Speech Prosody 2016
31 May - 3 Jun 2106, Boston, USA

Impact of prosodic structure and information density on vowel space size

Erika Schulz[1], *Yoon Mi Oh*[2], *Zofia Malisz*[1], *Bistra Andreeva*[1], *Bernd Möbius*[1]

[1]Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany
[2]Laboratoire Dynamique du Langage, Université de Lyon and CNRS, France
{eschulz; zmalisz; andreeva; moebius}@coli.uni-saarland.de, yoon-mi.oh@univ-lyon2.fr

Abstract

We investigated the influence of prosodic structure and information density on vowel space size. Vowels were measured in five languages from the BonnTempo corpus, French, German, Finnish, Czech, and Polish, each with three female and three male speakers. Speakers read the text at normal, slow, and fast speech rate. The Euclidean distance between vowel space midpoint and formant values for each speaker was used as a measure for vowel distinctiveness. The prosodic model consisted of prominence and boundary. Information density was calculated for each language using the surprisal of the biphone $X_n|X_{n-1}$. On average, there is a positive relationship between vowel space expansion and information density. Detailed analysis revealed that this relationship did not hold for Finnish, and was only weak for Polish. When vowel distinctiveness was modeled as a function of prosodic factors and information density in linear mixed effects models (LMM), only prosodic factors were significant in explaining the variance in vowel space expansion. All prosodic factors, except word boundary, showed significant positive results in LMM. Vowels were more distinct in stressed syllables, before a prosodic boundary and at normal and slow speech rate compared to fast speech.

Index Terms: vowel space, prosodic structure, speech rate, information density

1. Introduction

Vowel space size is influenced by several factors, such as sex [1], speaking style [2], language redundancy and prosodic structure [3]. Regarding its relationship with speech rate, there have been inconsistent findings. Some studies found strong vowel reduction as speech rate increases [4, 5] which was also reflected in the perception of speech tempo [6], while others only found minimal or no impact of speech rate on vowel space size [7, 8].

Fast speech rate is associated with overall shorter segment durations leading to reduced spectral characteristics of vowels [9]. This means that vowel formants move to a more central position in the vowel space. These effects have also been observed in German for naturally occurring differences in local speech rate [5]. Also, with varied intended speech rates the US English vowel space has been shown to be reduced in size with increasing speech rate. Vowel space size also added to speech intelligibility: normal and slow speech received higher ratings in intelligibility than fast speech [4].

Small differences in vowel quality were found between slow and fast speech of US English when the expansion of the vowel space was described as the Euclidean distance between the measured formant values and a neutral position of the formant track [8]. This neutral position is described by a uniform vocal tract [10] and is not based on the measured formant values. Spectral vowel reduction with increased tempo was neither found in the analysis of Dutch read speech from newspapers [7] nor in dynamic vowel measurements of Dutch [11]. The two latter studies have only investigated speech by a single individual.

Prosodic structure, in general, is thought to have an influence on vowel space size [12]. In Aylett and Turk's Smooth Signal Redundancy hypothesis [13, 3] language redundancy and acoustic redundancy show an inverse relationship which is mediated and implemented through prosodic structure. In their study on the influence of prosodic structure and information density on vowel characteristics in US English they investigated read speech from the Rhetorical corpus [3]. Their language redundancy model was designed using high, mid and low language redundancy based on unigram, bigram and trigram probabilities of syllables. The prosodic model defined prominence and prosodic boundaries. Results of the study showed that vowels were more centralized with increased language redundancy, vowel quality in prominent syllables was more distinct than in syllables that were not prominent, and spectral characteristics of vowels were also more distinct in syllables before prosodic boundaries than in syllables at word or no boundary.

While Aylett and Turk [3] made use of the language redundancy model described above there are alternative measures of information density, e. g. surprisal which is frequently used in psycholinguistic studies [14]. Surprisal $S(X_n)$ measures the surprise of encountering a linguistic unit X_n in a specific context X_c based on language models (LMs) which estimate the distribution of sequences of linguistic units in a language [15] (see eq. 1).

$$S(X_n) = -log_2 P(X_n|X_c) \tag{1}$$

The current study investigated the influence of information density and prosodic structure on vowel space size. French (FRA), German (DEU), Polish (POL), Czech (CES), and Finnish (FIN) were included in the analysis. Contrary to previous research [3], the current study used data with various intended speech rates. In addition, the language redundancy model and the measurement of vowel distinctiveness differed in this cross-linguistic approach from Aylett and Turk [3] to simplify comparisons between languages.

