

Language change in Japanese–English bilingual returnee children over the course of five years: Evidence from accent-rating

Second Language Research

1–26

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/02676583241230854

journals.sagepub.com/home/slr**Tim Joris Laméris** 

Leiden University, The Netherlands

Maki Kubota

UiT The Arctic University of Norway, Norway

Tanja Kupisch

University of Konstanz, Germany

Jennifer Cabrelli

University of Illinois at Chicago, USA

Neal Snape

Gunma Prefectural Women's University, Japan

Jason Rothman

UiT The Arctic University of Norway, Norway

Abstract

Few studies have examined global foreign accent (GFA) in bilingual children, and little is known about how GFA changes over time and what factors determine change. Here, we examine GFA trajectories in Japanese–English bilingual returnees (Japanese children who returned to Japan after having lived

*Tanja Kupisch is also affiliated with UiT The Arctic University of Norway, Norway. Jason Rothman is also affiliated with Universidad Nebrija, Spain.

Corresponding author:

Tim Joris Laméris, Leiden University, Postbus 9515, Leiden 2300 RA, The Netherlands.

Email: t.j.lameris@hum.leidenuniv.nl

in a majority English environment for several years). In two accent-rating tasks, first language (L1) speakers of English or Japanese rated returnee speech excerpts recorded at three time points over a five-year period. The ratings show a decrease in Japanese GFA one year after return to Japan, and an increase in English GFA, but only five years after return. These findings suggest rapid re-exposure effects of the L1 and relatively stable maintenance of the second language (L2). Changes varied by L2 English age of onset (AoO) and exposure to L2 English while abroad, suggesting a crucial role for these individual factors in transitory contexts such as returnee bilingualism.

Keywords

attrition, exposure, global accent, Japanese, returnees

1 Introduction

A global foreign accent (GFA)¹ refers to any perceived difference from a first language (L1) speaker, often resulting from influence from another language (Derwing and Munro, 2009). Such a perceived accent may result from any combination of objective differences in specific phonetic and phonological features. The present study seeks to shed light on accent trajectories in bilingual children in an effort to better understand how bilingual systems interact over time and across contexts. A primary question that underlies this line of inquiry regards how permeable accents are to change in childhood depending on factors such as residence in the first language (L1) versus second language (L2) environment, L2 age of onset (AoO), and relative exposure to the L1 and L2. One understudied bilingual scenario that promises to be richly informative in this regard is that of returnee bilingualism (for review, see Flores and Snape, 2021), where linguistic exposure to both languages, opportunity for (continued) use, and dominance between them can shift drastically and abruptly in childhood. Herein, we examine language change in Japanese–English returnee bilinguals over a period of five years upon return to their L1 Japanese environment, where ‘language change’ not only refers to the diachronic changes in the language system (Global Foreign Accent, here), but encompasses both development and attrition effects that we may see in L1 and L2 GFA of these returnees. To our knowledge, this is the first study to document accent trajectories in returnees’ L1 and L2 longitudinally. In doing so, these longitudinal data will uniquely inform the relationship between L1 and L2 accent across time and context.

Under a typical definition (e.g. Flores, 2010), returnees are children of immigrant families who spend a significant portion of their formative developmental years (school age) in a foreign majority language context, returning to their first language environment as older children or teenagers. In this context, ‘L1 environment’ has been defined either in terms of the country of birth, e.g. children born in Japan who moved abroad and returned to Japan (Kubota et al., 2020; Snape et al., 2014; Tomiyama, 1999), or the home country of their parents, e.g. children born in Germany to Portuguese parents who moved to their parents’ country of origin, Portugal (Flores, 2010, 2020); for Turkish returnees, see also Grasmuck and Hinze, 2016; Treffers-Daller et al., 2016. In the context of the present study, the returnees we work with are of the former type. During their time away from the L1 environment, such returnees are typically exposed to and educated in the majority language (ML) of the host country, where they form social networks with

speakers of the ML. In these cases, the ML is acquired as an early second language (L2) and may become the child's dominant language. At the same time, their native/first language (L1) becomes a heritage language (HL) (Montrul, 2016; Polinsky, 2018; Rothman, 2009). This typically entails significant reductions in exposure to and opportunities for use of the heritage L1 outside the home environment, potentially leading to changes to it. However, upon return to their birth country, the status of their two languages reverses. Their L1 heritage language again becomes the majority language, whereas their former majority L2 becomes a minority L2.

Given the context of language exposure during returnees' formative years, their trajectory provides a unique opportunity to investigate how the aforementioned change in the language environment (that is, re-immersion into the L1 and departure from the majority L2 environment) influences the development of the L1 and the L2, and how the trajectories of both interact over time. Previous work, including longitudinal studies on the same returnees as the present study (Kubota, 2019; Kubota et al., 2022), have shown that, over time, these returnees exhibit signs of L1 reversal – (developmental) effects of re-exposure to the former minority/heritage language (see Flores and Snape, 2021) – as well as L2 attrition in the domain of morphosyntax and lexical access (see also Taura, 2019). In the present study, we ask whether we can observe and unpack the aforementioned in the domain of global foreign accent (GFA).

To this end, we recorded speech samples from Japanese–English bilingual returnees in L1 Japanese and L2 English at three time points: (1) upon return to Japan from an English majority language environment; (2) one year after return; and (3) five years after return. We then conducted two accent-rating tasks in which L1 speakers of English and Japanese, respectively, rated speech samples from the three time points. Although longitudinal data are often absent in global accent studies, they are ideal in that they permit the documentation of changes in GFA over time using the participants' own data as a reference point. In addition to avoiding general issues of comparative fallacies in bilingualism research (Ortega, 2013; Rothman et al., 2022) this is particularly important in a context such as returnee bilingualism that is subject to substantial individual variation (for review, see, for example, Kubota, 2019). As we examine these developmental trajectories, we seek to uncover some of the sources of this variation. Specifically, to better understand the factors that determine the rate and degree of potential GFA change in Japanese and English, we model the effect of two individual factors posited to affect accent development in the L1 and L2 of returnees: Age of onset (AoO) of L2 English and exposure to L2 English (relative to L1 Japanese) while abroad.

II Background

I Global accent in bilingual children

Most existing studies on GFA in bilingual children are centered on heritage speaker bilinguals in the majority language (ML) setting, focusing on one of the two languages. The body of heritage bilingual GFA research suggests that GFA in the ML is modulated by age of first exposure and length of exposure (e.g. Asher and Garcia, 1969) and GFA in the HL is modulated by input quality and quantity (e.g. Wrembel et al., 2019). In the only

study to our knowledge to examine how age and experiential factors differentially affect both languages in bilingual children, Kupisch et al. (2021) found age effects in the accent ratings of Russian–German bilingual children in Germany. When speaking their ML (German), primary school children (7–9 years old) received lower GFA ratings than pre-school children (4–6 years old), while the opposite was true for the HL Russian: younger children were deemed less foreign-accented than the older children. Based on these results, the authors proposed that primary school entry (which, in their study, coincided with a steady increase in majority language exposure and use) might be a critical point for heritage language maintenance and dominance shift.

In what appears to be the only study to test GFA in returnees, Flores and Rato (2016) collected accent ratings on the Portuguese of 20 returnees who were raised in Germany by Portuguese parents and returned to Portugal between 11 and 29 years of age. The authors found that age of emigration to Germany (i.e. L2 AoO, corresponding to the age of departure from Portugal) was the only significant predictor of variability in L1 Portuguese GFA upon return to Portugal. Other factors, such as length of residence in the ML German environment, did not emerge as a significant predictor. The deterministic role of AoO of the L2 (German) suggests that children who left Portugal at a young age or were born in Germany (age of emigration ranges from zero to 7 years) and developed a foreign accent in their HL Portuguese will continue to have a foreign accent even if they are re-immersed in the HL later in life.

In sum, the above studies point to the critical role of age of onset (AoO) and language exposure to the majority L2 on accent trajectory in the L2, and potentially also in the (minority) L1. However, since none of the studies provided longitudinal data and, with only one study reporting data from both languages, we do not know how long it takes until we see potential effects of dominance shift in these early teenagers' productions. Bilingual returnees offer a unique opportunity to study the effects of dynamic, shifting language experience.

2 Factors that affect global foreign accent

Examining the effect of age of onset (AoO) on L2 acquisition has been a central focus in the L2 literature, as it provides insights into the key question of whether there are maturational constraints in language learning and whether the ultimate proficiency level will be different from that of a monolingual L1 speaker (Abrahamsson and Hyltenstam, 2009; Birdsong and Molis, 2001; Granena and Long, 2013; Rothman, 2008; Hartshorne et al., 2018). While it is not clear that GFA is an optimal domain to adjudicate between various proposals for what specifically underlies the significance of AoO in bilingual studies – e.g. global maturational constraints on acquisition/learning mechanisms,² specific physiological constraints on phonetic/phonological perception and/or articulation, degree of susceptibility to cross-linguistic influence and comparative opportunities for convergence given reduced input, a combination thereof or other variables – there is no denying that much research shows a robust correlation with AoO for GFA. That is, the later the AoO of the L2, the stronger the GFA is at the end state of the acquisition process (Abrahamsson and Hyltenstam, 2009; Flege et al., 1995, 1999).

