

The Perception of Polish Geminates:
A Study on Geminate Articulation and Social Meaning

Weronika Polakowska 13762796

BA thesis Linguistics

University of Amsterdam

Supervised by dr. Marijn van 't Veer

25 June 2024

Abstract

Geminates, or phonetically long consonants that can contrast with their short counterparts, are present in the Polish language. In addition to lengthening the consonant, the most common method of production, Polish geminates can also be rearticulated, which manifests as a separate pronunciation of each consonant. While the frequency of rearticulation varies on an inter-speaker basis, the motivation behind it is not yet known. Soukup's (2013) paradigm of social meaning construction has not been used in any prior research to examine Polish geminates. Specifically, it has not been explored whether listeners detect the difference in pronunciation and, if so, whether this difference has social implications related to standard Polish, such as higher socioeconomic background or intellect. The present study uses a matched-guise perception task in an effort to fill that theoretical gap. After listening to stimuli with one of the articulation methods, the participants were asked to rate the recordings using affective scales that indicated intellect, formality, and socioeconomic background.

The findings suggest that Polish speakers are able to discriminate between the two types of geminate production. Furthermore, rearticulated geminates were shown to carry social meanings representative of standard Polish, including higher intellect, education level, and socioeconomic status based on the scores from the ratings.

The purpose of this paper is to add to the body of knowledge on the Polish geminates and their behaviour, as well as offer a new perspective to the current literature on the subject, which primarily consists of production studies.

Acknowledgements

Words cannot express my gratitude to my supervisor dr. Marijn van 't Veer for providing his knowledge and valuable feedback. Many thanks to dr. Monique Flecken for offering help with the statistical analysis. Additionally, I am very grateful to Dirk Vet for creating the experiment and helping me with the audio recordings.

I am extremely thankful to my friends, whose support and insight were irreplaceable. Special thanks to Martyna, because of whom the idea for this paper was born, Ada, for her persistence and emotional support, and Pamir, for his editing help and late-night feedback sessions.

Lastly, I would like to thank my family for supporting my dreams of studying abroad. My academic endeavours would not have been possible without their support.

Table of contents

1. Introduction	4
1.1. Gemination in the languages of the world	4
1.2. Polish geminates	7
1.3. Types of geminate articulation	8
1.4. Standard Polish and its social status	11
1.5. Construction of social meaning	11
1.6. The current study	12
2. Methods	13
2.1. Stimuli	14
2.2. Participants & Procedure	15
2.2.1. Audio recordings	15
2.2.2. Perception task	18
3. Results.....	18
3.1. General results per method of articulation	19
3.2. Results per rating scale	20
3.3. Post hoc analysis: results per manner of articulation (affricates)	22
4. Discussion	25
4.1. Interpretation of results	25
4.2. Limitations	27
4.3. Future research	29
5. Conclusion	31
References	32
Appendices	36

1. Introduction

1.1. *Gemination in the languages of the world*

Geminates are considered to be phonetically long consonants which can create a phonemic contrast with their shorter—singleton—equivalent (Davis, 2011). Their length varies from one-and-a-half to three times the length of their singleton (Ladefoged & Maddieson, 1996) and the frequency of gemination depends on the features of the consonant; cross-linguistically, alveolar geminates are the most common (Thurgood, 1993). Although not as recurrent as long vowels—out of 451 languages, the UCLA Phonological Segment Inventory Database (UPSID) lists 12 with geminates and 51 with long vowels (Maddieson, 1984)—long consonants are observed in a variety of languages, such as Japanese, Italian, Arabic, Hungarian, and so on. Examples of the phonemic geminate–singleton contrast are shown in (1) and (2):

(1) Finnish minimal pair (Savinainen-Makkonen, 2007: 355)

kuka /kuka/ ‘who’

kukka /kukka/ ‘flower’

(2) Standard Arabic minimal pair (al-Tamimi *et al.*, 2010: 116)

سد *sad* /sad/ ‘water dam’

سدّ *sadd* /sadd/ ‘closed’

The primary cue to distinguish the geminate from its singleton is the difference in occlusion duration. This can be seen in Figure 1 for the *leki-lekki* ‘medicine–light’ word pair: the singleton /k/ is approximately 68 milliseconds long, while the geminate /k:/ is 228 milliseconds long. In some languages, the lengths of the preceding and following vowels act as additional cues (Rojczyk & Porzuczek, 2019a); for example in Japanese (Kawahara, 2006), the vowel preceding the long consonant gets lengthened, while in Hindi (Ohala *et al.*, 2007) that vowel gets shortened.

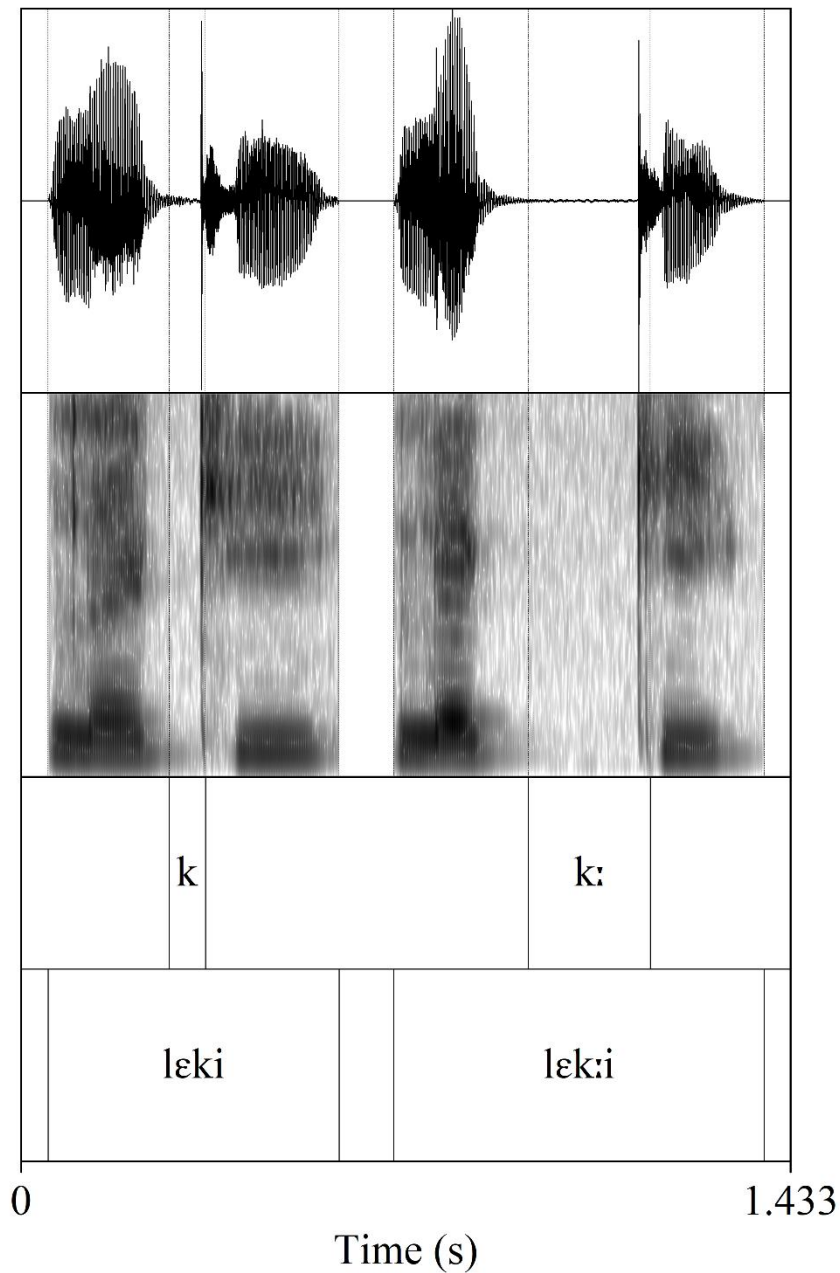


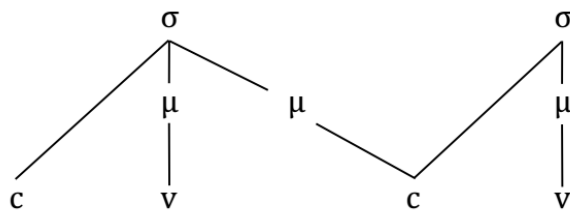
Figure 1. *The difference in occlusion duration between the geminate /k:/ and its singleton /k/.*

Geminate consonants most often occur in intervocalic positions, seldom observed not adjacent to a vowel (Pająk, 2009), and they are most frequent word-medially, although examples of word-final (e.g. Standard and Jordanian Arabic; al-Tamimi *et al.*, 2010) and word-initial (e.g. Kelantan Malay; Hamzah *et al.*, 2016) geminates do exist. They are divided into ‘true’ geminates, which are lexical and can form minimal

pairs with the singleton, and ‘fake’ geminates, which are either derived from assimilation processes at the morpheme boundary word-internally or concatenated across word boundaries (Oh & Redford, 2012).

From the standpoint of autosegmental phonology (Hayes, 1989), lexical geminates are considered to be symmetrically distributed across two heterosyllabic positions, belonging to one node or bundle of features, as shown in (3); this was previously attested synchronically in Italian and diachronically in the transition from Sanskrit to Pāli (Russo & Ulfsbjorninn, 2017).

