Bilingualism: Language and Cognition

cambridge.org/bil

Tuning in to the prosody of a novel language is easier without orthography

Kateřina Chládková[1,2], Václav Jonáš Podlipský[3], Lucie Jarůšková[1] and Šárka Šimáčková[3]

[1]Institute of Czech Language and Theory of Communication, Faculty of Arts, Charles University, Czech Republic; [2]Institute of Psychology, Czech Academy of Sciences, Czech Republic and [3]Department of English and American Studies, Faculty of Arts, Palacký University Olomouc, Czech Republic

Research Article

Cite this article: Chládková, K., Podlipský, V.J., Jarůšková, L. and Šimáčková, Š. (2025). Tuning in to the prosody of a novel language is easier without orthography. *Bilingualism: Language and Cognition*, 1–10 https://doi.org/10.1017/S1366728925000082

Received: 09 January 2024
Revised: 09 November 2024
Accepted: 15 November 2024

Keywords:
Czech; L2 learning; Māori; novel-language exposure; prosody; orthography

Corresponding author:
Kateřina Chládková;
Email: katerina.chladkova@ff.cuni.cz

This research article was awarded Open Data and Open Materials badges for transparent practices. See the Data Availability Statement for details.

Abstract

Mastering prosody is a different task for adults learning a second language and infants acquiring their first. While prosody crucially aids the process of L1 acquisition, for adult L2 learners it is often considerably challenging. Is it because of an age-related decline in the language-learning ability or because of unfavorable learning conditions? We investigated whether adults can auditorily sensitize to the prosody of a novel language, and whether such sensitization is affected by orthographic input. After 5 minutes of exposure to Māori, Czech listeners could reliably recognize this language in a post-test using low-pass filtered clips of Māori and Malay. Recognition accuracy was lower for participants exposed to the novel-language speech along with deep-orthography transcriptions or orthography with unfamiliar characters. Adults can thus attune to novel-language prosody, but orthography hampers this ability. Language-learning theories and applications may need to reconsider the consequences of providing orthographic input to beginning second-language learners.

Highlights

- After 5 minutes of exposure to a new language adults sensitise to the novel prosody.
- They recognize the new language, besides a similar one, in low-pass filtered recordings.
- Exposure to written forms of the novel speech can impair attunement to its prosody.

1. Introduction

At the start of the language learning journey, human fetuses and infants attune to the rhythm and melody of the ambient language without any apparent effort, very likely subconsciously, and with an outcome that seems perfect. After they begin tuning in to the global characteristics of the native language, infants come to learn the properties of the native-language segments (that is, vowels and consonants), word forms and meanings, and relations between words. Eventually, children become acquainted with how those segments and words are written. The journey is commonly traveled backward when learning subsequent languages later in life: learners read words and learn their meanings, and they learn how the words and sounds that constitute them are pronounced, but hardly ever do they fluently attune to the global rhythm and intonation of the language. Mastering the rhythm and melody of a second (or third or fourth) language is what older children and especially adults struggle with considerably. Why? Is it because they are past a sensitive window for learning prosody or is it because they never had a chance to experience the prosody the way first-language learners do? In the present study, we investigate whether adults display sensitivity to prosodic patterns of an unknown language and whether their prosodic sensitivity is modulated by the type of input they receive.

In the developmental literature, prosody is regarded as a window into the first language (Gleitman & Wanner 1982, Gervain et al. 2020). Before they are even born, humans have weeks to months of exposure to the primarily prosodic features of the ambient language. As shown by studies with fetuses and newborns, in this early period of development humans sensitize to the prosodic features in language-specific ways (DeCasper et al. 1994, Mampe et al. 2009, Granier-Deferre et al. 2011). The early attunement to native prosody then arguably helps children to recognize the language that they are learning, identify prominent chunks of speech, parse the continuous speech signal into words and syllables, or learn morphosyntactic relations in utterances (Johnson & Jusczyk 2001, Mandel et al. 1996, Gerken 1994, Gordon et al. 2015, Suppanen et al. 2019). Prosody opens the language window not only in the domain of perception. Language-specific prosodic patterns are manifested also in infants' own vocal productions: from the way in which newborn babies cry to how older infants babble and toddlers speak their first

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

CAMBRIDGE
UNIVERSITY PRESS

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

words (Mampe et al. 2009, Wermke et al. 2016, Levitt & Wang 1991, Hallé et al. 1991). Such prosodic bootstrapping seems to be characteristic of L1 acquisition (e.g., Gervain et al. 2020).

