PREDICTABILITY OF PLOSIVE REDUCTION FROM WRITTEN TEXT IN ESTONIAN

Liis Ermus[1,2], Meelis Mihkla[1]

[1]Institute of the Estonian Language, Estonia, [2]University of Tartu, Estonia
liis.ermus@eki.ee, meelis.mihkla@eki.ee

ABSTRACT

The article presents the results of the pilot study of research on coarticulation of short plosives in read Estonian. In it the authors try to elucidate correspondences between reduction patterns in speech that influence closure and burst of short plosives and features in written text. Analysis revealed extensive voicing of all plosives that was affected by segmental context and position. Burst phase was usually retained; reduction was more affected by positional parameters. Possible effect on speaking rate also occurred.

Keywords: plosives, coarticulation, allophonic variation, Estonian

1. INTRODUCTION

1.1. Coarticulation of plosives

Reduction due to coarticulation seems to be a universal tendency. Coarticulation is influenced by many factors such as speech situation, speech tempo, word frequency, prosodic structure of the utterance, phonetic environment, position of the sound in the word etc. [10, 17]. Reduction is commonly associated with everyday speech but also occurs in journalistic speech [3, 11].

Plosives differ from other consonants because their pronunciation cannot be viewed as static. Plosives consist of three phases: implosion – forming the closure, occlusion – holding the closure, and explosion or burst – opening the closure. Burst is characterised by voice onset time (VOT) – duration between the start of the explosion phase and the beginning of the voiced vibration or the re-emergence of harmonic vibration [5].

Voicing of plosives is often studied with other obstruents, which show similar patterns [6]. Voiceless plosives tend to get voiced in voiced environment and *vice versa*. More variation occurs in word-medial position and in languages that do not have voicing contrast [13]. In the case of partially voiced obstruents, voicing occurs mostly at the beginning of a sound, when voicing is carried on from the preceding sound, or at the end of a sound when voicing of the next sound starts during the closure [6].

Plosives can lose their characteristic burst phase. The loss of burst is usually observed in spontaneous speech and is mostly found in voiced plosives [7] in the word-medial position but also occurs in voiceless plosives, often accompanying voicing [9, 21].

1.2. Plosives in Estonian

Estonian has four plosive phonemes: bilabial /p/, alveolar /t/, palatalised alveolar /tʲ/ and velar /k/. Phonologically, Estonian plosives are voiceless and unaspirated; aspiration can occur utterance-finally. Word-medial and word-final plosives can occur in three quantity categories: short, long and overlong (spelt with letters *b*, *d*, *g*; *p*, *t*, *k*; and *pp*, *tt*, *kk* respectively) [2]. Word-initial plosives are similar in duration to short word-medial plosives [8] and in connected speech act similarly in voicing patterns [1, 9].

Pronunciation of plosives in connected speech in Estonian is very varied. In the Phonetic Corpus of Estonian Spontaneous Speech 44% of plosives are marked as voiced [20]. Increase of voicing in word-medial position has been noted already in [1]. Loss of burst has been found to occur in between 5.9% of tokens in read speech [19] and about 40% in spontaneous speech [9]. Velar /k/ stands out, with loss of burst in 25.9% tokens in read speech [19] and 58% in spontaneous speech [9]. Durations of both the word-initial and word-medial plosives tend to be significantly longer in contrastively accented words in read speech [19], but not in spontaneous speech [9].

1.2. Aims of this study

This research analyses the contextual dependence of allophones of Estonian plosives (voiceless vs. voiced, with burst vs. burstless) and estimates the strength of prediction of models for predicting allophonic variants based on textual features. The final purpose is to add allophonic variants of plosives to statistical-parametric models of Estonian text-to-speech synthesis (TTS) and therefore to TTS applications [16]. The development of TTS depends on consideration of allophonic variation, and on how the orthography of a language relates to its phonology [12, 18]. Taking the coarticulation of plosives into consideration and predicting their allophonic variation helps to enhance the variation and therefore naturalness of synthetic speech.

We formulated the following hypothesis based on our aims and previous research:

H: It is possible to predict plosive reduction in speech using features of written text. Research questions:

Q1: How do contextual and phrasal features influence the pronunciation of plosives in continuous speech?