(3) The syllable structure of a lexical geminate (adapted from Russo & Ulfsbjorninn, 2017: 165)



This distribution is further motivated by the fact that, although lexical geminates exhibit similar phonological behaviour to consonant clusters, they are resistant to the rules of epenthesis and cannot be split up like the latter (see 4):

(4) Geminate integrity in Hadhrami Arabic (Davis & Ragheb, 2014; adapted from Russo & Ulfsbjorninn, 2017: 166)

a. Consonant cluster

/gird/	[girid]	‘monkey’
/bint/	[binit]	‘girl’

b. Geminate

/rab:/ [rab:] *[rabib] ‘lord’

This is known as *geminate integrity* and demonstrates that phonological processes do not affect one part of the geminate (Russo & Ulfsbjorninn, 2017).

1.2. Polish geminates

Polish is one of the geminating languages. For the geminate–singleton contrast, calculated through the division of the duration of the geminate by the length of the single consonant, previous studies observed an overall ratio of 2.48 (Rojczyk & Porzuczek, 2019a). The individual consonant types also yielded different results (as seen in Table 1); the ratios varied from 2.00 for voiceless fricatives to 2.82 for nasals (Rojczyk & Porzuczek, 2019a). Furthermore, the authors observed no significant differences in duration for pre- and post-consonantal vowels.

Table 1

Polish geminate–consonant duration ratios across different manners of articulation

Consonant type	Singleton duration (ms)	Geminate duration (ms)	G/S ratio
Fricatives	112	224	2.00
Affricates	119	264	2.22
Plosives	88	233	2.65
Nasals	72	203	2.82

Note: Results adapted from Rojczyk & Porzuczek (2019a).

Polish geminates usually occur intervocalically and word-medially (as in 5), with rare examples in a word-initial (prevocalic) position like in (6) (Rojczyk & Porzuczek, 2019a).

(5) Geminates in word-medial context

panna	/panna/	‘maiden’
lekki	/lɛkki/	‘light’

(6) Geminates in word-initial context

ssak	/ssak/	‘mammal’
dźdźownica	/d͡z͡d͡zɔvnitsa/	‘earthworm’

The words in (5) and (6) show examples of lexical geminates—all of them with the exception of *dźdźownica* form a minimal pair with their shorter counterpart. In addition, there is also a variety of morphologically derived geminates in Polish, both word-internally (7) and across word boundaries (8).

(7) Geminates at morpheme boundary

wwozi	/v + vɔzi/	‘brings in’
bezsensowny	[bes + sensɔvni]	‘meaningless’

(8) Geminates at word boundary

ot tak	/ɔt + tak/	‘just like that’
w fotelu	[f + fotelu]	‘in an armchair’

According to a corpus study, sonorant geminates are the most common (Kozyra, 2008). While they only occur between vowels, obstruent geminates are attested in both aforementioned positions.

1.3. *Types of geminate articulation*

Although the single-articulation seems to be the more common method of production, Polish geminates can also be rearticulated. This type of pronunciation stands for each

consonant being produced separately, creating a disruption within the geminate (Rojczyk, 2022), which results in a perceptible release of the first consonant in the case of stops and affricates, or an insertion of a short unidentifiable vowel for nasals (Rojczyk & Porzuczek, 2019a). The spectrogram of this articulatory difference is shown in Figure 2. Rearticulation is considered the base of two-phase theory, which differentiates long consonants from geminates by assuming that the latter contains a syllable boundary—with the first consonant as a coda of the first syllable and the second as the onset of the following syllable (Lehiste *et al.*, 1973). However, the present study assumes no underlying difference between long consonants and geminates and uses the terms interchangeably; the type of articulation is specified instead.

Despite evidence of rearticulation having been found in the recordings of one speaker of Estonian in Lehiste *et al.* (1973), studies of other languages bore inconsistent results or did not find evidence of rearticulation at all (Ham, 2002). Polish appears to be a unique case of a language that shows more frequent rearticulation (Rojczyk & Porzuczek, 2019a). Rearticulation is in free variation with single-articulation (Rojczyk, 2022). Research suggests that the degree of rearticulation varies greatly on an inter-speaker basis, with some speakers only producing single-articulated geminates, while others rearticulated most of the time (Rojczyk & Porzuczek, 2019a; Thurgood & Demenko, 2003). Furthermore, the consonant group the geminate belongs to was found to have an impact on the rate of rearticulation—affricates were rearticulated most often (Thurgood, 2001; Thurgood & Demenko, 2003; Rojczyk & Porzuczek, 2019b), followed by stops, nasals, and fricatives (Rojczyk

& Porzuczek, 2019a). Rearticulation was believed to potentially enhance the geminate–singleton contrast, but no significant difference in perception accuracy was found between single- and rearticulated geminates (Rojczyk, 2022).

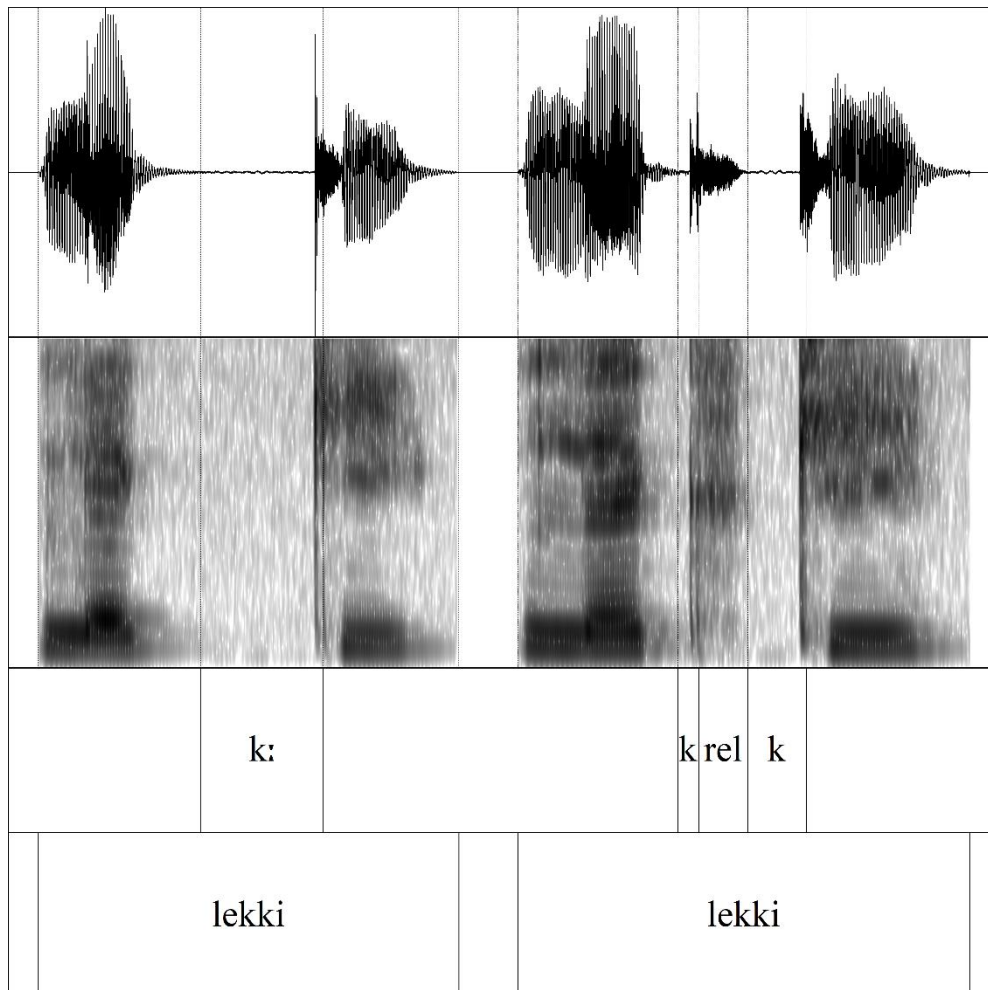


Figure 2. Two pronunciations of the geminate /k:/ in lekki ‘light’.
The ‘rel’ stands for the release of the first part of the geminate.

Although the between-speaker variability has been previously observed, not much is known about the motivation behind choosing a specific type of articulation. Other factors connected to the phonetic realisation of the Polish geminate are speech rate, lexical stratum, as well as hypo- and hypercorrection (Rojczyk & Porzuczek, 2014). The orthography of the language is also said to have a substantial impact on the pronunciation (Kozyra, 2008). Polish geminates are represented in spelling by a

sequence of two identical letters—thus, sometimes being referred to as double consonants. Due to the standard variety of Polish being based on the written language (Kućak, 2019), it is possible that the pronunciation that closely resembles spelling would be considered ‘more correct’.

1.4. Standard Polish and its social status

‘Correctness’ in Polish is a recurrent topic of debate—the language holds an important status in culture, politics, and society, while also being used as a tool to strengthen national identity (Kućak, 2019). As a result, there had been continuous efforts by the Rada Języka Polskiego (Polish Language Council) and other language experts to preserve the language in its ‘purest’ form and maintain a standard variety of the language. While prescriptive attitudes towards the language are now diminishing, ‘proper’ Polish still holds some social meaning—it denotes propriety, intelligence, and social class (Kućak, 2019). According to Markowski (1999), standard Polish should be “accepted by the vast majority of educated Poles, in particular by those who have learned the general Polish language at home and have completed more than secondary education, and who treat language as a value in itself” (p. 1702, translated verbatim from the original). Furthermore, Smakman (1999) suggested that Polish speakers associate standard Polish with highly educated people, specialists on the language, as well as radio and news presenters.

