INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

An acoustic analysis of /r/ in Tyrolean

Vincenzo Galatà, Lorenzo Spreafico, Alessandro Vietti, Constantijn Kaland

ALPs - Alpine Laboratory of Phonetics and Phonology, Free University of Bozen – Bolzano, Italy

{vincenzo.galata, lorenzo.spreafico, alessandro.vietti, constantijn.kaland} @unibz.it

Abstract

This paper offers a preliminary contribution to the phonetic description and acoustic characterization of /r/ allophony in Tyrolean dialect, an under-researched South Bavarian Dialect spoken in the North of Italy. The analysis of target words containing /r/ in different phonotactic contexts, produced by six Tyrolean female speakers, confirms the high degree of intra-speaker variation in the production of /r/ with a uvular place of articulation. The distributional analysis of the allophones in our sample shows a preference among all the speakers for a fricative manner of articulation followed by approximants and taps and, to a lesser extent, by trills (with a very small amount of vocalized variants). These results are in line with previous research in the South-Tyrolean community. Due to the high variability of rhotic sounds, we further investigate and report on some of their shared acoustic features such as duration across the different phonotactic contexts and Harmonics-to-Noise Ratio for the different allophones attested.

Index Terms: acoustic analysis, allophony, dialect, rhotic, Tyrolean

1. Introduction

South-Tyrol is a bilingual alpine region belonging to Italy that lies on the Germanic – (Italo) Romance dialectological border. South-Tyrol is described by [1] as a societal bilingualism where two linguistic communities coexist: (Bavarian) German and Italian.

According to the traditional dialectological literature ([2], [3]), the Tyrolean dialect spoken in the area under study (the small town of Meran and its neighbourhood) favours the realization of a uvular /r/ as in standard German (an in-depth description of /r/-sounds in Germany is offered in [4]). The data reported in the *Tirolischer Sprachatlas* [5] also show a relevant diatopic variation (with differences even within the same area): uvular phones prevail in *Burggrafenamt* (Meran), apical ones in the *Eisacktal* (Brixen) and in the Western *Pustertal* (Bruneck). The investigation by [6] on the *Bozner Deutsch* presents data accounting for uvular trill realizations with marginal productions of more fronted variants (similar tendencies are reported by [7] for Austrian German. Previous research by [8] on a Tyrolean speaker from Bolzano showed that the speaker only articulated back (uvular) /r/ variants when uttering both the target words in Tyrolean and in Italian. The most used uvular /r/ sound documented in this case for Tyrolean was the uvular fricative.

As reported in [9], the diatopic variation mentioned above is further complicated by additional sources of intra-speaker variation coming from the interaction along different dimensions (cognitive, phonetic, phonological and sociolinguistic).

Surprisingly, however, while systematic research on /r/ in other languages is abundant, the same is not true for the Tyrolean area (except [8], [9] and [10]). Apart from a few exceptions, the sources presented so far almost exclusively account for auditory investigations and no instrumental evidence is available as for other major languages (for German see [11]).

With its different manners and places of articulation, the class of rhotics is phonetically heterogeneous. It seems that the unique feature that unifies and represents the class of r-sounds is its orthographic rendition as “r” [12] or, as put forward by [13] in other terms, /r/-sounds can be considered as an agglomerate of parameter relations among phonetic and physical properties of /r/ forming a family of resemblances.

Our aim is to provide a preliminary acoustic description of /r/ allophony in Tyrolean, both in quantitative and qualitative terms. In the next paragraphs we will provide a detailed description of the methodology we adopted to describe the class of r-sounds. We will then report on the distribution of the /r/ variants in our sample and will then move to a more detailed description of its realizations according to their temporal cues. We conclude our analysis with an examination of the Harmonics-to-Noise Ratio. As reported by [14], fricatives have a significantly lower HNR than approximants: based on this measure we will indirectly objectify the goodness of the manual annotation (especially for voiced uvular fricatives and uvular approximants).

2. Materials and methods

2.1. The corpus

In order to study the allophony of /r/ we devised a list of real words in Tyrolean Dialect eliciting /r/ targets in onset and coda position, in simple syllables and in clusters, in word-initial, word-medial and word-final position. The list was prepared according to an in-depth scrutiny of all available Tyrolean dictionaries [15] and [16].

