

## RESEARCH ARTICLE

# Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates

Theresa Matzinger<sup>1,2\*</sup>, Nikolaus Ritt<sup>1</sup>, W. Tecumseh Fitch<sup>2\*</sup>**1** Department of English, University of Vienna, Vienna, Austria, **2** Department of Cognitive Biology, University of Vienna, Vienna, Austria\* [theresa.matzinger@univie.ac.at](mailto:theresa.matzinger@univie.ac.at) (TM); [tecumseh.fitch@univie.ac.at](mailto:tecumseh.fitch@univie.ac.at) (WTF)**OPEN ACCESS**

**Citation:** Matzinger T, Ritt N, Fitch WT (2020) Non-native speaker pause patterns closely correspond to those of native speakers at different speech rates. PLoS ONE 15(4): e0230710. <https://doi.org/10.1371/journal.pone.0230710>

**Editor:** Koji Miwa, Nagoya University, JAPAN

**Received:** July 19, 2019

**Accepted:** March 6, 2020

**Published:** April 3, 2020

**Copyright:** © 2020 Matzinger et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data as well as the Praat scripts and R scripts used for data processing and analysis can be accessed at the Open Science Framework: DOI [10.17605/OSF.IO/JC7EF](https://doi.org/10.17605/OSF.IO/JC7EF).

**Funding:** Preparation of this paper was supported by Austrian Science Fund (FWF; <https://www.fwf.ac.at>) DK Grant "Cognition & Communication" (#W1262-B29). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Abstract

When speaking a foreign language, non-native speakers can typically be readily identified by their accents. But which aspects of the speech signal determine such accents? Speech pauses occur in all languages but may nonetheless vary in different languages with regard to their duration, number or positions in the speech stream, and therefore are one potential contributor to foreign speech production. The aim of this study was therefore to investigate whether non-native speakers pause 'with a foreign accent'. We recorded native English speakers and non-native speakers of German or Serbo-Croatian with excellent English reading out an English text at three different speech rates, and analyzed their vocal output in terms of number, duration and location of pauses. Overall, all non-native speakers were identified by native raters as having non-native accents, but native and non-native speakers made pauses that were similarly long, and had similar ratios of pause time compared to total speaking time. Furthermore, all speakers changed their pausing behavior similarly at different speech rates. The only clear difference between native and non-native speakers was that the latter made more pauses than the native speakers. Thus, overall, pause patterns contributed little to the acoustic characteristics of speakers' non-native accents, when reading aloud. Non-native pause patterns might be acquired more easily than other aspects of pronunciation because pauses are perceptually salient and producing pauses is easy. Alternatively, general cognitive processing mechanisms such as attention, planning or memory may constrain pausing behavior, allowing speakers to transfer their native pause patterns to a second language without significant deviation. We conclude that pauses make a relatively minor contribution to the acoustic characteristics of non-native accents.

## Introduction

When speaking a foreign language, most non-native speakers can be readily identified by their accents, which are often distinctive depending on native language. To understand how different accents arise, and also to help second language learners to eliminate them to improve

intelligibility, it is essential to know which aspects of spoken language contribute to non-native speech production.

When studying foreign accents, it is crucial to distinguish between accents in production, i.e. atypical measurable acoustic characteristics of the speech signal, and accents in perception, i.e. listeners' identification of accents as non-native [1–3]. Although these factors are probably related, they are not identical, and the focus of the current study is solely on accents in production.

A crucial factor generating distinct foreign accents is that linguistic phenomena vary between languages, and speakers transfer language-specific features from one language to another (e.g. [4]). These language-typical features can range from pronunciation of individual phonemes (e.g. [5,6]) to suprasegmental prosodic features like intonation patterns (e.g. [7,8]). For example, native speakers of Japanese have difficulty distinguishing the phonemes /l/ and /r/ in foreign languages because Japanese has only one similar phoneme (e.g. [9]). Native German learners of English often substitute /d/ or /s/ for /ð/ or /θ/ (as in *this* or *thing*) because the latter two English sounds are not part of the German phoneme inventory [10]. German speakers also typically use a different pitch range than English, making native German speakers of English sound “bored” to English native speakers, whereas native English speakers often sound “over-excited” to German native speakers [11–13]. Such examples could be readily multiplied.

Certain linguistic features seem to contribute less to foreign speech production than others. These features include linguistic phenomena claimed to be language-universal [14] that are transferred to (or acquired easily in) a second language (L2). Such language-universal phenomena are rare, and have been suggested to mainly concern pragmatic, rather than phonological or grammatical, features [15–18].

The current study investigates whether L1 and L2 speakers differ in their pausing behavior when reading aloud, and hence whether different pausing patterns contribute to a ‘foreign accent’ in speech production. Pauses are particularly interesting candidates to investigate non-native accents for several reasons. Due to the natural need to breathe, pauses occur in all of the world’s languages. Still, not all pauses are determined by physiological needs, but have functions closely related to cognition: pauses help structure speech, determine speech tempo and rhythm, plan upcoming utterances, can add rhetorical emphasis, or structure turn-taking [19–23]. Such roles, reflected for example by the durations of pauses or by the positions of pauses in a stretch of speech, may be completely physiologically determined or may be realized similarly in different languages (e.g. [24–26,27] see S1 Table) and thus either transfer without change, or be easy to acquire, and thus not contribute to non-native speech production.

On the other hand, the different functions of pauses potentially make them subject to cross-linguistic variation (see S2 Table). For example, there is evidence that English speakers pause more frequently than French [28,29] or Turkish speakers [30], but less frequently than Spanish speakers [31]. Also, English speakers’ pauses are shorter than French speakers’ [28,32, but also 33] and Russian speakers’ [34], but longer than Italian speakers’ pauses [33]. Such cross-linguistic differences in pause patterns, if carried over to a foreign language, might contribute to non-native accents in speech production. The main aim of this study is to investigate these possibilities, to determine whether people speaking a non-native language pause with a foreign accent when reading aloud.

Previous studies on pausing behavior in non-native languages mostly concern fluency rather than foreign accent (e.g. [32,34–38]). Although a lack of fluency can contribute to recognizing speakers as non-native, fluency should be distinguished from accents, since proficient second language speakers might be highly fluent, but still possess a clear foreign accent [3,39–41]. Nonetheless, some tentative predictions can be drawn from fluency studies (see S3 Table): speakers tend to make more [32,42–47] and longer [32,46,47] (but also [32,42,43]) pauses when speaking their second language (L2) relative to their first language (L1). However, some

reports find associations between speakers' L1 and L2 pausing behavior. When comparing L2 speakers to L1 speakers of the target language, it seems that highly proficient speakers adhere more closely to native-like pause patterns, whereas less proficient speakers make more and longer pauses than L1 speakers (L1 Korean, L2 English 8, L1 Russian, L2 English 34; see [S4 Table](#)). In addition, L1 and L2 pause frequency are correlated: while, overall, speakers make more pauses in their L2 than in their L1, speakers' L2 pausing behavior can be predicted if their L1 pausing behavior is known [47]. Finally, studies on the perception of fluency and accent can lead to conclusions about how pauses influence accents. For example, acoustic measures of fluency (including pause incidence and duration) have been shown to be predictors of accent ratings [3]. Still, perception of foreign accent was only weakly correlated with acoustic fluency measures, which is why, overall, accentedness and fluency can be regarded as two separate, partially independent concepts [3].

An important methodological issue implies that these previous studies cannot be taken as a clear evidence that pauses contribute to foreign accents in production: an intrinsic lack of stimulus control during free speech. Because pauses serve multiple functions, they can be influenced by many different variables, such as speech genre, cognitive load, or syntactic complexity of the utterances [34,48]. These diverse influencing factors make it a challenge to tease out the contributions of pauses to foreign accents in spontaneous speech (e.g., picture description tasks or spontaneous monologues). Differences in pause patterns may arise not due to foreign transference, but due to different sentence structures employed or speech styles adopted by the speakers, which cannot be controlled in spontaneous speech. Additional factors such as communicative intent, personal speaking style, or emotional involvement with the speech task [34,47–49] are equally difficult to control for. Most previous studies on second language pausing controlled for some of these potential factors but, due to their focus on second language fluency, did not control enough factors to reliably attribute foreign accents to pause differences.

Here we introduce an experimental procedure which integrates and modifies methods from previous studies, safeguarding against potential confounds, to focus on the specific contribution of pauses to foreign accents in speech production. We compared the pausing behavior of L2 speakers of English to the pausing behavior of L1 speakers of English, reading out the same scripted text at different speech rates. We measured multiple characteristics of pauses: pause-to-utterance ratio (total pause time in relation to total speaking time), the mean duration of pauses, the number of pauses that speakers made, and the positions in the written text at which they occurred.

By having all speakers read the same written text, we could exclude the influence of morpho-syntactic factors (e.g. word length, word structure, and syntactic structure) that might underlie previous findings of language differences in pausing. Also, we reasoned that the cognitive load involved in reading a text is reduced compared to free speech. Thus, by using written text as a prompt, we aimed to limit the occurrence of vocal hesitations resulting from a lack of L2 fluency to distinguish fluency from foreign accent per se.

To examine whether unusual reading conditions affect pausing behavior, we investigated speakers' pausing performance at three different reading speeds (casual, slow and fast speech rate), aiming to disentangle the role of cognitive load on the realization of pauses. We reasoned that the cognitive load should be lowest in casual reading speed because speakers encounter this speech tempo frequently in daily life and therefore have more possibilities to acquire native-like pause patterns [50]. Accents might preferentially surface in unnatural reading conditions, particularly in speeded reading where the cognitive load is higher and speakers might fall back on cognitive mechanisms developed for their native language. Also, testing the same individuals at different reading speeds offers a within-individual manipulation, helping to address individual differences in speaking and/or pausing styles.