1.5. Construction of social meaning

According to previous studies, both speakers and listeners play an important role in constructing social meaning in relation to linguistic variation (Villarreal, 2018). A framework outlined in Soukup (2013) indicates that, to absorb a social meaning, the

listener needs to recognise the contrast between the linguistic forms and then assign contrasting social meanings to these forms. [Villarreal \(2018\)](#) conducted a listener experiment to examine whether Californians distinguish the California Vowel Shift (CVS) features from the other American English vowels, and if the social meanings previously assigned in production studies—such as whiteness, femininity, confidence, and an association with California—were also apparent in perception. The author found that the listeners did perceive the CVS to non-CVS contrast, and that they associated CVS features with California, sounding like a ‘Valley girl’, and with confidence, the latter especially for men. However, CVS features were secondary to attributes such as gender and ethnicity when it came to their judgement of the speakers. Nonetheless, it illustrated the need for perception studies to investigate the construction of social meaning from a different perspective.

With regard to Polish, there are no studies on the perception of the contrast between single- and rearticulated geminates, as well as on their potential social meanings.

1.6. The current study

Polish geminates can be phonetically realised through single- and rearticulation and there is great inter-speaker variability in the rate of rearticulation, with different frequencies observed in relation to consonant type. Other influential factors include speech rate and hypo- or hypercorrection. No significant difference was found between the types of articulation for perception accuracy ([Rojczyk, 2022](#)). However, no previous research has investigated Polish geminates from the perspective of [Soukup’s \(2013\)](#) framework of social meaning construction—whether listeners

perceive a difference in articulation and, if so, whether it carries any social implications, such as education level or intelligence.

The present study, partly inspired by the methodology seen in [Villarreal \(2018\)](#), aims to investigate these issues using a matched-guise task and ranking on affective scales which measure people's judgements on the presented stimuli. The research questions to be addressed in this paper are outlined below:

- 1) Do Polish speakers perceive the difference between single- and rearticulated geminates?
- 2) Is rearticulation associated with standard Polish by Polish speakers?

The study is of explorative nature due to the lack of previous research on the topic. It is hypothesised that if Polish speakers do recognise the contrast in geminate production, rearticulation will be perceived as a marker of 'proper' Polish due to its resemblance to the orthographic form ([Kozyra, 2008](#)), thus invoking the social implications of standard Polish, such as higher education level. Therefore, the prediction is that rearticulated geminates will be ranked higher in formality on the affective scales. Alternatively, if the difference is not perceived, both methods of articulation will be ranked equally on the affective scales and no social implications will be assumed.

2. Methods

To investigate the above questions, a perception task was conducted through an online survey. The methodology is partly modelled after [Villarreal \(2018\)](#), described in further detail above.

2.1. Stimuli

The stimuli consisted of 10 sentences made up of 15 to 22 words ($M = 18$). The Polish Web 2019 corpus ([plTenTen19](http://www.sketchengine.eu)¹), which is an all-purpose written corpus of 4.2 billion words containing Polish texts collected from the Internet in December 2019 and January 2020, was used to select the sentences. The sentences were emotionally neutral—this was determined by choosing sentences with no emotionally charged adjectives or phrases—and contained two words with geminates, with an equal number of lexical and morphologically derived geminates (as seen in [7b](#)). An effort was made to exclude loanwords, but for the sake of having a balance between the types of geminates, five of the selected lexical geminates were borrowings—i.e. *sawanna* ‘savannah’ from Spanish, *passa* ‘streak’ from French, *netto* ‘net’ from Italian ([Żmigrodzki, n.d.](#))—or foreign place names, e.g. Amman, Aleppo. The borrowings are not new to Polish, though—their dictionary entries date back at least to the late 19th and early 20th century ([Żmigrodzki, n.d.](#)). Overall, the sentences cover a variety of topics and have no overarching theme. One of the stimuli can be seen in ([9a](#)), with the words containing geminates transcribed in ([9b](#)). In addition, 10 filler sentences were used; these contained no words with geminates and no geminates at morpheme or word boundaries or resulting from assimilation. All the sentences can be seen in [Appendices I & II](#).

¹ Available at: <http://www.sketchengine.eu>.

(9) a. Stimulus example

<i>Olimpia</i>	<i>przełamała</i>	<i>fatalną</i>	<i>passę</i>	<i>7</i>	<i>meczów</i>
Olimpia	break.PST	terrible	streak	7	match.PL
<i>bez</i>	<i>zwycięstwa</i>	<i>oraz</i>	<i>4</i>	<i>meczów</i>	<i>bez</i>
without	win	and	4	match.PL	without
<i>wygranej</i>	<i>na</i>	<i>własnym</i>	<i>boisku.</i>		
win	on	own	field		

‘Olimpia broke the terrible streak of 7 matches without a win and 4 matches without a home win.’

b. Words with geminates and their derivation type

<i>passę</i> ‘streak’ (dative)	/passẽ/	lexical
<i>bez zwycięstwa</i> ‘without a win’	[bez + zvĩtɕɛ̃stfa]	morphological

The affricates /ts tɕ tʃ dʒ/, plosives /p t d k/, nasals /n m/, and fricatives /v s z/ were the geminates selected for the experiment, as the rearticulation of approximants /j l w/ and /r/ has not been previously attested in research.

2.2. Participants & Procedure

The experiment conducted for the purpose of this study required the recording of audio stimuli and the distribution of a survey.

2.2.1. Audio recordings

Five female native speakers of Polish were recruited at the University of Amsterdam (UvA) through convenience sampling to record the stimuli. They all lived in Poland until they reached adulthood and they studied in the Netherlands at the time of the experiment. They were informed about the goal of the study. The audio was captured

by a Neumann TLM103 cardioid condenser microphone with a custom-made amplifier at the UvA Speech Lab in Amsterdam in a sound-proof booth.

The sentences were shown to the participants on a piece of paper. The speakers were instructed to read the stimuli aloud to the microphone with one type of articulation at a time, as naturally as possible, to have each sentence recorded with single- and rearticulation. To make sure the pronunciation was consistent, they were directed to record the same sentence multiple times; thus, if necessary, the researcher could replace a deviate word with a consistent pronunciation from a different take after the recording session. Since, based on some pre-recording trials done by the researcher, it is less detrimental to the fluency of the audio to remove the second burst phase for plosives or the short inserted vowel for nasals (i.e. manipulate the rearticulated geminates into single-articulated ones), the recording of natural-sounding rearticulated geminates was prioritised. The aforementioned manipulation was performed once in the final dataset—the rearticulated /d/ in *oddech* ‘breath’ in Sentence 3 (full sentence in [Appendix I](#)) was modified to create the single-articulated stimulus for one of the speakers.

[Figure 3](#) shows the spectrogram of *dżdżysty* ‘rainy’—one of the two words with geminates recorded in Sentence 9 (full sentence in [Appendix I](#))—which includes a lexical geminate / $\widehat{d\zeta}$ / in word-initial position.

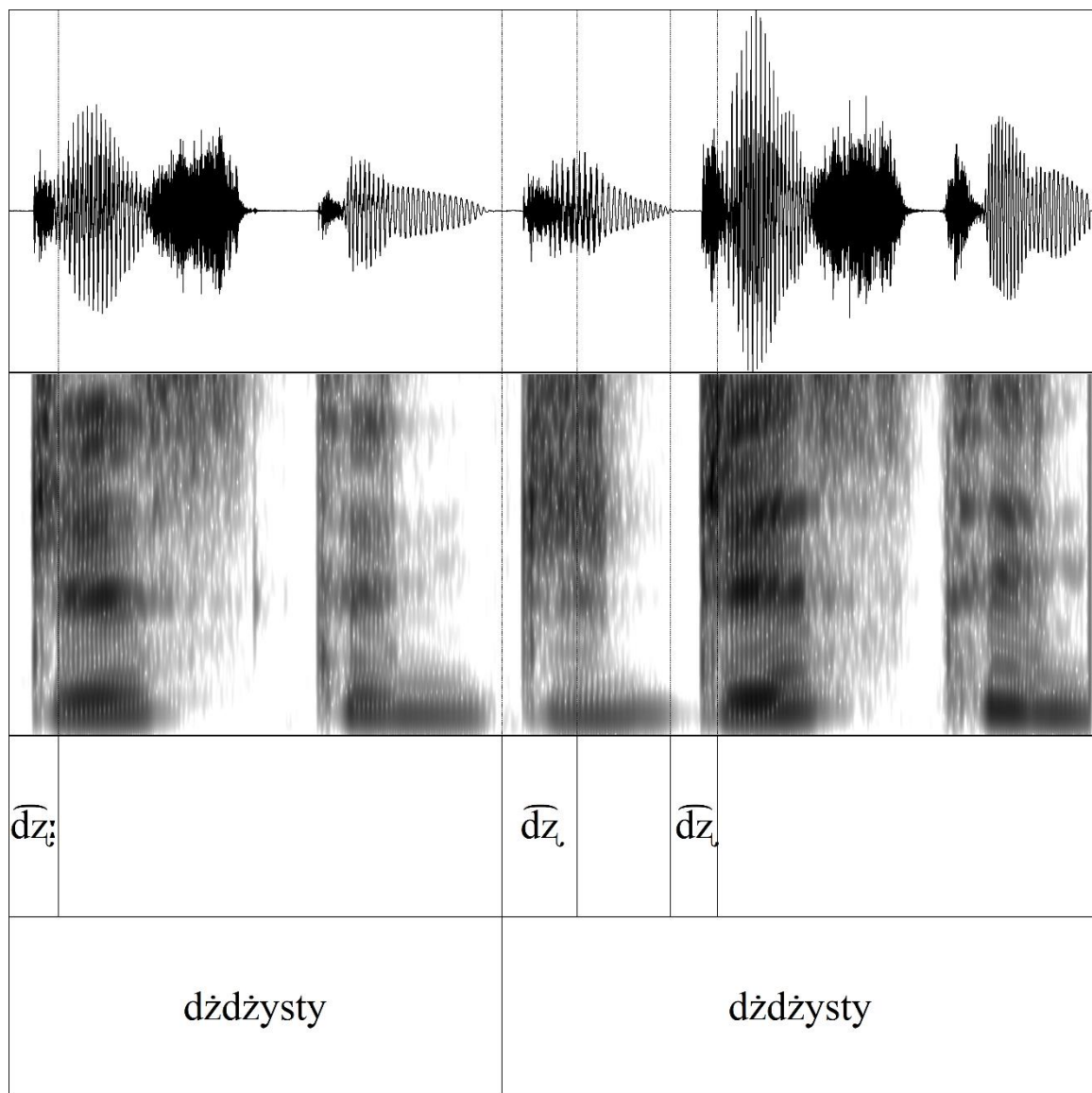


Figure 3. Spectrogram of dżdżysty ‘rainy’ from Sentence 9, pronounced by one of the speakers.
Full sentence can be seen in [Appendix I](#).

