INTERSPEECH 2009 BRIGHTON

Voicing profile of Polish sonorants: [r] in obstruent clusters

J. Sieczkowska[1], B. Möbius[1,2], A. Schweitzer[1], M. Walsh[1], G. Dogil[1]

[1] Institute for Natural Language Processing, University of Stuttgart
[2] Institute of Communication Sciences, University of Bonn

jagoda.sieczkowska@ims.uni-stuttgart.de, bernd.moebius@ims.uni-stuttgart.de,
antje.schweitzer@ims.uni-stuttgart.de, michael.walsh@ims.uni-stuttgart.de,
Grzegorz.Dogil@ims.uni-stuttgart.de

Abstract

This study aims at defining and analyzing voicing profile of Polish sonorant [r] showing the variability of its realizations depending on segmental and prosodic position. Voicing profile is defined as the frame-by-frame voicing status of a speech sound in continuous speech. Word-final devoicing of sonorants is shortly reviewed and analyzed in terms of the conducted corpus-based investigation. We used automatic tools to extract consonants' features, F0 values and obtain voicing profile. The results show that liquid [r] devoice word and syllable finally, particularly with left voiceless stop context.

Index Terms: sonorant, liquid, voicing, Polish, speech database.

1. Introduction

The feature [voice] is considered to be part of the set of the universal distinctive features [13]. Languages employ this feature when their consonants' types can be described in terms of presence or absence of voicing in the possible phonotactic contexts that are relevant for each language [13]. From the speech production perspective, voicing is defined as the presence of a periodic vocal fold vibration that produces a periodic excitation signal. Its acoustic consequences can be observed in the spectrum as the presence of low-frequency energy [6], [19].

Studies in voicing focus on voice onset time (VOT) [14], which is defined as the time interval between the stop release and the onset of voicing of the following vowel. Most languages are examined in terms of their obstruent voicing profile, while in Polish not only obstruents but also sonorants form oppositions in terms of voicing status: "Polish appears to display the common phenomenon of obstruent devoicing in word-final position, as well as voice assimilations in consonant clusters, both within words and at word junctures; additionally, sonorant consonants are devoiced in some positions [8]." Apart from the VOT analysis, a few other approaches to voicing analysis have been proposed: three category division into 'voiced', 'voiceless unaspirated' and 'voiceless aspirated' [1] or binary divisions based on articulatory features [2], [17]. This study is focused on the automatic analysis of temporal voicing values and their specification in context. Thus the description of Polish sonorants' voicing profiles is scalar rather than binary.

Sonorants are considered to be voiced as their production and cavity configuration includes spontaneous voicing [2], [9]. Spontaneous voicing refers to 'narrowing the air passage to a point where the rate of flow is reduced below the critical value needed for the Bernoulli effect to take place' following Chomsky and Halle's description [2]. "Sonorants are therefore articulated as voiced rather than voiceless unless some contrary gesture intervenes, such as spreading apart or tightly constricting the vocal folds so as to inhibit their vibration. Formally, spontaneous voicing can be expressed as the introduction into laryngealy unmarked sonorants of the glottal tension dimension (which by default implicates the gesture [slack]) [12]". Gussmann [8] notes that "Polish sonorants are voiced except for a position between voiceless consonants or after a voiceless obstruent before pause", like in krtań [krtaŋ] *larynx,* wiatr [vjatr̥] *wind,* kadr [katr̥] *frame* or rytm [rɨtm̥] *rhythm*. Sonorants' devoicing depends also on the dialect and other socio-phonetic features. For example if the sonorant's preceding obstruent preserves its voicing the word-final sonorant is voiced as well. In some cases a devoiced cluster might sound unnatural (like in wydm [wɨdm̥] *dune, gen. pl*), however, despite that, there are variants of sonorants that can only be devoiced, e.g. the ones in derivatives *–izm, -yzm* (marks-izm [markɕism̥] *Marxism*) [9].