Although there is consensus that AoO plays a central role in accent development (Flege, 1995), some studies underscore that it is not an *a priori* deterministic factor, not least as it is often confounded with other variables such as exposure, use, and length of residence (for an overview, see Jesney, 2004). Illustrative evidence in this vein comes from studies on L1 attrition (as defined more globally; see Schmid and Köpke, 2017), which have found that AoO interacts with experiential factors but is not necessarily the cause of increased GFA in the L1 of bilingual attriters. For example, evidence from L1-German (Hopp and Schmid, 2013) and L1-Turkish speakers (Karayayla and Schmid, 2019) who moved to a majority (L2) English environment has shown a significant, but not categorical, effect of L2 AoO on L1 accent. In Karayayla and Schmid, 2019, age interacts with external factors, particularly language exposure/engagement, in accounting for variation in GFA. In addition, a study by Yeni-Komshian et al. (2000) demonstrated that when controlling for L2 AoO (range 1–23 years) and length of residence, Korean–English immigrants in the USA with the highest English and lowest Korean pronunciation rating used English more often (and consequently less Korean) than the other bilinguals. Such findings lend support to the idea that age-related changes in GFA may be a consequence of how much competition there is between the two languages (MacWhinney, 1987, 2012, 2018) and to what extent L1 entrenchment has taken place (Steinkrauss and Schmid, 2017). That is, the more entrenched the L1 is as a function of more exposure, the less a speaker’s L1 phonological system is likely to be influenced by the L2, resulting in a more stable L1 system (see also Flege, 1987).

The effects of language exposure on GFA have both been present in adult and child bilinguals who have grown up as early or simultaneous bilinguals. Specifically, more HL exposure and/or use leads to less perceived GFA in the HL for adult (Kupisch et al., 2014; Lloyd-Smith et al., 2020) and child (Uzal et al., 2015; Wrembel et al., 2019) bilinguals. The results are mixed, however, in terms of whether the development of accent in the majority language and heritage language confers an inverse relationship (i.e. native-like accent in the ML comes with a cost of foreign-accentedness in HL and vice versa). While some studies show that stronger GFA in the ML correlates with higher use of the HL (Flege et al., 1997; Uzal et al., 2015), others found no relationship between GFA in ML and HL use (Lloyd-Smith et al., 2020).

In light of the research reviewed in this section, what might we predict for returnees? The cross-sectional returnee GFA data from Flores and Rato (2016) reviewed above suggest that departure from the majority L1 environment at an earlier age can yield reorganization of L1 speech production that cannot be further developed through re-immersion/re-exposure, given that only age of departure from the L1 environment (and not how long they have been back in the L1 environment) predicted their L1 GFA variability (see also de Leeuw et al., 2023, who show that earlier L2 AoO predicts L1 perceptual attrition). Despite the deterministic effect of age of onset, the returnees in Flores and Rato (2016) still performed better than L2 speakers (and closer to monolinguals’ GFA ratings), which corroborate the findings from studies with Spanish–English childhood overhearers (Au et al., 2002) and Korean–English heritage speakers (Oh et al., 2003) in which both bilingual groups performed better on phonemic tasks than their L2 control groups who had started learning Spanish/Korean in (young) adulthood.

With that said, it remains to be seen whether exposure modulates these L2 AoO effects on the HL within a longitudinal paradigm in which we can establish a clear reference point for each individual rather than probing for correlations at the group level. That is, can re-immersion experience in the L1 environment mitigate any age (of departure from L1 environment) effects for L1 GFA? Flores and Snape (2021: 362) further speculate that:

[Heritage speakers] with less exposure to their HL during migration may, in fact, benefit from the return and show stronger signs of accent change in their HL [in the direction of a monolingual baseline] after living for some time (back) in the home country.

The current study extends Flores and Rato's work by establishing an individual reference point of returnees' GFA (at the point of return to the L1 environment) in both languages, allowing us to truly isolate the effects of age-related factors (L2 English age of onset, i.e. departure from the L1 environment, henceforth: 'L2 English AoO') and exposure-related factors (exposure to majority L2 English in proportion to heritage L1 Japanese whilst away from the L1 environment, henceforth: 'Exposure to L2 English') on GFA in the L1 and L2, rather than examining these effects at a single point in time cross-sectionally.

3 *Speech features contributing to GFA in English and Japanese*

While a listener's perception of a GFA is subjective, this perception often correlates with objectively measurable acoustic properties of the L2 speech (Trofimovich and Baker, 2006; Wayland, 1997), although some studies fail to find such a correspondence (see Bergmann et al., 2016).

Although a detailed acoustic analysis into the phonological correlates of GFA was beyond the scope of this study, we examined if there were specific speech features (consonants, vowels, voice quality, rhythm, and intonation, here) that contributed to English and Japanese raters' perception of a strong GFA (see Section III). This examination was motivated by previous studies that suggest that the speech features contributing to GFA in English and Japanese may be language-specific, as we summarize hereunder.

In the contexts of L2 English spoken by L1 Japanese speakers and L2 Japanese spoken by L1 English speakers, previous research has isolated several features that are good candidates as contributors to GFA perception. For example, Riney and colleagues identified several features that correlate with perceived GFA in the L2 English of adult L1-Japanese speakers, including voice onset time (VOT) of plosives (Riney and Takagi, 1999) and realization of /l/ and /r/ liquids (Riney et al., 2000). In the only study to our knowledge to investigate the phonological correlates of perceived GFA in L2 Japanese spoken by L1 English speakers, Idemaru et al. (2019) found that among segmental (vowels and plosives) and suprasegmental features (rhythm, tone, and fluency), tone (operationalized by the proportion of pitch-accent errors) contributed most strongly to GFA.

These findings are intuitively plausible when considering the phonological systems of both languages. For instance, Japanese voiceless plosives typically have shorter VOTs than English (Riney et al., 2007), and Japanese does not have an English-like /l/-/r/ distinction, making the latter notoriously difficult to produce for L1-Japanese speakers

(Flege et al., 1995; Larson-Hall, 2006). With five vowels and 14 consonant phonemes, the segmental makeup of Japanese is relatively small compared to the larger segmental inventory of English (General American English has 24 consonant phonemes and 11 vowel phonemes), which may complicate production of English segments for Japanese speakers. By contrast, Japanese has phonemic vowel and consonant length, is mora-timed, and has lexical pitch which shapes phrase-level intonation, unlike English. These suprasegmental features of Japanese have been reported to be difficult to acquire for L2 learners of Japanese (Hirata, 2015).

Because of the different phonological makeup of both languages, it is logical to posit that the phonological correlates of a perceived GFA in English are not the same as in Japanese. In particular, segmental properties may have a greater weight on perceived GFA in English, whereas suprasegmental properties may have a greater weight on perceived GFA in Japanese. For instance, Komatsu and Kimoto (2008) reported that L1-Japanese raters predominantly attributed suprasegmental features (intonation) to perceived GFA in Japanese spoken by Japanese heritage speakers living in Brazil. Moreover, Riney et al. (2005) showed that even when rating foreign accent in a second language (English), L1-Japanese raters primarily used suprasegmental parameters to make perceptual judgments.

4 Research questions

Based on the above, we ask the following research questions:

- Research question 1: What are the trajectories of L1 Japanese and L2 English GFA ratings of returnees over three time points spanning five years, and what is the relationship between changes in L1 Japanese versus L2 English?
- Research question 2: What are the modulatory effects of L2 English AoO and exposure to L2 English on GFA trajectories over time?
- Research question 3: What speech features contribute to the perception of a strong GFA in English and Japanese, and how do they differ between English and Japanese raters?

5 Hypotheses

Based on the reviewed literature, our hypotheses and specific predictions for each research question are as follows:

- Hypothesis 1: Bilingual returnees' GFAs are dynamic and shift according to context in a mirror-image pattern. That is, in an L1 context, re-immersion yields a decrease in L1 GFA while L2 attrition yields an increase in the L2 GFA.
- Prediction 1: The pattern of English GFA changes will be the inverse of Japanese GFA changes: Over five years, L1 re-stabilization due to increased L1 exposure will yield a decrease in L1 Japanese GFA and a combination of L1 interference and a decrease in L2 exposure will yield an increase in L2 English GFA.

- Hypothesis 2: L1 and L2 maintenance are constrained by AoO as well as relative exposure to the L2.
- Prediction 2: L2 English AoO and relative exposure to English will predict individual differences in L1 Japanese and L2 English GFA ratings over time.
- Hypothesis 3: Segmental elements are the primary determiners of perceived English GFA; suprasegmental elements are the primary determiners of perceived Japanese GFA.
- Prediction 3: English raters will report greater reliance on segmental elements while Japanese raters will report greater reliance on suprasegmental elements to indicate a strong GFA.

III Methods

I Participants

a Speaker data: Returnee data. The University of Edinburgh Linguistics and English Language Ethics Committee (protocol number 11-1516/5) approved this study. Speech data from 17 returnees were used. The participants were recruited through Japan Overseas Educational Services (JOES). All returnees were born in Japan and had parents who were L1 speakers of Japanese. They had minimal exposure (i.e. language classes) to English before leaving Japan and acquired English after arrival to their new environment abroad. Eight participants lived in an environment where English is the dominant language of the society. While the other nine lived in China, France, or Germany, they all attended schools with English as the sole medium of instruction, and their parents reported that their children could not hold a conversation using the national language (e.g. Chinese, French, German, respectively) while they could do so in English.³ While abroad, the children continued to be exposed to Japanese at home. Upon their return to Japan, all children were enrolled in a Japanese school and were educated under the curriculum set by the Japanese Ministry of Education.