Contrary to its prime role in first language acquisition, prosody hardly serves as a gateway to the acquisition of a second language in adults. Naturally, adults never encounter the speech of a new language with primarily, or only, the prosodic cues available, as fetuses do when experiencing their native language(s) in the womb. This means that adults lack the opportunity to perceive L2 speech without the full segmental characteristics so that they could use prosody to bootstrap other linguistic knowledge the way infants acquiring their native language(s) do. What is more, prosody is what adults keep struggling with even at later stages of their second-language development, after they have learned a great part of the L2 grammar and vocabulary, possibly even more than with L2 segments, as reflected in comprehensibility, intelligibility, or accentedness ratings of adults' L2 speech (Anderson-Hsieh et al. 1992, Warren et al. 2009, Saito et al. 2016, see the review in Choi & Kang 2023). Some authors even claim that L2 prosody is unlearnable without explicit awareness and instruction (Chun & Levis 2020).

It might thus seem that beyond infancy or early childhood, humans can no longer access the mechanisms that enable them to acquire successfully the prosody of a new language. Previous research on L2 prosody acquisition suggests that while L2 prosody is learnable (Mennen 2004), the earlier L2 exposure begins and/or the more L2 exposure there is, the better can they establish new L2-like prosody differing sufficiently from their L1 prosody (Trofimovich & Baker 2006, 2007). Why is that? It could be due to the end of a sensitive period for efficient acquisition of the sound patterns of a second language, including its suprasegmental features (Long 1990). However, we are proposing an alternative explanation, one that considers the importance of the initial prosodic attunement in adult language learning. In the present experiment, we thus aim to test whether even adults can sensitize to the prosody of a new language through exposure and whether non-prosodic aspects of the input interfere with this sensitization.

One suspected hindering factor – orthography – is the focus of this study. Curiously, it is at about the age of 6 years – that is, the supposed end of the putative sensitive period for phonetic and phonological acquisition (Long 1990) – that language learners usually start gaining knowledge of orthography. The literature suggests that once children learn to read and write, they perform differently at novel word and grammar learning tasks. For instance, preliterate children better learn articles and nouns agreeing in gender, that is multi-word chunks, than nouns, probably because, unlike literate children, they are not biased by the visual salience of words in writing (Havron et al. 2018). A parallel orthography bias can be witnessed at the level of speech sound patterns: children who learn to read and write using an alphabetic system start to perceive – or at least think of – speech as a series of segments (Morais et al. 1979, Dehaene et al. 2010, Goetry et al. 2005). Getting acquainted with the orthographic representation of language will then very likely shift the metalinguistic attention to the segmental level of vowels and consonants and hence possibly hinder the perceptual sensitivity to the global (suprasegmental) prosodic patterns. Compared to small children, adults not only have a developed L1 system including its prosody, but they are also often literate; having access to orthographic segment-based input may thus be detrimental to adults' ability to attune to the suprasegmental-prosodic patterns of a language. Possibly, the lack of such initial prosodic attunement may hinder the further acquisition of the target language sound patterns as well as other, e.g., morphosyntactic, patterns.

The second-language speech learning literature has mostly researched the effects of orthography in the domain of segmental contrasts and focused on the congruence between sound-to-grapheme mapping in the first compared to the second language (Escudero et al. 2014, Bassetti 2017), including orthography effects on the acquisition of lexical tones (Mok et al. 2018) as well as position-dependent effects of orthography (Zhou & Hamann 2020). Studies have demonstrated that learning an L2 where the mapping between sounds and graphemes differs from L1 may hinder the acquisition of the L2 speech sound contrasts at hand. A notorious example is the English *ship-sheep* contrast where – for instance to an L1 Spanish learner – the orthography provides conflicting cues about the realization of the vowel sound (Escudero & Boersma, 2004). From the perspective of comprehending unfamiliar L2 speech, advanced L2 learners seem to benefit from encountering new L2 accents with L2 orthographic transcription, namely subtitles in the L2, but are hindered by encountering the new L2 accents with subtitles in the L1 (Mitterer & McQueen, 2009). It thus appears that in the case of advanced L2 users who are well familiar with the L2 orthography, the L2 spelling serves as a cue to phoneme and word identity and thus facilitates the understanding of L2 speech produced with accents unfamiliar to the learner. The facilitating effect of L2 subtitles is likely highly relevant at the segmental level where the familiar L2 spelling helps the learners decipher what the novel-accent vowel and consonant realizations stand for. Whether and how exposing *novice* learners to L2 orthography affects their ability to attune to the L2 *prosody*, in particular, has not been addressed yet.

