Frequency Effects on Vowel Reduction in Three Typologically Different Languages (Dutch, Finnish, Russian)

R.J.J.H. Van Son[1], *Olga Bolotova*[2], *Mietta Lennes*[3], *Louis C.W. Pols*[1]

[1]Institute of Phonetic Sciences/ACLC, University of Amsterdam, The Netherlands
[2]Department of Phonetics, State University of St Petersburg, Russian Federation
[3]Department of Speech Sciences, University of Helsinki, Finland
R.J.J.H.vanSon@uva.nl

Abstract

As a result of the cooperation in the Intas 915 project, annotated speech corpora have become available in three different languages for both read and spontaneous speech of some 4-5 male and 4-5 female speakers per language (6-10 minutes per speaker). These data have been used to study the effects of redundancy on acoustic vowel reduction, in terms of vowel duration, F_1-F_2 distance to a virtual target of reduction, spectral center of gravity, and vowel intensity. It was shown that in all three (typologically different) languages vowel redundancy increases acoustic reduction in the same way. The reduction of redundant vowels seems to be a language universal.

1. Introduction

Speech is commonly seen as an efficient communication channel: less speaking effort is spent on redundant than on informative items. A number of studies have shown that listeners identify redundant tokens better and that speakers take advantage of this by reducing any predictable items [1,5,6,9,11,13,15,17,20-23] (and references therein). For example, *nine* is pronounced more reduced in the proverb *A stitch in time saves nine* than in the neutral expression *The next number is nine* [13].

Although generally considered a universal feature of speech, the effects of redundancy have been studied for only a few languages, mainly English and Dutch. These languages are typologically related. Both are stress-timed languages and exhibit lexical stress with only limited morphological complexity. For instance, both have their syllable stress predominantly on the most “complex” (first) syllable [7]. It is not clear whether redundancy has similar consequences in other languages, although there is preliminary support for these effects to exist, e.g., in Finnish [11]. Therefore, in this study we will investigate the extent to which redundant vowels are reduced in three typologically unrelated languages: Dutch, Finnish, and Russian.

It has been found that phonemes in redundant words are more reduced in the languages investigated [1,5,6,9,11,13,15,17,20-23]. However, not all phonemes in a word are equally important for speech understanding. Therefore, if speech is really efficient, more “important” phonemes should be less reduced than more redundant phonemes. To measure this effect, a model of word recognition is needed to quantify the importance of a segment in speech communication.

To quantify the efficiency at the articulatory level, the effort invested in the “unit of articulation” must be matched against the importance of this unit. In this paper we take the vowel segment as the unit of articulation.

Theories of word recognition stress that word recognition is an incremental task that works on a phoneme by phoneme basis [14]. Often, words are recognized by their first syllable well before all phonemes have been processed [8]. In English and Dutch, the importance of the first syllable is reflected in the fact that lexical stress is predominantly on the first syllable of a word [7,8]. Finnish does not exhibit lexical word stress, but for any clearly pronounced or accented word, the primary word stress is regularly perceived on the first syllable. In Russian lexical stress can be on any syllable in a word, and is predominantly in the middle of a word. Grammatical information (e.g. noun case or verb form) is shown in a word ending. However, in connected speech word endings are often strongly reduced, because of the redundancy of the grammatical information.

The importance (or redundancy) of an individual vowel realization is measured in terms of its predictability, i.e., information content. This combines the global frequency of the word containing the vowel segment and the vowel's predictability after the preceding word-onset. The latter factor is the realization’s (incremental) contribution to word disambiguation in recognition. Therefore, next to the global word frequency, we use a model of word recognition with competition based on a frequency-sensitive incremental match of incoming phonemes in a lexicon [14,21,22,23]. To determine the vowel redundancy for Dutch, we had access to a word frequency list covering 39 million tokens (CELEX). For Russian a word form frequency list based on classical and modern texts corpora [12] was used. Since no sufficiently large spoken language corpora were available for Finnish, we calculated the corpus frequencies from transcripts of the speech material recorded for the Intas 915 project.

The Finnish language uses 8 different vowel qualities /A e i o u y @ ø/ [11]. All of these may occur phonologically as either long or short, or they may combine into diphthongs. Dutch uses 12 monophthongs, /i I e: E a: A O o: u y Y ø:/, the schwa, and three diphthongs /Ei @y Ou/ [17,22,23]. Russian uses six monophthongs, /i e a o u ɨ/. Vowels /o/ and /e/ appear only in stressed position, and /i a u ɨ/ undergo qualitative and quantitative reduction depending on the position with regard to lexical stress [4].