Q2: What features have most influence on reduction of the pronunciation of plosives?

2. MATERIAL

The recorded material consisted of four short news stories read alternately by male and female news announcers, lasting 6 minutes and 13 seconds in total. News texts were chosen because the final purpose is to use the results to enhance synthetic speech. As TTS applications are often used for reading news texts it seems reasonable to use similar texts in analysis and modelling.

News texts contain large amount of dates, numbers, foreign words and names and may contain long utterances. The median length of utterance in our sample was 8 words (range: 1-18).

For analysis we chose short plosives in all word positions adjacent to sonorants on both sides or to a sonorant and a pause. A total of 512 tokens were analysed, 262 from the male speaker and 250 from the female speaker. Average speaking rate was 5.1 syllables per second for the male speaker and 4.4 syl/sec for the female speaker.

The voiced and voiceless parts and burst phases of plosive tokens were annotated manually using Praat [4].

Features that can be obtained from written text were marked. They included position of the plosive in the word, position of the word in utterance (utterance corresponds to written sentence or sub-sentence), segmental context, parts of speech etc. Speaker was also included. Inclusion of parts of speech in the model was influenced by a study of the lexical prosody of Estonian [15] that showed that sounds were, on average, 10% longer in proper names than in verbs and 10% longer in verbs than in ordinals. This kind of change in speech tempo may also influence coarticulation.

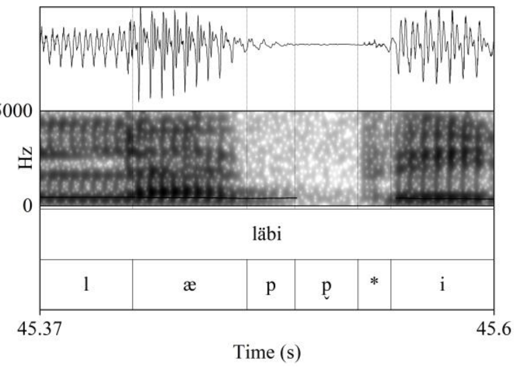
A summary of the features used is shown in Table 1.

3. RESULTS

3.1. Allophonic distribution

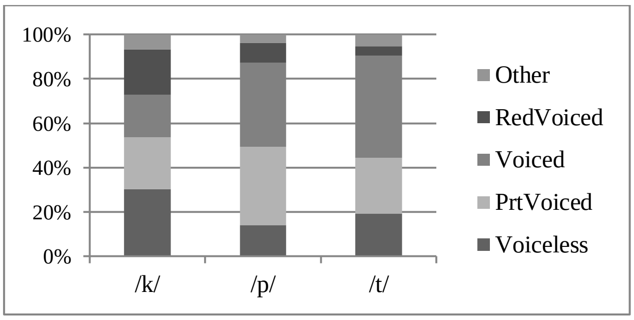
Four allophonic structures occurred frequently: fully voiceless with burst, partly voiced with burst (PrtVoiced), fully voiced with burst, fully voiced burstless (RedVoiced). Total loss and fricativisation also occurred in small frequencies. An example of a partly voiced /p/ is given in Figure 1.

Figure 1: Partly voiced allophone of /p/ in word *läbi* ‘through’. Asterisk * notes burst.



The distribution of allophones is shown in Figure 2.

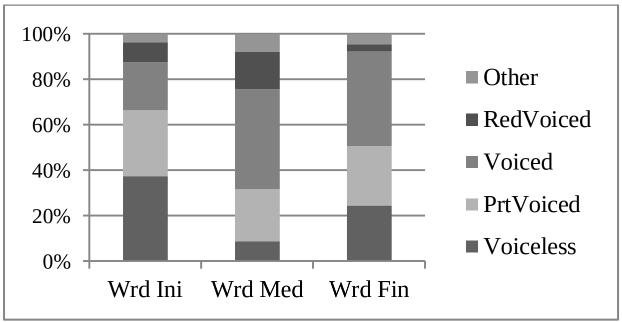
Figure 2: Distribution of allophones by phoneme.