The number of Polish syllables in the word is always the same as the number of the vowels in the word. Thus we can say that only the vowel can form a syllable nucleus. "The only other element that can appear in the rhyme is a sonorant consonant [8]". It has also been observed [8], that a single sonorant which separates two vowels will always be assigned to the following syllable, like in domy [do.my] *house, pl.* In modern Polish sonorants are admitted in the rhyme and in the word-final clusters of consonants like in łga[rstw] *lie, gen.pl.*[8]. Consonanants occurring in Polish word- and syllable-initial positions are said to have 'peculiar clustering possibilities' [8], where there are cases in which a sonorant can occur before an obstruent like in [mkn]ąć *to speed* or in four-consonantal sequences in monosyllabic words like in [drgn]ąć *to shudder* [8]. As devoicing concerns sonorants in word-final position, Gussmann [8] argues that it is dependent on syllable structure, because sonorants cannot be incorporated into a syllable structure that violates the Sonority Sequencing Principle. On the other hand, other studies, e.g. Rubach [20],[21] argued that it is not the syllable structure that licences devoicing but that the voice assimilation of the sonorants is governed by the linear adjacency of laryngeal nodes. It has been also pointed out in [20] that rules of voicing in Polish interfere with acquisition of foreign languages by Poles. Common errors occur due to

Final Devoicing (ex. [k] instead of [g] in *big*) and Spreading ([s] instead of [z] in *these people*) irrespective of their position in a prosodic structure [18]. Similarly Rubach [20] claims that sonorants which violate sonority relations are transparent to Final Devoicing and to Regressive Assimilation (e.g. Jędrka [jentr̥ka] *gen.sg*). Devoicing is also a matter of dialectal forms of Polish: Warsaw vs. Cracow form. Voice assimilation in obstruent clusters is the same in both dialects, voicing spreads in the Cracow dialect, but not in the Warsaw dialect [samɔχut rɔmana].
In our research we verify phonological hypotheses on sonorant voicing with the help of computational phonetic tools designed to operate on large phonetically labeled data base.

2. Method

2.1. Voicing profile

Our method differs from traditional acoustic measurements of voicing because the traditional approaches fail to provide sufficient information for sonorants occurring in different phonological environments. Thanks to the automatic tools which extract voicing profiles on a frame-by-frame basis in continuous speech, we can study all phonotactically possible contexts of sonorants. As shown in Figure 1, in the first step we used the ESPS get_F0 tool to obtain a frame-by-frame voicing decision at a sampling rate of 100 frames per second using an autocorrelation technique. The value of the voicing decision is almost always categorical (0 or 1) but see section 2.3. In addition, a list of Polish sonorants was defined: [j, l, w, m, n, ɲ, r].

2.2 Database

The analysis of sonorant consonants' profiles was conducted on the BOSS corpora designed for the Polish version of the BOSS unit selection speech synthesis system [4], [5], [15], [16]. It comprises 115min (3249 utterances) of speech recorded by a professional speaker during several recording sessions supervised by an expert phonetician. The corpus consists of several databases [4]: Base A: Phrases with the most frequent consonant structures. Polish has a number of complex consonant clusters. 258 consonant clusters of various types were used; Base B: All Polish diphones realized in 92 grammatically correct but semantically nonsense phrases; Base C: Phrases with CVC triphones (in non-sonorant voiced context and with various intonation patterns). 664 phrases were recorded for triphone coverage; Base D: Phrases with CVC triphones (in sonorant context and with various intonation patterns). The length of the 985 phrases varied from 6 to 14 syllables to provide full coverage of suprasegmental structures; Base E: Utterances with the 6000 most frequent Polish vocabulary items. 1109 sentences were recorded [4].
The computer coding conventions were drawn up in SAMPA for Polish and in the IPA alphabet [11]. Table below (Tab.1) shows the amount of analyzed sonorants.
The BOSS label files were converted to phone, syllable and word label files in ESPS format in a way that the utterances could be imported by the IMS German Festival TTS System [9]. We defined a new Polish SAMPA phone set for use with Festival and adjusted the BOSS phone labels accordingly.