As illustrated by studies referenced above, there is prior work on L2 prosody but this work has not focused on the initial prosodic attunement (and the factors affecting it). When asking whether adults can attune to novel-language prosody at all, one has to consider two aspects of second or foreign language learning. Firstly, adult second-language learners completely lack the initial prosody-only phase of the new language exposure (since they do not listen from the womb which would mask the frequencies above ~700–1000 Hz and thus also most segmental identities, making prosodic patterns stand out). Secondly, their chance of sensitizing to the prosodic features in the input they receive, and learning them along with their mental representations of L1 prosody, may be hindered also by premature exposure to the orthography of the target language. Considering how one can make the conditions of second language acquisition more similar to those of the first, it is more feasible and ecologically realistic to manipulate the latter aspect of L2 input (that is to remove orthography) rather than the former (that is, to try to block access to frequencies above approximately 1000 Hz in the speech input). In this experiment, we thus test whether adults can sensitize to an unfamiliar language through passive listening and whether their perceptual sensitization to the novel-language prosody is affected by orthographic input.

There are several ways in which orthography could interfere with auditory sensitization to prosody. Based on the proposal that learning to read and write using an alphabetic system promotes perceiving speech as a sequence of segments and/or paying greater attention to the segmental than the suprasegmental level, we hypothesize that (1) including alphabetic orthographic representation of the audio speech signal will attenuate perceptual sensitization to prosody. In line with previous findings showing that particularly non-transparent and L1-mismatching orthography hinders segmental learning (Escudero et al. 2014), and that adaptation to novel accents is hindered by language mismatch in audio

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

and subtitles (Mitterer & McQueen, 2009), we hypothesize that (2) deep orthography with less reliable mappings between sounds and letters will be more detrimental to prosodic sensitization than shallow orthography with better correspondence between sounds and letters. Lastly, (3) we hypothesize the smallest or no interference of orthography on prosodic sensitization when participants are exposed to an unfamiliar script, as they will be least likely to map the heard speech sounds to the unfamiliar letter shapes, potentially giving up trying to read along with listening, making this condition similar to the audio-only exposure.

To test the above hypotheses we exposed L1-Czech adults to 5 minutes of naturally-produced Māori speech (an audiobook) in one of four between-subject conditions: (a) audio only without any visual representation, (b) audio with transcription using shallow orthography, that is the original Māori Latin-alphabet-based text, (c) audio with transcription using deep orthography, and (d) audio with an unfamiliar script, namely Hebrew characters which, we reasoned, would not be recognized by most of our participants. After exposure, participants heard low-pass filtered excerpts spoken by different speakers either in the same language or in a different but related language, namely Malay, and had to indicate whether the excerpts were from the exposed language or not. Low-pass filtering with an 800 Hz cutoff and 100 Hz smoothing (i.e., gradual attenuation starting at 700 Hz, and not passing through frequencies of 900 Hz and higher) effectively removes most segmental information preserving the intonational and rhythmic dynamics (it also appears to be the point after which speech frequencies get attenuated as they pass into the womb, Granier-Deferre et al. 2011, Richards et al. 1992). If orthography hinders prosodic sensitization, we predict better outcomes in conditions (a) than in conditions (b) and (c). If deep orthography is particularly detrimental for sensitization to prosody we predict a smaller difference between (a) and (b) than between (a) and (c). If an unfamiliar script does not make participants hone in on segments at the cost of prosody, we predict the smallest or no difference between (a) and (d).