2. Method

2.1. Materials

A subset of the BonnTempo corpus [16] was analyzed with three female and three male speakers of FRA, DEU, FIN, CES, and POL. FIN was added to the BonnTempo corpus using the original instructions [16]. Speakers were given an excerpt of a novel in their native language, and were asked to familiarize themselves with the text. Next, speakers were recorded at what they considered to be reading at normal pace. Then, sub-

 doi: 10.21437/SpeechProsody.2016-72

Table 1: *Number of items and vowel identity per language.*

Language	Items	Vowels
FRA	689	/i, e, a, u/
DEU	825	/iː, ɪ, eː, ɛː, ɛ, aː, a, uː, ʊ/
FIN	1178	/i, eː, e, æ, æː, ɑː, ɑ, uː, u/
POL	790	/i, ɛ, a, u/
CES	1156	/iː, ɪ, ɛː, ɛ, aː, a, u/

Table 2: *Corpus and corpus size for language modeling.*

Language	Corpus	No. of tokens (M)
FRA	LEXIQUE 3.80	9.1
DEU	WebCelex	4.6
FIN	Finnish PAROLE	180
POL	Frequency dictionary	901
CES	Frequency dictionary	398

jects were asked to slow down, and to slow down even more. In a third step, fast speech rate was recorded asking speakers to speak fast, and speed up their speech rate until they considered they could not speed up any more. From these acceleration steps, normal speech rate, as well as the first steps of slow and fast speech rate were used for analysis.

The corpus was automatically segmented using SPPAS for French [17] and WebMaus [18] for all other languages. Automatic segmentation was manually verified by phonetic experts based on the beginnings and ends of vowels which were marked by clearly visible formant structure. Vowel phonemes were chosen to facilitate comparative analysis between the different languages in the corpus. If available in the data, tense and lax vowels in closed front, closed back, low and front mid position were used for analysis (see table 1). The total number of analyzed tokens was 4638.

In order to build a LM, large text corpora were collected for the five languages (see table 2). First, each corpus was cleaned by removing erroneous entries. Then, the data was phonetically transcribed into IPA and, if not provided, automatically syllabified. For French, Lexique 3.80 [19] was retrieved online, which provides phonetic transcription and syllabification. For German, the WebCelex corpus [20] includes syllabification, transcription, and stress assignment. For Finnish, the Finnish Parole Corpus [21] was acquired online. The data was automatically converted into IPA by the speech synthesizer eSpeak [22] and was automatically syllabified by a bash shell script. For Polish, a frequency dictionary derived from large-scale web corpora [23] was converted into IPA and syllabified by an automatic tool for transcription and syllabification [24]. For Czech, a frequency dictionary acquired from large-scale web corpora [23] was automatically transcribed by eSpeak and syllabified by a bash shell script.

2.2. Data analysis

F1 and F2 were measured at the temporal midpoint in vocalic nuclei. Formant analysis was conducted with the Burg algorithm in Praat [25] with a maximum of five formants, window size of 0.025 sec, pre-emphasis from 50 Hz, and a maximum formant threshold of 5000 Hz (male speakers) and 5500 Hz (female speakers). Formant values were cleaned and manually checked before speaker-dependent normalization was applied to control for differences in formant values due to sex or speaker [26]. As a measure for vowel distinctiveness, the Euclidean distance between midpoint of the vowel space and formant values for every vowel were calculated for each speaker [27]. The larger the distance between the vowel space midpoint and individual vowels gets, the more distinct is vowel quality. This measure is independent of differences in vowel inventory between the languages because it assumes that vowel distinctiveness is defined by vowel space expansion.

The prosodic model consisted of prominence and boundary. Prominence was a binary factor using primary lexical stress (stressed vs. unstressed) based on the BonnTempo corpus[1]. If monosyllabic, function words were counted as unstressed, whereas content words were identified as stressed. Boundary showed three factor levels: none, word boundary, and high likelihood of prosodic boundary. Vowels were counted to be the nucleus of a syllable with high likelihood before a prosodic boundary when a pause of at least 100 msec followed [3].

