# Different realizations of German underlying vs. derived diphthongs between two generations of Cantonese L1 speakers

Michelle Vuong, Marieke Einfeldt, Bettina Braun

Department of Linguistics, University of Konstanz, Germany
michelle.vuong@uni-konstanz.de, marieke.einfeldt@uni-konstanz.de, bettina.braun@uni-konstanz.de

## ABSTRACT

The theory of cross-linguistic influence predicts that languages influence each other, leading to phonetic realizations that differ between monolingual und bilingual speakers. This paper tested heritage speakers, examining the influence of Cantonese on the realization of two kinds of German diphthongs, underlying ([aɪ̯], [aʊ̯]) and derived ([iːɐ̯], [uːɐ̯], due to /ʁ/-vocalization), by first- and second-generation Cantonese speakers living in Germany (Mannheim area). A delayed imitation paradigm was used to reduce effects of orthography. We predicted that underlying diphthongs would result in smaller differences across generations than derived diphthongs. Formant tracks of first and second formants were analyzed using general additive mixed models. Results largely confirmed the predictions (significant differences between generations in F2 only for underlying diphthongs, but both for F1 and F2 for derived diphthongs). Second generation heritage language speakers were more similar to a German control group from the same region for some diphthongs.

**Keywords**: diphthongs, formants, Cantonese, German, general additive mixed models

## 1. INTRODUCTION

Heritage speakers (HSs) are early bilinguals who acquire two first languages (L1): the majority language (ML) of the society they live in and a minority language, also called heritage language (HL) [1]. HSs form a heterogenous group as they often differ from one another in terms of age of onset, proficiency, language use, etc. Most studies on HLs have investigated HSs' knowledge and performance in the HL [2–5]. Less research focused on the effects of the HL on the ML, and if it did, the ML was often English. This paper investigates the influence of Cantonese (HL or L1) on the ML German for Cantonese L1 speakers in Germany. While there are a number of studies on the influence of Cantonese on English (and vice versa) [5, 6, 7], most probably because of the language situation of Hong Kong (L1 Cantonese L2 English), the language combination Cantonese–German is largely underexplored.

To track the developmental trace of HSs, we compared two generations of Cantonese L1 speakers living in Germany, with a focus on diphthongs. Cantonese has a large inventory of vowels, with eight contrastive monophthongs and ten contrastive diphthongs (/ei̯, ai̯, aːi̯, ɔi̯, ui̯, au̯, aːu̯, iu̯, ou̯, œy̯/) [8]. The German inventory consists of fifteen contrastive monophthongs and three contrastive, closing, diphthongs (/aɪ̯, aʊ̯, ɔɪ̯/). We focus on two kinds of German diphthongs, those that are part of the phonological system of German (*underlying*, phonemic /aɪ̯/ and /aʊ̯/, which are more frequent than /ɔɪ̯/, cf. [9]) and those with a reversed vowel quality that are derived by a phonological rule of /ʁ/-vocalization. This rule changes underlying "r" in coda-position to [ɐ̯], e.g. "*vier*" [fiːɐ̯] 'four', which is diphthongized with the preceding vowel in many German dialects. We call this class *derived* diphthongs and test [iːɐ̯] and [uːɐ̯].

We hypothesize that the two generations of Cantonese L1 speakers realize the underlying diphthongs more similar to each other than the derived diphthongs. This hypothesis is based on two considerations: First, the closing diphthongs occurring in the underlying category (in which the second vowel target has a higher tongue position than the first, e.g. [aɪ̯]) are universally less marked than the opening diphthongs in the derived category (e.g. [iːɐ̯]). English and Cantonese, for instance, lack opening diphthongs. Research on the role of markedness in cross-linguistic influence has shown that markedness can affect rate and difficulty of acquisition: Generally, unmarked structures are acquired easier and faster (leading to positive cross-linguistic influence) than marked structures that tend to not be transferred at all [10]. Second, underlying diphthongs are represented in German orthography ([aɪ̯]: "ei", [aʊ̯]: "au", [ɔɪ̯]: "eu"), while derived diphthongs are not, which may influence language acquisition [11].

We analyzed the differences in the realization of the two kinds of German diphthongs between the two generations of Cantonese speakers using general additive mixed models (GAMMs). GAMMs allow a direct comparison between formant contours since they can model non-linear dependencies over time using smooth functions [12-17].