We tested non-native speakers with two different L1 backgrounds, namely German or Serbo-Croatian L1. Typologically, both English and German belong to the Germanic language family, and are stress-based languages [51], whereas Serbo-Croatian is a Slavic language [52] and is not stress-based [25]. This selection of L1 speakers thus would allow comparison of pausing by English native speakers with those of L2 speakers with an L1 background more (German) or less (Serbo-Croatian) similar to English. However, this comparison was not included in the final analysis due to a low sample size of Serbo-Croatian speakers.

Summarizing, we used a standardized procedure to compare pause patterns (pause-to-utterance ratio, pause duration, pause number, pause positions) of native speakers of English and two non-native English speaker groups reading out the same text at three different speech rates. Previous work leads to several hypotheses and predictions.

According to the *No Contribution* hypothesis, pauses do not contribute to non-native speech production. This is predicted if the acquisition of pause patterns is simple during language acquisition, because pauses are perceptually highly salient (Matzinger, Ritt, Fitch, in prep) and pausing is articulatorily simple (compared to the articulation of other vocal elements like vowels, consonants, or intonation). Alternatively, L2-typical pause patterns might result from similar pause patterns in the speakers' L1 being transferred to their L2, or even reflect language-universal pausing behavior. By the *No Contribution* hypothesis, accents are caused by non-native realizations of linguistic features other than pauses. These non-native realizations might for example concern phonemes, word stress patterns or prosody [2,39]. Besides that, non-nativeness might be signaled by atypical gestures or turn-taking behavior. The *No Contribution* hypothesis predicts no differences in the pausing behavior of native and non-native speakers of English: speakers should pause at the same syntactic positions, equally often and for similar durations as native speakers. If pauses can be acquired easily or are language-independent, this hypothesis also predicts no interactions between nativeness and reading speed: L2 speakers should perform similarly to L1 speakers whether reading fast or slowly.

Alternatively, the *Pause Contribution* hypothesis postulates that pauses contribute to foreign accents. This might be due to a higher cognitive load (e.g. processing or memory constraints [18,53]) when speaking an L2 [8,54,55]. Alternatively, differences in the typical pausing behavior of speakers' L1 and L2 may make the typical pause characteristics of the L2 difficult to attain, because speakers' native pausing pattern overrides the learned non-native pattern. The *Pause Contribution* hypothesis predicts that pause patterns of L2 speakers will differ significantly from pause patterns of L1 speakers. If differences result from a higher cognitive load when speaking the L2, there should be more and longer pauses in L2 speakers than in L1 speakers. Also, this predicts an interaction between nativeness and reading speed. Differences between L1 and L2 pausing should be smaller in casual reading speed than in unnatural reading conditions (i.e. fast or slow speech), which pose a higher cognitive load because they are encountered and practiced less frequently. L2 speakers might therefore have had fewer chances to acquire them in a native-like manner and might fall back on non-native patterns instead. Differences are predicted to be higher in fast than in slow reading aloud, because reading rapidly should be more cognitively challenging than reading slowly.

## Materials and methods

### Target languages and participants

We obtained speech samples from 41 participants of three different first languages: English native speakers (13 participants; 7f; mean age: 35.2) and non-native English speakers with German (18 participants; 10f; mean age: 29.5) and Serbo-Croatian (10 participants; 6f; mean age: 25.9) as their first languages. Participants were university students or staff recruited

individually at the University of Vienna. All non-native English participants were advanced learners of English, who did not have diagnosed reading or speaking difficulties, self-assessed themselves as being proficient in English (equivalent to CEFR level C1) and reported in post-experiment questionnaires (see [S1 Appendix](#)) to be concerned with English regularly both in the productive and receptive domain (e.g. in the university or work context, media exposure).

Nonetheless, and crucially, in a native language recognition test with our pool of speech samples, five English native speakers (m, mean age: 41.2) could still detect all non-native speakers due to their distinctive accents. This native language recognition test ensured that our non-native speakers qualified for the study: being identified as non-native in the accent recognition test suggests that certain features of native and non-native speech production differ. These might potentially include deviations in pause patterns.

The language recognition test was implemented using the software package PRAAT (Version 6.0.36, [56]). Raters listened to a speech sample of each participant reading out the target text (see below) in casual speech tempo. The task of the raters was to indicate if they believed the speakers to be native speakers of English, German or Serbo-Croatian. The raters controlled the timing, moving to the next speech sample as soon as they were sufficiently certain about their decision.

Although four German native speakers were misclassified as English native speakers by one or two of the raters each, the other raters correctly identified them as non-native. Also, one Serbo-Croatian native speaker was misclassified as an English native speaker by one of the raters, but correctly recognized as non-native by all other raters. The five raters correctly recognized all English native speakers, except that four English native speakers were classified as German L2 speakers by one rater each, and one of these English native speakers was additionally classified as a Serbo-Croatian L2 speaker by a second rater (see [S5 Table](#)). We concluded from these ratings that our speakers qualified for the analysis.

The study protocol was approved by the ethics board of the University of Vienna (reference number: #00333/00384). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

## Speech recordings

Speech samples were collected by recording participants reading out the English prose text *The boy who cried wolf* (see [S2 Appendix](#)), a fable frequently used for evaluating English pronunciation [57]. The recordings were made with a ZOOM Handy Recorder (H4n, ZOOM Corporation, Japan) either in a sound-proof room or in a quiet office environment. We recorded the participants reading out the text in three different speech tempi: fast, casually and slowly. Participants were instructed to read the text casually in the *casual* condition, to read as fast as they could in the *fast* condition and to read the text slowly (e.g., as if to a group of preschool children), in the *slow* condition. The order of the different tempo conditions was randomized for each participant. To elicit a natural reading style and minimize pauses resulting from hesitation due to unfamiliarity, participants were asked to read the text silently before recording in order to familiarize themselves with the text to avoid them being distracted by unfamiliar words or content during recording. In addition, before the actual recording, participants read the first sentence of the text aloud to make them comfortable reading aloud in the experimental setting, while the experimenter adjusted the signal recording levels.

## Measurements and analyses

For determining pauses in the recordings of participants reading aloud, a pause was defined as a period of silence with a minimal duration of 0.1 seconds, most likely occurring for breathing, rhythmic or pragmatic reasons (the choice of this threshold is explained in more detail in [Box 1](#)).

### Box 1. What is a pause?

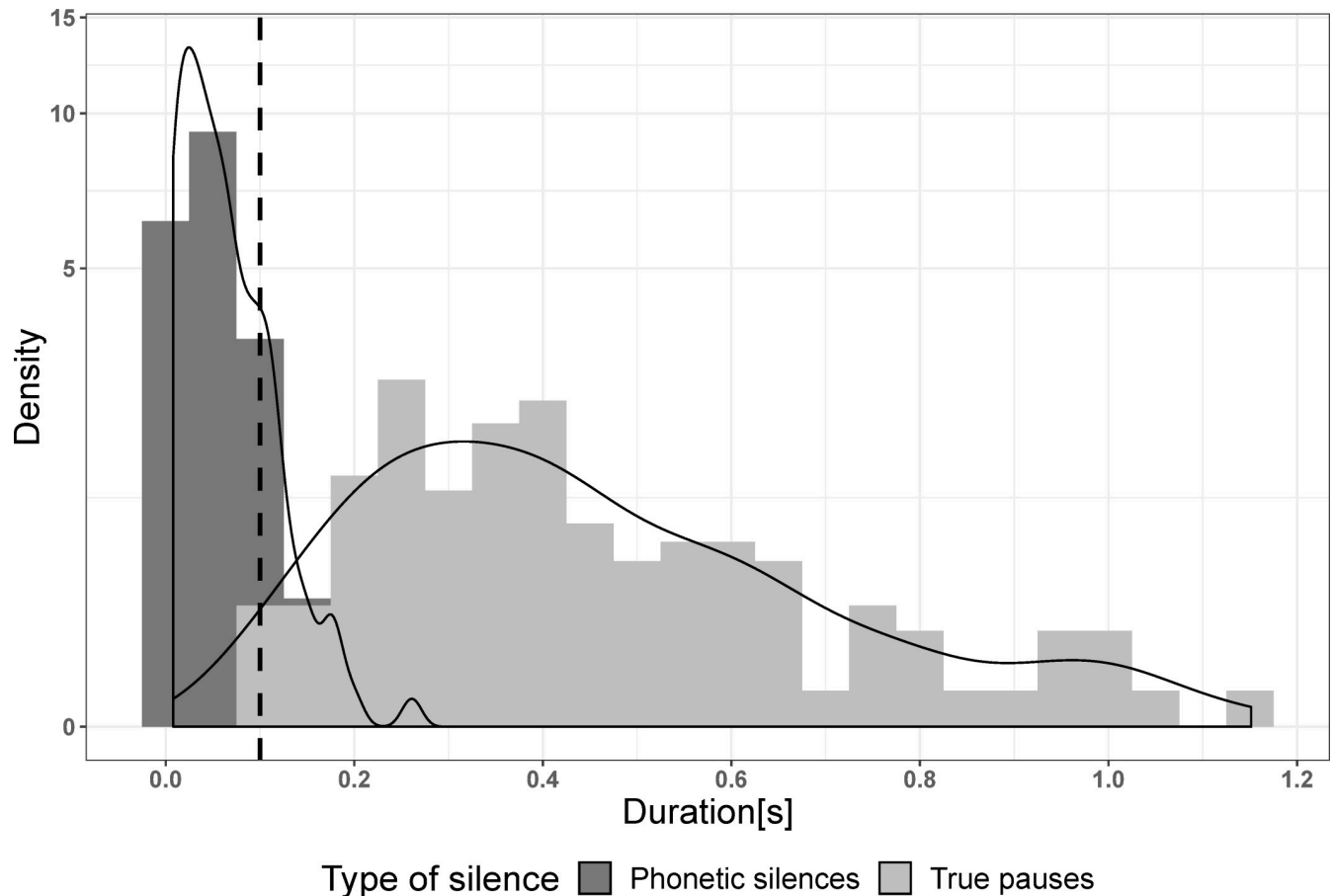
Previous studies differ considerably in what they consider the lower durational threshold for classifying silent intervals in speech recordings as pauses (reviewed in [37,80]). These threshold values start as low as 5 ms (e.g. [25]) and range up to values as high as 400 ms [45,53,81]. 200 ms is a popular threshold for pauses in L2 speech [37,80]. The choice of a particular lower durational threshold is often determined by the type of pauses investigated in a particular study. For example, studies concerned with pauses resulting from a lack of fluency in an L2 tend to choose longer durational thresholds than studies investigating pauses in L1 everyday conversation. For our purpose, it was essential to sample all pauses, without excluding pauses below a lower durational limit. Still, we could not automatically classify all silent intervals in our recordings as pauses because silent intervals can be of multiple origins.