While compiling the first list (Table 1), surrounding vowels (V) were restricted to /a, a:, i, i:, o, o:/; surrounding consonants (C) for /r/ in syllable onset (CRV) and coda (VRC) position were restricted to /t, d, k, g/. For /r/ in coda position words with /r/ + nasal or liquid were also included. The final list consisted of 69 words (#CRV, CRV, #RV, RV, CVR#, VR#, VRC#, VRCL#, VRCN#, VRL, VRL#, VRN#).

However, after recording a few participants, some minor changes were necessary to improve the list. In the second list (Table 1), the full set of surrounding plosive consonants (C) for /r/ in syllable onset (CRV) and coda (VRC) position was included (e.g. /t, d, k, g, p, b/). The difference on vowel duration for Tyrolean, present in the first list, was removed to reduce and balance the number of contexts. The contexts VRCL#, VRCN#,

http://dx.doi.org/10.21437/Interspeech.2016-434

VRL, VRL# and CVR# were removed as well because they were unbalanced for the vowel preceding /r/: only VRN# was kept. Two words were added containing /e/ in the VR# context in order to verify if, and to what extent, vocalization occurs as in standard German. The second list consisted of 69 words for Tyrolean (#CRV, #RV, CRV, VR#, RV, VRC#, VRN#).

In order to reduce the recording's session, the lists were kept as short as possible. Words with more /r/ targets were preferred with the aim to increase the number of observations in the sample (given in parentheses in Table 1). Notwithstanding the small differences within the two lists, these remain comparable.

Both lists have been double-checked by two native speakers of Tyrolean to ensure the words were recognizable and correctly uttered by the participants.

Table 1: *Distribution of the tokens in the two lists (extra /r/ targets from other words in parentheses).*

			wi	wm	wf	***total***
list 1	coda	cluster		19 (3)		*19 (3)*
		simple		1 (2)	8 (2)	*9 (4)*
	onset	cluster		36 (2)		*36 (2)*
		simple	6 (2)	10 (2)		*16 (4)*
		total	*6 (2)*	*66 (9)*	*8 (2)*	***80 (13)***
list 2	coda	cluster		16 (1)		*16 (1)*
		simple		(3)	5 (6)	*5 (9)*
	onset	cluster		36		*36*
		simple	3 (1)	9 (1)		*12 (2)*
		total	*3 (1)*	*61 (5)*	*5 (6)*	***69 (12)***

2.2. Participants

Six female speakers (M_{age} = 25 years, age range: 21-33), born and living in the area of Meran-Merano, a town in South-Tyrol (Northern Italy), were recruited as native speakers of Tyrolean Dialect and were asked to read the words in the above described list. None of the participants reported speech disorders and all had command of Standard German and Standard Italian at native-like level. The speakers were also assessed by means of a computer-based questionnaire [17] in order to define their social network and their use of Italian and Tyrolean Dialect (these data will not be used here).

2.3. Data collection

The acoustic recordings (22050Hz 16-bit mono PCM) were obtained by means of a Sennheiser ME2 microphone, placed at a distance of ca. 10 cm from the speaker's mouth, connected to a B1 Marantz PMD660. One recording session for each participant took place in a soundproof cabin with the participant sitting in front of a 21" monitor on which every single target word to be uttered was presented. Participants were instructed to read the words in Tyrolean dialect. They were allowed to familiarize with the targets words prior to the recording session.

It has to be noted that the audio recordings used here were collected as part of a broader project on "*The Articulatory Sociophonetics of Bilinguals in South-Tyrol: The Ultrasound Tongue Imaging Potential*". As such, the recordings of the speech samples from the participants were acquired via the Articulate Assistant Advanced (AAA) software package [18] running on a Desktop Workstation. Due to this specific recording set-up, the participants also wore a helmet stabilizing an ultrasound probe under their chin during the entire recording session (see [10] for further details).

The target words' list was administered via AAA to the participants: two times to speaker MRL and EVK who were presented the first list and three times to FRT, TAM, KAE and COM who were presented the second list, to gather at least two and three repetitions of the same target word, respectively.

2.4. Data preparation and annotation

The audio material consists of one *.wav file for each target word exported from AAA.