The audio files were converted into stereo during processing, and they were named after the method of articulation, stimulus number, and speaker number—G1 stood for single-articulation (SA) and G2 for rearticulation (RA), thus the file G2_6_3 was the RA stimulus 6 recorded by Speaker 3. The stimuli were divided into two trial groups of 10 items—Group 1 included the first five guises with SA and the other five guises with RA, while Group 2 covered the opposite.

2.2.2. Perception task

In addition, 38 participants took part in the perception task through an online survey. The listeners were all native speakers of Polish and they were recruited using the friend-to-friend sampling method. The information brochure provided in Polish along with the survey stated that the purpose of the experiment was to evaluate whether the speakers would be suitable for a job that required public speaking or storytelling. Similarly to Villarreal (2018), this was done to ensure that the listeners felt comfortable ranking and judging the recordings.

The survey was conducted through Experiment Designer (Vet, 2024). Each participant completed 10 trials from a randomly chosen trial group with the addition of five filler stimuli. The speaker for each stimulus per guise was randomised; as a result, every listener heard no more than two stimuli per speaker, each with a different type of articulation. After listening to each trial, the participant was asked to rank the speaker on four 7-point Likert scales: *formal–informal*, *suitable–not suitable* (for public speaking), *white collar–blue collar*, *more eloquent–less eloquent*. Because of the lack of previous research on the geminates in this context, the affective scales were instead selected based on the social meanings associated with standard Polish, such as the ones discussed in Smakman (1999) and Kułak (2019).

3. Results

Out of the 38 participants who took part in the listening experiment, 20 responses could be used for the analysis. One participant ranked only the first four stimuli, thus their responses were excluded. Furthermore, due to a technical error, two stimuli had to be excluded: G1_3_4 and G1_7_2. Overall, eight SA and 10 RA stimuli constituted

the basis of the analysis, along with the ratings from the four affective scales. In total, with responses to 10 stimuli per participant and on 4 scales and excluding the aforementioned SA sentences, 696 data points were analysed—328 for SA and 368 for RA. The analysis of the data was performed in R (R Core Team, 2023), and the code is available in [Appendix III](#).

3.1. General results per method of articulation

An exclusion criterion was used to eliminate potential outliers so that they do not hinder the performance of the model. Because the scales reflect similar social implications, i.e. a higher ranking in eloquence implies a higher likelihood of the speaker being a white-collar worker, the similarities in rankings across the scales were taken into consideration—if the difference between the lowest and highest of the ratings for a stimulus was higher than 4 (the mid-point of the Likert scale), the ranking was excluded. The general results of the experiment can be seen in [Table 2](#) and are illustrated in [Figure 4](#), in which the mean stands for the average rating across all scales.

Table 2

General results per method of articulation

Condition	Mean (points)	SD (points)
SA (<i>N</i> = 328)	3.30	1.54
RA (<i>N</i> = 368)	2.90	1.35

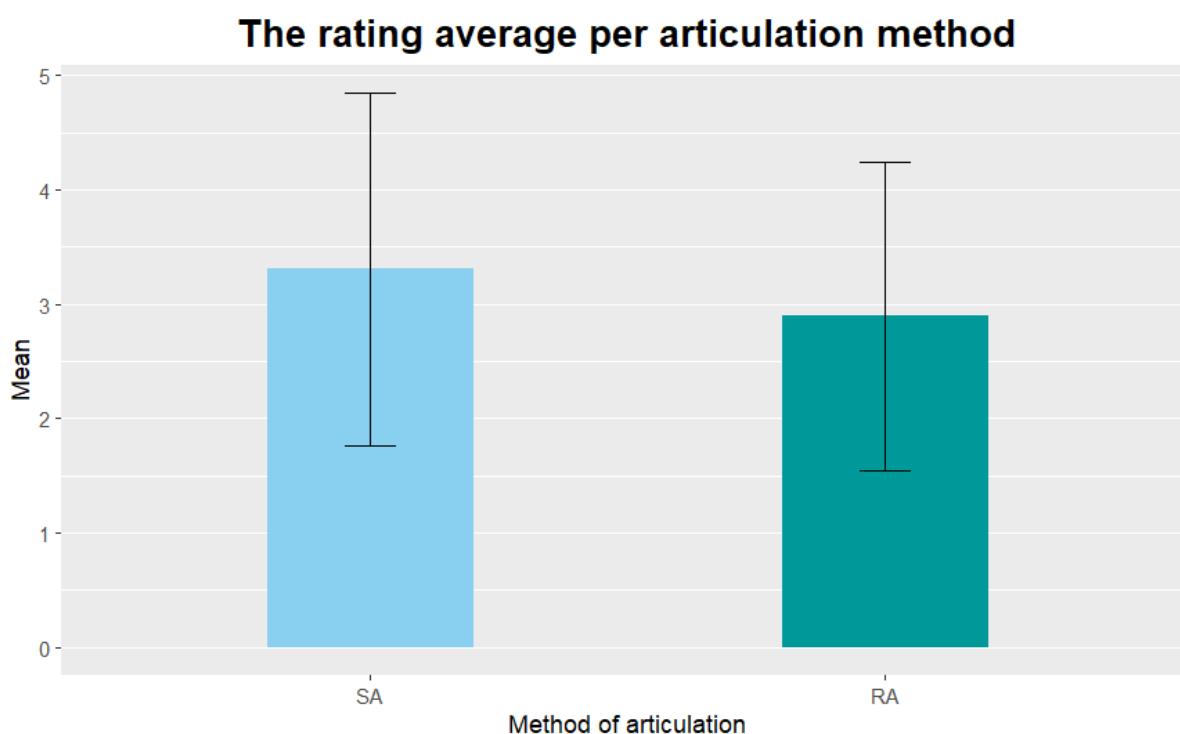


Figure 4. Results for rating average per method of articulation.
The error bars represent the standard deviation.

A linear model with the articulation type as the independent variable and the average affective scale ranking as the dependent variable shows that the SA stimuli were on average ranked 0.41 points higher than the RA stimuli (95% CI $[-0.84 \dots +0.02]$, $t = -1.88$, $p = .062$), which is not statistically significant. Based on this finding, it cannot be concluded with confidence whether Polish speakers perceive the contrast in geminate production or if it carries any social meaning.

3.2. Results per rating scale

In addition, the means for the individual ranking scales were calculated to compare between the two articulation types. As shown in [Table 3](#), the SA stimuli had a higher ranking average on all scales, and all of the means were higher than 3. In comparison, all the means for the RA stimuli were generally lower and below 3.

Table 3*Results per rating scale*

Rating scale	SA		RA	
	Mean (points)	SD (points)	Mean (points)	SD (points)
Scale 1	3.49	1.83	2.97	1.63
Scale 2	3.17	1.48	2.83	1.47
Scale 3	3.28	1.74	2.86	1.40
Scale 4	3.29	1.59	2.93	1.55

The findings of the linear model for each scale can be seen in [Table 4](#): the SA stimuli were ranked higher than the RA stimuli by 0.52 points on Scale 1 (*formal–informal*), 0.34 points on Scale 2 (*suitable–not suitable* (for public speaking)), 0.42 points on Scale 3 (*white collar–blue collar*), and 0.36 points on Scale 4 (*more eloquent–less eloquent*). These findings were only statistically significant in the case of Scale 1.

Table 4*Statistical results per rating scale*

	Estimate	CI (lower)	CI (higher)	t-value	p-value
Scale 1	– 0.52	0.001	1.039	– 1.98	.049*
Scale 2	– 0.34	– 0.787	0.098	– 1.54	.126
Scale 3	– 0.42	– 0.891	0.047	– 1.77	.078
Scale 4	– 0.36	– 0.829	0.113	– 1.50	.135

Note. The asterisk stands for significance ($p < .05$).

The general results for each ranking scale were plotted onto a bar chart with error bars and are summarised in [Figure 5](#). The rating with a significant difference in means, Scale 1 (*formal–informal*), is shown in full opacity.