Silent intervals in speech recordings can occur because of pauses that fit the definition for our analysis, i.e. silent intervals resulting from breathing, rhythmic, structural or pragmatic reasons (“true pauses”), but also because of holds in stop consonants (for example /p/, /t/ or /k/) or very low amplitude schwas or fricatives (for example /f/, /s/ or /h/), i.e. silent intervals that are not considered as pauses (“phonetic silences”). Our automatic pause detection algorithm should only detect the former pauses, but not the latter ones. In order to determine the lower durational limit for the automatic detection mechanism, we determined the threshold below which no more true pauses occurred. For that, we analyzed the recordings of 6 speakers (pseudo-randomized, we ensured that there were 2 speakers of each language, one male and one female, 2 recordings for each condition, and 3 speakers in the sound-proof room and in the office). We automatically detected all silent intervals longer than 0.001 seconds (threshold -35 dB, minimum silent interval: 0.001 s) and then manually determined whether the detected silences were phonetic silences or true pauses. We then evaluated the distributional pattern of silences.

In these analyses we found that 90.39% (CIs: 80.89 and 99.9%) of the phonetic silences ( $n = 338$ ) were shorter than 0.1 s, 9.29% (CIs: 0.44 and 18.14%) had durations between 0.1 and 0.2 s, and 0.32% (CIs: -0.5 and 1.13%) were longer than 0.2 s. In contrast, 93.14% (CIs: 86.09 and 100.1%) of the pauses ( $n = 132$ ) were longer than 0.2 s, 6.86% (CIs: -0.18 and 13.91%) had durations between 0.1 and 0.2 s, and no true pauses were shorter than 0.1 s (Fig 1, Table 1). This led us to the conclusion to choose 0.1 s as a lower threshold for the automatic annotation of pauses to exclude most phonetic silences and include all true pauses. The remaining phonetic silences were deleted manually.

Pause measurement was performed using the software package PRAAT (Version 6.0.36, [56]). For that purpose, pauses were automatically annotated (Annotate → To TextGrid (silences); guidelines for settings: Silence threshold: -35.0 dB, Minimum silent interval duration: 0.1 s, Minimum sounding interval duration: 0.1 s). Additionally, all pauses were checked visually (in the oscillogram and spectrogram) and acoustically, and adjusted manually, in order to remove rarely occurring incorrectly identified pauses (e.g. holds in plosives, low amplitude fricatives; see Box 1) or to insert pauses that had not been automatically detected (for example, because of breathing noise). With a PRAAT script, the total duration of each reading, the number of pauses and the duration of individual pauses in each reading were extracted. We included all true silent pauses (see Box 1). Although we intended to include





**Fig 1. Distribution of phonetic silences and true pauses.** Dashed line = threshold chosen for the subsequent automatic detection of pauses (0.1 s).

<https://doi.org/10.1371/journal.pone.0230710.g001>

**Table 1. Mean pause duration and mean proportion of short (< 0.1 s), medium (0.1 < x < 0.2 s) and long (> 0.2 s) phonetic silences and true pauses with respective low and high confidence intervals (CIs).**

	Duration [s]	Low CI	High CI	Short [%x]	Low CI	High CI	Medium [%]	Low CI	High CI	Long [%]	Low CI	High CI
Phonetic silences	0.047	0.033	0.061	90.39	80.89	99.9	9.29	0.44	18.14	0.32	-0.5	1.13
True pauses	0.404	0.301	0.506	0	0	0	6.86	-0.18	13.91	93.14	86.09	100.18

<https://doi.org/10.1371/journal.pone.0230710.t001>

filled pauses that contained a noise component resulting for example from breathing or from vocal hesitations such as “ehm” or “uh” [21,48,58,59], we did not find vocal hesitations in our data, most likely because participants read a scripted text and had familiarized themselves with the text before being recorded. Thus, the only filled pauses in our data are rarely occurring intervals containing obvious breathing noise.

We classified all pauses with regard to their position in the text. For each pause, we determined whether it occurred at a punctuation mark in the text (hereafter “marked pauses”; i.e. full stops, commas or quotation marks; no other punctuation marks occurred in the text), at an unmarked clause or phrase boundary (hereafter “unmarked pauses”; e.g. before a defining relative clause), or at any other position in the text. For the full text with the annotation of the pause categories see [S2 Appendix](#).

We used linear and logistic mixed effects models to investigate the influence of reading tempo, native language and in-text position on the realization of pauses. Preliminary analyses did not reveal differences between native German and native Serbo-Croatian non-native speakers of English, so we lumped these two groups together for our analyses. Thus our analyses compared two groups, namely native and non-native speakers of English (“nativeness” factor). Still, in our plots, we present the data for all three native languages separately, in order to allow visual comparisons between them.

For our model predictors reading tempo and nativeness, we used deviation coding [60]. Reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5).

To test whether total reading time, pause-to-utterance ratio and the duration of individual pauses were influenced by reading tempo and nativeness, we used linear mixed models [61] into which we entered these two predictors and an interaction term of the two as fixed effects. In order to reduce non-normality in the error structure of our models, the dependent variables total reading time and duration of individual pauses were log-transformed, because the optimal lambda for a Box-Cox transformation [62] was close to 0 in both cases, using the *boxcox* function of the *MASS* package [63]. The dependent variable pause-to-utterance ratio, which is proportional data bounded by 0 and 1, was logit-transformed, using the *logit* function of the *boot* package [64–67].

To investigate which factors influenced the probability of making a pause, we used a generalized linear mixed model [61] with binomial error structure and a logit link function. Each transition between two words represented a data point, and we determined for each of these transitions if a pause occurred there or not (similar to [38]). In total, this resulted in 26,456 data points. This number of data points can be explained as follows: 41 participants \* 3 reading tempi \* 215 word boundaries in the text = 26,445 data points. Additional 11 datapoints resulted from words that were not in the scripted text but that participants inserted spontaneously while reading. This yielded 26,456 data points in total, 2,750 of which were pauses. In this model, we included reading tempo, nativeness, in-text position and an interaction of reading tempo and nativeness as fixed effects.

All of our models included participant as a random intercept. For the three models testing the influence of reading tempo and nativeness on total reading time, pause-to-utterance ratio and the duration of individual pauses, our design, with one reading per tempo condition of each participant, did not allow us to accurately estimate random slopes. For the model testing the influence of reading tempo, nativeness and in-text position on the probability of making a pause, we ran an initial model that also included a random slope of position. However, this model did not converge. Thus, no random slopes are included in the models (but see [68–70]).

All models were fitted in R (version 3.5.1, [71]) and implemented in RStudio (version 1.1.456, [72]) using the *lmer* function of the *lme4* package [73].

For each linear mixed model, we visually inspected a qqplot and the residuals plotted against fitted values to check whether the assumptions of normally distributed and homogeneous residuals were fulfilled (using a function provided by Roger Mundry, Leipzig, Germany). These indicated no obvious deviations from normality or homoscedasticity.

Finally, we derived variance inflation factors (VIF, [74]) using the *vif* function of the R-package *car* applied to our models with the random effects excluded. They did not indicate collinearity to be an issue. We tested the significance of the respective full models as compared to the null models (comprising only the random intercept) by using a likelihood ratio test (R function *anova* with the argument *test* set to “Chisq”, [75,76]). In all cases, parameters were estimated using maximum likelihood (rather than Restricted Maximum Likelihood, [77]) in order to allow for likelihood ratio tests. To obtain p-values for the individual effects, we



conducted likelihood ratio tests comparing the full with respective reduced models ([69], R function *drop1*).

As indicators for the goodness-of-fit of our models, we follow [78] and report the marginal and conditional  $R^2$  for each full model. The marginal  $R^2$  ( $R^2_m$ ) reveals the variance explained by the entirety of the fixed effects, and the conditional  $R^2$  ( $R^2_c$ ) reveals the variance explained by the entirety of the fixed and random effects. Thus, these measures can be taken as indicators for the effect size for the full models. We calculated  $R^2_m$  and  $R^2_c$  using the *r.squaredGLMM* function from the *MuMIn* package [79].

