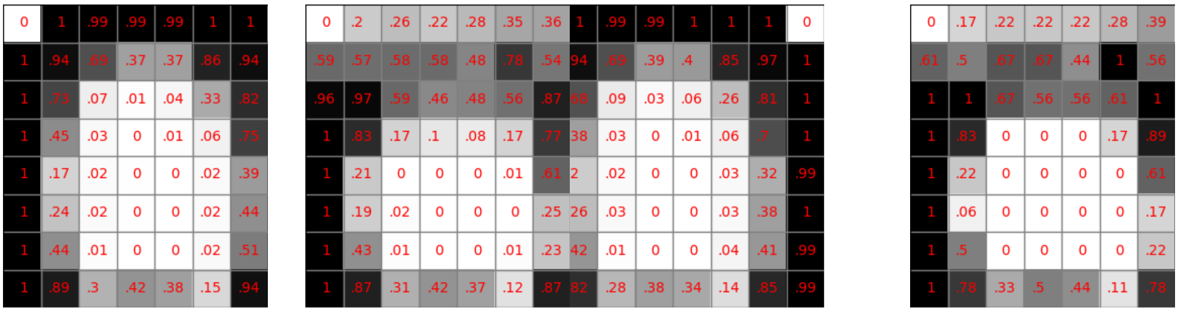


Table 7 summarizes the results for mean closure location for Polish /t d/ vs. /t͡ʂ d͡ʐ/. In that table, *all segments* was calculated as all occurrences of the phoneme in the header of the column as accounted for in the Polish section of the FEAMU corpus, *with closure* was calculated as the number of segments with at least one frame of full constriction anywhere in rows 1-5, *with closure %* was calculated as the percentage of segments with full constriction in rows 1-5 relative to all segments; *with continuous closure* was calculated as the number of segments with at least one frame of closure anywhere in rows 1-5 and with a closure in the frame that is in the middle between the first and the last frame with closure in the given segment; *with continuous closure %* was calculated as the percentage of segments with continuous closure among all segments; finally, *mean row with closure* was calculated as the average number of the row where closure occurred in the frame in the middle of the closure in the given segment.

Table 7. Mean row with closure used to identify the place of articulation in Polish /t d/ vs. /t͡ʂ d͡ʐ/

	t	t͡ʂ	d	d͡ʐ
all segments	4467	1174	2250	76
with closure	3522	963	1995	58
with closure %	78.84%	82.03%	88.67%	76.32%
with continuous closure	2270	392	1488	18
with continuous closure %	50.82%	33.39%	66.13%	23.68%
mean row with closure	1.11	2.32	1.15	2.31
SD	0.29	0.83	0.30	0.79
W-test	W = 798483, p = 0		W = 23566, p < 0.001	

Figure 8. Greyscale images from 4467 segments of /t/, 1174 segments of /t͡ʂ/, 2250 segments of /d/ and 76 segments of /d͡ʐ/ respectively



Two comparisons of independent samples were conducted for average row with closure for /t t͡ʂ/ vs. /d d͡ʐ/. The samples for /t/, /d/ and /t͡ʂ/ had standardized skewness and standardized kurtosis values outside the normal range, which invalidated tests comparing standard deviations; therefore, the Mann-Whitney W-test was used to compare the medians. For /t t͡ʂ/, W = 798483, P-value = 0, therefore the null hypothesis that medians of both samples are equal was rejected. Likewise, for /d d͡ʐ/, W = 23566, P-value = 0, which also showed a statistically significant difference between the medians at the 95% confidence level.

The results of the test above indicate that Polish /t d/ have a significantly different place of articulation from /t͡ʂ d͡ʐ/. The average closure of /t d/ at 1.11–1.15 (close to the 1st row, at 11–15% of the distance between the 1st and 2nd row) can be understood as a post-dental place of articulation of these phonemes. The average closure of /t͡ʂ d͡ʐ/ at 2.32–2.33 (behind the second row) indicates postalveolar articulation.

4.3. Flapping out

The movement of the tongue tip from a posterior to a more anterior position during the articulation of retroflex sounds is referred to as “flapping out” (Hamann 2003, 34). This feature has been observed to be different for different manners of articulation, e.g. at the onset of the following segment for fricatives, but during the closure phase for plosives. The gesture is claimed to be more extensive in the context of an open vowel like [a(ː)] than a close vowel like [i(ː)] (Simonsen et al., 2008, p. 387). Simonsen et al.’s EMA traces showed an average horizontal movement of about 1 mm in the Norwegian laminal alveolar plosives /t d/ but about 6.5 mm in the apical retroflex plosives /ʈ ɖ/ during the closure phase of these plosives. However, even the extensive movement of the tongue tip coil during the flapping out was not reflected in the EPG frames according to Simonsen et al. (2008, 399). In spite of that difficulty, we attempted to use EPG for detecting the movement of the tongue tip from a posterior to a more anterior position during the articulation of Polish affricates in their plosive phase.

