Specification in Context — Devoicing Processes in Polish, French, American English and German Sonorants

Jagoda Sieczkowska[1], Bernd Möbius[1,2], Grzegorz Dogil[1]

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

jagoda.sieczkowska, bernd.moebius, grzegorz.dogil@ims.uni-stuttgart.de

Abstract

This study investigates voicing properties of Polish, French, American English and German sonorant consonants, particularly rhotics. The analysis was conducted on four speech databases recorded by professional speakers. The term voicing profiles used in this article refers to the frame by frame voicing status of the sonorants, which was obtained by automatic measurements of fundamental frequency values and extraction of consonantal features. Results show resyllabification processes in Polish and French obstruent liquid clusters in word final positions, as well as contextual effects on devoicing in word initial and word medial American English and German obstruent sonorant clusters.

Index Terms: voicing, sonorant, rhotic, speech database, Polish, French, American English, German.

1. Introduction

From the perspective of speech production, voicing is the presence of periodic vocal fold vibration, which is spectrographically visible as low-frequency energy [16].

From the phonological point of view, the feature [voice] belongs to the universal set of features and most commonly corresponds to the phonetic correlates described above. In languages like Polish and French ('true voice' languages, [1]), [voice] serves as a contrastive feature in describing consonants, whereas Germanic languages, such as English and German ('aspirating languages' [1]), contrast in the feature [spread glottis] [2], [10], [15], [8] and many others. Studies concerning voicing patterns of consonants have mostly focused on stop consonants and voice onset time (VOT) measurements ([14], [9], [13] and many others), which when applied cross-linguistically, proved to be very successful.

Sonorant consonants are usually considered to be voiced since their production requires spontaneous voicing [2], [5]. In this study we examine temporal voicing values of the sonorants in Polish, French, American English and German (counted at each ten percent of their time duration), which undergo contextual and phonological process affecting their voicing status (see section 2.1).

Polish sonorants are voiced unless they occur between voiceless consonants or after voiceless obstruent before a pause [5], [6]. However, their devoicing depends not only on the phonotactic environment. As stated by Gussman [5], [6] and described in our previous study [20], Polish sonorant devoicing derives from the process of desyllabification. Gussmann [5] proposed that sonorants are only licensed for [voice] when they belong to a syllable, whereas in word-final position after a voiceless obstruent, a position in which they tend to be fully devoiced, sonorants are not incorporated into the syllable structure, as their would violate the Sonority Sequencing Principle, and thus desyllabified.

According to Dell's [3] description of consonant clusters occurring in different syllable positions, French sonorants in obstruent-sonorant clusters may occupy syllable and word final positions. In his comparison regarding English and French syllabicity, Tranel [24] claims that English liquids /l/ and /r/ are syllabic at the end of the word (forming a word-final consonant cluster, as in the word 'table' [teɪbl̩]), whereas French ones are not, e.g. 'table' [tabl]. Thus, English liquids are said to be able to play both the role of a vowel, constituting the syllable nucleus, and behave as a consonant, which is the part of the syllable margin, while French liquids hold only the consonantal role.

English sonorants are said to occur in word- initial and medial positions, both as singletons and in obstruent clusters [12]. Iverson and Salmons [8] notice devoicing of English sonorants when they are preceded by voiceless obstruent and explain the process as a result of sharing the feature [spread glottis] from the left-hand obstruent to the following sonorant. Similary, Möbius [16] describes devoicing of German sonorants when occurring in voiceless obstruent-sonorant clusters, pointing out an important role of the presence of syllable boundaries within those clusters.

Automatic voicing analysis described in this study, provides an insight into the temporal changes of the voicing values in four investigated languages and the diversity of factors influencing sonorant devoicing processes. The article is constructed as follows. Section 2 describes the method of with which the voicing profiles are obtained, as well as the speech databases used for this investigation. Section 3 demonstrates the results, which are followed by conclusion and discussion in Section 4.

2. Method

2.1 Automatic voicing extraction

The analysis of voicing in sonorant consonants follows the method described in [19]. It comprises extraction of voicing probabilities along with the contextual information about the investigated segment, i.e. its phonetic properties, phonotactic environment and utterance structure. As a first step, the ESPS get_F0 tool was used to obtain voicing information by analyzing the speech signal in frame-by-frame mode at a sampling rate of 100 frames per second, using an

26 – 30 September 2010, Makuhari, Chiba, Japan

autocorrelation technique. The voicing decision was always categorical (0 or 1). Using IMS German Festival speech synthesis tool [7] and phonetic annotation from the four speech databases [4],[17],[20],[21] we extracted the following features for each: type of consonant, its place of articulation, position within the word and syllable, as well as its voicing status at 10%, 20% … 90% of its duration. Results were averaged and plotted using R [18].