## Results

Our instructions successfully elicited three desired reading aloud tempi: the full model for the total reading time was clearly significant compared to the null model (likelihood ratio test:  $\chi^2 = 169.43$ ,  $df = 3$ ,  $p < 0.001$ , effect size for the full model:  $R^2_m = 0.72$ ,  $R^2_c = 0.82$ ). Specifically, there was an effect of reading tempo on total reading time (likelihood ratio test:  $\chi^2 = 163.67$ ,  $df = 1$ ,  $p < 0.001$ ), with reading time increasing from the fast to the slow condition. Furthermore, we found a significant main effect of nativeness on the total reading time (likelihood ratio test:  $\chi^2 = 10.21$ ,  $df = 1$ ,  $p = 0.001$ ) with the total reading duration being higher in non-native speakers than in native speakers of English. There was no significant interaction effect of reading tempo and nativeness, i.e. non-native speakers did not change their reading durations differently in fast and slow tempo compared to native speakers (Table 2; Fig 2A; random effects: S8 Table).

We next explored how reading tempo and nativeness influenced the realization of pauses. In particular, we investigated the proportion of the total reading time that was devoted to pauses (pause-to-utterance ratio; Fig 2B), and how long the individual pauses were (individual pause duration; Fig 2C) at the different reading tempi in native and non-native speakers. We also explored how many pauses speakers made (number of pauses; Fig 2D) and how the probabilities that speakers made pauses at certain positions in the text were influenced by reading tempo and nativeness.

### The effects of reading tempo and nativeness on the pause-to-utterance ratio

The full model for the pause-to-utterance ratio was significant compared to the null model (likelihood ratio test:  $\chi^2 = 138.08$ ,  $df = 3$ ,  $p < 0.001$ , effect size for the full model:  $R^2_m = 0.58$ ,

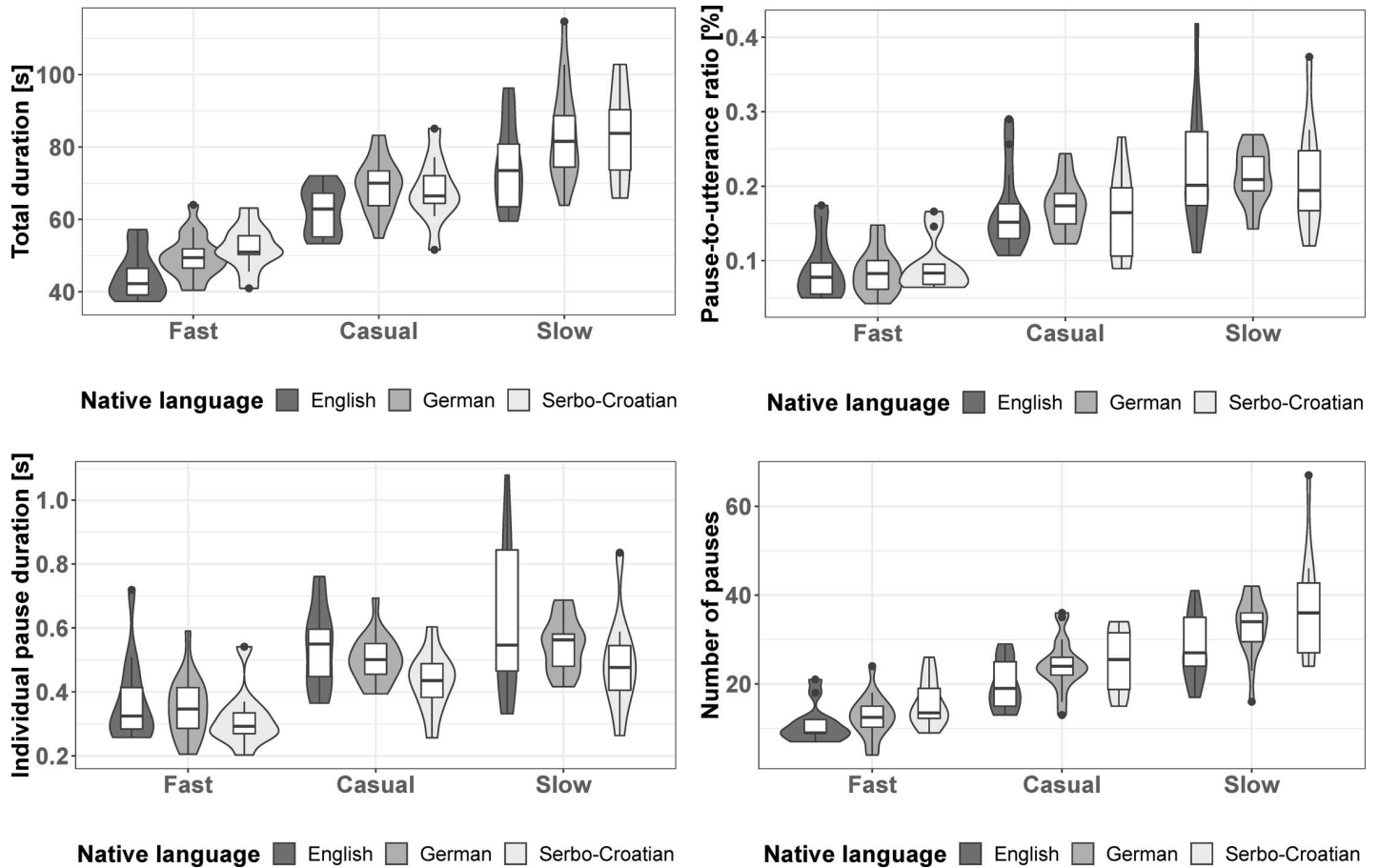
**Table 2. Results of the linear mixed model exploring the effects of reading tempo and nativeness on the total reading time (log-transformed).** The table reports estimated model coefficients, standard errors (SE) and lower and upper confidence intervals (CI),  $\chi^2$  values of likelihood ratio tests and respective degrees of freedom (df) and p-values (P).

Term	Estimate	SE	lower CI	upper CI	$\chi^2$	df	P
Intercept	4.07	0.03	4.02	4.13	a	a	a
Reading tempo <sup>b</sup>	0.51	0.04	0.43	0.60	163.67	1	< 0.001
NativenessNonNative <sup>b</sup>	0.11	0.03	0.05	0.18	10.21	1	0.001
Reading tempo:NativenessNonNative <sup>(2)</sup>	-0.02	0.05	-0.18	0.08	0.11	1	0.740

<sup>a</sup> not shown because of having a very limited interpretation

<sup>b</sup> Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5); the indicated tests were obtained from likelihood ratio tests comparing the full with a reduced model lacking reading tempo, nativeness and the interaction, respectively.

<https://doi.org/10.1371/journal.pone.0230710.t002>



**Fig 2. a)** total duration of the readings in seconds, **b)** pause-to-utterance ratio in %, **c)** duration of individual pauses in seconds and **d)** number of pauses in each condition for each native language. The violin plots show median values (horizontal black lines) with first and third quartiles (lower and upper end of boxes), minimum and maximum values limited to values no more than 1.5 IQRs distant from the respective end of the box (lower and upper end of vertical black lines) and outliers (black dots). The area around each box indicates the distribution of the data.

<https://doi.org/10.1371/journal.pone.0230710.g002>

$R^2_c = 0.80$ ). More specifically, reading tempo had a significant effect on the pause-to-utterance ratio (likelihood ratio test:  $\chi^2 = 137.94$ ,  $df = 1$ ,  $p < 0.001$ ). In contrast, we did not find a significant effect of nativeness on the pause-to-utterance ratio (likelihood ratio test:  $\chi^2 = 0.002$ ,  $df = 1$ ,  $p = 0.965$ ). Also, the effect of the interaction of reading tempo and nativeness was non-significant (likelihood ratio test:  $\chi^2 = 0.14$ ,  $df = 1$ ,  $p = 0.709$ ). Non-native speakers thus spent a similar amount of time on pauses as native speakers.

Averaged over native and non-native speakers, the mean pause-to-utterance ratio was 22.62% for slow, 14.35% for casual, and 8.76% for fast reading aloud (Table 3; Fig 2B; random effects: S9 Table). We used the following formulae for the back-transformation from the logit-transformed model estimates (Table 3):  $odds = \exp(-1.79 + 1.11 * \text{reading tempo} + 0.01 * \text{nativeness} - 0.05 * \text{reading tempo} * \text{nativeness})$ , and  $y = odds / (1 + odds)$ . Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5). The respective values for fast, casual and slow reading speeds were inserted into the formulae to calculate the mean pause-to-utterance ratios. To get values averaged for native and non-native speakers, we inserted 0 for nativeness.

**Table 3. Results of the linear mixed model exploring the effects of reading tempo and native language on pause-to-utterance ratio (logit-transformed).** The table reports estimated model coefficients, standard errors (SE) and lower and upper confidence intervals (CI),  $\chi^2$  values of likelihood ratio tests and respective degrees of freedom (df) and p-values (P).

Term	Estimate	SE	lower CI	upper CI	$\chi^2$	df	P
Intercept	-1.79	0.09	-1.96	-1.61	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>
Reading tempo <sup>b</sup>	1.11	0.10	0.91	1.32	137.94	1	< 0.001
NativenessNonNative <sup>b</sup>	0.01	0.10	-0.20	0.21	0.002	1	0.965
Reading tempo: NativenessNonNative <sup>(2)</sup>	-0.05	0.12	-0.29	0.20	0.14	1	0.709

<sup>a</sup> not shown because of having a very limited interpretation

<sup>b</sup> Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5); the indicated tests were obtained from likelihood ratio tests comparing the full with a reduced model lacking reading tempo, nativeness and the interaction, respectively.