Only affricates containing a single uninterrupted closure were selected. Each 62-cell frame of such a closure was reduced to an 8-number vector in which each number corresponded to each row of the frame; if all cells in a row of a frame were activated, the corresponding element of the vector was assigned the row number, otherwise it was assigned 0; e.g., if a frame contained a full closure in the rows 2 and 3, it was represented by the vector [0,2,3,0,0,0,0,0]. For each vector, we calculated the average of all values corresponding to the mean row where the centre of the closure occurred; in our example, the centre of the closure would be (2+3)/2 = 2.5, i.e., in the middle of the distance between the 2nd and 3rd row. For each frame of the closure, the mean row was calculated, e.g.; if the articulation of a phoneme was accompanied by frames whose closure vectors were [0,2,3,0,0,0,0,0], [0,0,3,4,0,0,0,0] – their centres would be [2.5, 3.5]. All the centres from a closure were then approximated to a line (henceforth referred to as *the slope of the closure*) using the least-squares method. If the slope of the closure was positive, like in the example above, it meant that the closure moved backwards. If the slope of the closure was negative, it meant that the closure moved forwards, which we assumed to be the best approximation of the flapping out gesture possible considering the fact that the EPG technique does not allow to track which part of the tongue is used to produce the closure. The more the number deviated from 0 towards 1 or -1, the faster the movement. The value of 0 indicated that the closure most likely stayed in the same place of articulation.

Table 8. Statistics for the slope of the closure in Polish retroflex /ʈʂ ɖʐ/ vs. Polish and English non-retroflex sounds; the greener the slope M, the more fronting the flap; the redder the slope M, the more backing the flap

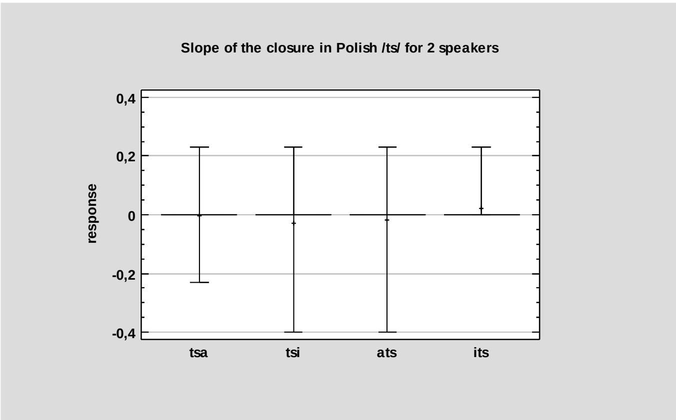
	Polish								English			
	t	t͡ʂ	t͡ɕ	ts	d	d͡ʐ	d͡ʑ	dz	t	tʃ	d	dʒ
all (a)	4468	1173	1173	1485	2250	76	658	472	2692	338	1706	435
count (c)	2040	475	275	281	1313	20	219	52	484	82	189	97
c/a	45.66%	40.49%	23.44%	18.92%	58.36%	26.32%	33.28%	11.02%	17.98%	24.26%	11.08%	22.30%
slope 0 (s)	1388	361	161	201	910	16	89	44	334	40	131	60
s/c	68.04%	76.00%	58.55%	71.53%	69.31%	80.00%	40.64%	84.62%	69.01%	48.78%	69.31%	61.86%
slope M	−0.0272	0.0033	0.0349	−0.0026	−0.0237	0.0482	0.0512	−0.0112	−0.0195	0.0128	−0.0261	0.0105
SD	0.0756	0.0884	0.1113	0.0795	0.0782	0.1084	0.1127	0.0350	0.0706	0.0840	0.0871	0.0865
move	front	back	back	front	front	back	back	front	front	back	front	back

The data do not confirm our prediction that /t͡ʂ d͡ʐ/ are accompanied by a flapping movement of the tip from a posterior position forward during their stop phrase. Polish and English laminal /t d/ included in the above table for reference display stronger forward movement than Polish /t͡ʂ/ and /d͡ʐ/. In fact, for /d͡ʐ/ the reverse direction seems to be more prevalent (M = 0.0482) while /t͡ʂ/ is relatively stable. This, however, does not mean that the fronting movement does not occur on the following vowel, as can be observed for the fricative retroflex phonemes in general (Hamann, 2003, p. 34) or that this movement is not produced during the fricative part of the affricate (c.f. Lorenc 2018, 143). In vowels and fricatives, which lack closure, calculating the negative slope to approximate the flapping out gesture would not be possible. It should also be noticed that in all of the above sounds we observe relatively large and non-homogeneous standard deviations, which makes the comparison of the means prone to error.

