RELEASE BURSTS VS. FORMANT TRANSITIONS IN POLISH STOP PLACE PERCEPTION

Grzegorz Aperliński, Geoffrey Schwartz

Adam Mickiewicz University in Poznań
gaperlinski@wa.amu.edu.pl, geoff@wa.amu.edu.pl

ABSTRACT

A perception study with acoustically manipulated tokens sought to establish the relative weight of noise vs. formant transition cues in the perception of stop place of articulation by Polish listeners, and to replicate previous findings from a crowdsourced experiment. Results suggest that noise cues are primary for Poles in stop place identification. It is argued that this finding is compatible with other aspects of Polish phonology, including the obligatory release of coda stops, and the relatively pure quality of vowels in the language.

Keywords: Place of articulation, Polish, speech perception, phonetics-phonology interface.

1. INTRODUCTION

Speech perception research has identified two primary types of acoustic cue for the identification of place of articulation of stop consonants. The spectral properties of the aperiodic noise associated with consonant release and frication constitute one type of cue, while the other consists of transitional patterns of formant movement on vowels adjacent to the consonantal constriction. In spectrographic displays, these acoustic properties are clearly separable entities. As such, they may be manipulated for research into the relative perceptual weight of noise vs. transitional cues to consonant place identification and discrimination of place contrasts.

Much of the published literature [1], [2], [3] on stop place perception has found that formant transitions play a more prominent role than burst noise for place perception. Wright [4] suggests that an explanation for these findings may be found in the fact that formant transitions are housed in a more robust periodic portion of the speech signal that maintains its perceptual robustness even in noisy listening conditions.

Due perhaps to the prominent place of formant transitions in the literature, few studies have investigated the relative weight of burst noise vs. transitions across languages. In other words, the studies cited above were all based on American English and implicitly assumed to apply to perception more generally. One exception is found in [5], which presents stop place perception data in Korean and English. In their study, formant transitions played a more significant role than noise bursts. However, there was a cue-language interaction by which formant transitions played a greater role in Korean than in English. Those authors attribute this finding to the fact that Korean contains a three-way laryngeal contrasts that is cued in part by burst amplitude and aspiration noise. They suggest therefore that the noise cues are somehow 'reserved' for the laryngeal contrast, forcing listeners to rely more heavily on formant transitions for place of identification. The fact that Korean leaves coda stops unreleased, forcing listeners to rely on transitions, is compatible with this interpretation. Thus, the relative weight of burst vs. transitional cues may be related to other aspects of a given language's phonology.

With regard to Polish, there is evidence to suggest that the relative weight of noise cues in place perception is greater than in other languages. In [6], [7], [8] it was shown that L1 Polish listeners identify sibilant contrasts at near ceiling levels on the basis of fricative noise spectra alone. This is in contrast to an earlier study [9], in which English-speaking listeners had trouble discriminating Polish /ʂ/ from /ɕ/ without the aid of formant transitions. With regard to stops in Polish, in [10] it was found that transitions played a smaller role in /p t k/ identification than burst spectra, while the opposite pattern was observed for English. The discrepancy in the role of formant transitions between the stop study and the work on sibilants stems from the fact that the stimuli in [10] did not contain palatal consonants, which have been found to have greater effects on vowel formants in Polish than other consonants [11].

The study in [10] was based on an on-line, or crowdsourced experiment. This paper attempts to replicate the results of the crowd-sourced study by means of a lab-based experiment. Thus, in addition to providing data on stop place perception in Polish, the results may inform the debate about crowdsourcing in speech perception (cf. [12]).

2. METHOD

2.1. Materials

The stimuli were created from tokens of monosyllabic words produced in a carrier phrase by a female native speaker of Polish. Six base CV sequences /pa ta ka pu tu ku/ were used to create a set of 15 stimuli of two types. In one type release bursts were removed, while in the other bursts from one place were spliced onto transitions from another. Table 1 summarizes the acoustic properties of the base syllables

Table 1: The acoustic properties of stimuli used.