2. Materials and Methods

2.1. Speech materials

Finnish and Russian speech was collected as part of the Intas 915 project [3]. The Dutch vowel realizations were selected from the IFAcorpus [10,19]. Spontaneous speech was recorded from laboratory dialogs (Russian and Finnish) or laboratory story telling and retelling (Dutch). Read speech was

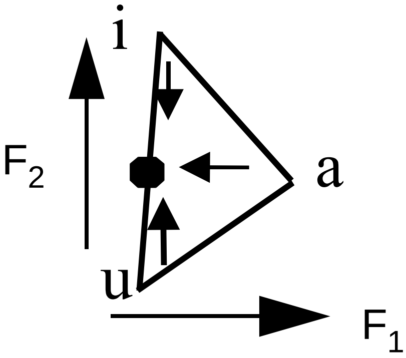

Figure 1. Definition of F_1/F_2 distance in vowel formant space with respect to a theoretical origin of reduction. The actual distances are calculated in semitones.

recorded from reading aloud transcripts of the speech [3,4,10,11,19]. As the Dutch recordings are not balanced over speakers, a combination of reading styles was selected to obtain a balanced sample of read speech, predominantly full text reading and isolated sentences. All speech was recorded directly to audio CD (Dutch) or DAT tape (Finnish and Russian). Three to five minutes per speaker and style were processed for these analyses. Speech recordings were manually segmented and labeled. Both Dutch and Finnish speech was first segmented automatically and then corrected by hand.

In principle, every word has exactly one syllable that may exhibit primary word stress. In languages with lexical stress this is often the most informative syllable [24]. Thus, the stressed syllable (in Finnish, the initial syllable) is well suited for making comparisons among different words and tokens. Also, there are more phonotactic restrictions on the structure of unstressed syllables in Finnish and Russian. Therefore, this study was done on stressed vowels only. These would be lexically stressed vowels for Dutch and Russian, and vowels in word-initial syllables for Finnish. Table 1 gives the number of vowel segments used in this study.

2.2. The importance of a segment

We use a measure of the position-dependent vowel contribution in distinguishing words given the preceding word-onset [20-23]. The lexical information I_L (in bits) of a segment s preceded by word onset [*onset*] is:

$$I_L = -\log_2(P(s|[onset]))$$
$$= -\log_2(P([onset]+s)/P([onset])) \qquad (1)$$

Table 1: Number of vowel monophthong segments from stressed syllables used for all languages and styles

Language	F/M	#	Vowels: #Read	#Spontaneous
Dutch	F	4	3490	2963
	M	4	3280	2604
Finnish	F	4	2173	1622
	M	4	594	1199
Russian	F	5	2389	2277
	M	5	2597	2384
Total		26	14523	13049

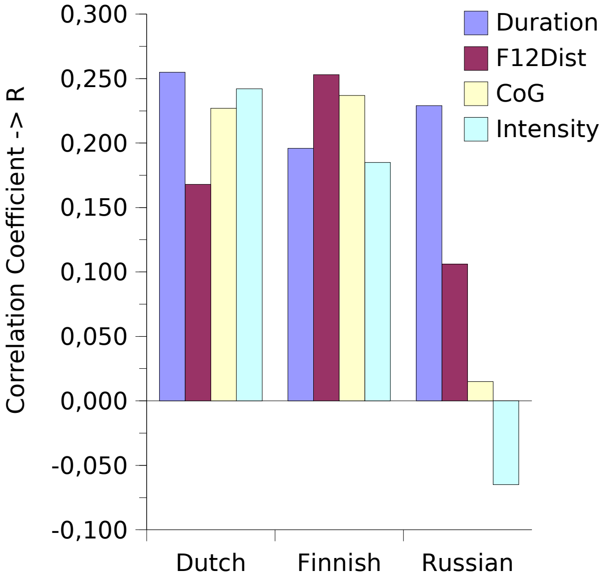

Figure 2. The correlation coefficient between global $-\log_2$(word frequency) and acoustic reduction for read speech. All correlation coefficients are significant ($p<=0.001$) except CoG for Russian. Note that for Intensity in Russian, $R=-0.065$, i.e., negative, and statistically significant ($p<=0.001$).