We checked for all speakers and for each speaker individually the degree of forward movement during the closure phase of /t͡ʂ d͡ʐ/ in Polish and English word-internal contexts before and after /a/ and /i/ following Simonsen et al. (2008, 399). They used EMA to detect a flapping movement during the closure phase of apical /ʈ/ before /a/ in Norwegian for one speaker.

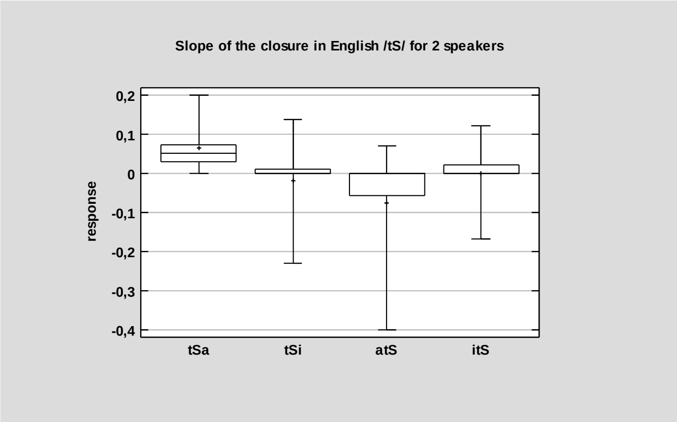
The choice of the phase of the affricate where the flapping out motion was tracked was justified by the fact that the stop portions of Polish affricates are considered apical, possibly partly sub-laminal, and therefore possibly retroflex (Keating 1991, 36, Wierzchowska 1980, 64; Żygis 2006, 105).

Figure 9. Slope of the closure in Polish /t͡ʂ/ for Polish speakers M1 and M2



In English, /ɑː ʌ aɪ aʊ/ were used for [a], and /i(ː) ɪ ɪə/ were used for [i]. Diphthongs were included in tʃ_ contexts only to compensate for the scarcity of data for monophthongs.

Figure 10. Slope of the closure in English /tʃ/ for two English speakers



The means and standard deviations from the box-and-whiskers plots above are presented in the right part of the table below.

Table 9. Slope of the closure in Polish /t͡ʂ/ in the context of /i ɨ a/. Slope of the closure in English /tʃ/ before /i(:) ɪ ɑ: ʌ aɪ aʊ/ or after /i(:) ɪ ɑ: ʌ/

	Polish				English			
	t͡ʂa	t͡ʂi	at͡ʂ	it͡ʂ	tʃa	tʃi	atʃ	itʃ
all (a)	124	227	103	178	28	48	38	68
count (c)	59	96	44	83	8	16	13	20
c/a	47.58%	42.29%	42.72%	46.63%	28.57%	33.33%	34.21%	29.41%
slope 0 (s)	47	72	36	70	1	9	7	11
s/c	79.66%	75.00%	81.82%	84.34%	12.50%	56.25%	53.85%	55.00%
M	–0.0030	–0.0155	–0.0078	0.0197	0.0636	–0.0200	–0.0763	0.0012
SD	0.0836	0.0808	0.0938	0.0584	0.0604	0.0952	0.1595	0.0635
move	front	front	front	back	back	front	front	back

Table 9 does not reveal a clear pattern of flapping out during the closure of Polish /t͡ʂ/, especially when contrasted with English. Indeed, in the context of /a/, the articulation is always fronted, which would confirm what (Simonsen et al., 2008, p. 387) observed in Norwegian, namely that the flapping out gesture is more extensive in the context of an open vowel like [a(:)] than a close vowel like [i(:)].

The results in the context of /i/, however, are ambiguous in Polish. Also, the magnitude of the flap-out in Polish is negligible when compared to the flap-out in English /atʃ/ (M = –0.0763), which is surprising considering the non-retroflex articulation of /tʃ/.

5. Conclusion

The results of the present study are equivocal in terms of providing clear evidence for the retroflexion of the Polish sounds under study.