The L2 English age of onset (AoO) ranged from 1 to 9.73 years ($M=5.14$, $SD=2.59$). Exposure was measured by The Bilingual Language Experience Calculator (Unsworth, 2013). This questionnaire quantifies language exposure via extensive questions about whom and how much the child spends time with on an average day in the week, which languages each person uses when addressing to the child and vice versa, and how much time the child spends on extra-curricular activities and in which languages. It also assesses when they started learning the languages and how long they have been using them. The returnees' mean proportion of exposure to English (relative to Japanese) during their stay abroad was 0.48 ($SD=0.14$), which means that they were exposed slightly less to English than to Japanese. An overview of the 17 returnee bilinguals and their demographics is provided in Table 1.

b Speaker data: 'Baseline' data. In addition to the returnees' speech samples, we included samples in the English rating task from 17 monolingual speakers of American English (Mean age=10.08, $SD=3.59$, range=5.67–18.00) and samples in the Japanese rating task from 14 monolingual speakers of Japanese (Mean age=9.01, $SD=0.96$, range=6.60–10.33) and three heritage speakers of Japanese, of whom two were dominant in German

Table 1. Information on the 17 returnee bilinguals.

	Mean	SD	Min	Max
Age at first session (years)	10.02	1.72	7.65	12.99
Age at second session (years)	11.03	1.48	8.68	14.01
Age at third session (years)	15.22	1.48	13.10	18.43
Length of residence abroad (years)	4.27	2.28	2.00	9.74
Age of onset to L2 English	5.14	2.59	1.00	9.73
Exposure to English (relative to Japanese) while abroad	0.48	0.14	0.31	0.85

and one in Norwegian, and who had impressionistically strong GFAs in Japanese (Mean age=9.23, SD=4.34, range=5.23–13.23). These groups were age-matched to the returnee sample (Age of three test sessions: M=12.09, SD=1.56, range=7.65–18.43).

We included these ‘baseline’ speech data to validate the reliability of the rating tasks, as we would expect monolingual speech to receive low foreign-accent ratings. The inclusion of the impressionistically strongly foreign accented Japanese heritage speech samples in the Japanese task was motivated by an observation by the authors and five raters in a pilot study that the returnees’ Japanese GFA was low overall. In such a case, raters may use the scale more critically and/or find the task monotonous (see Schmid and Hopp, 2014). Therefore, a few strongly foreign-accented Japanese samples were used to define the higher extreme of the scale and ensure that the raters could identify foreign-accented Japanese speech. We did not include additional L2 English samples (for instance, spoken by non-Japanese L1ers) because the returnees’ English was L2 English to begin with, and because our impression and that of five pilot raters was that the returnees’ English was overall more foreign-accented than their Japanese.

The inclusion of these baseline data also allowed us to validate that any observed change in the returnees’ foreign accent in Japanese (which we expect to decrease over time) is in fact a result of re-immersion into a Japanese majority environment, and not simply a concealed result of aging. It is possible that raters would perceive child-like speech in younger speakers to be indicative of a ‘foreign’ accent. To ensure that age did not strongly affect rater judgements, we examined whether within the monolingual data, there were notable differences in ratings according to age (see Section IV.1.b).

c Speech sample collection. Speech samples were taken from elicited narratives of a pictureless children’s book: *Frog on his own* (Mayer, 1973) for Japanese and *Frog, where are you?* (Mayer, 1969) for English narratives. Unless otherwise noted, all recordings were made using the second author’s laptop PC in a silent environment. For the ‘baseline’ samples, eight of the 17 English monolinguals were recorded in-person in the USA; the remaining nine recordings were taken from the CHILDES Frogs English ECSC corpus (Kallay and Redford, 2021) using the same wordless picture book as with the other English monolingual children. The Japanese monolingual children were all recorded in-person in Japan.

The returnees’ samples were recorded at three time points: time 1 (within a few weeks of the participants’ return); time 2 (a year after return); and time 3 (five years after return).

The language order of the narratives was counterbalanced across participants for all test sessions, and instructions were given in the respective languages by a Japanese–English bilingual researcher. Recordings took place at the participant’s home or JOES classrooms for the first and second test sessions while the third test session was conducted online via Zoom and audio was recorded via Audacity® on the researcher’s laptop (Audacity, n.d.).

d Raters. Forty-four L1 speakers of US American English (29 females, mean age=43.34, SD=7.57) and 46 L1 speakers of Japanese (six females, mean age=39.00, SD=6.81) rated the accent of the children in the respective languages. The English-speaking raters were recruited on *Prolific* (prolific.com) and the Japanese-speaking raters on *Lancers* (lancers.jp). All raters were L1 speakers born and raised in the respective country of origin and reported that they were parents who were familiar with child speech. Raters were paid a token fee for their participation.

2 Procedure

a Stimuli creation. Ten-second samples were extracted from the audio recordings to be used in the accent-rating task. Care was taken to ensure that the audio samples contained as much speech material as possible, that they did not contain clear grammatical errors or disfluencies, or start or end abruptly in the middle of an utterance. The samples were saved as .mp3 files and their average intensity was set to 70 dB, using the ‘scale intensity’ command in *Praat* (Boersma and Weenink, 2019). In total, the Japanese list comprised 68 sound samples (17 returnees × 3 time points and 17 baseline samples) and the English list comprised 83 samples (17 returnees × 3 time points + 17 baseline samples + 15 samples taken from three returnees used for a separate study).⁴

b Accent-rating task. The task was carried out online on *Gorilla.sc* (Anwyl-Irvine et al., 2020). Headphone checks (Woods et al., 2017) prior to the accent-rating task ensured that the raters were using headphones and were in a quiet environment.

Participants were provided with written instructions in their L1 and informed that they would be listening to short samples of child speech. They were asked to indicate how foreign-accented the children sounded on a scale of 1 (‘no foreign accent’) to 9 (‘very strong foreign accent’). ‘Foreign accent’ was described as ‘speech that contains features that would be unnatural for a native speaker of English or Japanese’. In light of findings that grammatical errors can influence accent ratings (e.g. McDermott, 1986) particularly with samples from L1 speakers (Hanulíková et al., 2012), the instructions emphasized that child speech is unlike adult speech, and that there could be small grammatical mistakes or disfluencies, but that these should not be considered when assigning a foreign accent rating. The instructions also emphasized that the raters should pay attention to foreign accent and disregard any regional or dialectal differences in English or Japanese.

The raters started with three practice items to familiarize themselves with the task format and the range of degree of accentedness represented in the task. The practice samples, which were from speakers not included in the main task, were of two returnee

speakers with impressionistically mild and strong foreign accents, respectively, and a monolingual speaker. After the practice, the raters were reminded in the written instructions to pay attention to foreign accent only, and that child speech differs from adult speech. At this point, they were also informed that, in some cases, they might be asked to indicate the feature(s) of the speech sample that contributed to their foreign accent choice. The set of features to choose from included vowels, consonants, intonation, rhythm, and voice quality. Each feature was briefly explained and illustrated in the instructions. Given that a pilot study indicated that presenting the follow-up question after each rating trial would make the task too tedious, and since we are most interested in features that contribute to the perception of a strong GFA, this follow-up question (which required the rater to indicate at least one feature) was limited to responses of ‘8’ or ‘9’ <Very strong foreign accent> on the scale.

In each trial, the 10-second sound fragment would play, after which the 9-point horizontal Likert scale was displayed with ‘1’ at the left extreme and ‘9’ at the right. Participants could replay the fragment once before selecting a point on the scale. In addition to the 68 or 83 speech samples, there were four attention checks, in which a rater would hear ‘This is an attention check, in the next window, please select [number] on the scale’. All raters passed these attention checks. The 68 or 83 sound samples and attention checks were presented in a fully randomized order per rater. There was a brief break halfway through the task, and raters averaged 22 minutes to complete it. After the task, the raters filled out a debriefing questionnaire about their language background and exposure to child speech and foreign accent and were given the opportunity to leave any comments they had regarding the task.

The online accent-rating tasks are available in the supplemental material.

c Statistical procedures. All analyses were performed in *R* 4.2.2 (R Core Team, 2022). Figures were generated with the *ggplot2* package (Wickham, 2016). We present descriptive statistics, and results from Bayesian inference address our research questions. Bayesian inference offers an alternative to frequentist analyses in that it includes a prior specification of assumed beliefs of a model parameter. The output of a Bayesian model is a posterior distribution, which contains updated model parameters after having been fitted on the data. This posterior distribution generates 95% Credible Intervals (CrIs), which indicate the range of parameter values within which one can be 95% certain that the true parameter value lies. The posterior distribution also generates maximum probabilities of direction (pd), which describe the probability that a parameter is positive or negative. Our choice for Bayesian inference in the present study was motivated by the observation that it enables the fitting of relatively complex models on relatively small data sets (Haendler et al., 2020), and that it is suitable for analysing Likert scale data points as a dependent variable (Douven, 2018).