Prior to the manual annotation, all files have been processed with the WebMAUSMultiple forced alignment (FA) webtool ([19], [20]) using the German language-dependent mode because of its affinity with the Tyrolean language under investigation here. This processing passage produced one Praat [21] *TextGrid* with one SAMPA *tier* providing a phone-level segmentation for each target word in each of the files. It may be worth mentioning here, that in a few cases the FA failed to identify and segment the word-final voiceless uvular fricative rhotic present in many target words. In other cases, the word-final /r/ tokens have been labelled by the FA tool with the SAMPA symbol "6" according to the German vocalization rules for word-final /r/. These cases have been fixed during the manual annotation carried out by one of the authors.

Starting from the FA segmentation, all /r/-segments were further refined and classified in Praat via auditory feedback through headphones and annotated via inspection of waveform and sonogram (with boundaries placed at zero crossing points in the waveform) on a tier *label* for the following features:

- Manner of articulation (MoA): trill (r), tap (t), approximant (a), fricative (f), vocalization (v), deletion (d). For undefined or problematic cases, mainly creaky productions, a question mark (?) was used.
- Place of articulation (PoA): uvular (u), alveolar (a). For vocalizations and deletions zero (0) was used.
- Voicing (±voice): voiced (v), voiceless (s). Ambiguous segments with partial voicing/devoicing were labelled with a question mark (mainly taps and trills).

Due to the high variability of the rhotic segments annotated and analysed, an additional *ac_label* tier (hierarchically dependent from the tier *label*) was added to the annotation with the aim of identifying and differentiating the phases or sub-segments making up the /r/ variants. Based on the inspection of waveform, sonogram and intensity curve, we identified and further segmented each r-sound according to three elements (a threefold label). The first label identifies the gesture of *aperture* and *constriction* characterizing each /r/ variant according to:

- *Aperture* (a), used to identify both 1) the *epenthetic* vowel fragment in some cases visible and audible before or after a tap and/or other r-sounds [22] and 2) the opening gesture following the constriction gesture in the production of a trill.
- *Constriction* (c), used to identify the stricture in the realization of r-sounds (tap, trill, fricative, approximant).
- Vocalizations and deletions were annotated separately.

The second label consists of a number starting from 1 to *n* to indicate the position of the annotated interval on tier *ac-label* with reference to the total number of intervals belonging to the annotation on Tier *label*. The number always started with 1 according to two hypothetic members of an *aperture* – *constriction* pair, where both are labelled as 1 (we say hypothetic since the pair may not exist due to the absence of the first *aperture*); continuing within the same r-sound, if other pairs are present, the number increases by $n + 1$. The third label added to the *ac_label* indicates whether there are other annotations flanking the segment labelled on tier *ac_label*: so we use *r* for an annotation on the right, *l* for an annotation on the left, *e* (empty) for no flanking annotations, *f* (full) for annotations on both sides. Further information available from the word list for each /r/ token (e.g. syllable type, position of /r/ within the syllable etc.) has been encoded on an extra *key* tier. Annotation and segmentation has been carried out manually by one of the authors, while another of them double-checked the data. An example of annotation is provided in Figure 1.

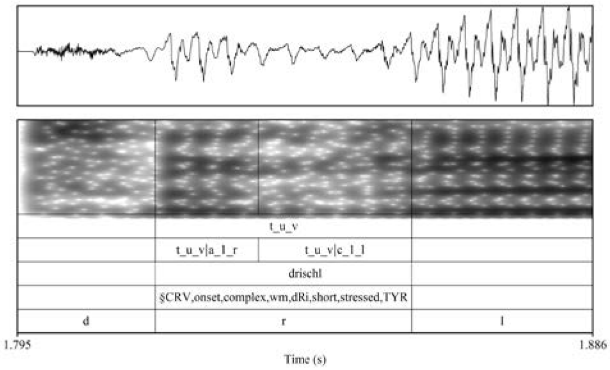

Figure 1: *A screenshot showing the annotation of a voiced uvular tap preceded by an epenthetic vowel.*

2.5. Acoustic and statistical analysis

Due to the high variability of rhotic sounds in general and due to the heterogeneous nature of the different allophones found in our data sample, for the present paper we selected measures that were comparable across the different variants. We further investigate and report on some of their shared acoustic correlates including duration and Harmonics-to-Noise Ratio (HNR) of the different produced variants leaving other measures for future work. In both cases we focus on the constriction phase of the segments (as described in section 2.4), excluding vocalizations and deletions from the analysis.

All measures have been extracted via Praat scripting. HNR has been computed using Praat's "To Harmonicity (cc)" object with standard settings (minimum frequency and number of periods per window, set to 100Hz and 1 respectively).