2.2 Databases

The Polish Speech Corpus [4] was designed for unit selection speech synthesis purposes and is the Text-to-Speech system, which bases on the Bonn Open Synthesis System (BOSS) [11], (originally aimed at German and Dutch speech). The database comprises 115min (3249 utterances) of speech recorded by a professional male speaker during several recording sessions supervised by an expert phonetician.

The French speech corpus is based on the data obtained thanks to the courtesy of the SVOX AG. Information included in the corpus files contained data with the duration of the phones, their syllabic, word and phrase position, as well as the binary voicing decision corresponding to each period of the duration of the phone. The data was based on the professional female speaker speech corpus designed for embedded automotive speech applications including speech recognition and text-to-speech solutions [21].

The Boston University Radio News Corpus [17] was the speech database used for the investigation of American English sonorants. This database contained recordings of FM radio news announcers associated with WBUR, a public radio station. The corpus was originally designed for the text-to-speech synthesis. It consists of two parts: radio news and lab news. The first part of the corpus embodies stories recorded during broadcast by seven professional speakers. 3 female and 4 male speakers were divided into two groups: A being speakers whose job is to read news live and B who normally pre-record and edit their stories. The second part of the corpus (lab news database) combines laboratory recordings of the same speakers reading stories previously read only by the B-type speakers. We decided to use the data from the F1A speaker, whose lab news materials are fully documented. The F1A radio news recordings are labeled only automatically (opposite to the laboratory news part, which was also manually corrected) and their analysis served as a comparison, which is not included in this article.

The MS corpus was designed for the purpose of the unit selection speech synthesis project, SmartKom [22] and served for the investigation conducted on German sonorants. It was designed to provide coverage not only for domain specific, but also for open domain output, where the entire language was the target. As a consequence, sentences for the corpus were selected from a large database in order to contain a maximum number of combinations of speech sounds and contexts in which they occur. The database thus comprises a full set of diphones, as well as rich combinations of phones and contexts in which they occur, including segmental context. Moreover, information about syllabic stress and syllable structure, positional and intonational factors is included. The database contains 17489 words, 56434 consonants embedded in 2301 sentences which last 160 minutes (for the sonorants' frequency of occurrence in the MS corpus see Table 1 below). The MS corpus was segmented automatically on the phone, syllable and word levels using HMM-based forced alignment [19] and was later manually corrected.

Table 1 shows the coverage of phones in each of the 4 speech databases.

Table 1. *Inventory and frequency of occurrence of Polish sonorants in the BOSS corpus, German sonorants in the MS corpus, French sonorants in the SVPOX corpus and American English sonorants in the Boston University radio and laboratory news corpus.*

Sonorats	Polish	German	French	Boston radio news	Boston lab news
[m]	4216	2708	2616	1658	179
[n]	7065	8285	2597	4145	447
[ŋ]	2540	669	2590	33	59
[l]	2600	3329	1893	2435	302
[r]	3394	2472	6982	3733	1559
[j]	4605	472	1890	663	71
[w]	3532		684	660	103

3. Experimental results

In the figures presented in this section, normalized time (percentage of the duration of the segment) is plotted on the x axis, whereas the voicing probability computed as the percentage of the phones undergoing de-/voicing (the number of exemplars in the corpus) is plotted on the y axis. Thus, the results show all sonorant exemplars in the corpora and their voicing probabilities over time.

Figure 1 illustrates the voicing probabilities of Polish sonorants in word initial position in a left voiceless obstruent context. Voicing values of all the sonorants apart from the liquid [r] decrease slightly, varying between 93% and 99%. Having observed similar minor devoicing tendencies in other sonorants in word final positions, we decided to focus only on the rhotics. Figure 2 displays voicing probabilities of the sonorant [r] in word final position as in 'wiatr' [vjatr̥] *wind*. As discussed in [20] the liquid [r] in this position is fully devoiced. Figure 3 shows the voicing probabilities of French sonorants in all word positions, whereas Figure 4 shows them in word final position, in both cases in a left voiceless obstruent context. Similar to the Polish results, it is the liquid [R] which demonstrates the biggest diversity, being also fully devoiced in word final position. Results (Fig. 5) of the investigation on German sonorants in a left voiceless obstruent context present the lowest voicing probabilities for the [R] liquid too, the extend of which depends on the presence vs. absence of the syllable boundary intervening between the sonorant and the obstruent (a phonological configuration which does not apply to word final obstruent-sonorant clusters in Polish or French). Figure 6 shows sonorants of American English in a left voiceless obstruent context in word initial and medial positions. Initial devoicing applies to the [w] and [m] segments and does not fall below 83% of exemplars.