We fitted our models using the *brms* package (Bürkner, 2018). Following common practice (Haendler et al., 2020; Vasisht et al., 2018), models were constructed using weakly informative priors, with prior specification in *brms* set as (0, 3) for ‘Intercept’; (0, 1) for ‘b’; (0, 0.1) for ‘sd’ priors and Lewandowski–Kuroiwicka–Joe (LKJ) distribution (2) for correlation priors (Coretta et al., 2022). Four sampling chains with 3,000 iterations each were run, with 1,500 warm-up iterations. Model diagnosis was carried out by

observing R_{hat} values (i.e. ensuring these were close to 1), effective sample size (ESS) values (ensuring these were at least $100 \times$ the number of sampling chains), and by inspecting posterior draws using the `pp_draws()` command of the *brms* package.

Our first model ('the accent-rating model') addresses research questions 1 and 2 and investigates the effects of time, language, and the two experiential factors (L2 AoO and L2 Exposure) on accent-rating. This model had the dependent variable *Response on Likert scale* (4,590 observations) and was fitted with a cumulative logit distribution (Douven, 2018). The model contained fixed effects for *Language* (English, Japanese; sum contrast-coded), *Time* (1, 2, 3; sum contrast-coded), *L2 AoO* (age of onset of L2 English; centered and scaled), *L2 Exposure* (Proportion of exposure to L2 English relative to L1 Japanese during stay abroad; centered and scaled), and three-way interactions with *Language*, *Time* and each of the two experiential factors, as well as the derived two-way interactions. The random effects structure contained a random intercept for *Subject* (Returnee) and a random intercept for *Rater*.⁵

The second model ('the feature model') addresses research question 3 and investigates the distribution of responses to the follow-up multiple-choice question ('Which feature(s) contributed to your strong accent rating?') that was presented to raters if they rated a sample with a high foreign accent rating (8 or 9). This follow-up question was presented 291 times (to 35 different raters) in the English group and 126 times in the Japanese group (to 27 different raters). To model the relative count of responses (vowels, consonants, intonation, rhythm, vowel quality) per question (417 observations) between the English and Japanese group, a model was fitted with a zero-inflated Poisson distribution (Winter and Bürkner, 2021). It contained *Language* (English, Japanese, sum contrast-coded) and *Response* (vowels, consonants, intonation, rhythm, vowel quality, sum-contrast coded) as fixed factors, and a two-way *Language:Response* interaction. The random-effects structure consisted of a by-response random slope for *Rater*.⁶

To investigate the nature of interactions in the models, planned comparisons were carried out using the *emmeans* package (Lenth, 2020). In Section IV, we highlight findings for which the 95% Credible Interval (CrI) of the effect estimates as provided by the posterior distribution did not contain zero, and for 95% CrIs that did contain zero but that had a relatively high maximum probability of direction (pd).⁷ We take such findings to be 'suggestions' of an effect (Nicenboim et al., 2018). For planned comparisons, we highlight findings for which zero was not included in the 95% highest posterior density (HPD) as calculated by the *emmeans* package. The complete statistical results (posterior distributions and multiple comparison tables) are reported in the supplemental material.

IV Results

I Descriptive statistics

a *Accent rating of returnees*. Internal consistency of the ratings in each language, as calculated by Cronbach's alpha in the *ltm* package (Rizopoulos, 2006), was 0.916 for the

Table 2. Accent rating per language and per time.

	Time 1	Time 2	Time 3
English	4.23 (1.50)	4.14 (1.19)	4.47 (1.44)
Japanese	4.18 (1.08)	3.82 (0.79)	3.53 (0.91)

Note. Values are means with SDs in brackets.

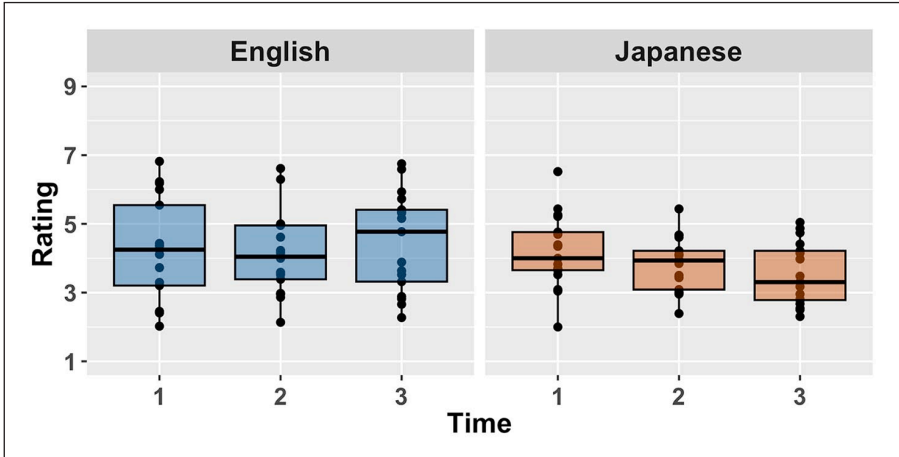


Figure 1. Boxplots showing perceived foreign accent (where 1 is ‘no accent’ and 9 a ‘very strong foreign accent’) for returnees per language and time.

Note. Dots represent individual speakers.

English and 0.908 for the Japanese ratings, indicating high internal consistency among both English and Japanese raters. Table 2 and Figure 1 show the ratings per language and per time. Overall, a subtle increase (higher values) in perceived GFA can be observed for the English speech over time, whereas a subtle decrease in perceived GFA (lower values) can be observed for the Japanese speech.

b *Accent rating of ‘baseline’ speakers.* The mean perceived GFA for the monolingual English children was 1.71 (SD=0.35), i.e. lower compared to the bilingual participants. For monolingual Japanese children it was 3.70 (SD=0.83), which was lower compared to the heritage Japanese (5.85; SD= 1.55) and the returnee participants at times 1 and 2. Baseline and returnee data are combined in Figure 2.

To verify whether age at the time of testing affected accent rating in the monolingual speakers (as we deemed it possible that raters would perceive child-like speech in younger speakers to be indicative of a ‘foreign’ accent, and that therefore any decrease in GFA, which we expected for the Japanese speech, could simply be an effect of aging), we fitted two separate models per language with *Age* as a fixed effect to predict the

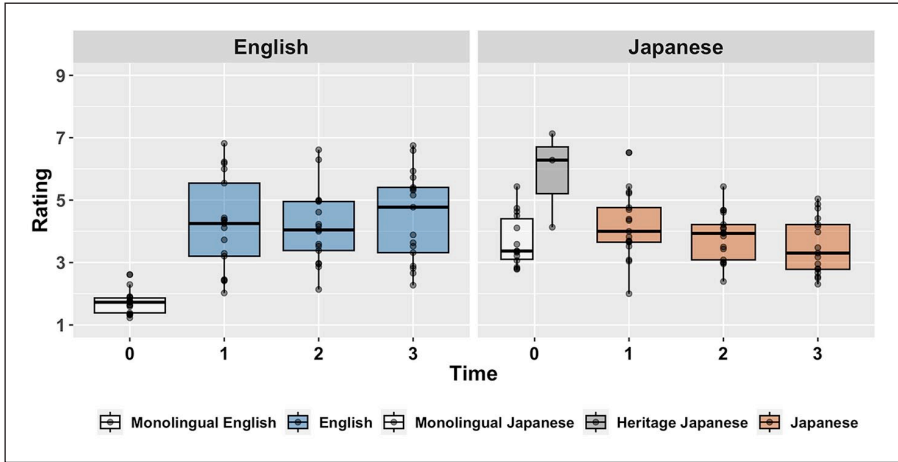


Figure 2. Boxplots showing perceived foreign accent for returnees and ‘baseline’ speakers (where 1 is ‘no accent’ and 9 a ‘very strong foreign accent’).
 Note. Dots represent individual speakers.

accent rating for the monolingual speech data.⁸ The posterior distribution suggested that *Age* affected accent ratings for American raters listening to monolingual English speech, with lower accent ratings given to older monolinguals, $b = -0.11$ ($-0.16, -0.05$). There was no suggestion that age affected accent ratings for Japanese raters listening to monolingual Japanese speech, with zero roughly halfway within the 95% Credible Interval, $b = -0.03$ ($-0.33, 0.27$).

2 Model results

In line with our research questions, we will first present model results from the accent-rating model regarding the effects and interactions of *Language* and *Time* on accent rating, as well as correlation coefficients between change in accent rating in English and Japanese (research question 1). We will then address the additional effects of *L2 English AoO* and *Exposure to L2 English* as found in the accent-rating model (research question 2), and finally present the results from the feature model (research question 3).

a RQ1: Effects of language and time point on GFA. Figure 3 shows the predicted GFA rating per time point and language for the returnee samples. The accent-rating model suggested an interaction between *Language* and *Time*, $b = -0.17$ ($-0.25, -0.11$). Multiple comparisons of time points per language suggested no change in English GFA rating from time 1 to 2, $b = -0.08$ ($-0.24, 0.12$) but an increase from time 2 to 3, $b = 0.31$ ($0.13, 0.50$). There was a decrease in Japanese GFA from time 1 to 2, $b = -0.31$ ($-0.48, -0.15$) and time 2 to 3, $b = -0.27$ ($-0.43, -0.10$). At time 3, Japanese GFAs were lower than English GFAs, $b = -0.79$ ($-1.17, -0.39$), unlike time 1, $b = 0.02$ ($-0.35, 0.44$) and time 2, $b = -0.21$ ($-0.50, 0.17$), for which there were no between-language differences.