Sonorants	Number
[r]	3394
[l]	2600
[m]	4216
[n]	7065
[ɲ]	2540
[j]	4605
[w]	3532

Table 1: *Number of analyzed sonorants form the BOSS corpus*

2.3. Feature extraction

We used the IMS German Festival synthesis system because it provides a wide range of information taken from the corpus about prosodic, syllable and segmental structure of the utterances [10]. For each segment we extracted the following features: consonant type, place of articulation, voicing structure of the segment, its preceding and following phone, word position, syllable structure, position in the syllable and voicing profile by extracting the voicing status at 10%, 20%.... 90% of its duration. If the time point of the phone in question happened to lie between two different voicing status values, its value was obtained by interpolation rather than by a categorical decision. Using R [17], the voicing profiles of phones in the following selected contexts were averaged and plotted:

all sonorants occurring in all C clusters
all sonorants occurring with voiceless and voiced stops as left context
[r] occurring with voiceless and voiced stops as left context
all sonorants occurring with voiceless and voiced fricatives as left context
[r] occurring with voiceless and voiced fricatives as left context
[ɲ] with voiceless fricative left context
[l] with voiceless stop left context

Thus voicing profiles are referred to as the frame-by-frame voicing status of the speech sounds. Voicing probability refers to the percentage of the exemplars in the corpus computed at a given temporal position [19].
In the figures below normalized time (percentage of time duration of the segment) is plotted on the x axis, whereas percentage of phones undergoing de-/voicing is plotted on the y axis. In this article only the main results on liquid [r] are presented to give an overview of voicing tendencies in Polish sonorants.

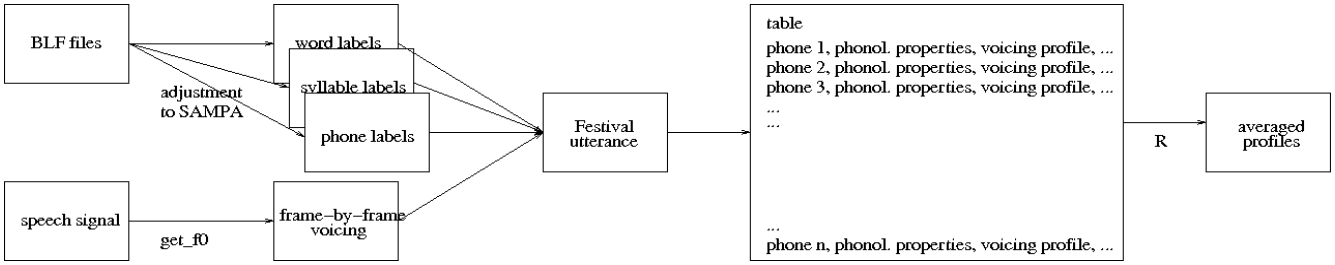

Figure 1: *Illustration of methodological steps of the study.*

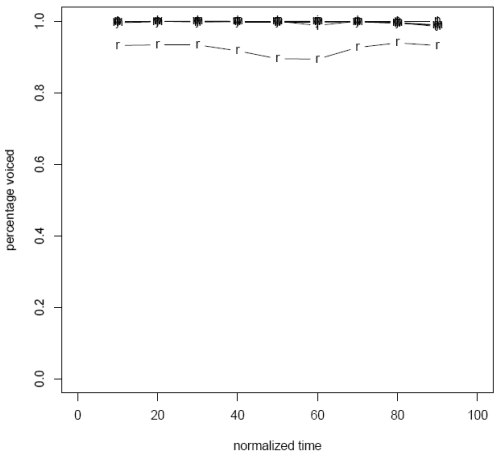

Figure 2: *The voicing profile of word- and syllable-final Polish sonorants in all-left-phoneme context.*

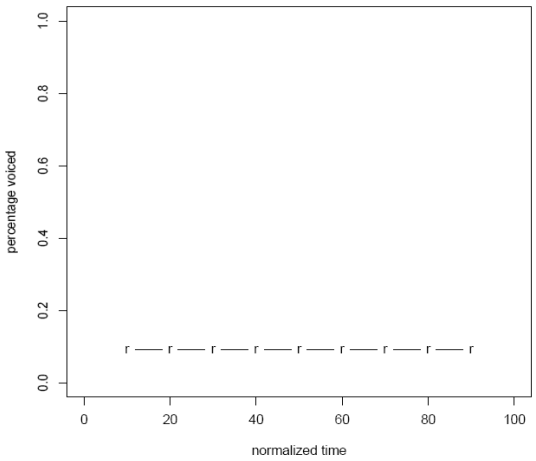

Figure 3*: The voicing profile of word- and syllable-final [r] in voiceless stop left context.*