As a measure of information density, surprisal values were estimated from a biphone LM with function and content words, taking the previous context into account. Contrary to Aylett and Turk [3], we chose a biphone LM due to the restricted number of different syllables in the BonnTempo corpus. In addition, the relationship between information density and phonetic structures is assumed to be better reflected by phoneme LMs [30]. The factor levels high, medium and low language redundancy [3] are not necessarily comparable across languages, in contrast to the continuous variable surprisal. Surprisal values were mean-centered.

3. Results

3.1. Correlation analysis

When averaged over all languages there was a positive relationship between vowel distinctiveness and surprisal of the biphone $X_n|X_{n-1}$ (r = 0.171, t(4636) = 11.808, p <.001). However, this relationship did not hold for all languages when each language was investigated individually. There was no significant correlation between information density and vowel distinctiveness in Finnish, and only a small, but significant positive relationship for these measures in Polish (see fig. 1).

In addition, the previously observed positive relationship between vowel distinctiveness and information density did not hold for all vowel identities in the corpus. In CES and DEU, all significant correlations described a positive relationship between the two variables for all vowel phonemes. Both languages also had the strongest positive correlation between information density and vowel distinctiveness (see fig. 1). The most consistent positive results were found for /a, aː/ and /i, iː, ɪ/, except for French /a/ (r = -0.163, p = 0.026) (see table 3).

3.2. Linear mixed effects model

We used R [31] and *lmerTest* [32] to perform a linear mixed effects analysis of the relationship between the dependent variable vowel distinctiveness and the fixed effects information density, vowel identity, speech rate, prosodic boundary, and prominence. Since the correlation analysis suggested that the effect of information density on vowel distinctiveness differs with vowel identity, the models also included the interaction of surprisal and vowel identity. Factor levels of vowel identity were expressed using deviation coding [33]. As random effects, we had

[1] In French, accent was marked on the last syllable of a phrase with a full vowel [28, 29].

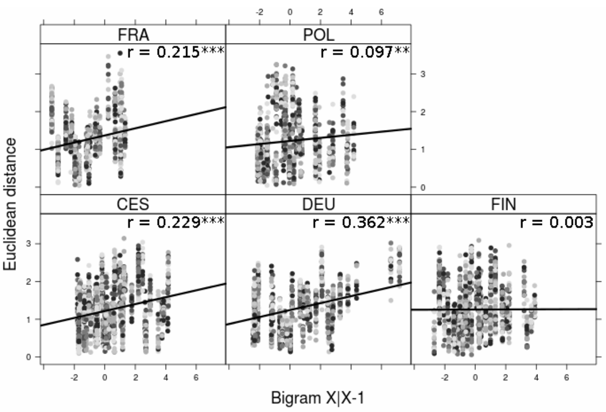

Figure 1: *Pearson's correlation of surprisal of the biphone $X_n|X_{n-1}$ and vowel distinctiveness for all languages (** p <0.01, *** p <0.001).*

Table 3: *Significant results of Pearson's correlation of surprisal of the biphone $X_n|X_{n-1}$ and vowel distinctiveness for each vowel identity (* p <0.05, ** p <0.01, *** p <0.001).*

LANG	V	*r*	LANG	V	*r*
FRA	i	0.251***	DEU	iː	0.336**
	e	-0.140*		ɪ	0.264***
	a	-0.163*		eː	0.564***
FIN	eː	-0.635***		aː	0.476***
	e	0.296**		a	0.261***
	æ	0.316***		uː	0.442**
	ɑ	0.135*		ʊ	0.229*
POL	ɛ	-0.267***	CES	ɪ	0.139*
	a	0.135**		ɛː	0.680***
	u	-0.256*		ɛ	0.241***
				u	0.848***

intercepts for word and language, as well as by-word random slope for the effect of information density. Visual inspection of residual plots did not show any obvious violation of the normality assumption or homoscedasticity. The model with the best fit was identified by likelihood ratio tests [34] (see fig. 2).