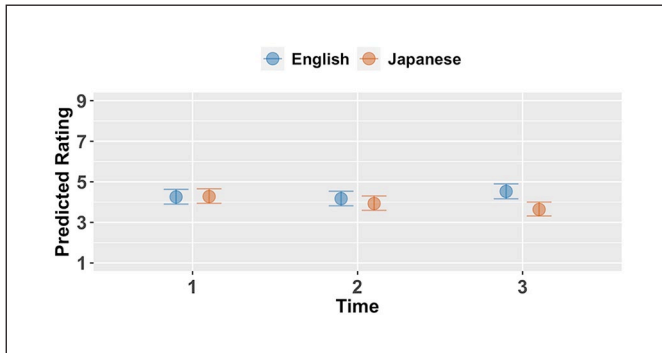


Figure 3. Predicted foreign accent rating per language and time point.

b RQ1: Correlation between change in English and Japanese accent rating. To determine whether changes in English GFA ratings correlated with changes in Japanese GFAs, we obtained the Pearson's coefficients for the point change in mean rating for each individual over time between the English and Japanese samples. There was no significant correlation between English and Japanese change from time 1 to time 2, $r(15) = -0.01, p = .976$; nor from time 2 to time 3, $r(15) = 0.13, p = .604$.

c RQ2: Effect of experiential factors on accent rating. Here, we examine the effect of L2 English AoO and L2 exposure on changes in GFA over time (research question 2). As for L2 English AoO, the accent-rating model revealed a moderate suggestion for a three-way interaction between *Language*, *Time*, and *L2 English AoO*, with zero included in the posterior distribution, $b = 0.05$ ($-0.02; 0.12$), and a probability of direction of 91.30%. Estimates for the effect of *L2 English AoO* per *Language* and *Time* should therefore be interpreted with caution. The estimates suggested that returnees with a later L2 English AoO had stronger GFAs in English for time 2, $b = 0.26$ ($0.04; 0.49$). The estimates further suggested that returnees with a later L2 English AoO English had milder GFAs in Japanese for time 1, $b = -0.42$ ($-0.64, -0.20$) and time 2, $b = -0.34$ ($-0.56; -0.12$). This is visualized in Figure 4. Overall, however, the accent-rating model revealed a two-way interaction between *Language* and *L2 English AoO*, $b = 0.25$ ($0.20; 0.30$). The estimates per language suggested that a later L2 English AoO led to stronger GFAs in English, $b = 0.16$ ($-0.02; 0.37$), with a probability of direction of 95.67%, and weaker GFAs in Japanese, $b = -0.32$ ($-0.51; -0.13$), averaged over the three time points.

As for L2 Exposure, the accent-rating model revealed a three-way interaction between *Language*, *Time*, and *L2 Exposure*, $b = -0.19$ ($-0.26; -0.12$). Estimates for the effect of English exposure per language and time point suggested that returnees with more exposure to English during their stay abroad had weaker GFAs in English for time 1, $b = -0.45$ ($-0.67; -0.24$) and time 3, $b = -0.29$ ($-0.26; -0.06$). In Japanese, the estimates suggested that returnees with more exposure to English were perceived as having stronger GFAs at time 1, $b = 0.26$ ($0.07; 0.48$), but not at times 2 and 3, with zero included in the 95% HPD. This is visualized in Figure 5.

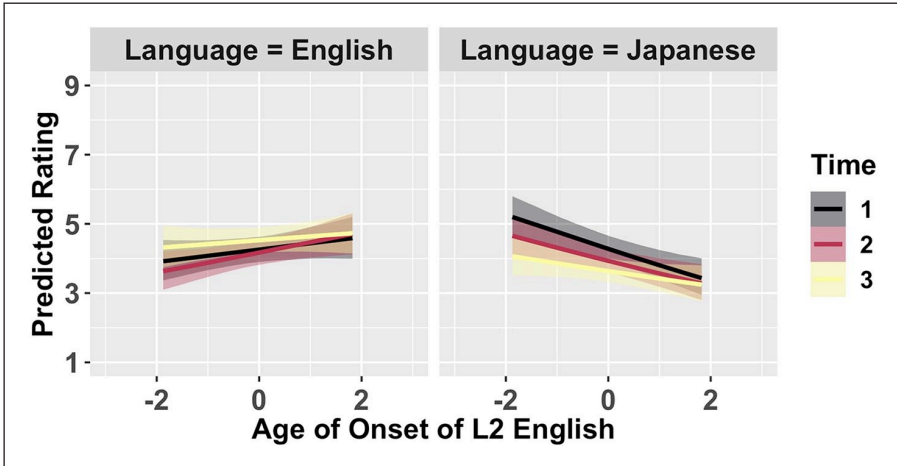


Figure 4. Predicted accent rating by second language (L2) age of onset (AoO), per time point and per language.

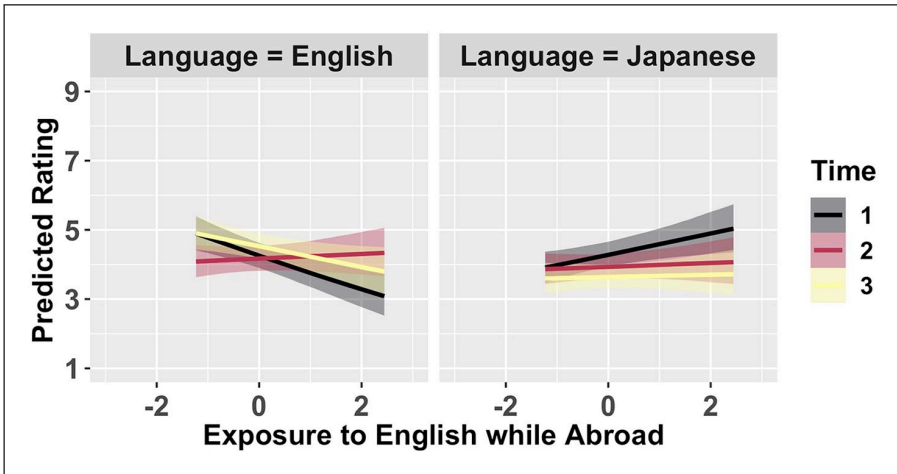


Figure 5. Predicted accent rating by second language (L2) Exposure, per time and per language.

d RQ3: Speech features contributing to perceived GFA. The feature model, which investigated the count of responses to the follow-up question ‘What feature(s) contributed to your strong accent-rating?’ between the English and Japanese groups suggested an *L1:Response* interaction, $b = -0.27$ ($-0.44, -0.11$). Multiple comparisons per response choice between languages suggested that English raters relatively more often responded with ‘vowels’ than did Japanese raters, $b = 0.64$ ($0.09, 1.19$). By contrast, Japanese raters relatively more often responded with ‘intonation’, $b = 0.41$ ($0.04, 0.78$), ‘rhythm’, $b = 0.71$

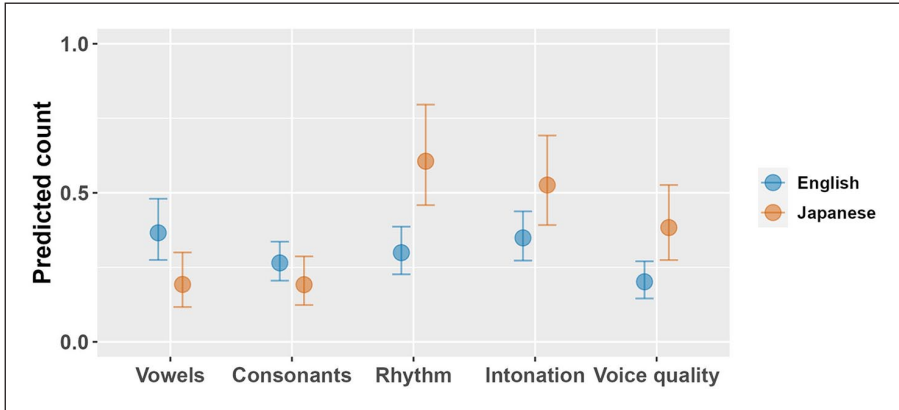


Figure 6. Predicted count of responses (averaged per question) to follow-up question ‘What feature(s) contributed to your strong accent rating?’, presented if raters indicated ‘8’ or ‘9’ on the foreign accent scale.

(0.35, 1.21) and ‘voice quality’, $b=0.64$ (0.20, 1.07) than did English raters. This is visualized in Figure 6.

V Discussion

1 Development and change in global accent

The first part of research question 1 asked what shape the trajectories of GFA ratings took in Japanese–English returnee bilinguals over a period of five years. The accent-rating data showed changes in GFA over time in opposite directions, with a continuous decline in Japanese GFA between one and five years after returning to Japan and an increase in English GFA between three and five years after returning, thus lending support to Hypothesis 1. It is important to note that some individuals exhibited patterns that differed from the aggregate data in the form of an increase in GFA in Japanese and/or a decrease in GFA in English. We discuss the factors that might explain the attested variability below in relation to research question 2.