Using a word frequency list with phoneme transcriptions, equation (1) is operationalized as:

$$I_L = -\log_2\{\mathrm{Freq}([onset]+s)/\mathrm{Freq}([onset]+x)\} \qquad (2)$$

where x is any segment and *Freq* the frequency of the corresponding matching words in the word list. Note that the global frequency of the parent word occurs on both sides of the division mark. This means that I_L is largely independent of the global word frequency but is slightly biased towards a negative correlation. We found that word frequency versus I_L indeed correlated negatively for Dutch and Russian ($R= -0.015$ and -0.006 respectively, not significant). However, we did find a rather large positive correlation for

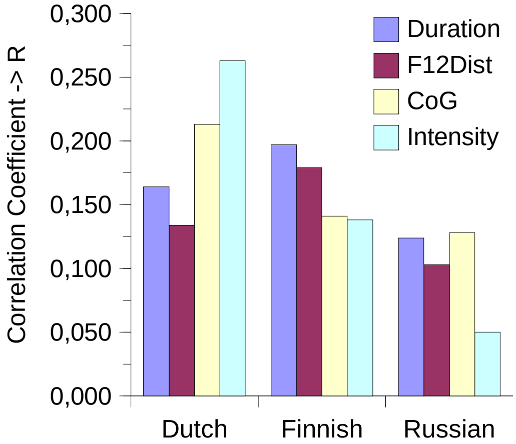

Figure 3. The correlation coefficient between global $-\log_2$(word frequency) and acoustic reduction for spontaneous speech. All correlation coefficients are significant ($p <= 0.001$).

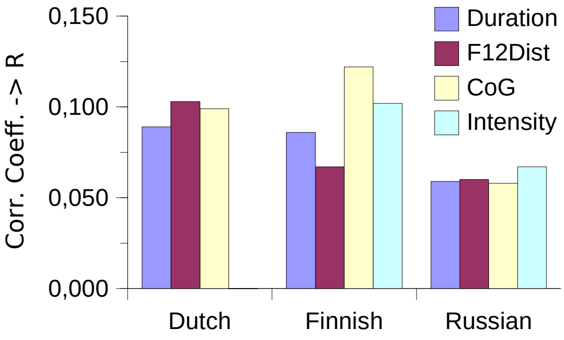

Figure 4. The correlation coefficient between segmental information content I_L and acoustic reduction for read speech. All correlation coefficients are significant (p<=0.001) except F12Dist for Finnish and Intensity for Dutch.

Finnish (R=0.357, p<0.001). This might indicate that for the Finnish corpus used, the stressed vowel tends to be instrumental in identifying the target word. This might be connected to the vowel harmony in Finnish words.

Dutch frequencies were calculated on a CELEX word-count list with normative transcriptions of Dutch, based on 39 million words. The word frequencies were estimated using a Katz smoothing on counts from 1-5 and an extrapolation based on Zipf's law. For Russian, a word forms frequency list was used [12], based on about a 40 million word corpus, which included texts in various styles. It should be noted that the word frequency distribution in our material (spontaneous speech) could considerably differ from the literary text. No external word-frequency lists were available for spoken Finnish. Therefore, the Finnish word form frequencies were based on transcripts of the Intas 915 recordings (approx. 30000 word forms).

As an illustration, the lexical information, I_L, is calculated for the vowel /o/ in the Dutch word /bom/ (*boom*, English *tree*, example from [20-23]).

• Token count of *boom* in CELEX	2227
• Word tokens starting with /bo/:	67,710 (1,172 CELEX entries)
• The same for /b./:	1,544,4831 (26,186 CELEX entries)

I_L=-log2(67710/15444831) = 4.5 (c.f. eq. (2))

Word realizations can differ from the lexical norm. For Dutch, the position of the realized phoneme in the normative lexical transcription was determined using a Dynamic Programming algorithm. Due to the complexity of rules converting a lexical word into an ideal phoneme string in Russian, a Dynamic Programming algorithm was used to link the acoustically realized phone with a lexical word. Thus while calculating the lexical information of a vowel segment I_L, the word onset frequency was defined by a lexical entry and not by a phonetic transcription. For Finnish, word and syllable boundaries were manually marked. Finnish orthography is close to phonemic structure. Therefore, the lexical normative transcription of the word-onset and phoneme identity were used to search the word-lists.