The proportion of voiceless tokens was greatest for /k/ with about a third of all tokens. At the same time /k/ also showed the greatest proportion, a fifth, of voiced burstless (RedVoiced) tokens. Majority of tokens for /p/ and /t/ were at least partly voiced but retained burst phase which was reduced in less than 10% of the tokens.

The distribution of allophones according to position in the word is shown in Figure 3. Almost 40% of tokens in word-initial position were voiceless. Partly voiced tokens occurred almost equally in all positions. Voiced tokens were in majority in word-medial and word-final positions. Word-medial position had the highest proportion of burstless tokens. Total loss occurred only in word-medial position.

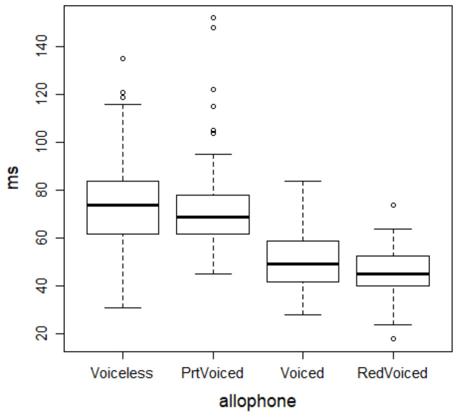
Figure 3: Distribution of allophones by position in the word.



3.2. Durations

Duration depended on plosive, the position of the plosive in the word and the realised allophone structure. Welch's one-way ANOVA tests were conducted to compare the duration differences.
The mean duration of /p/ was 66 ms; the mean duration of both /k/ and /t/ was 59 ms. /p/ lasted significantly longer than other plosives [F(1, 510)=13, p<0.01].

Figure 4: Durations of allophones



As Figure 4 shows, voiceless allophones had the longest duration (mean 75 ms) and burstless voiced allophones the shortest (46 ms). Two duration subgroups were identified: voiceless and partly voiced (mean durations over 70 ms) and both fully voiced allophones (mean durations 50 ms and 46 ms). The duration difference between the groups was statistically significant [F(1, 510)=307.65, p<0.001].

Word-initial and word-final tokens had similar mean durations, 68 ms and 69 ms respectively. The mean duration of word-medial tokens was much shorter, 51 ms, difference was statistically significant [F(1, 510)=109.5, p<0.01].

3.3. Prediction models

Binary logistic regression models were fit for every phoneme for voicing and for burst phase. Features obtainable from written text were used as predictors; they are listed in Table 1. The reference level was post-pausal (i.e. phrase and word initial) voiceless allophone with burst. For other predictors, most frequent values were chosen for reference. Optimal models were chosen by stepwise analysis. Classification ability of the models was evaluated using linear discriminant analysis run on the same data set.

Table 1: Features used in prediction models. Reference levels in models are in boldface.

Segmental features	Positional features	Other features
allophone (**voiceless**, voiced; **with burst**, burstless)	position of letter in word (**initial**, medial, final)	part of speech (reference **Substantive**)
preceding context (**pause**, vowel, sonorant consonant)	number of the letter-carrying syllable in the word from the **first**	foreign word (yes/**no**)
following context (pause, **vowel**, sonorant consonant)	position of the word in utterance (**initial**, medial, final)	compound word (yes/**no**)
		speaker (**F**, M)

3.3.1. Voicing

We were looking for parameters that influence the reduction of the closure phase, making it partially or fully voiced. Final models are presented in Table 2. No model could be fit for /p/ because no significant predictors emerged.

Table 2: Prediction models for voicing

Predictor	Estim.	Odds Ratio	p-value	Class. Ab.
M1: voicing of /t/				78%
Intercept	-0.9		0.133	
SonorantAfter	0.4	1.5	0.003	
PauseAfter	-2.3	0.1	<0.001	
Syllable	0.8	2.2	<0.001	
UtteranceFinal	-0.2	0.8	0.750	
UtteranceMedial	1.7	5.5	0.003	
M2: voicing of /k/				76%
Intercept	-3.5		<0.001	
SonorantBefore	3.3	26.7	0.003	
VowelBefore	3.6	35.2	0.002	
WordMedial	1.8	6.1	0.004	
WordFinal	16.2	>100	0.989	
Speaker M	0.8	2.3	0.040	

The models predicted greater likelihood of voicelessness for the reference level. The parameters with most influence on voicing were contextual and positional. Voicing was more frequent when the adjacent context was non-pausal. /k/ was more affected by preceding and /t/ by following context. /t/ was more likely to be voiced utterance-medially ($p=0.003$) and likelihood rose in non-initial syllables ($p<0.001$). Following pause significantly decreased the chance of voicing of /t/ ($p<0.001$). The probability of voicing in /k/ increased following voiced segments and word-medially ($p=0.004$). Speaker rose as significant predictor for /k/ ($p=0.04$)

The classification ability of the models as assessed by discriminant analysis was at least 76%, which is above average.