The increase in L2 English GFA between time point 2 (three years) and time point 3 (five years) is in line with previous L2 attrition work in other domains of grammar in these same returnee children and teenagers (e.g. morphosyntax and lexical access; Kubota, 2019; Kubota et al., 2022). Although L2 lexical and fluency effects have been reported after only half a year (Flores, 2015; Kuhberg, 1992; Tomiyama, 1999) and considerable changes in L2 morphosyntax can surface after approximately one year (Flores, 2010; Kubota et al., 2022; Snape et al., 2014; Tomiyama, 2000), speech production might be more resistant to change. For instance, Tomiyama (1999, 2000), who tracked the L2 English attrition trajectory of a Japanese returnee child over the course of 33 months, observed only four instances of mispronunciation in production data from a variety of tasks. Our results seem to suggest that it takes between two to five years until the L2 phonological system of this bilingual population starts to be affected.

L1 Japanese GFA ratings, by contrast, decreased substantially after only a year back in Japan. This suggests that effects of re-immersion in the former L1 environment on GFA surface earlier in the L1 than in the L2. A comparison of these GFA data with the same returnee sample's data in other domains of language (Kubota, 2019; Kubota et al., 2022) reveals a similar pattern for picture naming latencies, whereby latencies in L1 Japanese were shorter at time 2 than time 1 but L2 English latencies did not change. In contrast, Kubota et al. (2022) did not observe change in L1 Japanese in global linguistic measures including syntactic complexity, mean length of utterance, and lexical diversity. Taken together, these cross-domain findings in the returnees' L1 and L2 suggest that L1 re-immersion effects on L2 GFA take some time to set in, as with other domains of grammar. At the same time, L1 re-immersion yields comparatively rapid reduction of their L1 GFA. Such changes to the former HL/L1 are not visible in all aspects of the language tested and thus may be limited to global accent or to reaction-time based methods such as picture naming that require rapid integration of lexical information.

Of note here is that the difference in ratings between Japanese and English became greater with time. The returnees had similar GFA ratings in English and Japanese at time 1 (Japanese Mean=4.23; English Mean=4.18; difference=.05) but this difference increased at time 2 (Japanese Mean=3.82; English Mean=4.14; difference=.32) and became the greatest at time 3 (Japanese Mean=3.53; English Mean=4.47; difference=.94). This pattern indicates that the balance between the two languages (at least in terms of GFA) changes with increasing exposure to L1 Japanese such that balanced GFAs at the time of return become less balanced as L1 exposure increases and L2 exposure decreases. Kupisch et al. (2021) reported a similar trend in the GFAs of their bilingual heritage speaker sample, showing that, as age increased from preschool age to primary school age, GFA decreased in the ML (German) and increased in their HL (Russian). These mirror-image changes were attributed to relative exposure to the ML and HL once children start primary school, which comes with more formal (school type) ML, more social contacts in the ML context and, potentially, pressure to blend in with the majority.

Before we continue, we must note the difference in accent rating between the English and Japanese raters for the monolingual speech samples. Although English raters gave – as expected – very low foreign accent ratings to the monolingual English samples ($M=1.71$), the Japanese raters gave relatively high accent ratings to monolingual Japanese samples ($M=3.72$). Based on evidence of a relationship between rater familiarity with foreign-accented speech in their L1 and rating leniency (e.g. Hopp and Schmid, 2014; McDermott, 1986), we posit that the Japanese raters' lesser exposure to foreign-accented Japanese might explain the higher ratings of monolingual Japanese GFA compared to the ratings of English monolingual GFA. In the debriefing questionnaire, most American raters indicated regular exposure to foreign-accented English, either in their direct environment or via social media or television (29% of raters reported daily exposure, 40% reported weekly exposure, and 30% reported monthly or rare exposure). By contrast, none of the Japanese raters reported daily exposure to foreign-accented Japanese, 24% reported weekly exposure, and 76% indicated monthly or rare exposure. A second explanation for the discrepancy between English and Japanese ratings for monolingual samples is that – as we mentioned in Section III – the returnees' Japanese may not have been as strongly foreign-accented as their English, because English was

always their L2 and some returnees had later age of onset to English, while they were all exposed to Japanese from birth. Therefore, in absence of stark differences in global accent between monolinguals and returnee samples (observed more strongly for the English samples than the Japanese samples), the Japanese monolingual samples may have received higher accent ratings than the English monolinguals. Indeed, a study by Schmid and Hopp (2014) demonstrates exactly what we have speculated about above: (1) raters who are relatively unfamiliar with foreign accents tend to be stricter in their foreign-accent judgement and (2) the inclusion of more strongly foreign-accented samples lowers the overall foreign-accent judgements for non-native samples.

2 Relationship between GFA change in Japanese and English

The second part of research question 1 asked to what extent changes in GFA in Japanese would relate to changes in GFA in English. Although changes were observed in both languages over the five years, the correlation coefficients revealed that a decrease in perceived GFA in L1 Japanese was not necessarily accompanied by an increase in GFA in L2 English. This would suggest that there was no trade-off between L1 change and potential L2 attrition. Our longitudinal finding seems to contradict prior cross-sectional studies which demonstrated that more use of HL leads to stronger foreign accent in the ML (Flege et al., 1997; Uzal et al., 2015). Instead, our results appear to pattern with those of Fowler et al. (2008), whose study of VOT in simultaneous French–English bilinguals found that acquisition of one phonological system does not have to entail a decrease in the maintenance and development of the other system. However, since our study concerns a special group of bilinguals who have experienced multiple transitions between language environments, we need more studies that examine the relationship between L1 and L2 in a variety of populations, language combinations, and timing to uncover whether the processes of acquisition and attrition are mirrored in bilingual populations and under what conditions mirroring occurs.

3 Effects of experiential factors on accent development

In research question 2, we asked to what extent L2 English AoO (i.e. age of departure from the native L1 Japanese environment) and exposure to L2 English while abroad (relative to L1 Japanese) further shaped the trajectories of GFA in English and Japanese. As for the effect of AoO, we found that individuals who moved abroad and were exposed to a majority English-speaking environment at a relatively older age returned to Japan with greater GFAs in English than individuals who moved abroad at a younger age. These findings are in line with previous studies that find effects of L2 age of onset on L2 phonological development (e.g. Abrahamsson and Hyltenstam, 2009). Moreover, this effect of AoO was still evident at time 2, suggesting that earlier exposure to a majority L2 in life is beneficial to the maintenance of GFA at least a year after leaving the L2 dominant environment. However, this advantage appeared to attenuate by five years after return to the L1 environment.

Similarly, an earlier L2 English AoO (and by extension, an earlier departure from the L1 Japanese environment) led to a greater GFA in Japanese. As visualized in Figure 4,

the effect of AoO appeared to be particularly strong on GFA right after return to Japan. This is not surprising given that L2 AoO is one of the main predictors of degree of L1 attrition in speech perception (Ahn et al., 2017), global foreign accent (Hopp and Schmid, 2013; Karayayla and Schmid, 2019), and other aspects of grammar (Dragoy et al., 2019). Although the aforementioned studies tested the effect of L2 AoO at a single point in time (often when the participants had reached adulthood), a question that is central to our study is whether L2 AoO predicts individual variability in returnees' GFA at different points in time (the moment of return, one year after, and five years after) which all take place before young adulthood. We found moderate suggestion of a three-way interaction between language, time and AoO (probability of direction=91.30%), in which L2 AoO appeared to predict L1 Japanese GFA at time 1 and time 2, but not at time 3. The levelled slope at time 3 (Figure 4) suggests that after five years of intensive re-immersion in the L1 environment, the children (by then, early teenagers) were able to neutralize any effect of having left their majority L1 environment at an early age. These findings are in contrast to Flores and Rato (2016) who showed L2 AoO to be the sole significant predictor for L1 GFA (over length of residence/reintegration in the homeland). Our findings tentatively suggest that re-immersion in the L1 environment (i.e. 're-socialization') in late childhood/early teenage years can, with sufficient time, attenuate any effects of an early L2 AoO/age of departure from the L1 environment. This finding may not be surprising, given that all returnees in our study came back to Japan before puberty and thus were perhaps still in the phase in which they are able to re(develop) their accent over time. Our range of L2 AoO is in stark contrast to the returnees in Flores and Rato (2016) who returned to the homeland from puberty to adulthood (age 11–29 years). In order to disentangle the role that L2 AoO plays on GFA across the lifespan, we require a returnee population with much wider range of L2 AoO and examine whether those who returned to the L1 environment from puberty to adulthood can also improve their L1 GFA to the same extent as those who came back during childhood.

Our findings regarding the effects of language exposure corroborate previous work on the effect of exposure on GFA in the HL and the ML (Kupisch et al., 2014; Lloyd-Smith et al., 2020; Wrembel et al., 2019). The accent-rating model revealed that individuals with more exposure to English (relative to Japanese) during their stay abroad returned to Japan with less foreign-accented English than individuals who had less English exposure. Moreover, this effect was still present at time 3, five years onwards. This suggests that increased exposure to a majority L2 has long-lasting beneficial effects on L2 GFA, even after rather significant decreases in L2 exposure upon return to the L1 environment.