Surprisal of the biphone $X_n|X_{n-1}$ did not affect vowel distinctiveness significantly. Stress increased vowel distinctiveness significantly ($t(2562) = 6.511$, $p <0.001$) by 0.208 (*SD* = 0.032). Vowel space expanded by 0.031 (*SD* = 0.013) from fast to normal speech ($t(4274) = 2.482$, $p = 0.013$), and expanded even more by 0.115 (*SD* = 0.013) in slow speech compared to fast speech ($t(4285) = 9.150$, $p <0.001$). Vowels in syllables before prosodic boundaries are by 0.104 (*SD* = 0.034) more distinct than vowels in no boundary position ($t(3753) = 2.987$, $p = 0.003$). Significant results for vowel identity indicated differences in distance to the midpoint of the vowel space for different vowel categories. Front mid vowel /e/ (d = -0.685 (*SD* = 0.030)) and low vowel /a/ (d = -0.249 (*SD* = 0.024)) were less distant from the vowel space midpoint than all vowels on average, whereas high closed vowel /i/ was by 0.359 (*SD* = 0.031) more distant from the midpoint than the average ($t(2693) = 11.467$, $p <0.001$) (see fig. 2). The significant interactions between surprisal and vowel identity /a/ and /e/ showed that low mid vowels were by 0.157 (*SD* = 0.017, $t(2756) = 9.043$, $p <0.001$), and front mid vowels were by 0.076 (*SD* = 0.017, $t(3261) = 3.679$,

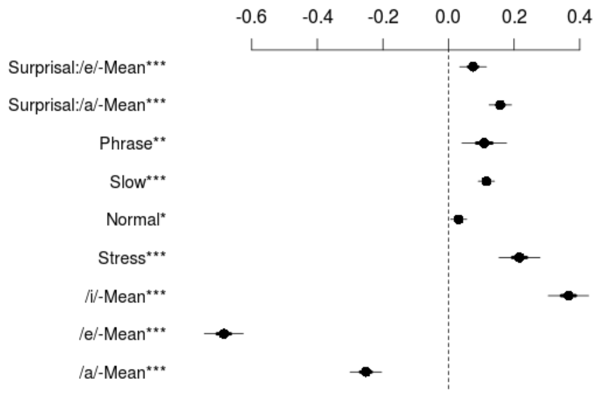

Figure 2: *Regression estimates for LMM with best fit (* p <0.05, ** p <0.01, *** p <0.001).*

$p <0.001$) more expanded under high surprisal than all other vowels on average.

Regarding the random structure of the model the intercept for language explained 0.013 (*SD* = 0.116) of the variance of the model, whereas intercept for word accounted for 0.299 (*SD* = 0.547) of the variance. By-word random slope for the effect of information density explained 0.167 (*SD* = 0.409) of the model variance correlating only slightly with each other (r = 0.29). The effect size of the entire model was ω_0^2 = 0.713.

4. Discussion

4.1. Correlation analysis

Averaged over all languages included in this paper there was a positive relationship between vowel distinctiveness and surprisal of the biphone $X_n|X_{n-1}$. This relationship did not hold when investigated for each language individually. The weak positive relationship between the two variables found in Polish can be due to the fact that there is (only slight) spectral vowel reduction in this language [35, 36]. This finding contrasted with previous observations on Polish as a language without spectral vowel reduction [9]. In Finnish, on the other hand, vowel reduction is realized by means of differences in duration [37]. Also, vowel quality in Finnish is morphophonemic. Front vowels (/æ, y, ø/) never appear with back vowels (/ɑ, u, o/) in the same lexeme [38] which is why vowel reduction would be detrimental to identifying lexeme boundaries in this language.

As visible in figures 3 and 4, differences in vowel space size at different speech rates are relatively small in these two languages. The polygon area of the Polish vowel space changes only slightly with tempo (A(slow) = 0.347 kHz^2, A(normal) = 0.335 kHz^2, A(fast) = 0.309 kHz^2). Similar small effects were observed for Finnish in the vowel space size under different speech rates (A(slow) = 0.648 kHz^2, A(normal) = 0.607 kHz^2, A(fast) = 0.522 kHz^2).

When vowel distinctiveness and information density were correlated for each vowel identity separately, it was observed that general observations for one language did not necessarily hold across languages. The two languages with the strongest correlation, DEU and CES, unsurprisingly only showed significant positive correlations between the two variables. Although there is no correlation between information density and vowel distinctiveness averaged over all vowels in FIN, the short open