3.3.2. Burst phase

We were looking for parameters that influence reduction of the burst phase. Final models are presented in Table 3. Again no model could be fit for /p/ as no significant predictors emerged.

Table 3: Prediction models for burst phase reduction

Predictor	Estim.	Odds Ratio	p-value	Class. Ab.
M3: burst of /t/				64%
Intercept	-3.2		<0.001	
WordFinal	1.6	4.8	0.080	
WordMedial	2.2	9.4	0.005	
Syllable	-0.7	0.5	0.037	
Speaker M	1.3	3.8	0.014	
M4: burst of /k/				73%
Intercept	-3.6		<0.001	
WordFinal	-15.2	>100	0.990	
WordMedial	1.7	5.6	<0.001	
Comp.Word	1.2	3.3	0.003	
Speaker M	2.0	7.4	<0.001	

Position of the plosive in the word had the biggest influence on burst phase. Loss of burst phase was more likely to occur in word-medial plosives ($p<0.01$ for both models). Occurrence of the plosive in non-initial syllables, on the other hand, decreased the possibility of burst reduction in /t/.

Speaker was a significant parameter, the male speaker was associated with more extensive reduction in bursts of both /t/ ($p=0.014$) and /k/ ($p<0.001$). Possible effect of word length emerged in M4: burst of /k/ was more likely to be reduced when appearing in compound words ($p=0.003$).

M4 had above average classification ability, but that of M3 was rather poor, at only 64%.

4. DISCUSSION AND CONCLUSIONS

Overall the results on allophonic distribution and behaviour of phonemes confirm our hypothesis. Analysis showed that even in a small sample there was variation in categories such as voiceless-voiced, burst-burstless. There was a large difference between the proportions of voiceless and voiced tokens, 23% and 77% respectively. The relative distribution of burst and burstless was 83% vs. 17%. These results suggest that it may be reasonable to add allophonic variables at least for voicing feature into speech synthesis phoneme list.

Prediction models contained rather similar features for both /t/ and /k/. For example the allophone is most likely to be reduced in the medial position, be it word or utterance medial. Voiced segmental context clearly had effect on voicing. Apart from that, /t/ was more influenced by utterance and /k/ by word level in voicing models. Bigger influence of utterance level on voicing of /t/ can be explained by context sensitivity. /t/ becomes voiced easily due to short durations and voiced tokens occur in all positions in the word. Voicelessness is preserved better adjacent to pauses i.e. on the utterance borders. Pronunciation of /k/ varies more depending on the position in the word so that has clearer influence. For burst phase word level effects were significant for both /t/ and /k/. Speaking rate may also play role in reduction amount, as speaker emerged as significant predictor in several models.

Classification ability of most models was above average. Poorer performance of M3 can be affected by very small amount of burstless tokens in /t/. It may be that no strong patterns emerged because of that.

Lastly, it should be remembered that although no statistically significant features appeared affecting reduction of /p/, it still showed regular voicing of closure phase in connected speech. Therefore it should not be excluded from future research and modelling.

The results should be treated with caution as the prediction models are based on data from only two speakers and some features may be idiolective. In future it might be worth examining donor voices individually and using individual characteristics for modelling. Models should also be tested on larger samples of spoken material, such as audiobooks, for validation.

5. ACKOWLEDGEMENTS

This research has been supported by the Centre of Excellence in Estonian Studies (CEES, European Regional Development Fund) and is related to research project IUT35-1 (Estonian Research Council).

REFERENCES

[1] Ariste, P. 1933. Eesti sulghäälikud k, p, t ja b, d, g. *Eesti Keel*. Nr. 3, 4, 73–82, 170–180.