## 2. EXPERIMENT

We collected productions using a delayed imitation paradigm with a 2500ms interval between stimulus and imitation. After this delay, the phonetic trace has decayed, and participants need to access phonological representations [18, 19].

### 2.2. Methods

### *2.2.1 Participants*

Thirty L1 speakers of Cantonese in Germany participated in the experiment. They were divided into four categories, split by gender and generation (Gen 1, Gen 2), see Table 1. Gen 1 included L1 speakers of Cantonese who were born in Vietnam (N = 11) or Hong Kong (N = 1); they all immigrated to Germany around puberty (mean age of immigration = 16.0, SD = 3.4 years). Gen 2 are HSs of Cantonese who were all born and raised in Germany by the Gen 1 participants. For them, German is the ML.

Gen 1 consists of six female (mean age = 54.0, SD = 2.1 years) and six male speakers (mean age = 58.0, SD = 2.4 years), while Gen 2 consists of seven female (mean age = 26.0, SD = 4.2 years) and eleven male speakers (mean age = 23.0, SD = 4.0 years). The majority of Gen 1 and Gen 2 participants were from Rhineland-Palatinate (N = 25). The remaining participants were from Baden-Württemberg (N = 3), Hesse (N = 1) and North Rhine-Westphalia (N = 1). The areas the participants came from all possess the process of coda /ʁ/-vocalization.

Each of the four speaker groups was matched with a baseline of two native speakers of German (GER 1, GER 2), who did not acquire a second language before the age of six and grew up in the same area, see Table 1. The GER 1 speakers had an average age of 54.3 years, SD = 1.9, the GER 2 speakers an average of 21.5 years, SD = 2.1. We could not control the educational status between Gen 1 and Gen 2 (higher degrees for Gen 2 than for Gen 1).

| Group | Female | Male | **Total** |
|---|---|---|---|
| Gen 1 | 6 | 6 | **12** |
| Gen 2 | 7 | 11 | **18** |
| GER 1 (control) | 2 | 2 | **4** |
| GER 2 (control) | 2 | 2 | **4** |
| **Total** | **17** | **21** | **38** |

**Table 1**: Distribution of participants across groups.

### *2.2.2 Material*

A total of 24 items were constructed, half containing the underlying diphthongs [aɪ̯] and [aʊ̯], half the derived diphthongs [iːɐ̯] and [uːɐ̯]. For each of the four diphthongs, six non-words (e.g. [ˈpaɪ̯pɐ]) were constructed and embedded in the same carrier sentence: *Er hat Peiper gesagt* (“He said Peiper”). The non-words were constructed to fit German and Cantonese phonotactics. Non-words were chosen over words to avoid effects of familiarity and lexical frequency, which may influence productions [17, 20]. The diphthongs were flanked by voiceless plosives (/p, t, k/) to ease segmentation.

The stimuli for imitation were recorded by a male speaker of GER 1. He read each of the 24 test sentences out loud three times. The production with the clearest pronunciation was chosen. Care was taken to have similar intonation contours across items (L+H* on target word, followed by low boundary tone L-%).

### *2.2.3 Procedure*

The only task for participants from GER 1 and 2 was reading the list of target sentences out loud, recorded directly onto an Acer Switch Alpha 12 computer using an AT2020USB+ microphone (16 Bit, 44.1 kHz). Participants from Gen 1 and Gen 2 first filled in a background questionnaire programmed via *SoSci Survey* [21], ran on an in-house server. They were asked for their age, year of immigration to Germany (if applicable), place of residence, and education. They were then instructed to rate their proficiency in Standard High German, the Palatinate (or other) dialect(s), Cantonese and Vietnamese. Lastly, the participants were asked which languages they speak at home and outside home and to what proportion. They could take part in a lottery for reimbursement.