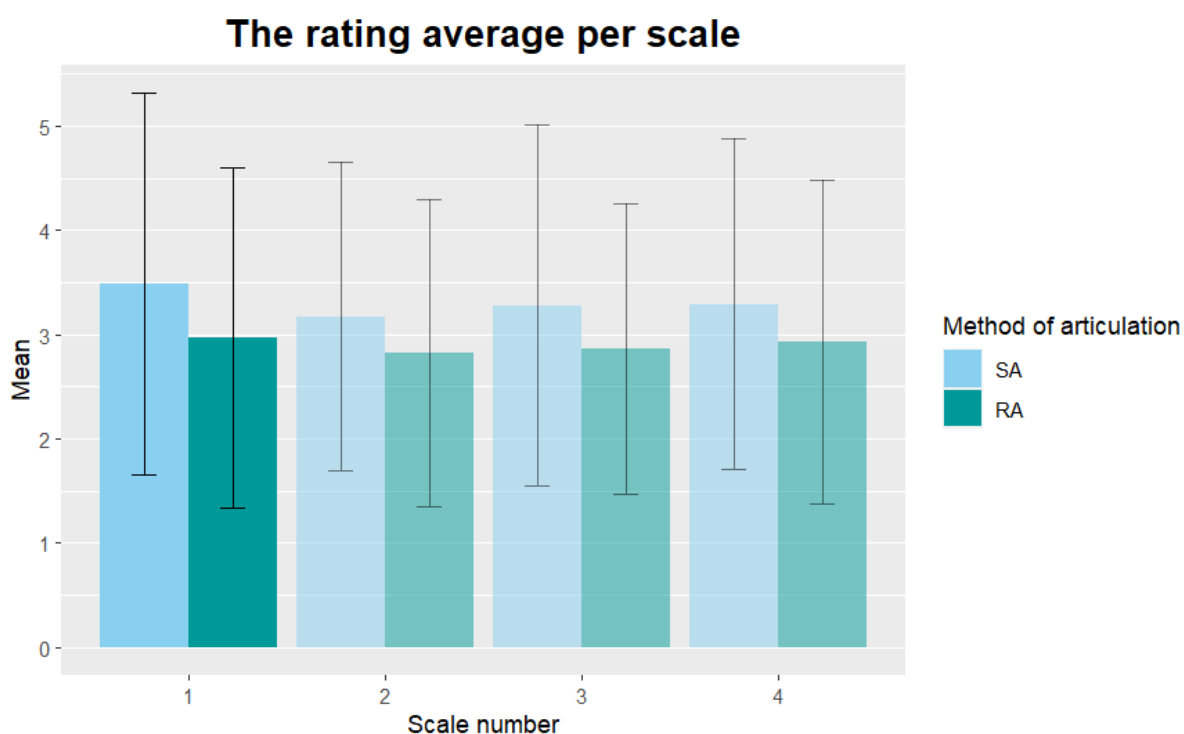


Figure 5. Results for rating average per scale. The error bars represent the standard deviation. The full opacity represents statistical significance.

In summary, the mean rankings on the affective scales ranged from 2.83 for Scale 2 to 2.97 for Scale 1 for RA, and from 3.17 for Scale 2 to 3.49 in Scale 1 for SA. None of the means for either articulation exceeded 4, which was the mid-point of the Likert scale. In terms of the estimate, the lowest difference between the types of articulation was found to be 0.34 for Scale 2 and the largest was 0.52 for Scale 1.

3.3. Post hoc analysis: results per manner of articulation (affricates)

In addition, a post hoc analysis was conducted to see whether the distribution of affricate geminates in the sentences had any impact on the rankings. This was motivated by the previous studies which found affricates to be the type of geminate that is rearticulated most frequently (see §1.3). This could potentially lead to single-articulated affricate geminates being ranked lower in formality. The overall results of this analysis were plotted in Figure 6.

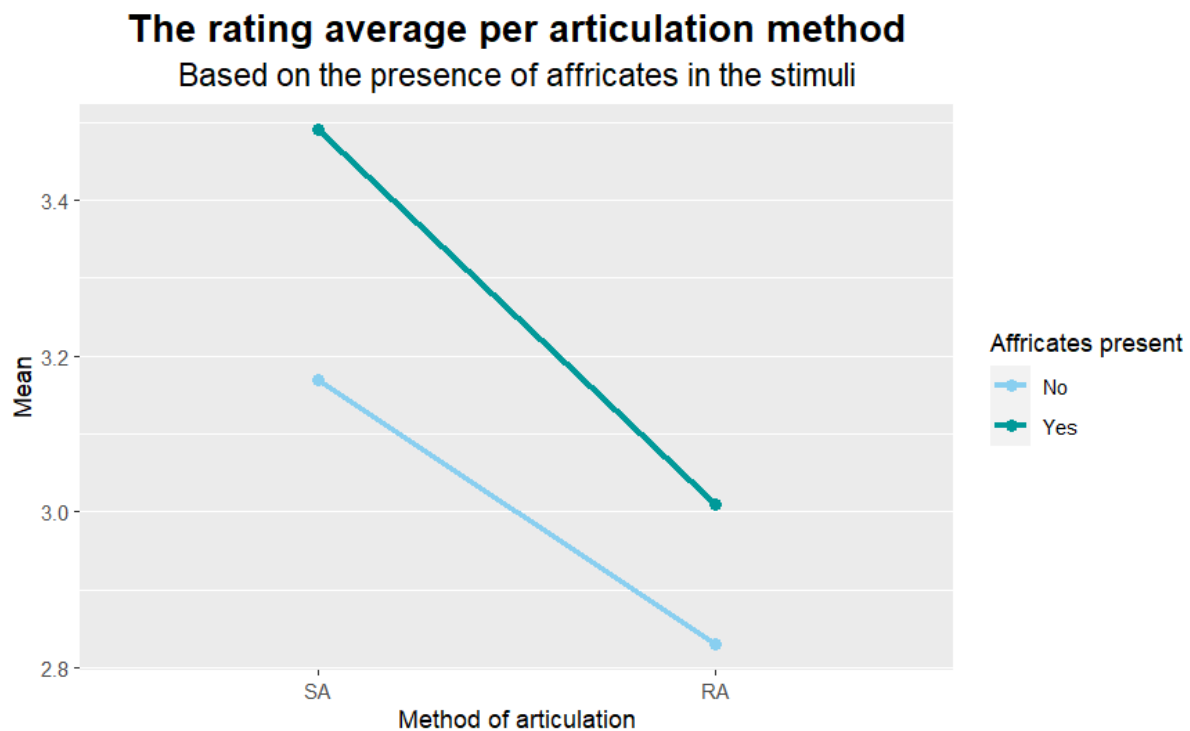


Figure 6. Overall results for rating average per method of articulation based on the presence of affricates in the stimuli.

As evident in the plot, the difference in the mean rankings for the SA and RA stimuli with affricates (3.49 v. 3.01) was larger than those without affricates (3.17 v. 2.83). Per type of articulation, the difference in the means was larger for SA (3.49 v. 3.17) in comparison to RA (3.01 v. 2.83). For both articulation types, stimuli without affricates were ranked lower on the scales (i.e. higher in formality). The linear regression with an interaction of articulation type and the presence of affricates yielded the following results, summarised in [Table 5](#).

Table 5

Statistical results per rating scale, divided by main effect and interaction

Condition	Effect	CI (lower)	CI (higher)	t-value	p-value
Articulation	− 0.41	− 0.856	+ 0.027	− 1.85	.067
Affricate	0.25	− 0.192	+ 0.693	1.12	.265
Articulation * Affricate	− 0.14	− 1.029	+ 0.740	− 0.32	.748

Overall, the results showed an effect of affricates—stimuli with affricates were ranked 0.25 points lower on the affective scales (CI $[-0.192 \dots +0.693]$, $t = 1.12$, $p = .265$). There was also an interaction between articulation and the presence of affricates (CI $[-1.029 \dots +0.740]$, $t = -0.32$, $p = .748$). The effect of the interaction for conditions without affricates is larger by 0.14 points compared to the ones with affricates.

Furthermore, [Table 6](#) shows the individual rating averages per scale, divided by condition: the type of articulation, the presence of affricates, and the interaction between the two factors.

Table 6

Statistical results per rating scale, divided by main effect and interaction

	Condition	Effect	<i>t</i> -value	<i>p</i> -value
Scale 1	Articulation	− 0.53	− 1.99	.049*
	Affricate	0.22	0.83	.410
	Articulation * Affricate	− 0.25	− 0.46	.647
Scale 2	Articulation	− 0.36	− 1.57	.118
	Affricate	0.05	0.24	.813
	Articulation * Affricate	− 0.19	− 0.42	.675
Scale 3	Articulation	− 0.42	− 1.73	.085
	Affricate	0.51	2.12	.036*
	Articulation * Affricate	− 0.20	− 0.41	.686
Scale 4	Articulation	− 0.34	− 1.40	.163
	Affricate	0.22	0.88	.379
	Articulation * Affricate	0.06	0.12	.908

Note. The asterisk in the *p*-value column stands for significance ($p < .05$).

Similarly to the results in the previous section, an effect of -0.53 points on the Likert scale was found for articulation in Scale 1 (*formal–informal*). Additionally, the

presence of affricates had a significant effect of 0.51 points for Scale 3 (*white collar–blue collar*); on average, stimuli with affricates were ranked 0.51 points higher than stimuli without (CI [0.017 ... 0.493], $t = 2.12$, $p = .036$). However, it can be observed in [Table 6](#) that the results for the other scales were not found significant. Moreover, all of the effects of the interaction between the type of articulation and the presence of affricates were not statistically significant.

4. Discussion

The purpose of this study was to investigate whether Polish speakers perceived the contrast between single-articulated and rearticulated geminates, and whether this perception is motivated by social implications related to the standard variety of Polish.