Token	VOT/Burst duration (ms)	F2/F3 at vowel onset (Hz)	Duration of transition (ms)
pa	15	1514/2832	30
ta	23	1858/3067	37
ka	31	1868/2566	39
pu	25	1029/2754	24
tu	30	1129/2900	28
ku	48	1337/2733	31

2.2. Participants and procedure

Thirty nine students at the Faculty of English at Adam Mickiewicz University in Poznań (UAM) participated in the experiment. The experiment was implemented in E-Prime and the Language and Communication Laboratory at the Faculty of English at UAM. The stimuli were presented in random order in two blocks of trials. For each trial, listeners were presented with a slide accompanied by a single repetition of the audio stimulus. The slide asked 'which consonant?' in Polish, and instructed participants to respond by tapping the appropriate key on a keyboard (*p*, *t*, *k*). An additional response, 'something else' was keyed by the letter *i*.

2.3. Data analysis

For the tokens without a burst, the means of the correct identification rates were calculated for the three places of articulation. The statistical significance of the differences between the correct identification rates for labial, coronal and dorsal tokens was assessed using the Friedman test, followed by a Post hoc analysis with Wilcoxon signed-rank tests conducted with a Bonferroni correction. The resulting figures were also compared to the ones in [10] using the Mann-Whitney U test.

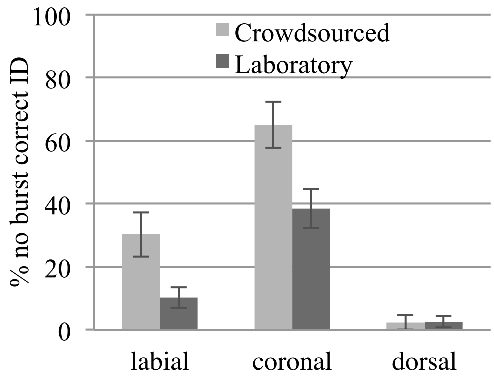
For the mismatched tokens the percentages of identifications matching each cue (% burst ID, % transition ID) were compared using the dependent-samples T-test. The means of these two variables were also compared with the corresponding values from [10] by means of the independent-samples T-test. Additionally, mean % burst ID and % transition ID were also calculated for each place of articulation and the differences between them were investigated using the Friedman test, followed by a post hoc analysis with Wilcoxon signed-rank tests conducted with a Bonferroni correction. Finally the mean response times were calculated for each burst condition and compared using a repeated measures one-way ANOVA, followed by post hoc tests using the Bonferroni correction.

3. RESULTS

3.1 Tokens without bursts

The results of the laboratory experiment are presented alongside the ones obtained in [10] to facilitate comparisons between the two experiments. Figure 1 illustrates the participants' performance on no-burst tokens.

Figure 1: The percentage of correct identification for no-burst tokens for the three POAs.

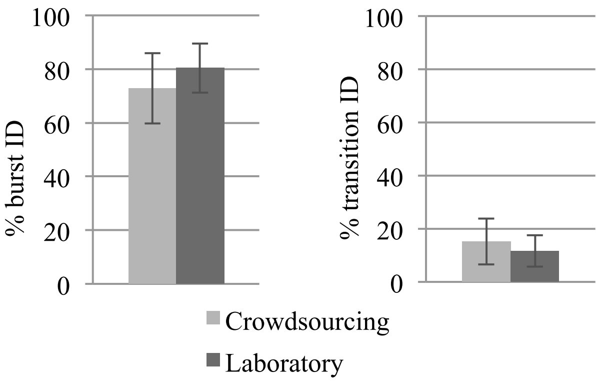


The participants performed worst with the dorsal tokens ($M = 2.56$, $SD = 2.33$), followed by the labial tokens ($M = 10.26$, $SD = 3.27$) and the coronal tokens ($M = 38.46$, $SD = 6.21$). The differences between these means were statistically significant ($\chi^2(2) = 28.31$, $p < .001$). A post-hoc analysis revealed a significant difference between the labial and coronal tokens ($Z = -3.42$, $p = .001$) as well as the coronal and dorsal tokens ($Z = -4.05$, $p < .001$). The difference between the labial and dorsal tokens was not significant ($Z = -2.12$, $p > .017$). The comparison of the means obtained in the current and previous experiment revealed that there was a statistically significant difference in the identification of the labial tokens ($U = 495.0$, $p = .001$) but not the coronal ($U = 658.5$, $p > .05$) or the dorsal ($U = 803.5$, $p > .05$) tokens.