<https://doi.org/10.1371/journal.pone.0230710.t003>

### The effects of reading tempo and nativeness on the duration of individual pauses

The full model for the individual pause durations was significant compared to the null model (likelihood ratio test:  $\chi^2 = 89.46$ ,  $df = 3$ ,  $p < 0.001$ , effect size for the full model:  $R^2_m = 0.37$ ,  $R^2_c = 0.73$ ). We found that there was a significant effect of reading tempo on the duration of individual pauses (likelihood ratio test:  $\chi^2 = 87.53$ ,  $df = 1$ ,  $p < 0.001$ ). Contrastingly, we did not find a significant effect of nativeness on pause duration (likelihood ratio test:  $\chi^2 = 1.73$ ,  $df = 1$ ,  $p = 0.188$ ). Likewise, the effect of the interaction of reading tempo and nativeness was non-significant (likelihood ratio test:  $\chi^2 = 0.21$ ,  $df = 1$ ,  $p = 0.651$ ), which indicates that native and non-native speakers of English altered the duration of their pauses similarly in different reading tempi.

The mean duration of individual pauses, averaged over native and non-native speakers, was 0.60 s in slow, 0.47 s in casual, and 0.37 s in fast reading aloud (Table 4; Fig 2C; random effects: S8 Table). We used the following formula for the back-transformation from the log-transformed model estimates (Table 4):  $y = \exp(-0.75 + 0.49 * \text{reading tempo} - 0.10 * \text{nativeness} - 0.04 * \text{reading tempo} * \text{nativeness})$ . Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5). The respective values for fast, casual and slow speech were inserted into the formula to calculate the mean durations. To get values averaged for native and non-native speakers, we inserted 0 for nativeness.

**Table 4. Results of the linear mixed model exploring the effects of reading tempo and native language on the duration of individual pauses (log-transformed).** The table reports estimated model coefficients, standard errors (SE) and lower and upper confidence intervals (CI),  $\chi^2$  values of likelihood ratio tests and respective degrees of freedom (df) and p-values (P).

Term	Estimate	SE	lower CI	upper CI	$\chi^2$	df	P
Intercept	-0.75	0.06	-0.87	-0.63	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>
Reading tempo <sup>b</sup>	0.49	0.07	0.36	0.62	87.53	1	< 0.001
NativenessNonNative <sup>b</sup>	-0.10	0.07	-0.24	0.05	1.73	1	0.188
Reading tempo: NativenessNonNative <sup>(2)</sup>	-0.04	0.08	-0.19	0.12	0.25	1	0.651

<sup>a</sup> not shown because of having a very limited interpretation

<sup>b</sup> Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5); the indicated tests were obtained from likelihood ratio tests comparing the full with a reduced model lacking reading tempo, nativeness and the interaction, respectively.

<https://doi.org/10.1371/journal.pone.0230710.t004>

**Table 5. Results of the logistic regression model exploring the effects of reading tempo, native language and in-text position on the occurrence frequency of pauses.** The table reports estimated model coefficients, standard errors (SE) and lower and upper confidence intervals (CI),  $\chi^2$  values of likelihood ratio tests and respective degrees of freedom (df) and p-values (P).

Term	Estimate	SE	lower CI	upper CI	$\chi^2$	df	P
Intercept	0.60	0.17	0.27	0.94	<sup>a</sup>	<sup>a</sup>	<sup>a</sup>
PositionUnmarkedPhraseBoundary <sup>b</sup>	-2.73	0.08	-2.89	-2.57	9759.82	1	< 0.001 <sup>c</sup>
PositionOtherWordBoundary <sup>b</sup>	-6.06	0.09	-6.25	-5.88			
Reading tempo <sup>b</sup>	2.20	0.15	1.91	2.48	906.75	1	< 0.001
NativenessNonNative <sup>b</sup>	0.54	0.20	0.14	0.94	7.05	1	0.008
Reading tempo: NativenessNonNative <sup>b</sup>	0.27	0.17	-0.06	0.60	2.59	1	0.11

<sup>a</sup> not shown because of having a very limited interpretation

<sup>b</sup> In-text position was dummy coded with the marked position being the respective reference category. Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5); the indicated tests were obtained from likelihood ratio tests comparing the full with a reduced model lacking in-text position, reading tempo, nativeness and the interaction between reading tempo and nativeness, respectively.

<sup>c</sup> This is the overall effect of in-text position on the occurrence frequency of pauses. Considering the differences between the individual levels, the model revealed that at unmarked phrase boundaries ( $z = -33.49$ ,  $p < 0.001$ ) and at other word boundaries ( $z = -65.17$ ,  $p < 0.001$ ), participants paused significantly less than at punctuation marks.

<https://doi.org/10.1371/journal.pone.0230710.t005>

## The effect of reading tempo, nativeness and position in the text on the occurrence frequency of pauses

The full model for the occurrence frequency of pauses was significant compared to the null model (likelihood ratio test:  $\chi^2 = 10186$ ,  $df = 5$ ,  $p < 0.001$ , effect size for the full model:  $R^2_m = 0.32$ ,  $R^2_c = 0.34$ ). We found significant effects of reading tempo (likelihood ratio test:  $\chi^2 = 906.75$ ,  $df = 1$ ,  $p < 0.001$ ), of nativeness (likelihood ratio test:  $\chi^2 = 7.05$ ,  $df = 1$ ,  $p = 0.008$ ) and of in-text position (likelihood ratio test:  $\chi^2 = 9759.82$ ,  $df = 1$ ,  $p < 0.001$ ) on the occurrence frequency of pauses. Non-native speakers made more pauses than native speakers, and people made more pauses the more slowly they read aloud. Also, people made more pauses at punctuation marks than at unmarked phrase boundaries and at other positions in the text. However, the effect of the interaction of reading tempo and nativeness was non-significant (likelihood ratio test:  $\chi^2 = 2.59$ ,  $df = 1$ ,  $p = 0.11$ ), which indicates that native and non-native speakers of English altered the occurrence frequency of their pauses similarly in different reading aloud tempi (Table 5). Regarding our random intercept of participant, the estimated standard deviation among participants was 0.56 (S9 Table). This is smaller than the magnitude of effects of in-text position and reading tempo, but similar to the magnitude of the effect of nativeness (cf. Table 5). This indicates that the influence of participant variability and nativeness are comparable.

Native and non-native speakers' predicted probabilities of making a pause in fast, casual and slow reading at punctuation marks, unmarked phrase boundaries and other word boundaries are given in Table 6. The corresponding partial effects of our model are shown in Fig 3. The overall numbers of pauses in each reading tempo and each native language are displayed in Fig 2D.

## Discussion

Our study evaluated whether pauses contributed to the speakers' non-native L2 production by examining whether native and non-native speakers of English differed in their pausing behavior when reading aloud at different speech rates. Although there were some rare misclassifications in our native language recognition test, overall, all non-native speakers were identified as

**Table 6. Predicted probabilities (in %) of making a pause in fast, casual or slow reading at punctuation marks, unmarked phrase boundaries or other positions in the text for native and non-native speakers of English.**

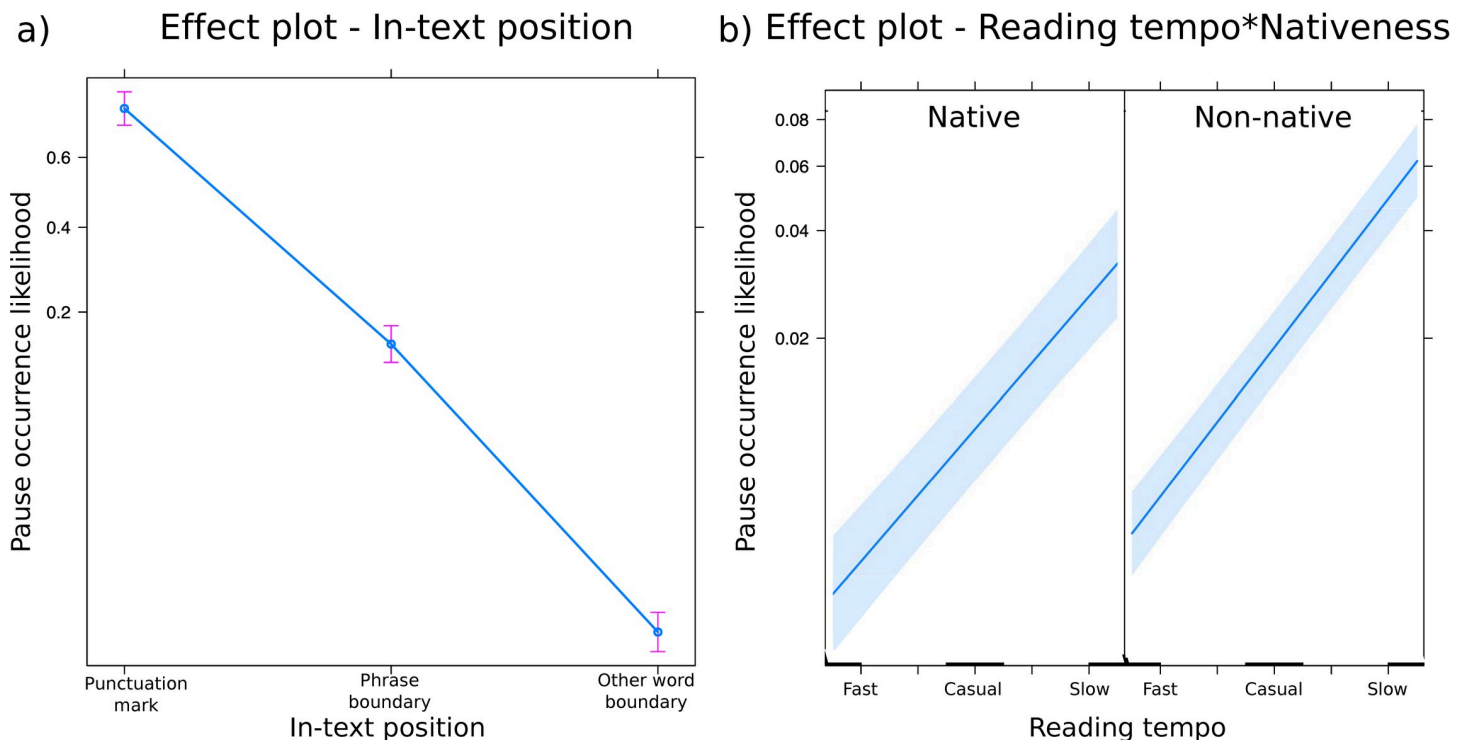
	Native <sup>a</sup>			Non-Native <sup>a</sup>		
	Fast	Casual	Slow	Fast	Casual	Slow
<b>Punctuation mark</b>	33.20	58.19	79.58	42.69	70.50	88.46
<b>Phrase boundary</b>	3.15	8.34	20.31	4.65	13.51	33.39
<b>Other word boundary</b>	0.12	0.32	0.90	0.17	0.55	1.76