4. Discussion and conclusions

This study presents devoicing processes of the sonorants in four languages and focuses on the analysis of rhotics, as their voicing probabilities vary to the largest extend. Polish trilled

[r] produced by brief, multiple touching and releasing of the tongue tip against the alveolum and French uvular trill [R] articulated with the trilling of the uvula, seems to demonstrate similar devoicing tendencies. Their complete devoicing in words like 'wiatr' [vjatr̥] *wind* or 'quatre' [katʀ̥] *four*, suggests similar phonological patterning, where, as suggested by [5] and [15] a sonorant which does not belong to the syllable structure, is unlicensed for [voice], i.e. desyllabified undergoing the process of Final Extrametricality.

On the other hand, German uvular [R] as well as English nasal [m] and approximant [w], tend to devoice in their initial phase of duration (an effect observed in much larger extent when the sonorant follows voiceless obstruent and not a vocalic segment). As stated by [8] devoicing of sonorants in obstruent-sonorant clusters in Germanic languages may be an effect of contextual influences, due to which the feature [spread glottis] is shared within the cluster, from the voiceless segment to the sonorant. This result seems to be supported additionally by the fact that German sonorants undergo devoicing to a larger extent when not separated by a syllable boundary from the preceding obstruent, as in the word 'knapp' [knap] *tight* (in contradiction to the ones which are separated by a syllable boundary, like in 'Stecknadel' [ʃtɛk.na:dəl] *pin*). A similar tendency was observed for the initial devoicing of American English [w] and [m], which was 10% stronger for clusters separated by a syllable boundary.

Our corpus-based study of devoicing processes demonstrates two main tendencies of the voicing behavior. Due to cross-linguistic differences and phonotactic constraints, it is the feature [voice] that licenses voicing of Polish and French sonorants and the feature [spread glottis], which influences devoicing in American English and German. Thus, devoicing processes in the four investigated languages are specified by the phonological and phonotactic contexts.

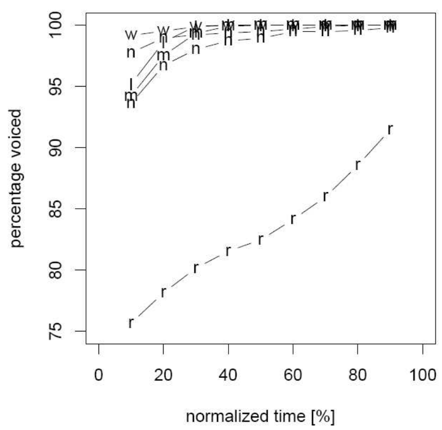

Figure 1: *Voicing profile of Polish word initial sonorants with left voiceless obstruent context.*

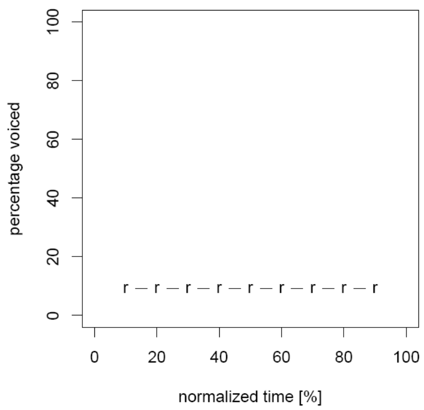

Figure 2: *Voicing profile of Polish word final [r] with left voiceless obstruent context [19].*

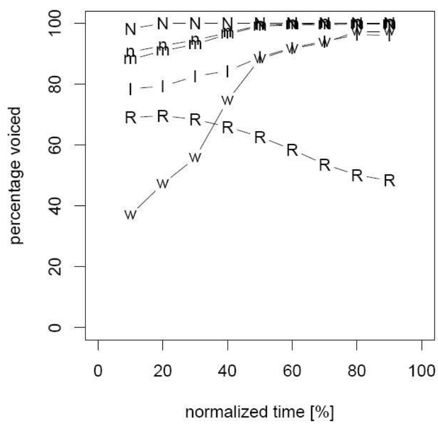

Figure 3: *Voicing profile of all French sonorants in all word positions with left general context.*

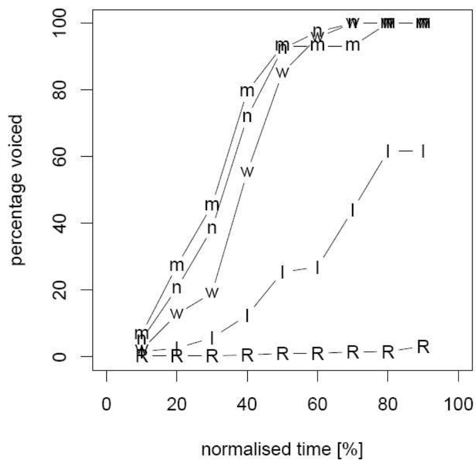

Figure 4: *Voicing profile of all French sonorants in word final position with left voiceless obstruent context.*