We do present evidence that /ʂ ʐ ʈʂ ɖʐ/ are posterior in their place of articulation when compared to /t d/. However, while posteriority is included in the literature as one of the criteria for retroflexion, it is not sufficient on its own. In other words, an articulation that is more posterior than alveolar is simply post-alveolar; and the standard IPA denotation of e.g. English /tʃ dʒ/ is just that, without an implication of retroflexion. Here, the Polish affricates display articulations behind the second row of electrodes, allowing for a relatively firm statement of the place of articulation exactly as post-alveolar.

The evidence provided by EPG for the concave/convex shape of the tongue is only indirect, and takes the form of the unidimensional Centrality Index. Here, our results are less clear than in the case of place of articulation. The use of a two-dimensional index is postulated to capture the difference between /ʈʂ ɖʐ/ and /tɕ dʑ/. However, it needs to be noted that, like posteriority, convex/concave posture of the tongue is not in itself a sufficient correlate of retroflexion. On its own, one could argue, a convex tongue shape is a strong predictor of palatality, but a concave or flat shape does not necessarily have to predict retroflexion.

Finally, there is evidence of a flapping-out motion in the Polish affricates under study; however, there is also evidence of the same movement in the English affricates that are not postulated as retroflex.

It seems that the criteria for retroflexion discussed in the Introduction need to be satisfied in parallel, since each individual one on its own is not sufficient for the recognition of an articulation as retroflex; this is true especially if curling up of the tongue tip is not to be included. If so, then our study provides only weak evidence in favour of recognizing Polish /ʈʂ ɖʐ/ as retroflex.

The novel contribution of the present study includes is the development and fabrication of a new Reading palate design. Thanks to the reduced thickness and increased wearing comfort, the palate is capable of providing new EPG data on a wide range of articulatory phenomena.

Acknowledgements

The authors would like to express their gratitude to Mikołaj Wypych and Michał Szczyszek for making their *phonetizer* available for creating the Polish phonetic dictionary; Valérie Mapelli and the ELRA for providing the ESPRIT-ACCOR data; Alan Wrench for providing insights about the interpretation of ESPRIT-ACCOR audio data; Jonathan Harrington for sharing Polish articulatory data; Paul Sharp and *Rose Medical Solutions Ltd* for cooperation.

Funding institution

The project this study is part of was funded by the Polish *National Science Centre* (grant no. 2013/11/B/HS2/03151).

References

ACCOR - English**.** (n.d.). Retrieved 20 May 2020, from https://catalogue.elra.info/en-us/repository/browse/ELRA-S0001/

Barry, M. C. (1992). Palatalisation, assimilation and gestural weakening in connected speech. *Speech Communication*, *11*(4–5), 393–400. https://doi.org/10.1016/0167-6393(92)90044-8

Bukmaier, V. (2017). *Zusammenhänge zwischen synchroner Variation und diachronem Lautwandel im Sibilantsystem des Polnischen* [Ludwig-Maximilians-Universität München]. http://nbn-resolving.de/urn:nbn:de:bvb:19-205298

Carreira-Perpiñán, M. A., & Renals, S. (1998). Dimensionality reduction of electropalatographic data using latent variable models. *Speech Commun.*, *26*, 259–282.

Cruttenden, A. (2014). *Gimson's Pronunciation of English* (0 ed.). Routledge. https://doi.org/10.4324/9780203784969

Dziubalska-Kołaczyk, K., Bogacka, A., Wypych, M., Pietrala, D., & Krynicki, G. (2006). PELT: An English language tutorial system for Polish speakers. In *ISCA Tutorial and Research Workshop on Multilingual Speech and Language Processing—Multiling 2006* (p. On CD).

Garofolo, J. S., Lamel, Lori F., Fisher, William M., Fiscus, Jonathan G., Pallett, David S., & Dahlgren, Nancy L. (1993). *Documentation for the DARPA TIMIT acoustic–phonetic continuous speech corpus—Tech. Rep.* CDROM.

Gibbon, F. E., & Nicolaidis, K. (1999). Palatography. In Hardcastle, William J. & Hewlett, Nigel (Eds.), *Coarticulation* (pp. 229–245). Cambridge University Press.

Gussmann, E. (2007). *The phonology of Polish*. Oxford University Press.

Guzik, K. M., & Harrington, J. (2007). The quantification of place of articulation assimilation in electropalatographic data using the similarity index (SI). *Advances in Speech Language Pathology*, *9*, 109–119.

Halle, M., & Stevens, K. (1989). The postalveolar fricatives of Polish. *Manuscript*.