<sup>a</sup>Values are derived from the estimates of our logistic regression model (see Table 5). We used the following formulae for the back-transformation from the logit-transformed model estimates:  $odds = \exp(0.60 - 2.73 * \text{unmarked phrase boundary} - 6.06 * \text{other word boundary} + 2.20 * \text{reading tempo} + 0.54 * \text{nativeness} + 0.27 * \text{reading tempo} * \text{nativeness})$ , and  $y = odds / (1 + odds)$ . In-text position was dummy coded, with marked boundaries serving as the reference category. Reading tempo and nativeness were deviation coded: reading tempo was coded as a continuous predictor (fast = -0.5, casual = 0, slow = +0.5), and nativeness was coded as a two-level factor (native = -0.5, non-native = +0.5). The respective values were inserted into the formulae to calculate the predicted probabilities of pause occurrence.

<https://doi.org/10.1371/journal.pone.0230710.t006>

non-native by native English raters, and thus had clearly recognizable non-native accents. However, our findings suggest that non-native pause patterns contributed little to the production of these non-native accents, at least for our relatively proficient speakers (see below). This supports the *No Contribution* hypothesis.

First, native and non-native speakers had similar pause-to-utterance ratios. Second, the durations of their pauses were similar (in line with 25,33, but in contrast to 46). Third, for native and non-native speakers, reading tempo had a similar significant influence on all of the pause characteristics measured, and there were no interactions between nativeness and



**Fig 3. a)** effect display of the significant main effect of in-text position, **b)** effect display of the significant main effects of reading tempo and nativeness. The effect of the interaction of reading tempo and nativeness was non-significant. The y-axis displays the probability of occurrence of a word after a pause. Error bars and shaded areas around the estimated effects represent 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0230710.g003>

reading rate. With an increasing reading tempo, native and non-native speakers made fewer and shorter pauses, and with a decreasing reading tempo, they made more and longer pauses. Thus, with changing reading tempo, native and non-native speakers altered both the duration and the occurrence likelihood of pauses in a highly similar way. During fast reading, native and non-native speakers' probabilities of making a pause were below 5% at unmarked phrase boundaries and below 0.2% at other unmarked positions in the text. Thus, almost only pauses at punctuation marks remained, suggesting that the visually salient punctuation marks help readers to structure their vocal output in similar ways.

Pause-to-utterance ratio changed with changing reading tempo (cf. similar findings in speakers of Dutch, French and Italian; [26]), indicating that in fast reading, pauses were reduced more compared to casual reading and in slow reading, pauses increased relative to casual reading. A potential explanation is that when reducing pauses, there is less information loss compared to reducing articulated sounds. If articulated sounds were deleted or shortened too much, words would be distorted to intelligibility and semantic content would be lost. Similarly, if segments were lengthened to an unnatural extent they would be difficult to produce and perceive. These factors apply much less to shortened or lengthened pauses, and reducing or increasing pause length is articulatorily simpler than varying the duration of vocal segments. This may explain the similar results regarding reading tempo for native and non-native speakers.

The only significant difference in pauses that may contribute to sounding foreign is the higher likelihood of having a pause in L2 speech: non-native speakers made more pauses than native speakers (in contrast to [25]). Although our non-native participants were highly proficient in English, they might still need more time for cognitive processing when speaking their L2 [46,82]. This might be reflected in their higher likelihood of occurrence of pauses (but interestingly not in longer durations of pauses). However, the reason speakers need more processing time for the foreign language might be rooted not in difficulties in adhering to target language pause patterns, but to other aspects of the L2, such as difficulties pronouncing L2-typical sounds. Further evidence that speakers need more time for cognitive processing in their L2 than in their L1 is that non-native speakers had longer reading durations than native speakers (cf. [83]). To rule out the possibility that non-native speakers have a slower reading speed in general, the reading durations in their respective first languages would need to be addressed in a follow-up study. Also, this would shed light on the influence of individual participants' reading proficiency on our results (see below).

There are several possible explanations why, overall, our non-native speakers did not appear to produce pauses 'with a foreign accent'. Pauses may be easier to acquire than other aspects of language because they are perceptually salient (Matzinger, Ritt, Fitch, in prep) and pausing is articulatorily easy, relative to phonemes or intonation patterns. Alternatively, pauses might not contribute to non-native speech production because the pause patterns in the speakers' L1 and L2 might be similar and speakers simply transfer their L1 pause patterns to their L2. Such a transfer might imply that pauses result from very general cognitive mechanisms [53] and thus have a more universal character than other aspects of language. However, to substantiate this hypothesis, non-native English speakers of a typologically diverse set of languages, including many other native languages than German and Serbo-Croatian, would need to be tested in a larger-scale study. If such L2 speakers still pause with a native-like accent when speaking English, this would be evidence for a language universal character of pause patterns.

If pauses result from basic cognitive mechanisms, L2 speakers should also pause with a non-native accent in other second languages than English.



Although our study controls for many aspects necessary to investigate foreign speech production, there are some aspects that it does not address, but that could be addressed in potential follow-up experiments. Our study tested the role of pauses of highly proficient L2 speakers. Certainly, L2 proficiency might have an influence on the realization of pauses [8,34]. L2-typical pause patterns might be acquired rapidly (and should thus not contribute to non-native accents in any proficiency level), slowly but still more quickly than e.g. phonemes (and should thus only contribute to non-native accents in beginners), or not at all (and should thus contribute to non-native accents in any proficiency level). Therefore, testing L2 speakers of different proficiency levels in a follow-up experiment might reveal more about the cognitive constraints associated with speaking an L2, which might contribute to non-native speech production. Our present finding that pauses contribute little to the acoustic peculiarities of L2 production would be further corroborated if a comparable study with less proficient non-native speakers yielded similar results.

Furthermore, we caution that our relatively small sample size of 41 participants, although clearly adequate to reveal multiple statistically significant effects, might potentially be inadequate to reveal more subtle differences of smaller effect size. Thus, like any null result, our “no difference” findings should be viewed with some caution. However, the considerable time and effort required to derive, annotate and manually check pause data (more than 26,000 data points of which more than 2,700 were pauses comprise our current dataset) would remain a challenge for gathering much larger samples of participants.

We only included academically educated participants, from which a similarly high level of reading proficiency could be assumed. Testing people with a high reading proficiency might have contributed to the fact that, overall, we found similar pause patterns in native and non-native speakers. Highly proficient readers might be able to override challenges in foreign speech production when reading, but not when speaking freely (see below). Less proficient readers might not be able to mask difficulties in non-native speech production when reading, which might result in different pause patterns between native and non-native speakers. A potential follow-up study could explicitly test participants’ reading proficiency and include it as a predictor for pause patterns.

The text that the participants read out in our study contained punctuation marks. Punctuation marks are salient visual cues that might prime participants’ pause patterns [84]. We used a text containing punctuation marks to make the procedure as close to real-life reading situations as possible. How much punctuation contributes to adopting the pause patterns of an L2 in read speech could easily be tested in a follow-up study with texts without punctuation marks. Similar results using texts without punctuation marks could evaluate the possibility that the realization of pauses is determined by basic cognitive processes, because in such cases participants would not be primed by visual cues.

Results of our study on read texts might not be entirely transferrable to spontaneous speech, because pauses in reading aloud might also reflect reading difficulties or spelling-to-sound difficulties. In contrast, pauses in spontaneous speaking might reflect difficulties in conceptualizing the message or in linguistic formulation. Still, we argue that testing speakers during reading aloud has ecological relevance because it occurs in several real-life contexts, such as when reading aloud to children or in the (language) classroom. Especially during second language learning, reading aloud is a commonly practiced exercise [85]. Furthermore, reading is ideal for our purposes since it captures difficulties in articulation which are highly relevant for non-native speech production. Further, reading a complete story also captures pauses in longer stretches of speech, as opposed to reading or repeating isolated sentences (e.g. [8]). Thus, we feel that our paradigm adds a rigorous and useful new method to the existing literature on spontaneous speech.

One crucial point that our procedure cannot address is whether it is a transfer from L1 pause patterns that either hinders (because L1 and L2 pause patterns are different) or facilitates (because L1 and L2 pause patterns are similar) the acquisition of L2 typical pause patterns. Testing similarities or differences of different native languages' pause patterns is almost impossible: even if speakers of different L1s are tested on similar tasks, such as reading out texts matched in syntax and content, translations can never be fully identical in syntactic structure or small nuances of content. Such minute differences might already shape pause patterns in a way that makes it difficult to determine these potential differences' contribution to L2 accents. Nonetheless, testing the same speakers in their L1 and L2 would be useful to address the effect of individual speaking style on pause patterns [47].