The main part of the study was the delayed imitation experiment programmed using the software ***Presentation®*** (Neurobehavioral Systems), installed on an HP NB ProBook 650 G2. The majority of participants (N = 26) was tested in a quiet environment in a private setting in presence of the first author. Participants’ imitations were recorded using an AT2020USB+ microphone. Four participants took part remotely as they were not available for testing on-site. In those cases, a video of the experiment was screenshared and imitations were recorded on smartphone microphones. This process was monitored remotely by the first author. Participants were instructed to imitate the utterance in their usual voice. Four practice trials familiarized the participants with the task. The main experiment consisted of 24 trials. Each trial started with a 500ms sine tone played at 300Hz, followed by 1500ms silence. Next, the sentence was played three times with 1500ms silence between each time. This was followed by 2000ms silence and a 500ms sine tone played randomly at either 150Hz or 450Hz to avoid predictability and to overwrite acoustic traces, as in [22, 23]. Participants then pressed “A” on the laptop keyboard to record themselves and “L” to stop the recording. After stopping the recording, participants were automatically directed to the next trial and the recording was saved onto the laptop (16 Bit, 44.1 kHz). The order of trials was randomized.

### *2.2.4 Data Treatment and Statistical Analysis*

First, all 912 sound files (38 participants x 24 imitations) were annotated semi-automatically using *Web-MAUS* [24]. Next, diphthong boundaries were manually corrected using a wide-band spectrogram and following standard segmentation criteria [25]. We used *emuR* [26] to create a database combining all annotated items and merged the database with the participants’ meta data. The function get_trackdata() extracted the formants (separately for male and female participants); these were time-normalized to get 10 formant points per diphthong. The formants were converted to bark [27].

The basic model included a smooth over normalized time and smooth terms for participants and items [28, 29]. Next, we included smooth terms for the factors *gender* and *generation* and kept them if this improved the model fit. Model comparison was done using the function compareML() [13, 15, 30]. Since there is a strong autocorrelation between subsequent formant points, we used the autocorrelation parameter *rho* to correct for this. *Rho* was determined by the acf_resid()-function [30]. Models were checked using gam.check() to see if the number of knots was appropriate and the residuals normally distributed. All F2 models were re-run with the scaled-t-distribution (family=”scat”), because the default gaussian distribution resulted in a highly tailed distribution of residuals [14]. Difference plots (across generations) are used to determine and visualize the parts of the diphthong in which significant differences occurred.

### 2.3 Results

The smooth terms for gender and generation always improved the model fit. Model comparisons showed no significant improvement for adding an interaction smooth between generation and gender. Figure 1 shows the averaged difference curve (solid line) for F1 and F2 for the **underlying diphthongs** ([aɪ̯] top panels, [aʊ̯] bottom panels) between Gen 1 and Gen 2. The difference curve shows the subtraction of the Gen 1 contour from the Gen 2 contour, revealing the parts in which they significantly differ from each other (areas in which the gray shading of the 95% Confidence Interval (CI) excludes 0). These areas are additionally marked by the vertical red lines. F1 of [aɪ̯] and [aʊ̯] showed no significant differences, while F2 of [aɪ̯] is significantly different from 0.17 to 0.28 and from 0.61 to 1 of the normalized time. Main significant differences in F2 of [aʊ̯] are from 0 to 0.08 and from 0.66 to 0.85 normalized time.

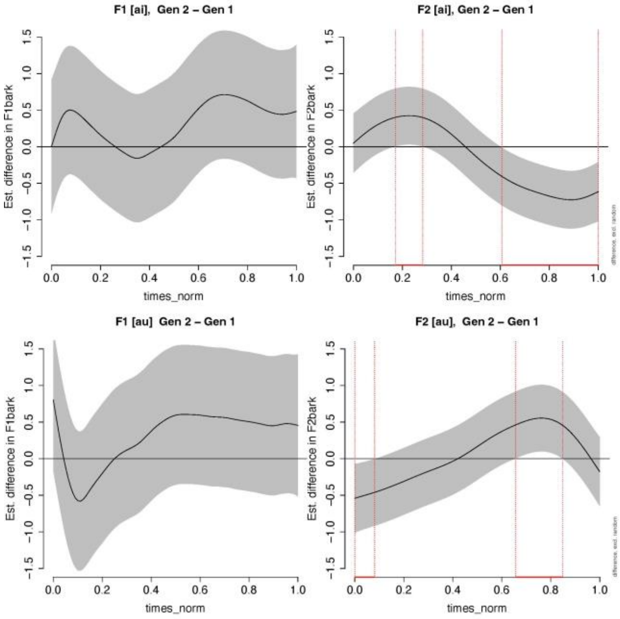

**Figure 1**: Estimated differences between Gen 2 and Gen 1 for F1 and F2 of [aɪ̯] (top panels) and [aʊ̯] (bottom panels). Red areas show significant differences.