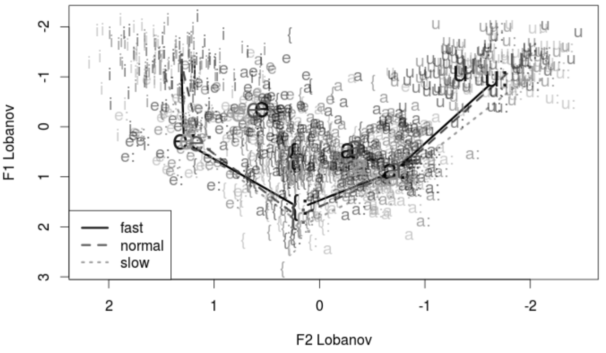

Figure 3: *Normalised Finnish vowel space at different speech rates.*

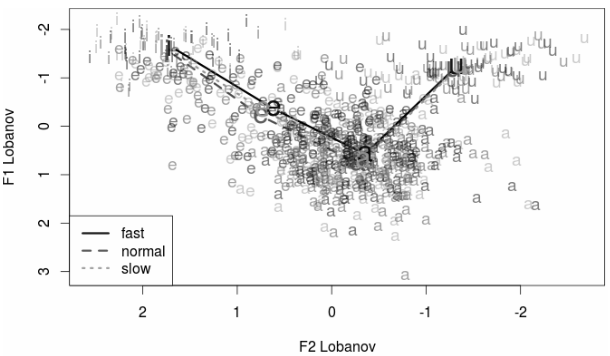

Figure 4: *Normalised Polish vowel space at different speech rates.*

mid vowels /æ, ɑ/ and the short front mid vowel /e/ showed a significant positive relationship. In FRA and POL, the overall positive result of correlations could not be replicated for all vowel qualities. These findings were in line with Aylett and Turk [3] who observed that their language redundancy model did not explain the variance in all investigated US English vowels. However, it should be noted that there was only a small number of items per vowel identity. Interestingly, vowel identities /i/ and /a/ showed the most consistent positive results in correlation analysis. These vowel phonemes also denote the maximum distance in perceptual contrast [39].

4.2. Linear mixed effects model

Contrary to previous findings [3], results of the LMM did not show a significant effect of information density on vowel distinctiveness. However, we found significant effects for all prosodic factors in the model, except for word boundary. The higher the intended speech rate got, the more reduced was vowel quality. This finding is in line with previous work which found a connection between speech rate and vowel space size [4, 5, 6]. Also, this study used the actual vowel space midpoint for each speaker to measure vowel distinctiveness and not an abstract neutral position [8], facilitating a more realistic estimation of vowel distinctiveness at different speech rates. Regarding prominence stressed vowels were more distinct in their vowel quality than unstressed vowels, as previously observed [9, 8]. Also, vowels in syllables before prosodic boundaries were more distinct in their spectral characteristics than vowels in syllables at no boundaries [3]. Prominence was the strongest predictor of vowel distinctiveness in this model.

Significant results for the fixed effect vowel identity indicated that vowels differed in their distance to the vowel space midpoint. Front mid vowels and low vowels were less distant from vowel space midpoint than all vowels on average, while closed front vowels were more distant than the average. This finding might be due to the fact that the vowel space midpoint was lowered towards the direction of vowel identities /e/ and /a/ because they were more numerous in the corpus (n = 3078) than closed vowels /i/ and /u/ (n = 1560). Open mid vowels and to a lesser extent also front mid vowels expanded more than all other vowel identities on average under high surprisal. This interaction could be explained by jaw undershoot for open vowels which are highly predictable from their preceding context, thereby reducing articulatory effort.

Despite the correlation analysis which revealed that information density and vowel distinctiveness did not correlate positively in all languages, the intercept for language did only explain a small amount of variance in the LMM (0.013 (*SD* = 0.116)). This finding was due to the fact that this positive relationship held for the majority of the languages investigated in this study. Also, LMM with language as fixed factor did not yield a significantly better log likelihood (*log(L)* = -2188.6) than the model with language as intercept (*log(L)* = -2187.5; $\chi^2(3) = 2.266$, $p = 0.519$).

Vowel formants have been reported to be context-dependent [9, 40]. However, this additional factor was not included in the current study because of the relatively small number of investigated items. In order to minimize the contextual influence, vowel formants were measured at the temporal midpoint of the vowel. Assumed vowel targets with a small degree of spectral change occur close to the midpoint of the vowel [9, 41], but c.f. [42].

Further development of this research will involve alternative measures of vowel distinctiveness as well as a revision of the prosodic model. In particular, we are interested in investigating the relationship between vowel space expansion and realised prominence.