2.3. Acoustic reduction

Acoustic reduction was measured on duration, formant values (F_1 and F_2), the spectral Center of Gravity (CoG, c.f., [16]), and the sound power (Intensity). These values were calculated from the temporal mid point of each vowel segment. Only monophthongs were used. The position in vowel formant space was also used. The values of the F_1 and F_2 (in semitones) were combined, see Figure 1, as the distance to a virtual target of reduction, determined for each speaker separately as a point with an F_1 midway between the /i/ and the /u/ and an F_2 of the /a/, measured in read speech. Reduction of a vowel results in a shorter distance to this virtual point in vowel space (Figure 1). All measurements were done using the Praat program [2].

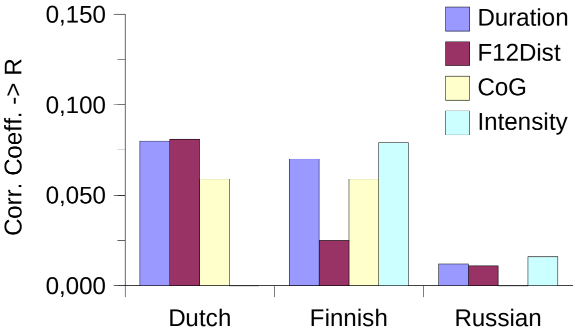

Figure 5. The correlation coefficient between segmental information content I_L and acoustic reduction for spontaneous speech. Correlation coefficients for Duration and F12Dist for Dutch and Intensity for Finnish are statistically significant (p <= 0.001), the others are not statistically significant (p > 0.001)

2.4. Statistics

To account for the large number of tests performed in this study a Bonferroni correction was applied. A $p <= 0.001$ was chosen as the level of statistical significance.

To quantify the relation between redundancy and reduction, the probabilities of finding a certain segment were recalculated into information (bits) as -$\log_2$(Probability). This information was correlated with the acoustic measures of reduction. As these acoustic measures are all highly dependent on speaker, speaking style and vowel identity, all values were normalized for these three factors. Therefore, the data were divided into separate sub-samples for each language, speaker, speaking style, and vowel. In each sub-sample, values were normalized to zero mean and a standard deviation of 1. Two degrees of freedom were subtracted for each sub-sample [20-23].

3. Results

Figure 2 and 3 give the strength of the correlation between the word redundancy, i.e., -$\log_2$(word-frequency), and the measure of acoustic reduction. There is a clear positive correlation between these two factors. It shows that stressed vowels are less reduced if they are part of words with a lower frequency. This is found for all three languages and all measures of acoustic vowel reduction, except for CoG and Intensity in read Russian speech. The correlation coefficients are rather small, i.e., R <= 0.25, explaining less than 6% of the variance. However, the underlying data is very noisy, and large measurement and sampling errors may be expected for both word-frequencies and the level of acoustic reduction. Furthermore, redundancy is only one of many factors affecting pronunciation. As all other prosodic and articulatory aspects are ignored, the consistency found in the correlation coefficients is rather surprising.

Figures 4 and 5 indicate the correlation coefficients between the segmental information content (I_L) and measures of acoustic reduction. The correlation coefficients are much lower than those for word-redundancy. For read speech (figure 4), they are mostly statistically significant. For spontaneous speech, they are mostly statistically *not* significant (figure 5).

Overall, the correlations are stronger for *read* speech than for *spontaneous* speech. This indicates that in read speech the lexical factors that we use to determine redundancy are more important than in spontaneous speech.

4. Discussion and Conclusions

Our results clearly indicate that there is a consistent correlation between redundancy at the *word level*, i.e., word-frequency, and acoustic vowel reduction. The correlation strengths are comparable in all three languages, which is remarkable given that these languages have quite different morphological and rhythmic structures.

A similar reduction effect was independently found at the *segmental level* (I_L, figures 4 and 5), but here the picture is much less consistent. To some extent, the weaker correlations can be attributed to the fact that the calculation of the segmental redundancy (I_L) involves dividing two large sums of rare word counts, which is inherently noisy.

We can conclude that all three languages respond to lexical and segmental redundancy by increasing acoustic reduction in the same way. We investigated four properties of vowels: duration, formant space, spectral Center of Gravity (first spectral moment), and intensity (acoustic power). All of these measures were reduced for the stressed vowels in redundant words and segments. Our study indicates that the link between vowel reduction and redundancy might be a language universal.

5. Acknowledgments

This research is part of the Intas 915 project. The work of Mietta Lennes has been funded by the Academy of Finland (project 53623). The IFAcorpus was made possible by grant 355-75-001 of the Netherlands Organization for Scientific Research (NWO). The IFAcorpus is licensed under the GNU GPL by the Dutch Language Union. We wish to thank the participants of Intas 915 in both St. Petersburg and Helsinki for their valuable assistance in the labeling work of the Finnish and Russian speech materials.