For the statistical analysis, the measures of duration and HNR in the constriction phase were analysed by fitting two linear mixed-effects models in R [23] using lme4 [24] for model fitting. Deviations from homoscedasticity or normality have been checked by visual inspection of residual plots. P-values for each model were obtained by likelihood ratio tests of the full model against the reduced model. In each model, subject and prompt were included as random factors.

3. Results

3.1. Allophonic distribution of /r/

As shown in Table 2, vocalizations and deletions (counted together) are differently present in all the speakers. For COM a possible interference from Italian is found: she produces four of the tokens with an alveolar /r/ (namely [ɾ]). The rest of the tokens is clearly produced by the participants with a uvular place of articulation as expected for Tyrolean Dialect ([2], [3]).

Table 2: *Number of /r/ tokens per participant divided by place of articulation (PoA).*

	PoA			word	n°	*tot.*
	voc/del	alv	uvu	list	reps.	*tokens*
COM	15	4	224	2	3	*243*
EVK	17	-	169	1	2	*186*
FRT	3	-	240	2	3	*243*
KAE	6	-	237	2	3	*243*
MRL	5	-	181	1	2	*186*
TAM	-	-	243	2	3	*243*

The preliminary acoustic-auditory labelling process identified five possible /r/-variants differently distributed among the six speakers: in this order, we find fricatives, approximants, taps, dorso-uvular trills and some sparse more vocalized variants. For all the allophones, except for the approximants, both voiced and voiceless allophones are present.

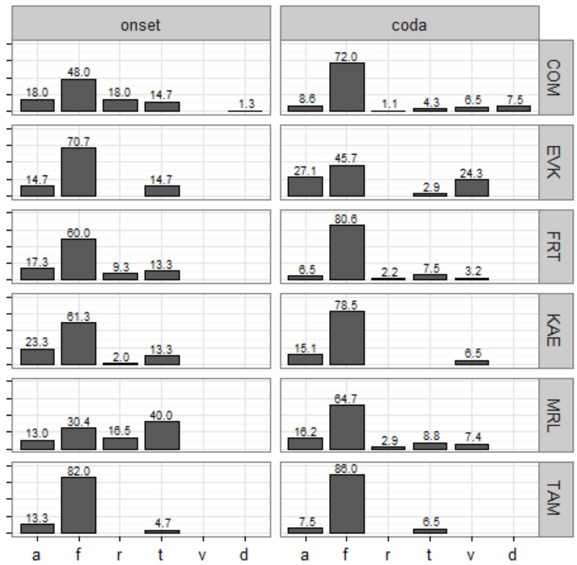

Figure 2: *Percent distributions of the /r/ tokens per subject grouped according to onset and coda position.*

Moving into more detail in Figure 2, despite the different proportions of realized allophones, some tendencies emerge:

- The uvular fricative (f) is the most frequent and the most context-independent variant appearing in both coda and onset position. This is also the dominant realization as voiceless fricative in word-final (wf) position.
- /r/-vocalization (v) and deletion (d) occur only in coda position.
- The uvular approximants (a), trills (r) and taps (t) are more frequent in onset position and in word-medial (wm) position.
- Uvular trills (r) never appear in word-final position.

In the realization of voiced uvular allophones (mainly in onset position) for all the speakers, the insertion of a vocalic fragment (annotated in our sample as an *aperture* gesture) is often attested (with tap undergoing to this phenomenon more frequently than the other allophones, see example in Figure 1).

3.2. Acoustic analysis

3.2.1. Duration

To analyse the temporal patterns of the constriction gestures for the uvular variants in our group of Tyrolean speakers, we focused on the word-medial context for which both onset and coda position were represented in our data (see Table 1). A linear mixed-effect model was fitted for duration of the constriction as dependent variable. Variant (4 levels: approximant, fricative, tap and trill) and position (2 levels: onset and coda) were modelled as fixed factors (with interaction), while subject and prompt were included as random factors.

As reported in Table 3, the fitted model (χ^2 (3) = 54.91, < .001) reveals significant interaction effects for variant and syllable position for fricatives which have a longer *constriction* phase in coda (67.83 ± 2.26 SE) as compared to onset position (47.88 ± 1.99 SE). No significant difference is found between approximants, taps and trill in onset vs. coda position. No significant differences are found between approximants and fricatives in onset position.