4.1. Interpretation of results

Overall, the results found that stimuli with rearticulated geminates were judged higher in formality on all the affective scales, which suggests that this type of pronunciation is considered more ‘proper’ by Polish speakers. This is in line with the hypotheses, which stated that there will be a difference in perception for single- and rearticulation, and that due to the relation between rearticulation and the written form of Polish geminates ([Kozyra, 2008](#)), this type of pronunciation will carry social meanings similar to standard Polish. Furthermore, these findings corroborate the framework of social meaning construction presented in [Soukup \(2013\)](#).

The ranking averages were found significant for Scale 1, which judged the level of formality—the significance in this case could be explained by how straightforward the scale was, but it could also suggest that when judging the correctness of someone’s speech, the participants first took note of how formal the person sounded. On the

other hand, the differences in the ratings on the other three scales— Scale 2 (*suitable–not suitable* (for public speaking)), Scale 3 (*white collar–blue collar*), and Scale 4 (*more eloquent–less eloquent*)—were not found significant. This is somewhat unexpected, as previous findings suggest that standard Polish is associated with social cues, such as higher education ([Markowski, 2009](#)) and higher social class ([Kulak, 2019](#)), as well as being primarily considered to be the variety used by news and radio presenters ([Smakman, 1999](#)).

The scale with the lowest difference in rankings for both types of articulation, Scale 2, judged the speaker's suitability for public speaking, and it could have been the vagueness of that description that might have been confusing for the participants. Perhaps phrasing it differently, e.g. along the lines of 'could work as a news presenter', would have made the purpose of the scale more clear to the listeners without explicitly stating the purpose of the experiment.

The additional post hoc analysis was performed on the basis of previous findings on the frequency of rearticulation, which indicated affricates to be the consonant type that was rearticulated most often ([Thurgood, 2001](#); [Thurgood & Demenko, 2003](#); [Rojczyk & Porzuczek, 2019b](#)). Because of this, it could be assumed that upon hearing the affricate geminate produced with single-articulation, the listener would consider it 'unnatural' and as a consequence, rate it lower in formality on the scales. The results corroborated this assumption, as the SA stimuli with affricates were overall ranked lower in formality than the RA stimuli with affricates. Additionally, significance was found for the presence of affricates in Scale 3 (*white collar–blue collar*)—test items with affricates were ranked higher in formality than test

items without them. However, this finding is difficult to interpret due to it being the only instance of significance in this part of the analysis. There is also no previous research, to the best of the author's knowledge, that would connect the consonant category of Polish affricates with standard Polish or, in the case of this scale, socioeconomic background.

Another challenge arises when considering that the results for the overall rating averages for each articulation type and the individual means for each rating scale were not significant, and no significance was found for the interaction between the articulation method and the presence of affricates. In summary, while the overall findings for this criterion may well reflect the differences in perception of the two types of articulation for the participants of this study, they cannot be generalised across the entire population of Polish speakers.

4.2. Limitations

Despite being present, the differences in ranking between single- and rearticulation were not substantial and, in most instances, they were not statistically significant—the null hypothesis can be rejected only for a small minority of the results. Additionally, none of the means were higher than 4, the mid-point of the Likert scale, therefore the results were confined to the high-formality end of the scale and neither type of articulation showed a distinct relation to lower intelligence or socioeconomic status.

The lack of statistical significance for some of the affective scale ratings also casts doubt on the suggestion that, based on these results, Polish speakers can be said to distinguish between single- and rearticulation. In a similar fashion to [Rojczyk](#)

(2022), which found no significant difference in the perception accuracy for both single- and rearticulated geminates, the ratings in the present study are too similar to warrant a definitive conclusion.

Along with that, the results were based on a rather small subject sample due to an error during the distribution of the experiment—the survey could not be accessed on mobile devices, which could be assumed to have discouraged potential participants. The two technical errors regarding the stimuli further limited the analysis and created an imbalance in the number of the tested SA and RA stimuli.

Although the stimuli were controlled for geminate articulation, other idiolectal or dialectal features that could not be associated with the standard variety of Polish were not taken into account. This could potentially influence the rankings, as the participants may have been guided by other phonetic cues to make their judgements on the stimuli, such as the pronunciation of nasal vowels. According to Gussmann (2007), the word-final /ɔ̃/ vowel is realised as [ɔw̃], an oral vowel followed by a labial-velar nasalised glide. However, the realisation of this segment in speech can also vary among [ɔm], [ɔw], and [ɔ], the first of which is especially common in the Greater Poland area (Każmierski & Szlandrowicz, 2020). These nonstandard variants could evoke connotations of dialectal speech or ‘incorrect’ Polish (Dunaj, 2006), which go against the principles of the standard language. More importantly for the present study, it could result in the participants judging the stimuli to be less formal or less ‘correct’.

4.3. *Future research*

On the basis of the aforementioned gaps in the present research, the following aspects could be improved and elaborated on in future studies.

In general, testing a larger participant group would create a more well-rounded sample of the studied population. Due to the lack of literature on the Polish geminates and their articulation, follow-up studies are needed, particularly to further investigate the reasons for the distribution of single- and rearticulation. The next step to obtaining a complete understanding of the potential social meanings carried by the articulation of the geminates is a production study which could shed light on the hereby collected data, and examine whether the listeners' perceptions and the speakers' productions converge.

Since the ability of Polish speakers to recognise the contrast in articulation is yet to be fully corroborated by data and was not the main focus of the present study, it should also be examined in more detail, possibly by means of a perception experiment such as an ABX discrimination task, in which A and B would stand for single- and rearticulation, respectively, and the participants would be asked to match X to one of them.

The methodology used in this study, partly inspired by [Villarreal \(2018\)](#), could be applicable to future research of similar nature, provided that some improvements are made to the preparation of the test items. Firstly, including a warm-up 'politeness' check before the participants could proceed with the test items could serve as an explanation of their later judgements. One method that would give more control over

the production of the stimuli would be to use simple carrier phrases (as in 10) to reduce individual differences to a minimum.

(10) Example carrier sentence

To się wymawia “WORD” po polsku.

‘This is pronounced “WORD” in Polish.’

Following this format, it would be easier to have the sentences fully recorded in standard Polish and even segments frequently varying in pronunciation, e.g. nasal vowels, could be accounted for. Outside of the present application of the method, the matched-guise task can also be beneficial for other sociolinguistic studies or perception experiments.

As mentioned above, features representative of the Polish dialects, as opposed to the standard variety, could influence the way Polish speakers judge speech and correctness. Therefore, including it in the experiment—perhaps by asking the listeners to point to where they think the speaker is from on a map (as in Villarreal, 2018)—could verify how impactful phonetic dialectal cues or geographic location are in the case of geminate production and perception.

Lastly, it would be interesting to investigate whether gender is an influential factor with regard to the distribution of rearticulation, as previous sociolinguistic research has consistently found that men use nonstandard forms more frequently than women, and that women adapt the new forms more often than men (Labov, 1990). This poses an interesting question in the case of Polish geminates, as rearticulation is

technically both the nonstandard, due to it being the less common articulation method, and the new variant.

5. Conclusion

This research investigated the perception and social meanings of Polish geminate production—single-articulation and rearticulation—through the means of a matched-guise perception task. The participants were asked to listen to stimuli with one method of articulation and rank the recordings on affective scales which indicated socioeconomic background, formality, and intelligence.

The results suggest that Polish speakers can distinguish between the two types of geminate articulation. Additionally, based on affective scale ratings, it was found that rearticulated geminate consonants carry social meanings indicative of standard Polish, such as higher intelligence, education level, and socioeconomic status. The additional analysis that was conducted on the interaction between articulation and the presence of affricates did not lead to a definitive conclusion.

This study aims to contribute towards a deeper understanding of the Polish geminates and their behaviour, as well as provide a new angle to the already existing literature on the topic, which consists mostly of research on production. To the author's knowledge, it is the first attempt not only at examining the geminates from the standpoint of perceptual and social meaning, but also investigating the speaker's motivation behind choosing the type of articulation.

References

- Davis, S. (2011). Geminate. In *The Blackwell Companion to Phonology* (M. Oostendorp, C. J. Ewen, E. Hume & K. Rice, Eds.)
<https://doi.org/10.1002/9781444335262.wbctp0037>.
- Davis, S., & Ragheb, M. (2014). Geminate representation in Arabic. *Perspectives on Arabic linguistics*, 25, 3–19.
- Dunaj, B. (2006). Zasady poprawnej wymowy polskiej. *Język Polski*, 3, 161–172.
- Gussmann, E. (2023). *The phonology of Polish* (1st ed.). Oxford University Press.
<https://doi.org/10.1093/oso/9780199267477.001.0001>.
- Ham, W. (2002). *Phonetic and Phonological Aspects of Geminate Timing* (1st ed.). Routledge. <https://doi.org/10.4324/9781315023755>.
- Hamzah, M. H., Fletcher, J., Hajek, J. (2016). Closure duration as an acoustic correlate of the word-initial singleton/geminate consonant contrast in Kelantan Malay. *Journal of Phonetics*, 58, 135–151. [10.1016/j.wocn.2016.08.002](https://doi.org/10.1016/j.wocn.2016.08.002).
- Hayes, B. (1989). Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry*, 20(2), 253–306. <http://www.jstor.org/stable/4178626>.
- Kawahara, S. (2006). A Faithfulness Ranking Projected from a Perceptibility Scale: The case of [+ Voice] in Japanese. *Language*, 82(3), 536–574.
<http://www.jstor.org/stable/4490204>.
- Kaźmierski, K., & Szlandrowicz, M. (2020). Word-final /ɕ/ in Greater Poland Polish: A Cumulative Context Effect? *Research in Language*, 18(4), 381–394.

- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography* 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>.
- Kozyra, A. (2008). Geminate in Slavonic Languages. *LingVaria: Półrocznik Wydziału Polonistyki UJ poświęcony zagadnieniom języka i językoznawstwa*, 1, 257–266.
- Kuław, K. (2019). Linguistic Prescriptivism in Present-Day Poland: The normative attitudes of the speakers of Polish. In Brunn, S., Kehrein, R. (Eds.) *Handbook of the Changing World Language Map*. Springer, Cham. https://doi.org/10.1007/978-3-319-73400-2_120-1.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2(2), 205–254. doi:10.1017/S0954394500000338.
- Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford: Blackwell Publishers.
- Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Markowski, A. (1999). *Nowy słownik poprawnej polszczyzny*. Warszawa.
- Oh, G. E., & Redford, M. A. (2012). The production and phonetic representation of fake geminates in English. *Journal of Phonetics* 40, 82–91.
- Ohala, M., Solé, M. J., & Beddor, P. (2007). Experimental methods in the study of Hindi geminate consonants. *Experimental approaches to phonology*, 351–368.
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

- Rojczyk, A., & Porzuczek, A. (2014). Acoustic properties of nasal geminates in Polish. *Crossing phonetics-phonology lines*, 347–364.
- Rojczyk, A., & Porzuczek, A. (2019a). Durational properties of Polish geminate consonants. *The Journal of the Acoustical Society of America*, 146(6), 4171–4182.
- Rojczyk, A., & Porzuczek, A. (2019b). Rearticulated geminates are not sequences of two identical sounds: Evidence from Polish affricate geminates. In *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3671–3675). Canberra: Australasian Speech Science and Technology Association Inc.
- Rojczyk, A. (2022). The perception of rearticulated and single-articulated geminates in Polish. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique*, 67(1–2), 118–126. doi:10.1017/cnj.2022.5.
- Russo, M., & Ulfsbjorninn, S. (2017). Breaking the symmetry of geminates in diachrony and synchrony. *Papers in Historical Phonology*, 2, 164–202.
- Smakman, D. (1999). Criteria to define the standard language: A survey among Polish students. *Roczniki humanistyczne*, 47(6), 63–85.
- Soukup, B. (2013). Austrian dialect as a metonymic device: A cognitive sociolinguistic investigation of Speaker Design and its perceptual implications. *Journal of Pragmatics*, 52, 72–82.
- al-Tamimi, F., Abu-Abbas, K., & Tarawnah, R. (2010). Jordanian Arabic Final Geminates: An experimental clinical phonetic study. *Poznań Studies in Contemporary Linguistics*, 46(2), 111–125. <https://doi.org/10.2478/v10010-010-0006-6>.

- Thurgood, E. (2001). The phonetic realizations of phonologically geminate affricates in Polish: The long and the short of it. *Speech and Language Technology*, 5, 9–19.
- Thurgood, E., & Demenko, G. (2003). Phonetic realizations of Polish geminate affricates (M. J. Solé, D. Recasens & J. Romero, Eds.). In *Proceedings of the 15th International Congress of Phonetic Sciences*, 1895–1898.
- Savinainen-Makkonen, T. (2007). Geminate template: A model for first Finnish words. *First Language*, 27(4), 347–359. <https://doi.org/10.1177/0142723707081728>.
- Villarreal, D. (2018). The construction of social meaning: A matched-guise investigation of the California Vowel Shift. *Journal of English Linguistics*, 46(1), 52–78.
- Vet, D. J. (2024). *Experiment Designer (ED): IDE for coding scientific experiments in Java/JavaScript*. <https://www.fon.hum.uva.nl/dirk/ed.php>.
- Żmigrodzki, P. (ed). (n.d.). *Wielki słownik języka polskiego* (The great dictionary of Polish). <https://wsjp.pl/> (Accessed: 26 April 2024).

Appendices

I. Stimuli sentences with their English translation, and with the geminates and their derivation type per stimulus

No.	Sentence	Translation	Geminate	Lexical (L)/ morphological (M)
1.	Wysoki komfort zapewniają miękkie i wygodne siedziska, dodatkowy plus to unoszone, regulowane zagłówki, które ustawiać można w wygodnej dla siebie pozycji.	High comfort is ensured by the soft and comfortable seats, an additional plus being the liftable, adjustable headrests, which can be set in a comfortable position.	k	L
			v	M
2.	Lekka trema jest zrozumiała, w końcu nie codziennie mamy okazję rozmawiać z pisarzami, których nazwiska padają na nieformalnej noblowskiej giełdzie.	The slight stage fright is understandable; after all, it is not every day that we have the opportunity to talk to writers whose names are being put forward in the informal Nobel debate.	k	L
			n	M
3.	Oddech w tym momencie jest mocno przyspieszony z charakterystycznym świstem lub furczeniem oraz wydłużoną fazą wydechową.	Breathing at this point is strongly accelerated with a characteristic wheezing or whirring sound and a prolonged expiratory phase.	d	M
			m	M
4.	Niepodzielnymi władcami sawanny są ssaki , którym najlepiej spośród wszystkich zwierząt udało się dopasować do warunków naturalnych otoczenia.	The undisputed rulers of the savannah are the mammals, which have succeeded best of all animals in adapting to the natural conditions of their surroundings.	n	L
			s	L
5.	Zaznaczyła, że nic co osiągnęła w życiu nie było	She pointed out that nothing she has achieved	(ts)	M

	jej dane za darmo i nie jest to zasługa jej urody.	in life has been given to her for free and it is not due to her beauty.	t	M
6.	Olimpia przełamała fatalną passę 7 meczów bez zwycięstwa oraz 4 meczów bez wygranej na własnym boisku.	Olimpia broke the terrible streak of 7 matches without a win and 4 matches without a home win.	s	L
			z	M
7.	Aby pomiar metodą bioimpedancji był jak najbardziej poprawny podczas wizyt porannych najlepiej być na czczo lub przynajmniej 4-5 godzin po ostatnim posiłku.	For the bioimpedance measurement to be as correct as possible during morning visits, it is best to be fasting or at least 4-5 hours after your last meal.	n	M
			t̥ʂ	L
8.	Linia kolejowa prowadzi ze Sztambułu w Turcji do Aleppo i z Damaszku do Ammanu w Jordanii.	The railway line runs from Istanbul in Turkey to Aleppo and from Damascus to Amman in Jordan.	p	L
			m	L
9.	Jesienną porą w dżdżysty dzień znów zawitaliśmy w urocze okolice z pogranicza Warmii-Mazur i Mazowsza.	On a rainy autumn day, we once again visited the charming regions on the border between Warmia-Mazury and Masovia.	n	M
			ḍʒ	L
10.	W Brukseli słyszać ciągle głosy, które przypominają, że wiele europejskich państw należy do beneficjentów netto funduszy UE.	There are constant voices heard in Brussels reminding us that many European countries are net beneficiaries of EU funds.	t̥ʂ	M
			t	L

II. Filler sentences with their English translation

No.	Sentence	Translation
1.	Ludzie chętniej się gościli, świąteczną babką częstowano dziesięć dni po wypieku i nadal była wyborna.	People were more willing to be hosted, the holiday cake was served ten days after baking and was still delicious.
2.	Dzięki naszym wskazówkom łatwiej będzie Ci uporać się z domową usterką, czy też większym przedsięwzięciem remontowo - budowlanym.	Our tips will make it easier for you to deal with a fault in the house or a major renovation and construction project.

3.	Do tej pory uznawane były tylko kwalifikacje tych polskich pielęgniarek, które ukończyły studia licencjackie bądź magisterskie.	Until now, only those Polish nurses with a bachelor's or master's degree have had their qualifications recognised.
4.	Dodatkowe wejście do strefy audytorium zlokalizowane jest w zachodniej części działki, prowadzi do niego ciąg pieszy.	The additional entrance to the auditorium area is located in the western part of the plot, with a pedestrian walkway leading to it.
5.	Lingwistyka stosowana być może nie należy do grona najpopularniejszych kierunków, ale i tak dosyć często widnieje w ofertach dydaktycznych.	Applied linguistics may not be one of the most popular courses, but it is still quite frequently featured in course offers.
6.	Wakacje w słonecznej Italii czy Hiszpanii po przyjęciu wspólnotowej waluty z pewnością będą tańsze.	Holidays in sunny Italy or Spain will certainly be cheaper after the adoption of the common currency.
7.	Większość filmów przygotowanych przez organizatorów prawdopodobnie nigdy nie wejdzie do polskich kin, a są wśród nich produkcje z RPA, Estonii, Iranu i Holandii.	Most of the films prepared by the organisers will probably never be released in Polish cinemas, and they include productions from South Africa, Estonia, Iran and the Netherlands.
8.	Każdy, kto współpracuje z wieloma osobami nad projektem, który korzysta z obszernej, często zmieniającej się bazy danych, stanie nie raz przed problemem scalania zmian.	Anyone working with a number of people on a project that uses a large, frequently changing database will face the problem of merging changes more than once.
9.	Jak długo święta wiara ożywia polską krew, stać będzie Polska stara, bo każdy Polak lew.	lit. As long as holy faith revives Polish blood, Poland of old will stand, for every Pole is a lion.
10.	Jedyne, co zrobił w tej sytuacji nurek, to nabrał powietrza, spodziewając się, że wieloryb ze zdobyczą zanurkuje.	All the diver did in this situation was inhale, expecting the whale to dive with its prey.