2. Method

2.1. Participants

A total of 221 students from Charles University, Prague, Czechia, and Palacký University Olomouc, Czechia, participated for course credit, being assigned randomly to one of the four conditions. They were native speakers of Czech (all were monolingually raised in Czech) with self-reported normal hearing and normal or corrected-to-normal vision. Forty-seven participants were excluded because, at the end of the experiment, they indicated they could identify or probably identify the exposure language (in some cases motivated by the audio or Maori script, in some cases by the Hebrew script). We excluded all participants who indicated that they thought they had possibly identified the language, even those whose guess was actually incorrect, as their belief might still have influenced their choices on the post-test. The remaining 174 participants were distributed across the 4 conditions as follows: 43 in condition a (audio only), 44 in condition b (shallow orthography), 47 in condition c (deep orthography), and 40 in condition d (unfamiliar script). The participants' mean age was 22.29 (sd = 3.96); 145 identified as women, 28 as men, and 1 as non-binary. The experiment was approved by the Ethics Committee of the Institute of Psychology, Czech Academy of Sciences.

2.2. Materials: target language

Our participants' L1 was Czech, all spoke English as an L2, and many knew or learned other Indo-European languages. As the target language of exposure for the present experiment, we chose Māori, a Malayo-Polynesian language from the Austronesian family, because it fulfilled the following conditions: (1) it is an unfamiliar and very likely an unencountered language for the population represented by our participant sample, (2) it is phonotactically relatively simple and segmentally rather similar to, or simpler than, the L1 of our participants, (3) there are high-quality audio materials available, and (4) there is a closely related language for which comparable high-quality audio materials are available as well. As the competitor language for the test phase, we selected Malay, also a Malayo-Polynesian language from the Austronesian family. The audio materials from which we extracted our stimuli were Harry Potter audiobooks (Rowling 2022).

As for the auditory features that could be preserved in the low-pass filtered post-test stimuli, Māori has 5 vowel qualities /i e a o u/, and a two-way contrast in vowel length which seems to be preserved only in the low-vowel /a/-/aː/ pair. Malay has 6 vowels /i e a o u ə/ and does not contrast vowel length. Phonotactically, Māori contains open syllables, permitting onsetless syllable structure. While Malay permits coda consonants, neither of the two languages permits consonant clusters. Rhythmically, both Māori and Malay seem to be in-between on the scale of stress-timed versus syllable-timed languages, patterning with the former on some rhythm metrics and with the latter on other metrics (Clynes & Deterding 2011, Wan 2012, Harlow et al. 2009, Maclagan et al. 2009). For the rhythm-metrics values published for Māori and Malay, see the ranges in brackets in Table 1.

2.3. Stimuli

The exposure materials were excerpts from the Māori audiobook that were recorded by a female speaker. We selected three parts from different chapters of the audiobook that conformed to the following criteria: they were continuous speech streams of mainly descriptive style with a relatively stable, natural prosody, not containing direct speech (dialogues between characters), and they did not contain internationally known English names or expressions. The three parts from the different chapters were concatenated, and the total duration of the exposure material was 5 min and 17 s altogether.

For the test phase, we selected 3 excerpts from the Māori audiobook (different from those in the exposure stimuli) that were recorded by a male speaker and 3 excerpts from the online Malay audiobook recorded by a female speaker. We ensured there were no identifiable character names in the post-test materials, such as Hare ("Harry") that occurred several times in the exposure materials. The six clips lasted between 4 and 6 s each. Whereas the exposure material was naturally produced, each of the post-test stimuli was low-pass filtered using Praat (Boersma & Weenink 1992–2023) with a cut-off frequency of 800 Hz and a smoothing setting of 100 Hz, which means that frequencies below 700 Hz were preserved, those above 900 Hz were completely filtered out, and those in between were gradually more and more reduced in intensity. This filtering procedure effectively reduced the cues to spectral information above 700 Hz and removed them completely at 901 Hz and above: while some vowels' first formants could be discernible in such signal, it is unlikely that the heavily impoverished vowel