5. Summary

The results of this study showed that the relationship between information density and vowel distinctiveness depended on the language and vowel identity under investigation. Averaged over all languages there was a positive relationship between the two variables: Vowel distinctiveness increased with increasing surprisal. However, this relationship could not be observed when vowel distinctiveness was modeled as a function of prosodic factors and information density. Here, only prosodic factors had an influence on vowel space size. This was in contrast to previous research for US English [3]. In addition, this study added support to the literature that there is indeed a relationship between speech rate variation and differences in vowel space size.

6. Acknowledgements

This research was funded by the German Research Foundation (DFG) as part of SFB 1102 'Information Density and Linguistic Encoding' at Saarland University.

7. References

[1] A. P. Simpson and C. Ericsdotter, "Sex-specific differences in f0 and vowel space," in *Proceedings of XVIth ICPhS*, 2007, pp. 933–936.

[2] A. R. Bradlow, N. Kraus, and E. Hayes, “Speaking clearly for children with learning disabilities: Sentence perception in noise,” *Journal of Speech, Language, and Hearing Research*, vol. 46, pp. 80–97, 2003.

[3] M. Aylett and A. Turk, “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei,” *Journal of the Acoustical Society of America*, vol. 119, pp. 3048–3058, 2006.

[4] G. S. Turner, K. Tjaden, and G. Weismer, “The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis,” *Journal of Speech and Hearing Research*, vol. 38, pp. 1001–1013, 1995.

[5] B. Weiss, “Rate dependent vowel reduction in German,” in *Proceedings of the 12th SPECOM*, Moscow, 2007.

[6] M. Weirich and A. P. Simpson, “Differences in acoustic vowel space and the perception of speech tempo,” *Journal of Phonetics*, vol. 43, pp. 1–10, 2014.

[7] R. J. J. H. van Son and L. C. W. Pols, “Formant frequencies of Dutch vowels in a text, read at normal and fast rate,” *Journal of the Acoustical Society of America*, vol. 88, pp. 1683–1693, 1990.

[8] M. Fourakis, “Tempo, stress, and vowel reduction in American English,” *Journal of the Acoustical Society of America*, vol. 90, no. 4, pp. 1816–1827, 1991.

[9] B. Lindblom, “Spectroraphic study of vowel reduction,” *Journal of the Acoustical Society of America*, vol. 35, no. 11, pp. 1773–1781, November 1963.

[10] G. Fant, *Acoustic theory of speech production*. The Hague, Netherlands: Mouton, 1970, vol. 2.

[11] L. C. W. Pols and R. J. J. H. van Son, “Acoustics and perception of dynamic vowel segments,” *Speech Communication*, vol. 13, pp. 135–147, 1993.

[12] D. R. van Bergem, “Acoustic vowel reduction as a function of sentence accent, word stress, and word class,” *Speech Communication*, vol. 12, pp. 1–23, 1993.

[13] M. Aylett and A. Turk, “The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech,” *Language and Speech*, vol. 47, no. 1, pp. 31–56, 2004.

[14] A. Frank and T. F. Jaeger, “Speaking rationally: Uniform information density as an optimal strategy for language production,” in *CogSci 2008*. Washington, DC, USA: Cognitive Science Society, 23-26 July 2008, pp. 939–944.

[15] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press, 1999.

[16] V. Dellwo, I. Steiner, B. Aschenberner, J. Dankovicova, and P. Wagner, “BonnTempo-corpus and BonnTempo-tools: a database for the study of speech rhythm and rate,” in *Interspeech 2004*, 2004, pp. 777–780.

[17] B. Bigi. (2013) SPPAS - Automatic Annotation of Speech. Banque de données parole et langage (SLDR/ORTOLANG).

[18] T. Kisler, F. Schiel, and H. Sloetjes, “Signal processing via web services: the use case WebMAUS,” in *Digital Humanities 2012*, Hamburg, Germany, 2012.

[19] B. New, C. Pallier, L. Ferrand, and R. Matos, “Une base de données lexicales du français contemporain sur internet: LEXIQUE 3.80,” *LAnnée Psychologique*, vol. 101, pp. 447–462, 2001. [Online]. Available: http://www.lexique.org

[20] Max Planck Institute for Psycholinguistics. Webcelex. Retrieved on March 18, 2013 and on August 6, 2014. [Online]. Available: http://celex.mpi.nl

[21] Department of General Linguistics. (1996–1998) Finnish parole corpus. University of Helsinki AND Institute for the Languages of Finland. [Online]. Available: http://kaino.kotus.fi/sanat/taajuuslista/parole.php

[22] J. Duddington. eSpeak text to speech. Retrieved on 1 February 2015. [Online]. Available: http://espeak.sourceforge.net/

[23] A. Zséder, G. Recski, D. Varga, and A. Kornai, “Rapid creation of large-scale corpora and frequency dictionaries,” in *LREC 2012*, 2012.