3. Results

Figure 2 shows all Polish sonorants in word- and syllable-final position in all-left-phoneme contexts. We decided to focus on the sonorant with most variable voicing - [r], as our data show that other sonorants are categorically voiced in all contexts. After further specification of the phoneme context it has been observed (Figure 3) that liquid [r] in word- and syllable-final position preceded by a voiceless stop like in the word 'wiatr' [vjatr̥] *wind,* undergoes devoicing in almost all the exemplars in our corpus. Figure 4 shows the voicing profile of word- and syllable-initial Polish sonorants in all-left-phoneme context. The same tendency is observed – only liquid [r] undergoes devoicing to a sizeable degree. Figure 5 shows liquid [r] in word- and syllable-initial position preceded by a voiceless stop, like in the word 'trawa' [trava] *grass.* In this context, [r] is devoiced in more exemplars than in general all-right-phoneme contexts, similarly for word- and syllable-final position, however here the amount of devoiced exemplars is higher and has a value of slightly more than 80%.

From the example of [r] in word- and syllable-final context preceded by all phonemes (Fig.2), we have seen that its minimum standard deviation value is 0.23 while the maximum value is 0.3. The standard deviation values for [r] in word- and syllable-initial positions with left voiceless stop context (Fig. 5) are min. 0.22 and max. 0.4 respectively. That is why we do not include further information about the standard deviation values of [r] exemplars in the other contexts, as it demonstrates even higher stability.

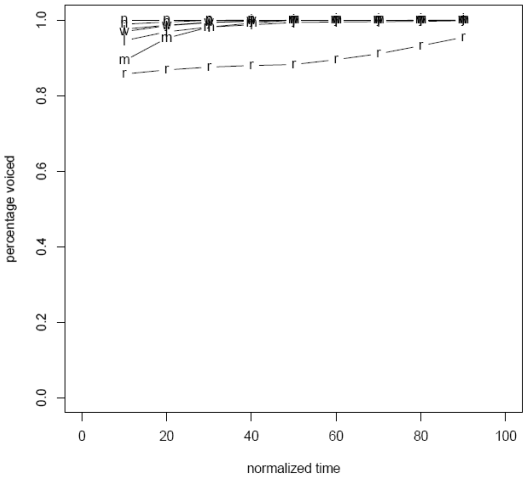

Figure 4*: The voicing profile of word- and syllable-initial Polish sonorants in all-left-phoneme context.*

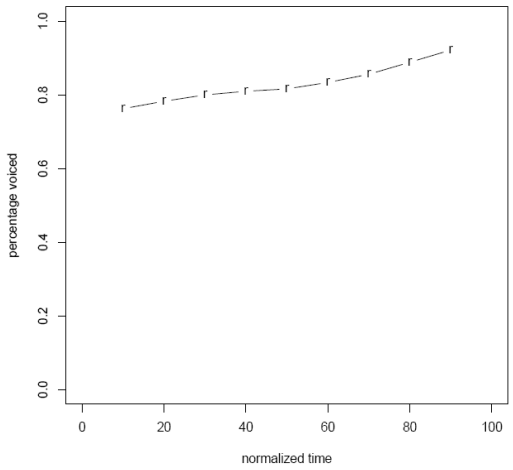

Figure 5*: The voicing profile of word- and syllable-initial [r] in left voiceless stop context.*

4. Discussion and Conclusion

Polish [r] is articulated by producing a trill, which is formed by an egressive pulmonic airstream that causes the vibration of the articulators resulting in multiple, brief and fast touching of the tongue tip against the alveolum and its release [7]. According to previous studies, this movement can occur as much as 25, 30 or even 40 times per second [7]. Both liquid voiced [r] and liquid voiceless [r̥] are articulated at the same place in the oral cavity, although it is observed that in fast speech there exists an allophone of [r] which is no longer a trill but resembles more a stop because of only one-time-touching of the articulators. In this study we analyzed word/syllable initial and final Polish sonorants in general left context (all-phonemes) and following voiceless stops. We observed that, almost exclusively, sonorant [r] undergoes devoicing in those contexts in the data we analyzed. The most interesting example is the case of [r] following voiceless stops, which is almost completely devoiced word- and syllable-finally. Obstruent-liquid clusters are a peculiarity of the Polish language and have thus received many phonological descriptions [3], [8], [20], [21]. The word-final liquid which follows an obstruent has to be accommodated into a phonological string by special abstract devices like extrametricality, segment-government or an empty nucleus. Such a special phonological status licenses phonetic behavior of the odd sonorant. Our data-driven computational procedures of voicing profile extraction demonstrate that both the phonological status and prosodic position of the liquid influence its phonetic properties. However, word- and syllable-final position of the sonorant [r] governs its voicing profile more than its phonological environment, because in this case we observed that the devoicing is much stronger than in the initial position, despite the same phonological context.