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

Table 1. Selected acoustic properties of the exposure and test stimuli. The ranges in brackets are values published for different datasets in prior studies for Māori (Maclagan et al. 2009) and Malay (Wan 2012). The values for the various rhythm metrics provided here should be taken only as indications of some of the rhythm properties of the two languages. It is seen that our Malay samples have lower nPVI and rPVI but higher varcoC than our Maori test and exposure materials. However, if listeners discriminate between the languages, we do not conclude that they do so on the basis of rhythm alone (as intonation patterns were also present) or on the basis of these particular rhythm metrics: a separate rhythm-cue-weighting experiment would need to address that. Explanation of abbreviations: %V is the percentage of vocalic intervals; varcoC is the percentage that the standard variation of the consonantal interval duration takes up of the mean duration of the consonantal intervals (Dellwo 2006), rPVI and nPVI are the raw and rate-normalized, respectively, Pairwise Variability Indices, that is, the average differences between consecutive consonantal and vocalic intervals (Grabe & Low 2002)

		Exposure Māori	Test Māori (all 3 stimuli together)	Test Malay (all 3 stimuli together)
Duration (s)		317	15	15
F0 (Mel)	0.1 quantile	129	114	147
	Median	168	142	168
	0.9 quantile	216	183	220
	Mean	171	147	177
	St. deviation	33	29	32
	Range	193	133	180
Rhythm metrics	%V	53	54	54 (46–54)
	VarcoC	43	45	47
	nPVI (V intervals)	56 (39–57)	48 (39–57)	45 (35–55)
	rPVI (C intervals)	49 (37–47)	47 (37–47)	42 (20–47)

quality cues would drive the discriminability of Malay and Māori whose vowel inventories are quite similar.

We used different speakers for the exposure and test phase and intentionally selected a male speaker (that is different sex than exposure materials) for the same-language test trials and a female speaker (that is the same sex as exposure materials) for the different-language test trials. As seen in Table 1, the voice quality (as measured by F0) of the exposure materials was more similar to the different-language test trials than to the same-language test trials. Therefore, if participants rated the stimuli based on acoustic voice properties, they would be below chance with their classification of *language similarity*. On the other hand, Table 1 also shows that the prosodic rhythm properties of the exposure stimuli are more similar to the same language than to the different-language test trials. If the listeners' decisions were cued by these linguistic rhythm properties, they would be more likely to correctly identify the exposure language at the test.

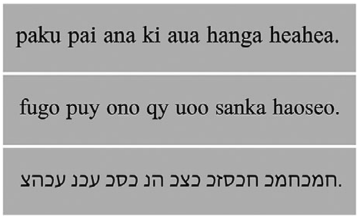
Table 1 shows information about the melodic and rhythmic properties of the exposure and test stimuli. F0 was measured in Praat (using Praat's Filtered autocorrelation method with standard settings and manually correcting octave jumps due to non-modal phonation). The rhythm measurements were based on manual labeling of the vocalic and consonantal intervals. For the exposure recording, the first 30 s and the last 30 s were used; for the test stimuli, the entire original unfiltered recordings were used. Durations of the labeled intervals were measured by a Praat script and submitted to the calculations of the rhythm metrics (according to Grabe & Low 2002 and Dellwo 2006).

The auditory stimuli in the exposure phase were presented in four conditions. Condition (a) contained only audio; condition (b) paired the audio stimuli with the actual Māori Latin-alphabet-based orthography, which is a shallow orthography with a transparent and straightforward mapping between phonemes and graphemes; condition (c) paired the audio stimuli with an artificially modified, deep, orthography, where each phoneme was, each time it occurred and each time differently, represented by one randomly selected letter out of 2 or 3 Latin alphabet letters preselected as possible spellings of the sound (one of them being the original Māori spelling but removing any diacritics); condition (d) presented the audio stimuli with transcriptions using an unfamiliar script (Hebrew) with the unfamiliar placement of characters (right-to-left within words which were placed left-to-right). Conditions (b) through (d) thus represented three levels of orthographic complexity, ranging from the simplest, most transparent phoneme-to-grapheme mapping in condition (b) to a highly complex, unfamiliar, and not trivially retrievable phoneme-to-grapheme mapping in condition (d). Examples of what participants saw on the screen in the three orthography conditions are shown in Figure 1.