[24] A. Zeldes. (2008–2014) Automatic phonetic transcription and syllable analysis. Georgetown University. [Online]. Available: http://corpling.uis.georgetown.edu/amir/phon.php

[25] P. Boersma and D. Weenink. (2015) Praat: doing phonetics by computer [computer program]. version 5.4.22. [Online]. Available: http://www.praat.org/

[26] B. M. Lobanov, “Classification of Russian vowels spoken by different speakers,” *Journal of the Acoustical Society of America*, vol. 49, pp. 606–608, 1971.

[27] N. Amir and O. Amir, “Novel measures for vowel reduction,” in *ICPhS XVI*, Saarbrücken, 2007, pp. 849–852.

[28] S.-A. Jun and C. Fougeron, “A phonological model of french intonation,” in *Intonation*, ser. Text, Speech and Language Technology, A. Botinis, Ed. Springer Netherlands, 2000, vol. 15, pp. 209–242. [Online]. Available: http://dx.doi.org/10.1007/978-94-011-4317-2_10

[29] C. Féry, “Final compression in French as a phrasal phenomenon,” in *Perspectives on linguistic structure and context: Studies in honour of Knud Lambrecht*, S. Katz Bourns and L. L. Myers, Eds. Amsterdam: John Benjamins, 2014, pp. 133–156.

[30] Y. M. Oh, C. Coupe, E. Marsico, and F. Pellegrino, “Bridging phonological system and lexicon: Insights from a corpus study of functional load,” *Journal of Phonetics*, vol. 53, pp. 153–176, 2015.

[31] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2015. [Online]. Available: http://www.R-project.org/

[32] A. Kuznetsova, P. Bruun Brockhoff, and R. Haubo Bojesen Christensen, *lmerTest: Tests in Linear Mixed Effects Models*, 2014, r package version 2.0-20. [Online]. Available: http://CRAN.R-project.org/package=lmerTest

[33] M. Kuhn, contributions from Steve Weston, J. Wing, J. Forester, and T. Thaler, *contrast: A collection of contrast methods*, 2013, r package version 0.19. [Online]. Available: http://CRAN.R-project.org/package=contrast

[34] A. Zeileis and T. Hothorn, “Diagnostic checking in regression relationships,” *R News*, vol. 2, no. 3, pp. 7–10, 2002.

[35] P. M. Nowak, “Vowel reduction in Polish,” Ph.D. dissertation, University of California, Berkeley, 2006.

[36] W. J. Barry and B. Andreeva, “Cross-language similarities and differences in spontaneous speech patterns,” *Journal of the International Phonetic Association*, vol. 31, no. 1, pp. 51–66, 2001.

[37] K. Suomi, J. H. Toivanen, and R. Ylitalo, *Finnish sound structure. Phonetics, phonology, phonotactics and prosody*. Uolo University Press, 2008.

[38] R. Bertram, A. Pollatsek, and J. Hyönä, “Morphological parsing and the use of segmentation cues in reading Finnish compounds,” *Journal of Memory and Language*, vol. 51, pp. 325–345, 2004.

[39] J. Liljencrants and B. Lindblom, “Numerical simulation of vowel quality systems: The role of perceptual contrast,” *Language*, vol. 48, no. 4, pp. 839–862, 1972.

[40] J. Hillenbrand and M. J. Clark, “Effects of consonant environment on vowel formant patterns,” *Journal of the Acoustical Society of America*, vol. 109, no. 2, pp. 748–763, 2001.

[41] K. N. Stevens and A. S. House, “Perturbation of vowel articulations by consonantal context: An acoustical study,” *Journal of Speech, Language, and Hearing Research*, vol. 6, pp. 111–128, 1963.

[42] J. Olive, J. van Santen, B. Möbius, and C. Shih, “Synthesis,” in *Multilingual Text-to-Speech Synthesis: The Bell Labs Approach*, R. Sproat, Ed. Dordrecht: Kluwer, 1998, ch. 7, pp. 191–228.