## Conclusions and future work

We asked native and non-native speakers to read the same English text, thus excluding potential L1-specific morpho-syntactic factors from influencing the vocal output. We found that speakers inserted pauses into the reading stream in similar ways and at similar locations, and changed this pattern in similar ways at different reading tempi, regardless of their native language. The only difference between pause patterns in native and non-native speakers was that the non-native speakers made more pauses than the native speakers. This might reflect cognitive processing constraints in an L2 that result from other aspects of the L2 than pausing behavior per se. Overall, we conclude that in reading aloud, the influence of nativeness on the realization of pauses is marginal, suggesting that pauses play little role in the production of foreign accents in this context

## Supporting information

**S1 Table. Results of cross-linguistic studies suggesting that the numbers and durations of pauses in different languages are similar.**

(DOCX)

**S2 Table. Results of cross-linguistic studies suggesting that the numbers and durations of pauses in different languages are different.**

(DOCX)

**S3 Table. Results of studies on the numbers and durations of pauses during L2 speech.**

Results in the table concern comparisons between pauses in speakers' L2s and these speakers' L1s (as opposed to L1 speakers of the target L2).

(DOCX)

**S4 Table. Results of studies on the numbers and durations of pauses during L2 speech.**

Results in the table concern comparisons between pauses in speakers' L2s and L1 speakers of the target L2.

(DOCX)

**S5 Table. Native language recognition ratings.**

(DOCX)

**S6 Table. Estimated variance components and standard deviations for the random intercept of participant of the full model exploring the effects of reading tempo, and nativeness on the total reading time.**

(DOCX)

**S7 Table. Estimated variance components and standard deviations for the random intercept of participant of the full model exploring the effects of reading tempo and nativeness on pause-to-utterance ratio.**

(DOCX)

**S8 Table. Estimated variance components and standard deviations for the random intercept of participant of the full model exploring the effects of reading tempo, and nativeness on the duration of individual pauses.**

(DOCX)

**S9 Table. Estimated variance components and standard deviations for the random intercept of participant of the full model exploring the effects of reading tempo, nativeness and in-text position on the occurrence frequency of pauses.**

(DOCX)

**S1 Appendix. Post-experiment questionnaire.**

(DOCX)

**S2 Appendix. Annotated text *The boy who cried wolf*.**

(DOCX)

## Acknowledgments

We thank Kamil Kaźmierski, Andreas Baumann, and two anonymous reviewers for helpful comments on previous versions, and Klaus Hofmann and Magdalena Schwarz for helpful discussions.

## Author Contributions

**Conceptualization:** Theresa Matzinger, W. Tecumseh Fitch.

**Investigation:** Theresa Matzinger.

**Methodology:** Theresa Matzinger, W. Tecumseh Fitch.

**Supervision:** Nikolaus Ritt, W. Tecumseh Fitch.

**Writing – original draft:** Theresa Matzinger.

**Writing – review & editing:** Nikolaus Ritt, W. Tecumseh Fitch.

## References

1. Anderson-Hsieh J, Johnson R, Koehler K. The Relationship Between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Lang Learn.* 1992; 42(4):529–55.
2. Idemaru K, Wei P, Gubbins L. Acoustic Sources of Accent in Second Language Japanese Speech. *Lang Speech.* 2019; 62(2):333–57. <https://doi.org/10.1177/0023830918773118> PMID: 29764295
3. Pinget AF, Bosker HR, Quené H, de Jong NH. Native speakers' perceptions of fluency and accent in L2 speech. *Lang Test.* 2014; 31(3):349–65.
4. Bransford JD, Brown AL, Cocking RR. How people learn—brain, mind, experience, and school. Washington D.C.: National Academy Press; 1999.
5. Major RC. Foreign Accent—The Ontogeny and Phylogeny of Second Language Phonology. Mahwah: Lawrence Erlbaum Associates; 2001.
6. Flege JE. Second Language Speech Learning: Theory, Findings, and Problems. In: Strange W, editor. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research.* Timonium, MD: York Press; 1995. p. 233–77.

7. Chun DM. Discourse intonation in L2: From theory and research to practice. Amsterdam: John Benjamins Publishing; 2002.
8. Trofimovich P, Baker W. Learning Second Language Suprasegmentals: Effect of L2 Experience on Prosody and Fluency Characteristics of L2 Speech. *Stud Second Lang Acquis*. 2006; 28(1):1–30.
9. Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, et al. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. 2003; 87:B47–57. [https://doi.org/10.1016/s0010-0277\(02\)00198-1](https://doi.org/10.1016/s0010-0277(02)00198-1) PMID: 12499111
10. König E, Gast V. Understanding English-German contrasts. 3rd ed. Berlin: Erich Schmidt Verlag; 2012.
11. Gibbon D. Intonation in German. In: Hirst D, Di Cristo A, editors. *Intonation systems: a survey of twenty languages*. Cambridge: Cambridge University Press; 1999. p. 78–95.
12. Mennen I, Schaeffler F, Dickie C. Second Language Acquisition of Pitch Range in German Learners of English. *Stud Second Lang Acquis*. 2014; 36(2):303–29.
13. Mennen I, Schaeffler F, Docherty G. Cross-language differences in fundamental frequency range: A comparison of English and German. *J Acoust Soc Am*. 2012; 131(3):2249–60. <https://doi.org/10.1121/1.3681950> PMID: 22423720
14. Hockett CF. The Origin of Speech. *Sci Am*. 1960; 203:88–111. <https://doi.org/10.1038/scientificamerican1260-88>
15. Levinson SC. Pragmatics, Universals in. In: Hogan PC, editor. *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press; 2011. p. 654–7.
16. Stivers T, Enfield NJ, Brown P, Englert C, Hayashi M, Heinemann T, et al. Universals and cultural variation in turn-taking in conversation. *Proc Natl Acad Sci*. 2009; 106(26):10587–92. <https://doi.org/10.1073/pnas.0903616106> PMID: 19553212
17. Evans N, Levinson SC. The myth of language universals: Language diversity and its importance for cognitive science. *Behav Brain Sci*. 2009;(32):429–92.
18. Seifart F, Strunk J, Danielsen S, Hartmann I, Pakendorf B, Wichmann S, et al. Nouns slow down speech across structurally and culturally diverse languages. *Proc Natl Acad Sci [Internet]*. 2018; 115(22):5720–5. Available from: <http://www.pnas.org/lookup/doi/10.1073/pnas.1800708115> PMID: 29760059
19. Goldman-Eisler F. *Psycholinguistics: experiments in spontaneous speech*. London: Academic Press; 1968.
20. Nooteboom S. The Prosody of Speech: Melody and Rhythm. In: Hardcastle WJ, Laver J, Gibbon FE, editors. *The handbook of phonetic sciences*. Oxford: Blackwell; 1997. p. 640–73.
21. Oliveira M. The role of pause occurrence and pause duration in the signaling of narrative structure. *Adv Nat Lang Process [Internet]*. 2002;43–51. Available from: [http://link.springer.com/chapter/10.1007/3-540-45433-0\\_7](http://link.springer.com/chapter/10.1007/3-540-45433-0_7)
22. Swerts M. Prosodic features at discourse boundaries of different strength. *J Acoust Soc Am*. 1997; 101(1):514–21. <https://doi.org/10.1121/1.418114> PMID: 9000742
23. Yang X, Shen X, Li W, Yang Y. How listeners weight acoustic cues to intonational phrase boundaries. *PLoS One*. 2014; 9(7):1–9.
24. Huensch A, Tracy-Ventura N. Understanding second language fluency behavior: the effects of individual differences in first language fluency, cross-linguistic differences, and proficiency over time. *Appl Psycholinguist*. 2016;1–31.
25. Smiljanic R, Bradlow AR. Production and perception of clear speech in Croatian and English. *J Acoust Soc Am*. 2005; 118:1677–88. <https://doi.org/10.1121/1.2000788> PMID: 16240826
26. Demol M, Verhelst W, Verhoeve P. The duration of speech pauses in a multilingual environment. *Proc Annu Conf Int Speech Commun Assoc INTERSPEECH*. 2007; 1(1):117–20.
27. Yang L. Duration and pauses as boundary-markers in speech: a cross-linguistic study. In: *Proceedings of Interspeech 2007 [Internet]*. 2007. p. 458–61. Available from: <http://www.speech.kth.se/prod/publications/files/100444.pdf>
28. Grosjean FE, Deschamps A. Analyse contrastive des variables temporelles de l'anglais et du français: vitesse de parole et variables composantes, phénomènes d'hésitation. *Phonetica*. 1975; 31:144–84.
29. Holmes VM. A crosslinguistic comparison of the production of utterances in discourse. *Cognition*. 1995; 54:169–207. [https://doi.org/10.1016/0010-0277\(94\)00635-x](https://doi.org/10.1016/0010-0277(94)00635-x) PMID: 7874876
30. De Jong NH, Steinel MP, Florijn A, Schoonen R, Hulstijn JH. Linguistic skills and speaking fluency in a second language. *Appl Psycholinguist*. 2013; 34(5):893–916.
31. Johnson TH, O'Connell DC, Sabin EJ. Temporal analysis of English and Spanish narratives. *Bull Psychon Soc*. 1979; 13(6):347–50.