[2] Asu, E. L. and Teras, P. 2009. Estonian. *Journal of the International Phonetic Association* 39(3), 367–372. DOI:https://doi.org/10.1017/S002510030999017X.

[3] Blom, J. N., Ejstrup, M., Hopmann, D. N. 2018. The effects of phonetic reduction on actual and perceived comprehension by news audiences. *Journalism Studies* 19(5), 745–763. DOI:https://doi.org/10.1080/1461670X.2016.1215256.

[4] Boersma, P., Weenink, D. 2018. *Praat: doing phonetics by computer* [Computer program].

[5] Cho, T., Ladefoged, P. 1999. Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 27(2), 207–229. DOI:https://doi.org/http://dx.doi.org/10.1006/jpho.1999.0094.

[6] Davidson, L. 2016. Variability in the implementation of voicing in American English obstruents. *Journal of Phonetics* 54, 35–50. DOI:https://doi.org/https://doi.org/10.1016/j.wocn.2015.09.003.

[7] Duez, D. 1995. On spontaneous French speech: Aspects of the reduction and contextual assimilation of voiced plosives. *Journal of Phonetics* 23, 407–427.

[8] Eek, A., Meister, E. 1996. Eesti sõnaalguliste sulghäälikute akustika ja tajumine. *Keel ja Kirjandus* 3–5, 164–170, 241–253, 314–321.

[9] Ermus, L. 2017. Eesti keele lühikeste klusiilide hääldusе variatsioon ja seda mõjutavad tegurid. *Mäetagused* 68, 27–52. DOI:https://doi.org/doi.org/10.7592/MT2017.68.ermus.

[10] Ernestus, M., Hanique, I., Verboom, E. 2015. The effect of speech situation on the occurrence of reduced word pronunciation variants. *Journal of Phonetics* 48, 60–75. DOI:https://doi.org/https://doi.org/10.1016/j.wocn.2014.08.001.

[11] Hallé, P. A., Adda-Decker, M. 2007. Voicing assimilation in journalistic speech. *Proc. 16th ICPhS* Saarbrücken, 493–496.

[12] de Jesus Aguiar Pontes, J., Furui, S. 2010. Predicting the phonetic realizations of word-final consonants in context – A challenge for French grapheme-to-phoneme converters. *Speech Communication* 52(10), 847–862. DOI:https://doi.org/10.1016/j.specom.2010.06.007.

[13] Keating, P. A., Linker, W., Huffmann, M. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11, 277–290.

[14] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W. J., Marchal, A. (eds), *Speech Production and Speech Modelling*. Netherlands: Springer, 403–439.

[15] Mihkla, M. 2007. Morphological and syntactic factors in predicting segmental durations for Estonian text-to-speech synthesis. *Proc. 16th ICPhS* Saarbrücken, 2209–2212.

[16] Mihkla, M., Hein, I., Kiissel, I. 2018. Self-reading texts and books. *Human Language Technologies – The Baltic Perspective.* 79–87. IOS Press. DOI:https://doi.org/10.3233/978-1-61499-912-6-79.

[17] Schuppler, B., van Dommelen, W. A., Koreman, J., Ernestus, M. 2012. How linguistic and probabilistic properties of a word affect the realization of its final /t/: Studies at the phonemic and sub-phonemic level. *Journal of Phonetics* 40, 4, 595–607. DOI:https://doi.org/10.1016/j.wocn.2012.05.004.

[18] Somers, H. 2005. Faking it: Synthetic Text-to-speech Synthesis for Under-resourced Languages – Experimental Design. *Proc. Australasian Language Technology Workshop 2005* Sydney, 71–77.

[19] Suomi, K., Meister, E. 2012. A preliminary comparison of Estonian and Finnish plosives. *Linguistica Uralica* 3, 187-193. DOI:https://doi.org/10.3176/lu.2012.3.04.

[20] Teras, P. 2018. The phonetic variation of short intervocalic /h/ in Estonian. *Proc. 9th International Conference on Speech Prosody* Poznań. 883–887.

[21] Torreira, F., Ernestus, M. 2011. Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology* 2, 331–353. DOI:https://doi.org/10.1515/labphon.2011.012.