Whereas increased exposure to English while abroad reduced the returnees' perceived GFAs in L2 English, it increased their perceived GFAs in L1 Japanese. Individuals with more exposure to English (and thus less exposure to Japanese) returned to Japan with stronger GFAs in Japanese than individuals with less English and more Japanese exposure. However, this effect was only present at time 1, a few weeks after return. This suggests a rapid decrease in Japanese GFA despite limited exposure to their L1 (primarily in the home environment) while abroad. Taken together, our findings show that (1) environmental transitions from an L1 to an L2-dominant environment at a younger age and (2) more exposure to the L2 (versus L1) contribute to lesser foreign-accented speech in the L2 but greater foreign-accented speech in the L1. While the positive effect of greater L2

exposure in the majority L2 environment prevails even after five years upon return to the L1 environment, the disadvantageous effect of leaving the L1 environment at a younger age and receiving less L1 exposure diminishes after the returnees have spent some time (at least three years, herein) in their L1 environment.

4 Phonological features contributing to global accents

Finally, in research question 3 we asked what phonological features contributed to the raters' perception of a foreign accent by presenting follow-up questions in case raters selected '8' or '9' (very strong foreign accent). We emphasize here that, because these follow-up questions were only presented in these specific conditions, they are not representative of the entire rating task. Nevertheless, the distributions of responses (Figure 6) suggest an interesting difference in the features that listeners attributed to perceived GFA in Japanese and in English. Specifically, whereas English raters indicated that both segmental and suprasegmental features contributed to their perception of a foreign accent, Japanese raters appeared to chiefly list suprasegmental features (rhythm, intonation, and voice quality), and they did this more often than did English raters. This latter finding is in accord with studies that suggest that mainly suprasegmental features contribute to GFA perception in Japanese (Idemaru et al., 2019; Komatsu and Kimoto, 2008; Riney et al., 2005). These differences in the make-up of perceived GFA between English and Japanese may have to do with the phonological systems of the respective languages. As discussed in the introduction, Japanese has a relatively small segmental inventory but has a relatively complex suprasegmental system (Hirata, 2015). Thus, when asked to rate a 'foreign accent', Japanese listeners might attend most to deviations in these suprasegmental cues. By contrast, English listeners might rely on a greater ensemble of segmental and suprasegmental cues to detect a foreign accent. This cross-linguistic difference could also (at least partially) explain the discrepancy in the Japanese versus English monolingual accent rating, as discussed earlier. Such cross-linguistic differences in accent perception warrant future work that investigates the exact acoustic and phonetic correlates of global accent and how these differ across languages.

VI Conclusions

This study sheds new light on longitudinal changes in the domain of speech production by examining the development of GFA in Japanese–English bilingual returnees over the course of five years after return to Japan. Results from an accent-rating task confirmed that the returnees exhibited changes in accent ratings over time for both their L1 Japanese and L2 English, evidenced by a rapid and steady decrease in GFA for Japanese and an increase in GFA for English five years after return. We further observed that L2 AoO and L2 exposure during the stay abroad shape the trajectories of GFA in the long run after return to the L1 environment, and that the features that contribute to the perception of a GFA may differ cross-linguistically.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Jason Rothman, Maki Kubota, and Tanja Kupisch were funded via the Tromsø Forskningsstiftelse (Tromsø Research Foundation) Grant No. A43484: Heritage-bilingual Linguistic Proficiency in their Native Grammar (HeLPiNG) (2019–2023) as well as the AcqVA Auora Center grant.

ORCID iD

Tim Joris Laméris  <https://orcid.org/0000-0002-1365-3022>

Supplemental material

Supplemental materials for this article (data and codes; model output; online experiment) are available online at <https://osf.io/68eb4/>.

Notes

1. Since not all accents are foreign, and L1 accents which are regional or diastratic may be mistaken as foreign, some authors have used the more neutral term ‘global accent’ in the context of early bilingualism research (see Kupisch et al., 2014).
2. For a discussion on the lack of evidence from neuroscience for loss of brain plasticity and its implications for attributing/mapping so-called ‘maturational effects’ in observable age-related linguistic differences directly to the mind/brain, see DeLuca et al., 2019.
3. As pointed out by a reviewer, country of residence (English-speaking or non-English speaking) could have affected L2 English GFA, despite the fact that these children all attended English-language schools. We considered the potential effect of country of residence (English-dominant or non-English dominant) by running a model that – in addition to our factors of interest (Language, Time, and L2 AoO and L2 Exposure) – contained a factor for country of residence (English-dominant or non-English dominant). This model revealed largely the same results as the model without this factor. Crucially, the model did not reveal differences in L2 English GFA between returnees who had lived in an English-speaking country and those that did not. The model results are provided in the supplemental materials (code) for reference.
4. These 15 speech samples were of three returnee children for whom English speech data were collected following the same procedure as the rest of the returnees, but for whom the time points differed (five time points spanning two years). For this reason, they were not analysed together with the other speech data.
5. Model structure: accent rating \sim language \times time \times L2 AoO + language \times time \times L2 exposure + (1|subject) + (1|rater).
6. Model structure: count \sim language \times response + (1 + response |rater).
7. Note that we do not report explicitly report pds for 95% CrIs that do not contain zero since the pds in those cases are naturally (very close to) 100%.
8. Model structure: accent rating \sim age + (1|subject) + (1|rater). Prior specification and model diagnosis was identical as described in ‘Statistical Procedures’.

References

- Abrahamsson N and Hyltenstam K (2009) Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59: 249–306.
- Ahn S, Chang CB, DeKeyser R, and Lee-Ellis S (2017) Age effects in first language attrition: Speech perception by Korean–English bilinguals. *Language Learning* 67: 694–733.

- Anwyl-Irvine AL, Massonnié J, Flitton A, Kirkham N, and Evershed JK (2020) Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52: 388–407.
- Asher JJ and Garcia R (1969) The optimal age to learn a foreign language. *The Modern Language Journal* 53: 334–41.
- Au TK, Knightly LM, Jun S-A, and Oh JS (2002) Overhearing a language during childhood. *Psychological Science* 13: 238–43.
- Audacity Team (n.d.) *Audacity(R)*. Available at: <https://www.audacityteam.org> (accessed January 2024).
- Bergmann C, Nota A, Sprenger SA, and Schmid MS (2016) L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics* 58: 71–86.
- Birdsong D and Molis M (2001) On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language* 44: 235–49.
- Boersma P and Weenink D (2019) *Praat: doing phonetics by computer: 6.0.48* [computer program]. Available at: <http://www.praat.org> (accessed January 2024).
- Bürkner P-C (2018) Advanced Bayesian multilevel modeling with the R package *brms*. *The R Journal* 10: 395–411.
- Coretta S, Casillas JV, and Roettger T (2022). Learn Bayesian analysis for the speech sciences. Available at: <https://learnb4ss.github.io/learnB4SS/index.html> (accessed January 2024).
- de Leeuw E, Kapia E, and Lewis S (2023) Sound change in Albanian monolinguals and Albanian–English sequential bilingual returnees in Tirana, Albania. *Languages* 8: 80.
- DeLuca V, Miller D, Pliatsikas C, and Rothman J (2019) Brain adaptations and neurological indices of processing in adult second language acquisition. In: Schwieter JW (ed.) *The handbook of the neuroscience of multilingualism*. Chichester: Wiley, pp. 170–96.
- Derwing TM and Munro MJ (2009) Putting accent in its place: Rethinking obstacles to communication. *Language Teaching* 42: 476–90.
- Douven I (2018) A Bayesian perspective on Likert scales and central tendency. *Psychonomic Bulletin and Review* 25: 1203–11.
- Dragoy O, Virfel E, Yurchenko A, and Bastiaanse R (2019) Aspect and tense attrition in Russian–German bilingual speakers. *International Journal of Bilingualism* 23: 275–95.
- Flege JE (1987) The production of ‘new’ and ‘similar’ phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15: 47–65.
- Flege JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press, pp. 233–77.
- Flege JE, Frieda EM, and Nozawa T (1997) Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25: 169–86.
- Flege JE, Munro MJ, and MacKay IRA (1995) Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97: 3125–34.
- Flege JE, Yeni-Komshian GH, and Liu S (1999) Age constraints on second-language acquisition. *Journal of Memory and Language* 41: 78–104.
- Flores C (2010) The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition* 13: 533–46.
- Flores C (2015) Losing a language in childhood: A longitudinal case study on language attrition. *Journal of Child Language* 42: 562–90.
- Flores C (2020) Attrition and reactivation of a childhood language: The case of returnee heritage speakers. *Language Learning* 70: 85–121.
- Flores C and Rato A (2016) Global accent in the Portuguese speech of heritage returnees. *Heritage Language Journal* 13: 161–83.