Table 3: *Coefficient and SE estimates, Satterthwaite approximated degrees of freedom and t statistics predicting duration of the constriction phase for the variant and /r/ position (with interaction). Ref. levels: var = a; r_pos = onset.*

	Estimate	SE	df	*t* value	Pr(>|t|)
(Intercept)	45.541	2.201	21.1	20.688	< .001
varf	2.341	1.675	1062.6	1.397	n.s.
varr	-31.320	2.037	1124.9	-15.379	< .001
vart	-24.818	2.006	1110.1	-12.373	< .001
r_poscoda	1.369	2.688	652.6	0.509	n.s.
varf:r_poscoda	18.576	3.074	655.3	6.043	< .001
varr:r_poscoda	-8.229	5.196	1101.0	-1.584	n.s.
vart:r_poscoda	-1.986	4.352	1100.3	-0.456	n.s.

3.2.2. Harmonics-to-Noise Ratio

In order to test for significant differences in HNR for the /r/-variants that were more often the most difficult to distinguish during the annotation passage described in section 2.4 (e.g. approximants vs. voiced fricatives) and to verify if HNR may be used to discriminate them, we performed a linear mixed-effects analysis. For the present analysis all the tokens labelled as partially voiced or unvoiced were discarded by selecting only those variants being labelled as fully voiced (+) or fully voiceless (-). However, this data selection procedure significantly reduced the number of tap and trill variants and for this reason we decided to include them in the model without voicing distinction. The linear mixed-effect model was fitted with HNR for the *constriction* phase as dependent variable, variant as fixed factor (5 levels: approximant, voiced fricative, voiceless fricative, tap and trill), while subject and prompt were included as random factors.

The fitted model (χ^2 (4) = 501.27, $p < .001$) for the estimated HNR is presented in Table 4. The approximant is the variant with the highest HNR value as one would expect (approximants have a vowel-like shape, exhibit a formant structure and are more periodic). Compared to the approximant the fricative variant lowers HNR by about 6.44 dB ± 0.49 (SE) in its voiced realization and by about 11.23 dB ± 0.42 (SE) in its voiceless realization.

Table 4: *Coefficient and SE estimates, Satterthwaite approximated degrees of freedom and t statistics predicting HNR in the constriction phase for variant. Ref. level: var = a+.*

	Estimate	SE	df	*t* value	Pr (>|t|)
(Intercept)	14.59	0.508	17.6	28.74	< .001
varf-	-11.23	0.417	647.9	-26.91	< .001
varf+	-6.44	0.491	878.8	-13.14	< .001
varr	-7.50	0.523	881.9	-14.32	< .001
vart	-7.75	0.484	879.0	-16.03	< .001

4. Preliminary conclusions

The work presented here provides a preliminary contribution to the phonetic description of the inventory of r-sounds in Tyrolean speakers.

As confirmed by the sample of speakers analysed so far, the uvular place of articulation is the preferred realization as already documented by previous impressionistic reports on Tyrolean ([2], [3]).

Despite the dominance of back variants in Tyrolean, as well as the amount of the uvular fricatives realized context-independently (appearing in both onset and coda position), our data show a high intra- and inter-speaker variability that needs to be taken into account for a better understanding of the /r/ allophony. To this end, we plan to further explore this issue of intra- and inter-speaker variability with the help of the information gathered from the questionnaire on the participant's social network that accounts for his/her use of Italian and Tyrolean Dialect.

Concerning the acoustic results, we examined the r-tokens according to their *constriction* phase. We only mentioned the insertion of vocalic fragments (*apertures*) independently from the /r/ variant, leaving this aspect to further investigation. We also showed that, with reference to their *constriction* phase, the various allophones are grouped according to different temporal cues and that these are partially dependent on the phonotactic context with the uvular fricative standing out as the most variable segment. Finally, the HNR values proved to be a good approach to distinguish among uvular approximants and voiced uvular fricatives.

Future work will include the addition of more speakers to the sample in order to better account for both intra- and inter-speaker variation. We plan to test and analyse more acoustic features and to compare our results with those that are becoming available for the same speakers from an articulatory investigation via the Ultrasound Tongue Imaging technique.

5. Acknowledgements

This study is financially supported by the Provincia Autonoma di Bolzano – Alto Adige, Ripartizione allo studio, Università e ricerca scientifica 2013-16 "The Articulatory Sociophonetics of Bilinguals in South-Tyrol: The Ultrasound Tongue Imaging Potential". The authors would like to thank all participants to the study for their effort and collaboration and three anonymous reviewers for their insightful comments on the paper.