2.4. Procedure

The participants were tested individually or in small groups in a quiet room using a desktop computer and circumaural headphones; the experiment was implemented in Praat, using a Demo window script (Boersma & Weenink 1992–2023). Before the experiment, the participants were told that they would hear a new language for 5 minutes and were asked to listen attentively. They were told they might also see the language in writing on the screen, in which case they were asked to pay attention to the screen as well. They were told that after the exposure they would be asked a few simple questions about this language. An experimenter made sure participants were paying attention to the auditory and visual stimulation throughout the session. Prior to exposure, the volume was adjusted to a comfortable level for each participant individually.

Immediately after the exposure ended, a screen with information about the post-test appeared and the participants commenced with a click. Each of the six trials presented one of the test stimuli, audio only for all participants, in random order with the question of whether the audio clip came from the language the participant had heard and then a forced choice between 'yes' and 'no'. At the very end, the participants were asked to choose whether or not they thought they had recognized the exposure language and, if so, type in which language they thought it was.



paku pai ana ki aua hanga heahea.

fugo puy ono qy uoo sanka haoseo.

צהכע נכע כסכ הנ כצכ כזסכח חמכחמכ.

Figure 1. Example subtitling of a segment of the exposure audio [ˈpakʉˈpai̯ʲanakʰi̥ʲˈawa ˈhaŋaˈhea̯hea̯] in the different orthography conditions. The first panel shows condition (b) which used subtitles in the original Māori shallow orthography, the middle panel shows condition (c) which uses deep-orthography subtitles, and the last panel shows condition (d) using a script unfamiliar to the participants. The audio-only condition (a) displayed a plain grey screen throughout the experiment.

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

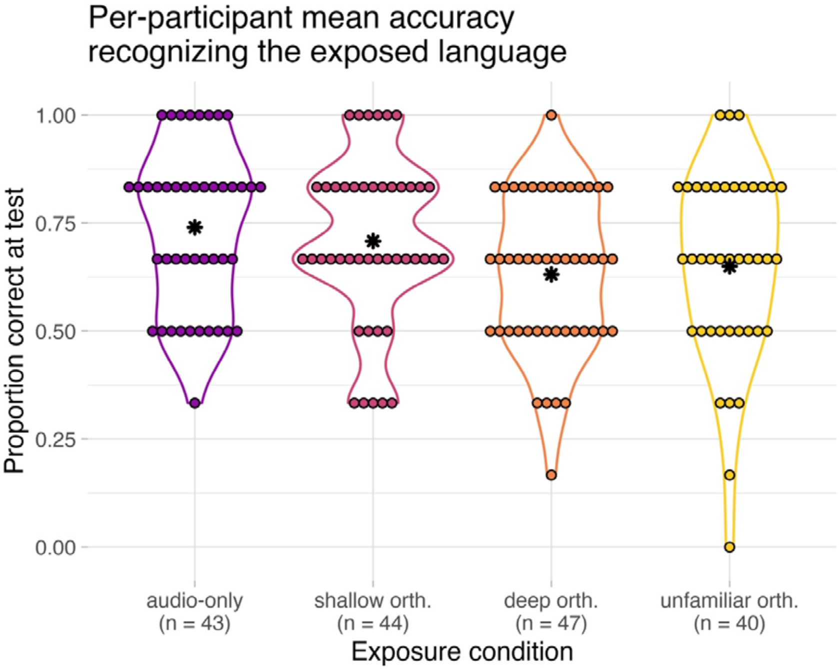


Figure 2. Stacked dot plots and overlaid violin plots of the proportions of correctly recognizing the exposed and rejecting the competitor language in each exposure condition. Colored dots show per-participant proportions correct (recognitions and rejections together), and black asterisks show group means. Numbers of participants per condition are given in parentheses.

3. Results

To provide a glimpse of the data before statistical modeling, Figure 2 plots per-participant overall accuracy at post-test in recognizing the exposed and rejecting the competitor language. The figure suggests that the accuracy was higher in the audio-only condition than in the deep and unfamiliar orthography condition. The accuracy seems in-between for the shallow orthography condition.

To test this statistically, the raw binomial scores from the post-test were analyzed with a logistic mixed-effects regression model, using the function glmer() of the *lmer* package in R (Bates et al. 2015, R Core Team 2021). The modeled variable was response accuracy (that is, correct identification of the exposed and correct rejection of the competitor language). The fixed factors were Condition (4 levels, treatment contrasts with audio-only as the reference level) and Trial number (numeric, mean-centered). The modeled random effects were intercepts per participant and per item (six test trial identities). Means and confidence intervals were estimated using the function ggemmeans() in the R package *ggeffects* (Lüdecke et al. 2020).