- Flores C and Snape N (2021) Language attrition and heritage language reversal in returnees. In: Montrul S and Polinsky M (eds) *The Cambridge handbook of heritage languages and linguistics*. Cambridge: Cambridge University Press, pp. 351–72.
- Fowler CA, Sramko V, Ostry DJ, Rowland SA, and Hallé P (2008) Cross language phonetic influences on the speech of French–English bilinguals. *Journal of Phonetics* 36: 649–63.
- Granena G and Long MH (2013) Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29: 311–43.
- Grasmuck S and Hinze AM (2016) Transnational heritage migrants in Istanbul: Second-generation Turk-American and Turk-German ‘Returnees’ in their parents’ homeland. *Journal of Ethnic and Migration Studies* 42: 1959–76.
- Haendler Y, Lassotta R, Adelt A, et al. (2020) *Bayesian analysis as alternative to frequentist methods: A demonstration with data from language-impaired children’s relative clause processing*. In: Brown MM and Kohut A (eds) *Proceedings of the 44th Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, pp. 168–181.
- Hanulíková A, van Alphen PM, van Goch MM, and Weber A (2012) When one person’s mistake is another’s standard usage: The effect of foreign accent on syntactic processing. *Journal of Cognitive Neuroscience* 24: 878–87.
- Hartshorne JK, Tenenbaum JB, and Pinker S (2018) A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition* 177: 263–277.
- Hirata Y (2015) 18 L2 phonetics and phonology. In: Kubozono H (ed.) *Handbook of Japanese Phonetics and Phonology*. Berlin: De Gruyter Mouton, pp. 720–62.
- Hopp H and Schmid M (2013) Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 34: 361–94.
- Idemaru K, Wei P, and Gubbins L (2019) Acoustic sources of accent in second language Japanese speech. *Language and Speech* 62: 333–57.
- Jesney K (2004) *The use of global foreign accent rating in studies of L2 acquisition*. Calgary, AB: University of Calgary.
- Kallay JE and Redford MA (2021) Clause-initial AND usage in a cross-sectional and longitudinal corpus of school-age children’s narratives. *Journal of Child Language* 48: 88–109.
- Karayayla T and Schmid MS (2019) First language attrition as a function of age at onset of bilingualism: First language attainment of Turkish–English bilinguals in the United Kingdom. *Language Learning* 69: 106–42.
- Komatsu M and Kimoto FT (2008) Foreign accent of Brazilian Portuguese and Japanese: Perceptual evaluation of modified and unmodified speech. In: Paulmann S, Garrido-Válesquez Schmidt P, Pell MD, and Kotz SA (eds) *Proceedings of the 4th International Conference on Speech Prosody, SP 2008*. International Speech Communication Association Archive, pp. 767–770.
- Kubota M (2019) *Language change in bilingual returnee children: Mutual effects of bilingual experience and cognition*. Unpublished PhD dissertation, University of Edinburgh, Edinburgh, UK.
- Kubota M, Chevalier N, and Sorace A (2020) How bilingual experience and executive control influence development in language control among bilingual children. *Developmental Science* 23: e12865.
- Kubota M, Chondrogianni V, Clark AS, and Rothman J (2022) Linguistic consequences of toing and froing: Factors that modulate narrative development in bilingual returnee children. *International Journal of Bilingual Education and Bilingualism* 25: 2363–81.
- Kuhberg H (1992) Longitudinal L2-attrition versus L2-acquisition, in three Turkish children-empirical findings. *Interlanguage Studies Bulletin (Utrecht)* 8: 138–54.

- Kupisch T, Barton D, Hailer K, et al. (2014) Foreign accent in adult simultaneous bilinguals. *Heritage Language Journal* 11: 123–50.
- Kupisch T, Kolb N, Rodina Y, and Urek O (2021) Foreign accent in pre- and primary school heritage bilinguals. *Languages* 6: 96.
- Larson-Hall J (2006) What does more time buy you? Another look at the effects of long-term residence on production accuracy of English /inverted r/ and /l/ by Japanese speakers. *Language and Speech* 49: 521–48.
- Lenth R (2020) *emmeans: Estimated Marginal Means, aka Least-Squares Means*. Available at: <https://cran.r-project.org/package=emmeans> (accessed January 2024).
- Lloyd-Smith A, Einfeldt M, and 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.
- MacWhinney B (1987) The competition model. In: MacWhinney B (ed.) *Mechanisms of language acquisition*. Mahwah, NJ: Lawrence Erlbaum, pp. 249–308.
- MacWhinney B (1997) Second language acquisition and the competition model. In: de Groot AMB and Kroll JF (eds) *Tutorials in bilingualism: Psycholinguistic perspectives*. Mahwah, NJ: Lawrence Erlbaum, pp. 113–42.
- MacWhinney B (2012) The logic of the unified model. In: Gass S and Mackey A (eds) *The Routledge handbook of second language acquisition*. Abingdon: Routledge, pp. 211–27.
- MacWhinney B (2018) A unified model of first and second language learning. In: Hickmann M, Veneziani E, and Jisa H (eds) *Sources of variation in first language acquisition: Languages, contexts, and learners*. Amsterdam: John Benjamins, pp. 287–312.
- Mayer M (1969) *Frog, where are you?* New York: Dial Press.
- Mayer M (1973) *Frog on his own*. New York: Dial Press.
- McDermott WLC (1986) *The scalability of degrees of foreign accent*. Unpublished PhD dissertation, Cornell University, Ithaca, NY, USA.
- Montrul S (2016) Age of onset of bilingualism effects and availability of input in first language attrition. In Nicoladis E and Montanari S (eds) *Bilingualism across the Lifespan: Factors moderating language proficiency*. Washington, DC: American Psychological Association, pp. 141–61.
- Nicenboim B, Vasisht S, Engelmann F, and Suckow K (2018) Exploratory and confirmatory analyses in sentence processing: A case study of number interference in German. *Cognitive Science* 42: 1075–100.
- Oh JS, Jun S-A, Knightly LM, and Au TK (2003) Holding on to childhood language memory. *Cognition* 86: B53–64.
- Ortega L (2013) SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning* 63: 1–24.
- Polinsky M (2018) Bilingual children and adult heritage speakers: The range of comparison. *International Journal of Bilingualism* 22: 547–63.
- R Core Team (2022) *R: A language and environment for statistical computing* [software]. Vienna: R Foundation for Statistical Computing. Available at: <https://www.r-project.org> (accessed January 2024).
- Riney TJ, Takada M, and Ota M (2000) Segmentals and global foreign accent: The Japanese flap in EFL. *TESOL Quarterly* 34: 711–37.
- Riney TJ and Takagi N (1999) Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49: 275–302.
- Riney TJ, Takagi N, and Inutsuka K (2005) Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly* 39: 441–66.

- Riney TJ, Takagi N, Ota K, and Uchida Y (2007) The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics* 35: 439–43.
- Rizopoulos D (2006) Ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software* 17: 1–25.
- Rothman J (2008) Why all counter-evidence to the critical period hypothesis in second language acquisition is not equal or problematic. *Language and Linguistics Compass* 2: 1063–88.
- Rothman J (2009) Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism* 13: 155–63.
- Rothman J, Bayram F, DeLuca V, et al. (2022) Monolingual comparative normativity in bilingualism research is out of ‘control’: Arguments and alternatives. *Applied Psycholinguistics*. Advance online publication. <http://doi.org/10.1017/S0142716422000315>
- Schmid MS and Hopp H (2014) Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects. *Language Testing* 31: 367–88.
- Schmid MS and Köpke B (2017) The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism* 7: 637–67.
- Snape N, Matthews J, Hirakawa M, Hirakawa Y, and Hosoi H (2014) Aspect in L2 English. *EUROSLA Yearbook* 14: 79–110.
- Steinkrauss R and Schmid MS (2017) Entrenchment and language attrition. In: Schmid H-J (ed.) *Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Washington, DC: American Psychological Association, pp. 367–83.
- Taura H (2019) Attrition studies on Japanese returnees. In: Schmid MS and Köpke B (eds) *The Oxford handbook of language attrition*. Oxford: Oxford University Press, pp. 390–402.
- Tomiya M (1999) The first stage of second language attrition: A case study of a Japanese returnee. In: Hansen L (ed.) *Second language attrition in Japanese contexts*. Oxford: Oxford University Press, pp. 59–79.
- Tomiya M (2000) Child second language attrition: a longitudinal case study. *Applied Linguistics* 21: 304–32.
- Treffers-Daller J, Daller M, Furman R, and Rothman J (2016) Ultimate attainment in the use of collocations among heritage speakers of Turkish in Germany and Turkish–German returnees. *Bilingualism: Language and Cognition* 19: 504–19.
- Trofimovich P and Baker W (2006) Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1–30.
- Unsworth S (2013) Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition* 16: 86–110.
- Uzal M, Peltonen T, Huottilainen M, and Aaltonen O (2015) Degree of perceived accent in Finnish as a second language for Turkish children born in Finland. *Language Learning* 65: 477–503.
- Vasishth S, Nicenboim B, Beckman ME, Li F, and Kong EJ (2018) Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of Phonetics* 71: 147–61.
- Wayland R (1997) Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied Linguistics* 18: 345–73.
- Wickham H (2016) *ggplot2: Elegant graphics for data analysis*. Cham: Springer.
- Woods KJP, Siegel MH, Traer J, and McDermott JH (2017) Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, and Psychophysics* 79: 2064–72.
- Wrembel M, Marecka M, Szewczyk J, and Otwinowska A (2019) The predictors of foreign-accentedness in the home language of Polish–English bilingual children. *Bilingualism: Language and Cognition* 22: 383–400.
- Yeni-Komshian GH, Flege JE, and Liu S (2000) Pronunciation proficiency in the first and second languages of Korean–English bilinguals. *Bilingualism: Language and Cognition* 3: 131–49.