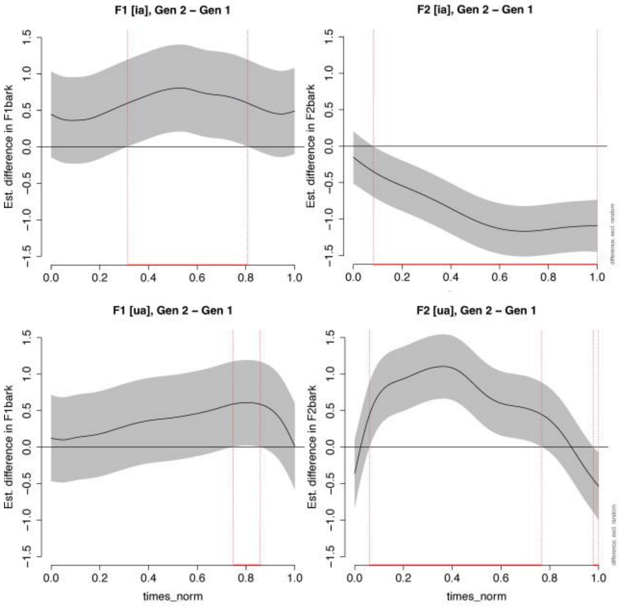

**Figure 2**: Estimated differences between Gen 2 and Gen 1 for F1 and F2 of [iːɐ̯] (top panels) and [uːɐ̯] (bottom panels). Red areas show significant differences.

Figure 2 displays the averaged difference curve between Gen 1 and Gen 2 for F1 and F2 of the **derived diphthongs** ([iːɐ̯] top panels, [uːɐ̯] bottom panels). [iːɐ̯] and [uːɐ̯] show significant differences in both F1 and F2: from 0.31 to 0.81 for F1 and 0.08 to 1 for F2 of [iːɐ̯] and from 0.75 to 0.86 for F1 and from 0.06 to 0.77 and 0.98 to 1 for F2 of [uːɐ̯]. Comparison between Figures 1 and 2 suggest that F2 differences between generations are considerably larger in derived than in underlying diphthongs.

Figure 3 compares the L1 Cantonese speakers to the monolingual German control group (GER, merging data from GER 1 and GER 2). This comparison may be informative to investigate whether the Gen 2 speakers are closer to the German speakers than Gen 1 speakers (due to the small sample of GER speakers, a statistical analysis with GAMMs could not be conducted). Squares with dotted lines highlight formant trajectories, for which Gen 1 speakers were closer to the German controls than Gen 2 speakers (n = 3), squares with dashed lines highlight formant trajectories for which Gen 2 speakers are closer to German controls (n = 5). For the other formant trajectories, no pattern could be discerned. This exploratory analysis suggests that Gen 2 speakers were closer to the monolingual Germans than Gen 1.

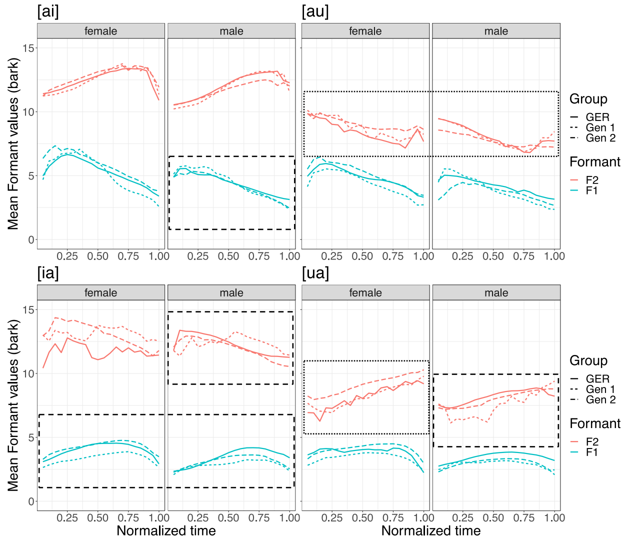

**Figure 3:** F1 and F2 contours of Gen 1, Gen 2 and GER for female and male for [aɪ̯] (top left panel), [aʊ̯] (top right), [iːɐ̯] (bottom left) and [uːɐ̯] (bottom right).