5. Acknowledgements

This study is part of the research project SFB 732 A2 funded by the German National Science Foundation (DFG).

6. References

[1] Abercrombie, D. 'Elements of General Phonetics'. Chicago Aldine Atherton, Inc. 1967.
[2] Chomsky, N., Halle, M. "The sound pattern of English'. Harper and Row. NY. 1968.
[3] Cyran, E. "Sound pattern of Polish: phonotactic paradoxes at the right edge of words". Studies in Polish Linguistics. Vol 2, p. 61-89. 2005.
[4] Demenko, G., Bachan J., Möbius, B., Klessa, K., Szymański, M., Grocholewski, S. "Development and Evaluation of Polish Speech Corpus for Unit Selection Speech Synthesis Systems". Interspeech Proceedings 2008.
[5] Demenko, G., Möbius, B., Klessa, K. "The design of Polish Speech Corpus for Unit Selection Speech Synthesis." In press (Language Technology vol.11). 2008.
[6] Dogil, G. et al. "Voicing", Special Issue of Folia Linguistica XXXVIII/1-2. Mouton de Gruyter. Berlin. 2004.
[7] Dukiewicz, L., Sawicka, I. "Fonetyka i fonologia". Wydawnictwo Instytutu Jezyka Polskiego PAN. Krakow. 1995.
[8] Gussmann, E. "Resyllabification and Delinking: The Case of Polish Voicing". Linguistic Inquiry 23:29-56. 1992.
[9] Gussmann, E. The Phonology of Polish. Oxford University Press. 2007
[10] IMS German Festival Home Page. http://www.cstr.ed.ac.uk/projects/festival/. 2009.
[11] The International Phonetic Association (IPA) Homepage. http://www.arts.gla.ac.uk/ipa/ipa.html, accessed on 13 Apr 2008.
[12] Iverson, G., Sang-Cheol Ahn. "English voicing in dimensional theory". Language Sciences 29. 2007.
[13] Jessen, M. "Phonetics and Phonology of Tense and Lax Obstruents in German". John Benjamins Publishing Company. Amsterdam/Philadelphia. 1998.
[14] Keating, P. 'A phonetic study of a voicing contrast in Polish'. PhD Thesis. Brown University. 1980.
[15] Klabbers, E., Stöber, K. "Creation of Speech Corpora for the Multilingual Bonn Open Synthesis System." ISCA ITRW. Scotland. 2001.
[16] Klabbers, E., Stöber, K., Veldhuis, R., Wagner, P., Breuer, S. "Speech Synthesis Development Made Easy: The Bonn Open Synthesis System". Eurospeech Conference. Scandinavia. 2001.
[17] Lisker, L., Ambramson, A. 'A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements'. Word 20, 384-422. 1964.
[18] Lombardi, L. "Laryngeal Features and Laryngeal Neutralization, Ph.D. dissertation. University of Massachusetts, Amherst. 1991.
[19] Möbius, B. "Corpus-based investigations on the phonetics of consonant voicing". Folia Linguistica 38 (1-2), 5-26. 2004.
[20] R project. The R Project for Statistical Computing. (http://www.R-project.org/]. 2001.
[21] Rubach, J. "Nonsyllabic Analysis of Voice Assimilation in Polish". Linguistic Inquiry 27:69-100. 1996.
[22] Rubach, J. "Prevocalic Faithfulness". Phonology, Vol.25 Number 3. Cambridge University Press. 2008.
[23] Shih, Ch., Möbius, B., Narasimhan, B. „Contextual effects on consonant voicing profiles: a cross-linguistic study". Proceedings of the 14th ICPhS Conference . San Francisco. 1999.