6. References

[1] Aylett, M., *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech*, PhD thesis, University of Edinburgh, 190 pp, 1999.

[2] Boersma, P.P.G. & Weenink, D. http://www.praat.org.

[3] Bondarko L.V., Iivonen A., Pols L.C.W., de Silva V., “Common and language dependent phonetic differences between read and spontaneous speech in Russian, Finnish and Dutch”, *Proc. ICPhS’03*, Barcelona, 2977-2980, 2003.

[4] Bolotova O. “On some phonetic features of spontaneous speech and reading in Russian”, *Proc. ICPhS’03*, Barcelona, 913-916, 2003.

[5] Cutler, A., “Speaking for listening”, in A. Allport, D. McKay, W. Prinz & E. Scheerer (eds.) *Language perception and production, London*; Academic Press, 23-40, 1987.

[6] Cutler, A., “Spoken word recognition and production”, in J.L. Miller & P.D. Eimas (eds.) *Speech, Language, and Communication. Handbook of Perception and Cognition, 11*, Academic Press, Inc, 97-136, 1995.

[7] Cutler, A. & Carter, D.M., “The predominance of strong initial syllables in English vocabulary”, *Computer Speech and Language* 2, 133-142, 1987.

[8] Cutler A., “The comparative perspective on spoken-language processing”, *Speech Communication* 21, 3-15, 1997.

[9] Fowler, C.A., “Differential shortening of repeated content words in various communicative contexts”, *Language and Speech* 31, 307-319, 1988.

[10] IFAcorpus, http://www.fon.hum.uva.nl/IFAcorpus.

[11] Lennes, M., “On the expected variability of vowel quality in Finnish informal dialogue”, *Proc. ICPhS’03*, Barcelona, Vol. 3, 2985-2988, 2003.

[12] Sharov S.A. Word frequency list for Russian. http://www.artint.ru/projects/frqlist.asp.

[13] Lieberman, P., “Some effects of semantic and grammatical context on the production and perception of speech”, *Language and Speech* 6, 172-187, 1963.

[14] Norris D., McQueen J.M. & Cutler A., “Merging information in speech recognition: Feedback is never necessary”, *Behavioral and Brain Sciences* 23, 299-325, 2000.

[15] Van Son, R.J.J.H., Koopmans-van Beinum, F.J. & Pols, L.C.W., “Efficiency as an organizing factor in natural speech”, *Proc. ICSLP'98*, Sydney, 2375-2378, 1998.

[16] Van Son, R.J.J.H. & Pols, L.C.W., “An acoustic description of consonant reduction”, *Speech Communication* 28, 125-140, 1999.

[17] Van Son, R.J.J.H. & Pols, L.C.W., “Effects of stress and lexical structure on speech efficiency” *Proc. Eurospeech'99*, Budapest, 439-442, 1999.

[18] Van Son, R.J.J.H. & Pols, L.C.W., “Perisegmental speech improves consonant and vowel identification”, *Speech Communication* 29, 1-22, 1999.

[19] Van Son, R.J.J.H., Binnenpoorte, D., Van den Heuvel, H. & Pols, L.C.W., “The IFAcorpus: a phonemically segmented Dutch open source speech database”, *Proc. Eurospeech’01*, Aalborg, Vol. 3, 2051-2054, 2001.

[20] Van Son, R.J.J.H. & Pols, L.C.W., “Evidence for efficiency in vowel production”, *Proc. ICSLP’02*, Denver, Vol I, 37-40, 2002.

[21] Van Son, R.J.J.H. & Pols, L.C.W., “An acoustic model of communicative efficiency in consonants and vowels taking into account context distinctiveness”, *Proc. ICPhS’03*, Barcelona, Vol. 2, 2141-2144, 2003.

[22] Van Son, R.J.J.H. & Pols, L.C.W., “Information structure and efficiency in speech production”, *Proc. Eurospeech’03*, Geneva, Vol. 1, 769-772 2003.

[23] Van Son, R.J.J.H. & Pols, L.C.W., “How efficient is speech?”, *Proceedings of Institute of Phonetic Sciences, University of Amsterdam* 25, 171-184, 2003.

[24] Zue, V.W., “The use of speech knowledge in automatic speech recognition”, *Proc. IEEE* 73, 1602-1616, 1985.