3.2 Mismatched tokens

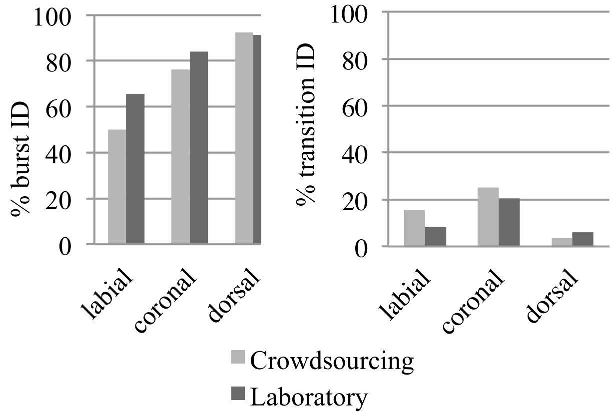
The mean percentages of the identification of mismatched tokens are presented in Figure 2, and broken down by cue (burst or transition) in Figure 3. Polish listeners relied more on the burst cue (M = 80.34, SD = 9.11) than the transition cue (M = 11.65, SD = 5.91) to make place judgments. As in [10], the difference between the mean burst identification rate and the mean transition identification rate was statistically significant, $t(38) = 32.15$, $p < .001$. When these values were compared with the ones from the previous experiment (burst: M = 72.87, SD = 13.12; transition M = 14.92, SD = 8.64) statistical significance was attested for the differences in the burst identification rates, $t(80) = -2.96$, $p = .004$, but not for the differences in the transition identification rates, $t(80) = 1.98$, $p > .05$.

Figure 2: The percentage of identification of mismatched tokens based on the burst (left) and transition (right) cues.



The analysis confirmed that there was a statistically significant difference between the places of articulation both in the case of % burst ID, $\chi^2(2) = 28.31$, $p < .001$, as well as % transition ID, $\chi^2(2) = 27.87$, $p < .001$, which is also consistent with the results in [10]. For % burst ID, post-hoc analysis indicated that there was a statistically significant difference between labial and coronal POA ($Z = -4.07$, $p < .001$) as well as between labial and dorsal POA ($Z = -4.70$, $p < .001$). The difference between coronal and dorsal POA was not statistically significant ($Z = -1.79$, $p > .05$). These results suggest that the participants were more likely to identify a token based on the burst noise if it had a coronal or dorsal burst noise. For % transition ID, post-hoc analysis revealed a statistically significant difference only between labial and coronal POA ($Z = -3.72$, $p < .001$). No statistical significance was found for the differences between labial and dorsal POA ($Z = -.72$, $p > .05$) as well as for coronal and dorsal POA ($Z = -1.73$, $p > .05$).

Figure 3: Identification of mismatched tokens based on burst (left) and transition (right) cues.



3.3 Response times

To provide further insight into the relative weight of burst noise vs. transitions, mean response times for the four burst conditions were calculated. A one way ANOVA revealed a main effect of burst type ($F(3, 102) = 36.39$, $p < 0.001$). Post-hoc tests indicated that the mean response time for tokens without burst noise (M = 2154.66, SD = 530.28) was statistically significantly different from response times for tokens with labial (M = 1630.17, SD = 498.66), coronal (M = 1623.76, SD = 392.77) and dorsal (M = 1499.35, SD = 430.06) bursts. These results suggest that the absence of burst noise significantly slowed down the participant's responses compared to the mismatched tokens.

4. DISCUSSION

The results of the present lab-based experiment replicate the earlier findings that noise bursts play a more significant role than formant transitions in the identification of place of articulation of stop consonants in Polish. Considering the prominence of formant transitions in previous published works on place perception, this result may be seen as surprising. However, most of that literature has been based upon studies of English. Since one's L1 background has been shown to affect perception, there is no reason to expect equal weightings across languages. Certain aspects of Polish phonology may lead to the expectation of noise-based place perception of consonants. In what follows, we shall briefly discuss these features and how they relate to the phonetics of stop consonants.

The first relevant fact is that with the exception of those appearing in homorganic clusters, post-vocalic stops in Polish are always produced with an audible release. This is in contrast to English, in which coda stops are frequently left unreleased. If a language permits unreleased stops, it may be assumed that listeners may recover place of articulation on the basis of the formant transition alone. Thus, according to the H&H theory of Lindblom [13], speakers may spare the effort of producing final release bursts, yet still maintain 'sufficient discriminability'. Since in Polish coda stops are produced with an obligatory release burst, sufficient discriminability is not guaranteed on the basis of formant transitions alone. It may therefore be expected that noise should play a primary role in consonant place identification.