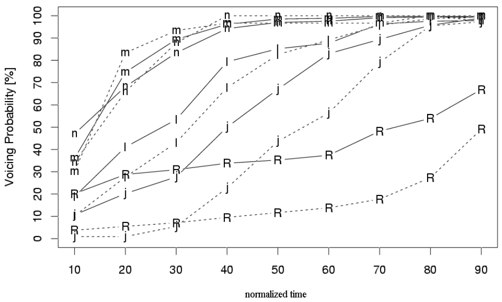

Figure 5: *Voicing profile of German sonorants in word initial and medial positions with left voiceless obstruent context within one syllable (dashed lines) and separated by a syllable boundary (solid lines) [16].*

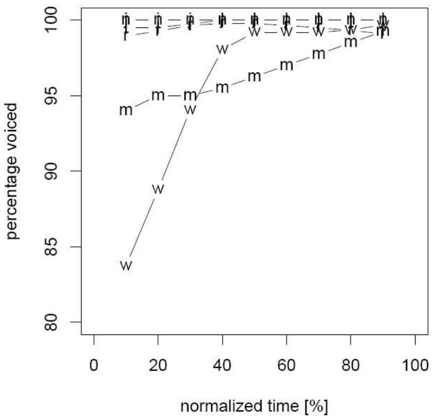

Figure 6: *Voicing profile of American English sonorants (lab news corpus) in the word initial and medial positions with left voiceless obstruent context.*

5. Acknowledgements

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

6. References

[1] Beckmann, J., Jessen, M., Ringen, C. "German Intervocalic Stops Differ from Stops in True Voice Languages: Empirical Evidence". Talk presented at the Berkley Linguistics Roundtable, University of California, Berkley. 2010.

[2] Chomsky, N. Halle M. "The sound pattern of English. New York". Harper and Row. 1968.

[3] Dell, F. "Consonant clusters and phonological syllables in French". Lingua 95. Pp. 5-26. 1995.

[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". Proceedings of Interspeech, 2008.

[5] Gussmann, E." The Phonology of Polish". Oxford University Press. 2007.

[6] Gussmann, E. "Resyllabification and Delinking: The Case of Polish Voicing." Linguistic Inquiry 23: 29–56. 1992.

[7] IMS German Festival Home Page. http://www.cstr.ed.ac.uk/projects/festival.

[8] Iverson, K., Salmons J.C. "Aspiration and Larnygeal Representation in Germanic". Phonology, Vo.12, No.3, pp. 369-396. 1995.

[9] Keating, P. "Phonetic and Phonological Representation of Stop Consonant Voicing." Language 60/2, 286–319.

[10] Kingston, J. Diehl R. L." Phonetic Knowledge." Language. 70/3. 1994.

[11] Klabbers, E. Stöber K. "Creation of Speech Corpora for the Multilingual Bonn Open Synthesis System". In: ISCA ITRW. 2001.

[12] Kreidler, Ch W. "The pronounciation of English. A course book". Blackwell Publishing. 2004.

[13] Laddefoged, P., Maddieson, I. "The Sounds of the World's Languages". Blackwell Publishing. 1996.

[14] Lisker, L., Abramson, A. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements." Word 20, 384-422. 1964.

[15] Lombardi, L. "Laryngeal Neutralization and Syllable Wellformedness". Natural Language and Linguistic Theory Vol. 13, No. 1, pp. 39-74. 1995.

[16] Möbius, B. "Corpus-based investigations on the phonetics of consonant voicing". Folia Linguistica 38 (1-2), 5-26. 2004.

[17] Ostendorf, M., Price, P.J., Shattuck-Hufnagel, S. "The Boston University Radio News Corpus ". 1995.

[18] R project. The R project for Statistical Computing. (http://www.R-project.org/].2001

[19] Rapp, S. "Automatic Phonemic Transcription and Linguistic Annotation from Known Text with Hidden Markov Models: An Aligner for German." Proceedings of ELSNET Goes East and IMACS Workshop 'Integration of Language and Speech in Academia and Industry' (Moscow, Russia). 1995.

[20] Sieczkowska, J., Möbius, B., Schweitzer, A., Walsh, M., Dogil, G. "Voicing profile of Polish sonorants: [r] in obstruent clusters". Proceedings of Interspeech, 2009 (Brighton), 2319-2322.

[21] SVOX AG. SVOX French Corpus. 2009.

[22] SmartKom. "Das Leitprojekt SmartKom: Dialogische Mensch-Technik-Interaktion durch koordinierte Analyse und Generierung multipler Modalitäten". 2003.

[23] Tranel, B. "The sounds of French." Cambridge university press. 1987