32. Trouvain J, Fauth C, Möbius B. Breath and Non-breath Pauses in Fluent and Disfluent Phases of German and French L1 and L2 Read Speech. In: Speech Prosody (SP8) [Internet]. 2016. p. 31–5. Available from: [http://www.ifcasl.org/docs/Trouvain\\_Fauth\\_Moebius\\_2016.pdf](http://www.ifcasl.org/docs/Trouvain_Fauth_Moebius_2016.pdf)
33. Campione E, Véronis J. A large-scale multilingual study of pause duration. Speech Prosody 2002 Proc 1st Int Conf Speech Prosody [Internet]. 2002;199–202. Available from: [http://www.isca-speech.org/archive/sp2002/sp02\\_199.html](http://www.isca-speech.org/archive/sp2002/sp02_199.html)
34. Riazantseva A. Second Language Proficiency and Pausing: A Study of Russian Speakers of English. Stud Second Lang Acquis [Internet]. 2001; 23(4):497–526. Available from: [http://www.journals.cambridge.org/abstract\\_S027226310100403X](http://www.journals.cambridge.org/abstract_S027226310100403X)
35. Derwing TM, Munro MJ, Thomson RI, Rossiter MJ. The relationship between L1 fluency and L2 fluency development. Stud Second Lang Acquis. 2009; 31(4):533–57.
36. Rose R. Temporal variables in first and second language speech and perception of fluency. ICPHS 2015 Proc 18th Int Congr Phonetic Sci [Internet]. 2015;1–5. Available from: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0405.pdf>
37. De Jong NH, Bosker HR. Choosing a threshold for silent pauses to measure second language fluency. DiSS 2013 Proc 6th Work Disfluency Spontaneous Speech [Internet]. 2013;17–20. Available from: [http://www.isca-speech.org/archive/diss\\_2013/diss6\\_017.html](http://www.isca-speech.org/archive/diss_2013/diss6_017.html)
38. De Jong NH. Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency. Int Rev Appl Linguist Lang Teach. 2016; 54(2):113–32.
39. Derwing T, Munro M. Accent, intelligibility, and comprehensibility: Evidence from Four L1s. Stud Second Lang Acquis. 1997; 19(1):1–16.
40. Derwing TM, Munro MJ, Thomson RI. A longitudinal study of ESL learners' fluency and comprehensibility development. Appl Linguist. 2008; 29(3):359–80.
41. Thomson RI. Fluency. In: Reed M, Levis JM, editors. The Handbook of English Pronunciation. Chichester: John Wiley & Sons; 2018. p. 209–26.
42. Raupach M. Temporal variables in first and second language speech and perception of fluency. In: Dechert HW, Raupach M, editors. Temporal variables in speech: Studies in honour of Frieda Goldman-Eisler [Internet]. The Hague: Mouton; 1980. p. 263–70. Available from: <https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0405.pdf>
43. Deschamps A. The syntactical distribution of pauses in English spoken as a second language by French students. In: Dechert HW, Raupach M, editors. Temporal Variables in Speech: Studies in honour of Frieda Goldman-Eisler. The Hague: Mouton; 1980. p. 255–62.
44. Isarankura S. Variability of Pause Patterns in English Read Speech of Thai EFL Learners. J Educ Soc Res [Internet]. 2013; 3(7):346–54. Available from: <http://www.mcser.org/journal/index.php/jesr/article/view/971>
45. Tavakoli P. Pausing patterns: Differences between L2 learners and native speakers. ELT J. 2011; 65(1):71–9.
46. Kolly M-J, Leemann A, Boula P, Mareüil D, Dellwo V. Speaker-idiosyncrasy in pausing behavior: evidence from a cross-linguistic study. Proc Int Congr Phonetic Sci 2015 Glas. 2015;1–5.
47. De Jong NH, Groenhout R, Schoonen R, Hulstijn JH. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. Appl Psycholinguist. 2015; 36(2):223–43.
48. Fletcher J. The Prosody of Speech: Timing and Rhythm. In: Hardcastle WJ, Laver J, Gibbon FE, editors. The handbook of phonetic sciences. 2nd ed. Hoboken: Wiley-Blackwell; 2010. p. 523–602.
49. Duez D. Silent and non-silent pauses in three speech styles. Lang Speech. 1982; 25(1):11–28.
50. Bybee J. Frequency of Use and the Organization of Language. Oxford: Oxford University Press; 2007.
51. Nespor M, Shukla M, Mehler J. Stress-timed vs. syllable-timed languages. In: Oostendorp M Van, Ewen CJ, Hume E, Rice K, editors. The Blackwell companion to phonology. Malden, MA: John Wiley & Sons; 2011. p. 1–13.
52. Comrie B, editor. The World's Major Languages. 2nd ed. London: Routledge; 2009.
53. Segalowitz N. Cognitive Bases of Second Language Fluency. New York and London: Routledge; 2010.
54. Bilá M, Džambová A. A preliminary study on the function of silent pauses in L1 and L2 speakers of English and German. Brno Stud English. 2011; 37(1):21–39.
55. Cenoz J. Pauses and hesitation phenomena in second language production. In: ITL: Review of Applied Linguistics. 2000. p. 53–69.
56. Boersma P, Weenik D. Praat: doing phonetics by computer [Internet]. 2017. Available from: <http://www.praat.org/>

57. Deterding D. The North Wind versus a Wolf: short texts for the description and measurement of English pronunciation. *J Int Phon Assoc* [Internet]. 2006; 36(2):187–96. Available from: <https://www.cambridge.org/core/journals/journal-of-the-international-phonetic-association/article/div-classtithe-north-wind-versus-a-wolf-short-texts-for-the-description-and-measurement-of-english-pronunciation/div/984AC3D9FB1F625823E523D2E428B1BE>
58. Zellner B. Pauses and the Temporal Structure of Speech. In: Keller E, editor. *Fundamentals of speech synthesis and speech recognition*. Chichester: John Wiley; 1994. p. 41–62.
59. Shriberg E. To 'errrr' is human: Ecology and acoustics of speech disfluencies. *J Int Phon Assoc*. 2001; 31(1):153–69.
60. Alkharusi H. *Categorical Variables in Regression Analysis: A Comparison of Dummy and Effect Coding*. *Int J Educ*. 2012; 4(2):202.
61. Baayen RH. *Analyzing linguistic data*. Cambridge: Cambridge University Press; 2008.
62. Box GE, Cox DR. An analysis of transformations revisited, rebutted. *J R Stat Soc Ser B*. 1964; 26(2):211–52.
63. Venables WN, Ripley BD. *Modern applied statistics with S*. 4th ed. New York: Springer; 2002.
64. Baum CF. Stata tip 63: Modeling proportions. *Stata J*. 2008; 8(2):299–303.
65. Chen K, Cheng Y, Berkout O, Lindhiem O. Analyzing Proportion Scores as Outcomes for Prevention Trials: a Statistical Primer. *Prev Sci*. 2017; 18(3):312–21. <https://doi.org/10.1007/s1121-016-0643-6> PMID: 26960687
66. Lesaffre E, Rizopoulos D, Tsonaka R. The logistic transform for bounded outcome scores. *Biostatistics*. 2007; 8(1):72–85. <https://doi.org/10.1093/biostatistics/kxj034> PMID: 16597671
67. Canty A, Ripley BD. *boot: Bootstrap R (S-Plus) Functions*. 2017.
68. Schielzeth H, Forstmeier W. Conclusions beyond support: Overconfident estimates in mixed models. *Behav Ecol*. 2009; 20(2):416–20. <https://doi.org/10.1093/beheco/arn145> PMID: 19461866
69. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem Lang*. 2013; 68(3):255–78.
70. Aarts E, Dolan C V., Verhage M, Van der Sluis S. Multilevel analysis quantifies variation in the experimental effect while optimizing power and preventing false positives. *BMC Neurosci*. 2015; 16(1):1–15.
71. R Development Core Team. *R: A Language and Environment for Statistical Computing* [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2018. Available from: <http://www.r-project.org/>
72. RStudioTeam. *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc.; 2018.
73. Bates D, Mächler M, Bolker BM, Walker SC. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw*. 2015; 67(1).
74. Field A, Miles J, Field Z. *Discovering Statistics Using R*. International Statistical Review. Los Angeles: SAGE; 2012.
75. Dobson AJ, Barnett AG. *An introduction to generalized linear models*. 4th editio. Boca Raton: CRC Press; 2018.
76. Forstmeier W, Schielzeth H. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav Ecol Sociobiol*. 2011; 65:47–55. <https://doi.org/10.1007/s00265-010-1038-5> PMID: 21297852
77. Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol*. 2008; 24(3):127–35.
78. Nakagawa S, Schielzeth H. A general and simple method for obtaining R<sup>2</sup> from generalized linear mixed-effects models. *Methods Ecol Evol*. 2013; 4(2):133–42.
79. Bartón K. *MuMIn: Multi-Model Inference*. R package version 1.42.1 [Internet]. 2018. Available from: <https://cran.r-project.org/package=MuMIn>
80. Kahng J. Exploring Utterance and Cognitive Fluency of L1 and L2 English Speakers: Temporal Measures and Stimulated Recall. *Lang Learn*. 2014; 64(4):809–54.
81. Derwing TM, Rossiter MJ, Munro MJ, Thomson RI. Second Language Fluency: Judgements on different tasks. *Lang Learn*. 2004; 54(4):655–79.
82. Grosjean FE. Temporal variables within and between languages. In: Dechert HW, Raupach M, editors. *Towards a Cross-Linguistic Assessment of Speech Production*. Bern: Peter Lang; 1980. p. 39–53.
83. Munro MJ, Derwing TM. Modeling perceptions of the accentedness and comprehensibility of L2 speech: the role of speaking rate. *Stud Second Lang Acquis*. 2001; 23:451–68.
84. Janiszewski C, Wyer RS. Content and process priming: A review. *J Consum Psychol* [Internet]. 2014; 24(1):96–118. Available from: <http://dx.doi.org/10.1016/j.jcps.2013.05.006>
85. Gibson S. Reading aloud: A useful learning tool? *ELT J*. 2008; 62(1):29–36.