Aspects of the Polish vowel system are also compatible with the findings of the current study. Unlike English in which vowels undergo numerous diphthongization processes, Polish vowels are relatively pure in quality. When vowels are pure in quality, less of their duration is available to listeners as CV transitions to aid in consonant identification. Thus, it may be expected that formant transitions might play a lesser role in place perception in languages with relatively stable vowel systems. As a further consequence, the greater role of noise in stop place perception may serve as a guard against consonant lenition processes, which are largely unattested in Polish. By contrast in languages with more dynamic vowel systems such as English or Danish, consonant lenition is much more common.

With regard to specific places of articulation, the results revealed some interesting patterns. First, the likelihood of identification based on noise bursts was dependent to some degree on place of articulation. In particular, labial bursts were least likely to serve as the basis for identification, while dorsal bursts were the most likely to do so. This finding may be explained by the fact that VOT and burst noise associated with labials is typically weakest among the three major places, while dorsals showed longer VOT. These tendencies were also observed in the experimental stimuli. The other notable pattern was that tokens identified on the basis of transitions, both with or without bursts, were apparently more likely to be heard as coronal. This finding should not be expected on the basis of the acoustic properties of the stimuli – the coronal transitions were not more acoustically prominent than those of the other places with respect to CV transition duration or F2/F3 onset frequencies. It is therefore possible that listeners were biased toward a coronal response, which may be claimed to support the notion that coronal is the unmarked place of articulation.

Finally, a comparison of the crowdsourced and laboratory results shows that the dominance of burst percepts was greater in the laboratory study. From the point of view of perceptual robustness, formant transitions have been found to be more easily recoverable in poor listening conditions [3]. It may be hypothesized that the listeners in the laboratory study, in which the listening conditions may be assumed to have been better, were able to focus their attention more closely on the noise cues than those in the online experiment.

6. REFERENCES

[1] Kewley-Port, D., Pisoni, D.B., Studdert-Kennedy, M. 1983. Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *J. Acoust. Soc. Am.* 73, 1779–1793.

[2] Walley, A.C., Carrell, T.D. 1983. Onset spectra and formant transitions in the adult's and child's perception of place. *J. Acoust. Soc. Am.* 73, 1011-1022.

[3] Wright, R. 2001. Perceptual cues in contrast maintenance. In: Hume, E., Johnson, K. (eds), *The role of speech perception in phonology*. San Diego: Academic Press, 251-277.

[4] Wright, R. 2004. Perceptual cue robustness and phonotactic constraints. In Hayes, B. Kirchner, R., Steriade, D. (eds), *Phonetically based phonology*. Cambridge: Cambridge University Press, 34-57.

[5] Hume, E., Johnson, K, Seo, M., Tsardenelis, G. 1999. A cross-linguistic study of stop place perception. *Proc. 14th ICPhS San Francisco*, 2069-2072.

[6] Żygis, M., Hamann, S. 2003. Perceptual and acoustic cues of Polish coronal fricatives. Proc. 15th ICPhS, Barcelona, 395-398.

[7]Nowak, P., 2006. The role of vowel transition and frication noise in the perception of Polish sibilants. *Journal of Phonetics* 34, 139-152.

[8] Żygis, M., Padgett, J., 2010. A perceptual study of Polish fricatives, and its implications for historical sound change. *Journal of Phonetics* 38, 207-226.

[9] Lisker, L. 2001. Hearing the Polish sibilants [s ś s]: phonetic and auditory judments. In: Grønnum, N., Rischel, J. (eds.), *To honour Eli Fischer-Jørgensen*. Travaux du cercle linguistique de Copenhague XXXI. Copenhagen: Reitzel, 226-238.

[10] Schwartz, G., Aperliński, G., 2014. The phonology of CV transitions. In Cyran, E., Szpyra-Kozłowska, J., (eds), *Crossing phonetics-phonology lines*. Newcastle: Cambridge Scholars Publishing, 277-298.

[11] Nowak, P., 2006. Vowel reduction in Polish. Ph.D. dissertation. University of California at Berkeley.

[12] Cooke, M., Barker, J., Garcia-Lecumberri, M.L. 2013. Crowdsourcing in speech perception. In: Eskanazi, M. et al. (eds), *Crowdsourcing for speech processing*. Sussex: Wiley, 137-172.

[13] Lindblom, B. 1990. Explaining phonetic variation: a sketch of the H&H theory. In Hardcastle, W. Marchal, A. (eds), *Speech Production and Speech Modelling*. The Netherlands: Kluwer Academic, 403-439.