III. R Code used in the analysis

```

```{r}
loads libraries
library(readr)
library(tidyverse)
library(dplyr)
library(psych)
library(pastecs)
library(ggplot2)
options(scipen=999)
```

```{r}
reads file
results <- read_csv("results_proc.csv")

```

```

...

Preprocessing data
```{r}
# removes errors from analysis
results <- results[!(results$soundFile %in% c('G1_3_4','G1_7_2')),]

# calculates means, min, max
results <- results[!(results$condition %in% 'F'),] %>% rowwise() %>%
  mutate(ratingAvg = mean(c_across(rating1:rating4)),
         min = min(c_across(rating1:rating4)),
         max = max(c_across(rating1:rating4)))
...

```{r}
calculates difference
results$difference <- results$max - results$min

excludes difference higher than 3
results <- results[!(results$difference>3),]
...

```{r}
# sets 'condition' as factor
results$condition <- as.factor(results$condition)
...

```{r}
marks sentences with affricates
results$affr <- with(results, ifelse(sentence == '5'|sentence == '7'|
 sentence == '9'|sentence == '10', 1, 0))
...

Overall statistical results per condition
```{r}
# linear model
contrast <- cbind (c(-1/2, +1/2))
colnames(contrast) <- c("-G1+G2")
contrasts(results$condition) <- contrast
modelResults <- lm(ratingAvg ~ condition, results)
summary(modelResults)

# confidence intervals
confint(modelResults)
...

```{r}
creates separate dataset for each condition

SA
resultsG1 <- subset(results, condition=='G1') %>% describe()
resultsG1$condition = 'G1'

RA
resultsG2 <- subset(results, condition=='G2') %>% describe()
resultsG2$condition = 'G2'
...

```{r}

```

```

# creates dataframe with statistical results per condition (taken from subsets
above)
results_stat <- data.frame(condition=c('G1','G2'),
                           mean=c(3.307927, 2.896739),
                           sd=c(1.5412531, 1.3501893))
...

## Plot: results per condition
```{r}
overall results per condition bar plot
ggplot (results_stat, aes(x = condition, y = mean, fill = condition)) +
 geom_bar(stat="identity",position="dodge", width = 0.4) +
 geom_errorbar(aes(ymin=mean-sd, ymax=mean+sd), width=.1) +
 scale_fill_manual(values=c("#89CFF0", "#009999")) +
 scale_x_discrete(labels=c('SA', 'RA')) +
 labs(title = 'The rating average per articulation method',
 x = 'Method of articulation',
 y = 'Mean') + theme(legend.position='none') +
 theme(panel.grid.major.x = element_blank(),
 panel.grid.minor.x = element_blank()) +
 theme(plot.title = element_text(color="black", face="bold", size=16, hjust=0.5))
...

Statistical results per scale
```{r}
# creates dataframe with statistical results (taken from the subsets, significance
taken from linear model)

# per rating scale
results_rating <- data.frame(condition=c('G1','G1','G1','G1','G2','G2','G2','G2'),
                             rating=c(1,2,3,4,1,2,3,4),
                             mean=c(3.487805,3.170732,3.280488,3.292683,
                                     2.967391,2.826087,2.858696,2.934783),
                             sd=c(1.8341338,1.4807812,1.7375617,1.5907744,
                                   1.6337854,1.4721495,1.3952945,1.5534743),
                             significance=c(1,0,0,0,1,0,0,0))
...

```{r}
sets 'condition' column as factor
results_rating$condition <- as.factor(results_rating$condition)
...

ggplot(results_, aes(x=affr, y=mean, group=condition, colour=condition)) +
geom_line() + geom_point() + scale_colour_manual(values=c("navyblue", "steelblue"))
Plot: results per scale

```{r}
# results per rating bar plot
ggplot (results_rating, aes(x = rating, y = mean, fill = condition,
                           alpha = as.factor(significance))) +
  geom_bar(stat="identity", position="dodge") +
  geom_errorbar(aes(ymin=mean-sd, ymax=mean+sd), position = position_dodge(0.9),
               width = 0.25) +
  scale_y_continuous(breaks=c(0,1,2,3,4,5,6)) +
  scale_fill_manual(values=c("#89CFF0", "#009999"), name = 'Method of articulation',
                   labels = c('SA', 'RA')) +
  labs(title = 'The rating average per scale',
       x = 'Scale number',

```

```

        y = 'Mean') + theme(panel.grid.major.x = element_blank(),
                             panel.grid.minor.x = element_blank()) +
        theme(plot.title = element_text(color="black", face="bold", size=14, hjust=0.5))
+
    scale_alpha_manual(values = c(0.5, 1), guide = 'none')
...

# Linear models per scale
```{r}
rating 1 statistics
rat1 <- results %>% select(subject, rating1, condition, sentence, affr)

linear model
contrast <- cbind (c(-1/2, +1/2))
colnames(contrast) <- c("-G1+G2")
contrasts(rat1$condition) <- contrast
modelRat1 <- lm(rating1 ~ condition, rat1)
summary(modelRat1)

confidence intervals
confint(modelRat1)
...

```{r}
# rating 2 statistics
rat2 <- results %>% select(subject, rating2, condition, sentence, affr)

# linear model
contrast <- cbind (c(-1/2, +1/2))
colnames(contrast) <- c("-G1+G2")
contrasts(rat2$condition) <- contrast
modelRat2 <- lm(rating2 ~ condition, rat2)
summary(modelRat2)

# confidence intervals
confint(modelRat2)
...

```{r}
rating 3 statistics
rat3 <- results %>% select(subject, rating3, condition, sentence, affr)

linear model
contrast <- cbind (c(-1/2, +1/2))
colnames(contrast) <- c("-G1+G2")
contrasts(rat3$condition) <- contrast
modelRat3 <- lm(rating3 ~ condition, rat3)
summary(modelRat3)

confidence intervals
confint(modelRat3)
...

```{r}
# rating 4 statistics
rat4 <- results %>% select(subject, rating4, condition, sentence, affr)

# linear model
contrast <- cbind (c(-1/2, +1/2))

```

```

colnames(contrast) <- c("-G1+G2")
contrasts(rat4$condition) <- contrast
modelRat4 <- lm(rating4 ~ condition, rat4)
summary(modelRat4)

# confidence intervals
confint(modelRat4)
```

Statistical results: affricates
```{r}

results$affr <- as.factor(results$affr)

# descriptive statistics divided by condition and affricates
describeBy(results, list(results$condition, results$affr))
```

```{r}
# creates dataframe of descriptive statistics per affricate and condition (taken
from subset above)
results_affr <- data.frame(condition=c('G1','G1','G2','G2'),
                           affr=c(0,1,0,1),
                           mean=c(3.17,3.49,2.83,3.01),
                           sd=c(1.52,1.58,1.27,1.48))

results_affr$affr <- as.factor(results_affr$affr)
```

Plot: overall results per condition and affricates
```{r}
ggplot(results_affr, aes(x=condition, y=mean, group=affr, colour=affr)) +
  geom_line(lwd=1.2) + geom_point(size=2.2) +
  scale_x_discrete(labels=c('SA','RA')) +
  labs(title = 'The rating average per articulation method',
       subtitle = 'Based on the presence of affricates in the stimuli',
       x = 'Method of articulation',
       y = 'Mean',
       colour = "Affricates present") +
  scale_color_manual(labels = c("No", "Yes"),
                    values = c("#89CFF0", "#009999")) +
  theme(panel.grid.major.x = element_blank(),
        panel.grid.minor.x = element_blank()) +
  theme(plot.title = element_text(color="black", face="bold", size=16, hjust=0.5),
        plot.subtitle = element_text(size = 14, hjust=0.5))
```

```{r}
# results affricate statistics

# setting contrasts
contrast <- cbind(c(-1/2, +1/2))
colnames(contrast) <- c("-noAff+yesAff")
contrasts(results$affr) <- contrast
colnames(contrast) <- c("-G1+G2")
contrasts(results$condition) <- contrast

# linear model
modelAffr <- lm(ratingAvg ~ condition * affr, results)

```

```

summary(modelAffr)

# confidence intervals
confint(modelAffr)
```

Linear models per rating scale: affricates
```{r}
# sets 'affricate' column as factor in each rating's dataframe
rat1$affr <- as.factor(rat1$affr)
rat2$affr <- as.factor(rat2$affr)
rat3$affr <- as.factor(rat3$affr)
rat4$affr <- as.factor(rat4$affr)
```

rating 1 affricate
contrasts(rat1$affr) <- contrast
modelRat1Affr <- lm(rating1 ~ condition * affr, rat1)
summary(modelRat1Affr)

confidence intervals
confint(modelRat1Affr)
```

# rating 2 affricate
contrasts(rat2$affr) <- contrast
modelRat2Affr <- lm(rating2 ~ condition * affr, rat2)
summary(modelRat2Affr)

# confidence intervals
confint(modelRat2Affr)
```

rating 3 affricate
contrasts(rat3$affr) <- contrast
modelRat3Affr <- lm(rating3 ~ condition * affr, rat3)
summary(modelRat3Affr)

confidence intervals
confint(modelRat3Affr)
```

# rating 4 affricate
contrasts(rat4$affr) <- contrast
modelRat4Affr <- lm(rating4 ~ condition * affr, rat4)
summary(modelRat4Affr)

# confidence intervals
confint(modelRat4Affr)
```

```