## 3. DISCUSSION

We examined inter-generational differences between Cantonese L1 speakers of German in their realizations of German underlying ([aɪ̯], [aʊ̯]) vs. derived ([iːɐ̯], [uːɐ̯]) diphthongs. We found significant differences between generations for both kinds of diphthongs in the realization of F2 (which mostly represents the horizontal tongue position, front-back). Moreover, there were fewer differences for the underlying diphthongs (differences only in F2 trajectory) than for the derived diphthongs (differences in both F1 and F2 trajectories). Furthermore, differences in F2 were larger and spanned a longer time interval of the diphthong for the derived compared to the underlying diphthongs. Comparison between Gen 1, Gen 2 and the GER baseline showed that Gen 2 was more similar to GER than Gen 1 in more cases, but there were too little data for statistical comparison.

Our findings support the hypothesis that the two generations of L1 Cantonese speakers in Germany realize the underlying diphthongs more similar to each other than the derived ones. This hypothesis was based on the fact that the closing diphthongs ([aɪ̯], [aʊ̯]) are less marked than the opening diphthongs ([iːɐ̯], [uːɐ̯]). Moreover, the German derived opening diphthongs [iːɐ̯], [uːɐ̯] are generally marked and Cantonese does not have opening diphthongs at all (and does not possess /ʁ/-vocalization). At the same time, the German underlying, closing diphthongs [aɪ̯], [aʊ̯] may be similar to the Cantonese diphthongs /ai̯, aːi̯, au̯, aːu̯/, possibly leading to positive influence from Cantonese to German.

The research hypothesis was also based on the fact that underlying diphthongs are represented in German orthography ([aɪ̯]: “ei”, [aʊ̯]: “au”), whereas the grapheme-phoneme correspondence of derived diphthongs is more opaque (e.g. [iːɐ̯]: “ier”, “ir”, “ihr”). Research on the effect of orthography on L2 production has shown that L2 orthographic forms influence the acquisition of L2 phonology. For instance, L2 speakers of American English pronounce the flap /ɾ/ as [t] when spelled with <t> (*beauty*) and as [d] when spelled with <d> (*lady*) [31].

From a perception perspective, many of the derived diphthongs ([iːɐ̯], [uːɐ̯]) by Gen 1 participants sounded like monophthongs ([i] and [u] respectively). This observation during annotation would need to be confirmed in a proper perception experiment. This would yield a more complete picture for the question how foreign language learners of German (here Gen 1) internalize and process /ʁ/-vocalization. Conceivably, differences in acquisition may also be expected for other phonological rules which might have implycations for acquisition theories and teaching.

## 4. CONCLUSION

Two generations of Cantonese L1 speakers of German were tested on the realization of two kinds of German diphthongs (underlying vs. derived). Results showed fewer and smaller differences between generations for the underlying, closing diphthongs ([aɪ̯], [aʊ̯]) than for the derived, opening diphthongs.

## 5. ACKNOWLEDGEMENTS

We thank Oleksiy Bobrov for programming the experiment and Bernhard Brehmer for advice and literature suggestions on heritage languages.

## 6. REFERENCES

[1] Kan, R. T. Y. 2021. Phonological Production in Young Speakers of Cantonese as a Heritage Language. *Language and speech* 64, 73-97.

[2] Kupisch, T. et al. 2014. Foreign Accent in Adult Simultaneous Bilinguals. *Heritage language journal* 11, 123–150.

[3] Lloyd-Smith, A., Einfeldt, M., Kupisch, T. 2020. Italian-German bilinguals: The effects of heritage language use on accent in early-acquired languages. *International Journal of Bilingualism* 24, 289–304.

[4] Cummings Ruiz, L. D. 2019. North Midland /u/-fronting and Its Effects on Heritage Speakers of Spanish. *Proc. 19th ICPhS* Melbourne, 1099–1103.

[5] Tse, H. 2016. Variation and change in Toronto heritage Cantonese. *Asia-Pacific Language Variation* 2, 124–156.

[6] Tse, H. 2019. *Beyond the Monolingual Core and out into the Wild: A Variationist Study of Early Bilingualism and Sound Change in Toronto Heritage Cantonese*. Dissertation, University of Pittsburgh, Pittsburgh.