Hamann, S. (2003). *The Phonetics and Phonology of Retroflexes*. LOT.

Hardcastle, W. J., Gibbon, F., & Nicolaidis, K. (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics*, *19*(3), 251–266. https://doi.org/10.1016/S0095-4470(19)30343-2

Harrington, J. (2010). *Phonetic analysis of speech corpora*. Wiley-Blackwell.

Jassem, W. (2003). Illustrations of the IPA: Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107.

Jones, W., & Hardcastle, W. J. (1995). New developments in EPG3 Software. *International Journal of Language & Communication Disorders*, *30*(2), 183–192. https://doi.org/10.3109/13682829509082529

Keating, P. A. (1991). CORONAL PLACES OF ARTICULATION. In C. Paradis & J.-F. Prunet (Eds.), *The Special Status of Coronals: Internal and External Evidence* (pp. 29–48). Academic Press. https://doi.org/10.1016/B978-0-12-544966-3.50008-7

Kelly, S., Main, A., Manley, G., & McLean, C. (2000). Electropalatography and the Linguagraph system. *Medical Engineering & Physics*, *22*(1), 47–58. https://doi.org/10.1016/S1350-4533(00)00007-2

Lorenc, A. (2018). Articulatory Characteristics of Polish Retroflex Sibilants. Analysis Using Electromagnetic Articulography. *Logopedia*, *47*, 131–149.

Lorenc, A. (2020). Charakterystyka artykulacyjna polskich sybilantów retrofleksyjnych. Badanie z wykorzystaniem artykulografii elektromagnetycznej. *Logopedia*, *47*(2), 157–176.

MacKenzie, L., & Turton, D. (2020). Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard*, *6*(s1). https://doi.org/10.1515/lingvan-2018-0061

Marchal, A., & Hardcastle, W. J. (1993). Accor: Instrumentation and Database for the Cross-Language Study of Coarticulation. *Language and Speech*, *36*(2–3), 137–153. https://doi.org/10.1177/002383099303600302

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal Forced Aligner [Computer program] Version 0.9.0.* http://montrealcorpustools.github.io/Montreal-Forced-Aligner

***MOCHA-TIMIT*.** (n.d.). Retrieved 20 May 2020, from http://www.cstr.ed.ac.uk/research/projects/artic/mocha.html

Moen, I., Hanne, & Simonsen, H. (2011). *The Norwegian Retroflex Fricative*. 17–21.

***MRC Psycholinguistic Database: Machine Usable Dictionary.* (2.00).** (1997). [Computer software]. https://websites.psychology.uwa.edu.au/school/mrcdatabase/mrc2.html

Pompino-Marschall, B., & Żygis, M. (2003). *Surface Palatalization of Polish Bilabial Stops: Articulation and Acoustics*.

Rubach, J. (1984). *Cyclic and lexical phonology: The structure of Polish*. DE GRUYTER. https://doi.org/10.1515/9783111392837

Simonsen, H. G., Moen, I., & Cowen, S. (2008). Norwegian retroflex stops in a cross linguistic perspective. *Journal of Phonetics*, *36*(2), 385–405. https://doi.org/10.1016/j.wocn.2008.01.001

Śliwińska-Kowalska, M., Kotyło, P., & Sol, S. S. (2013). Opracowanie list zdaniowych testu rozumienia zdań w szumie dla języka polskiego. *Otorynolaryngologia*, *2013/12*(4), 176–182.

Szczyszek, M. (2013). *O porozumiewaniu się międzyludzkim: Wariant mówiony języka polskiej wspólnoty komunikatywnej : słowotwórstwo, leksyka, składnia*. Wydawnictwo 'Rys'.

Wierzchowska, B. (1980). *Fonetyka i fonologia języka polskiego*. Zakład Narodowy im. Ossolińskich.

Winkelmann, R., Harrington, J., & Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, *45*(Supplement C), 392–410. https://doi.org/10.1016/j.csl.2017.01.002

Wrench, A. A. (2007). Advances in EPG palate design. *Advances in Speech Language Pathology*, *9*(1), 3–12. https://doi.org/10.1080/14417040601123676

Wrench, A. A., & Hardcastle, W. J. (2000). A multichannel articulatory database and its application for automatic speech recognition. In *5th Seminar on Speech Production: Models and Data* (pp. 305–308).

Wypych, M. (2011). *Układ rozpoznający struktury intonacyjne w sygnale mowy*.

Żygis, M. (2006). *(Non)Retroflexivity of Slavic Affricates and Its Motivation. Evidence from Polish and Czech*.