6. References

[1] A. Mioni, “La standardizzazione fonetico-fonologica a Padova e a Bolzano (stile di lettura),” in *L'italiano regionale*, M. Cortelazzo and A. Mioni, Eds. Roma: Bulzoni, 1990, pp. 193–208.

[2] E. Kranzmayer, *Historische Lautgeographie des gesamtbairischen Dialektraumes*. Graz – Köln: Böhlaus, 1956.

[3] B. Alber, “Aspetti fonologici del mòcheno,” in *Introduzione alla linguistica del mòcheno*, E. Bidese and F. Cognola, Eds. Torino: Rosenberg & Sellier, 2013, pp. 15–35.

[4] R. Wiese, “The Unity and Variation of (german) /r/,” in *Zeitschrift für Dialektologie und Linguistik*, vol. 70, no. 1, H. Van de Velde and R. Van Hout, Eds. 2003, pp. 25–43.

[5] K. K. Klein and L. E. Schmitt, *Tirolischer Sprachatlas*. Tyrolia Verlag, 1969.

[6] L. Tonelli, *Regionale Umgangssprachen*. Unipress, 2002.

[7] C. Ulbrich and H. Ulbrich, “The Realisation of / r / in Swiss German and Austrian German,” in *ICPhS XVI*, 2007, no. August, pp. 1761–1764.

[8] L. Spreafico and A. Vietti, “/r/ in un parlante bilingue: un'indagine articolatoria basata su dati ecografici,” in *Proceedings of the 7° AISV AISV national conference*, 2011, pp. 248–259.

[9] A. Vietti, “Language contact and sociophonetic variation,” in *Sociophonetics, at the crossroads of speech variation, processing and communication*, 2012, pp. 76–79.

[10] A. Vietti, L. Spreafico, and V. Galatà, “An Ultrasound Study on the Phonetic Allophony of Tyrolean /R/,” in *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*, 2015, pp. 1–4.

[11] A. P. Simpson, “Phonetik und Phonologie des deutschen r,” in *Phonetische Datenbanken des Deutschen in der empirischen Sprachforschung und der phonologischen Theoriebildung*, AIPUK 33., K. J. Kohler, Ed. 1998, pp. 105–150.

[12] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. 1995, Wiley-Blackwell, 1995.

[13] M. Lindau, “The story of /r/,” in *Phonetic Linguistics, Essays in Honor of Peter Ladefoged*, V. Fromkin, Ed. Academic Press Inc., 1985, pp. 157–168.

[14] S. Jaworski and E. Gillian, “On the phonetic instability of the Polish rhotic /r/,” *Poznań Stud. Contemp. Linguist.*, vol. 47, no. 2, pp. 380–398, 2011.

[15] J. Schatz, *Wörterbuch der Tiroler Mundarten*. Innsbruck: Wagner, 1956.

[16] H. Haller and F. Lanthaler, *Passeier Wörterbuch. Wörter - Ausdrücke – Beispiele*. St. Martin in Passeier: Verlag Passeier, 2004.

[17] C. McCarthy, “EgoNet (Egocentric Network Study Software).” 2011.

[18] Articulate Instruments Ltd, “Articulate Assistant Advanced User Guide: Version 2.15.” Articulate Instruments Ltd, Edinburgh, UK, 2015.

[19] F. Schiel, “Automatic phonetic transcription of nonprompted speech,” in *Proceedings of ICPhS*, 1999, pp. 607–610.

[20] T. Kisler, F. Schiel, and H. Sloetjes, “Signal processing via web services: the use case WebMAUS,” in *Proceedings Digital Humanities 2012, Hamburg, Germany*, pp. 30–34, 2012.

[21] P. Boersma and D. Weenink, “Praat: doing phonetics by computer [Computer program],” *version*, vol. 5, no. 1, p. 32, 2010.

[22] S. Jannedy, “Rate effects on german unstressed syllables,” *OSU Work. Pap. Linguist.*, vol. 44, pp. 105–124, 1994.

[23] R Core Team, “R: A language and environment for statistical computing”, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/, 2016.

[24] D. Bates, M. Maechler, B. Bolker, S.Walker, “Fitting Linear Mixed-Effects Models Using lme4”, *Journal of Statistical Software*, 67(1), pp. 1-48, 2015.