[7] Soo, R., Monahan, P. J. 2022. Phonetic and Lexical Encoding of Tone in Cantonese Heritage Speakers. *Language and speech*.

[8] Chan, A. Y. W., Li, D. C. S. 2000. English and Cantonese Phonology in Contrast: Explaining Cantonese ESL Learners' English Pronunciation Problems. *Language, Culture and Curriculum* 13, 67–85.

[9] Pätzold, M., Simpson, A. P. 1997. *Acoustic analysis of German vowels in the Kiel Corpus of Read Speech*. University of Kiel.

[10] Jarvis, S., Pavlenko, A. 2010. *Crosslinguistic influence in language and cognition*. London, New York: Routledge.

[11] Mitterer, H. 2021. *The role of orthography in learning a second language: Evidence from Maltese English*. Malta Chamber of Scientists.

[12] Baayen, R. H., van Rij, J., de Cat, C., Wood, S. 2018. Autocorrelated Errors in Experimental Data in the Language Sciences: Some Solutions Offered by Generalized Additive Mixed Models. In: D. Speelman, K. Heylen, D. Geeraerts (eds.) *Mixed-Effects Regression Models in Linguistics*, Springer International Publishing, 49–69.

[13] Wieling, M. 2018. Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics* 70, 86–116.

[14] van Rij, J. , Hendriks, P., van Rijn, H., Baayen, R. H., Wood, S. N. 2019. Analyzing the Time Course of Pupillometric Data. *Trends in hearing* 23.

[15] Sóskuthy, M. 2021. Evaluating generalised additive mixed modelling strategies for dynamic speech analysis. *Journal of Phonetics* 84.

[16] Voeten, C. C. 2021. Regional variation in ongoing sound change: The case of the Dutch diphthongs. *Journal of Linguistic Geography* 9, 162–177.

[17] Beaman, K. V., Tomaschek, F. 2021. Loss of Historical Phonetic Contrast across the Lifespan. in Buchstaller, I., Beaman, K. V. (eds.) *Language Variation and Language Change Across the Lifespan*, Routledge, 209–234.

[18] Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

[19] Baddeley, A. D. (2003). Working memory and language: An overview. *J. Commun. Disord.* 36, 189–208. doi: 10.1016/S0021-9924(03)00019-4.

[20] Hay, J., Jannedy, S., Mendoza-Denton, N. 1999. Oprah and /ay/: Lexical frequency, referee design and style. *Proc. 14th ICPhS* San Francisco, 1389–1392.

[21] Leiner, D. J. 2019. *SoSci Survey* (Version 3.2.55 [Computer software] Available at https:// www.soscisurvey.de.

[22] Zahner-Ritter, K., Einfeldt, M., Wochner, D., James, A., Dehé, N., Braun, B. 2022. Three Kinds of Rising-Falling Contours in German wh-Questions: Evidence From Form and Function. *Frontiers in Communication* 7.

[23] Zahner-Ritter, K., Zhao, T., Einfeldt, M. Braun, B. 2022. How experience with tone in the native language affects the L2 acquisition of pitch accents. *Frontiers in psychology* 13.

[24] Kisler, T., Reichel, U., Schiel, F. 2017. Multilingual processing of speech via web services. *Computer speech & language* 45, 326–347.

[25] Turk, A., Nakai, S., Sugahara, M. 2006. Acoustic Segment Durations in Prosodic Research: A Practical Guide. In: *Methods in empirical prosody research*, New York, Berlin, 1–28.

[26] Winkelmann, R., Jaensch, K., Cassidy, S., Harrington, J. 2021. *emuR: Main Package of the EMU Speech Database Management System*. R package version 2.3.0.

[27] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88, 97–100, 1990.

[28] Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 3–36.

[29] Wood, S. N. 2017. *Generalized additive models: an introduction with R*, 2nd ed. London, New York, Boca Raton: CRC Press, Taylor & Francis Group.

[30] van Rij, J., Wieling, M., Baayen, R., van Rijn, H. 2022. *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs*. R package version 2.4.1.

[31] Vokic, G. 2011. When alphabets collide: Alphabetic first-language speakers' approach to speech production in an alphabetic second language. *Second language research* 27, 391–417.