Table 2 shows the fixed-effects model output; Table 3 and Figure 3 show the estimated marginal means per condition. First, the analysis showed that the estimated logit intercept was reliably higher than zero, indicating that in the audio condition, participants were above chance in accurately identifying the exposed language and rejecting the competitor. Next, the slope for the condition was significant for two of the three contrasts: for audio-only versus deep orthography and for audio-only versus an unfamiliar script. Comparisons of the estimated means reveal that the audio-only condition yielded significantly higher post-test scores than the deep-orthography condition by 11% and than the unfamiliar-script condition by 9%. The numerical difference between the audio-only and the shallow-orthography condition

Table 3. Estimated probabilities of recognizing the exposed and rejecting the competitor language at post-test, means, and 95% confidence intervals

Condition	Mean	95% CI lower	95% CI upper
Audio-only	0.75	0.66	0.83
Shallow orthography	0.72	0.62	0.80
Deep orthography	0.64	0.53	0.73
Unfamiliar script	0.66	0.55	0.75

Table 2. Fixed-effects model summary with 95% confidence intervals of the estimated logits

	Estimate	SE	95% CI lower	95% CI upper	*z* value	*p*
Intercept	1.115	0.241	0.605	1.642	4.623	<0.001
Cond. (shallow ortho.)	−0.172	0.211	−0.592	0.245	−0.815	0.415
Cond. (deep ortho.)	−0.545	0.203	−0.954	−0.146	−2.677	0.007
Cond. (unfam. script)	−0.458	0.212	−0.882	−0.041	−2.161	0.031
Trial number	0.068	0.040	−0.011	0.147	1.683	0.092

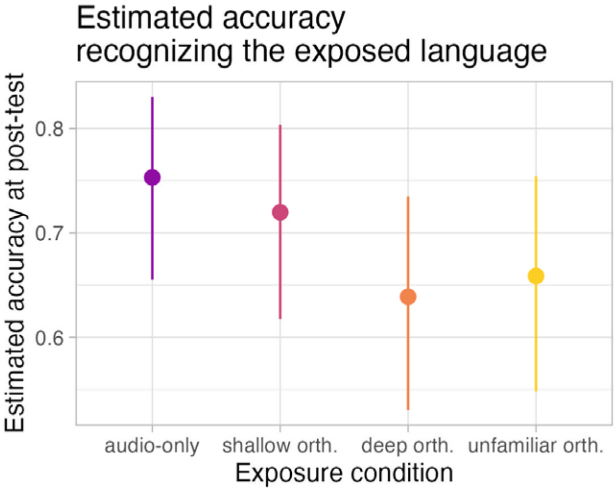


Figure 3. Estimated marginal means and 95% confidence intervals of accuracy in recognizing the exposed and rejecting the competitor language at post-test.

trended in the same direction (by 3%) but was non-significant. Despite the lower scores in some of the orthography conditions, performance was still above chance (that is, above 0.5 probability), as shown by the lower bounds of the confidence intervals given in Table 3.

As for the size of the effects across the three contrasts of condition, pairwise comparisons of mean effects sizes and their confidence intervals in Table 2 show that while deep orthography and unfamiliar script yielded effects of comparable size, the effect of shallow orthography seems numerically smaller than the effect of deep orthography, but this apparent difference is not statistically significant (as the 95% c.i.s overlap with the means of the contrasting level).

4. Discussion

We investigated whether adults are able to sensitize to the prosody of a new unfamiliar language during a 5-minute first encounter and whether this ability is modulated by orthographic input. Adult native speakers of Czech listened to an audiobook in Māori, either without any visual input, or along with a transcription using shallow and familiar (Latin-alphabet-based Māori) orthography, or using deep and familiar (Latin-alphabet-based) orthography, or using unfamiliar (Hebrew-character-based) orthography. Attunement to the novel-language prosody was tested immediately after exposure by a language recognition task with low-pass filtered utterances in Māori and Malay (recorded by different speakers than the exposure materials). The low-pass filtering approximated the quality of speech input that fetuses receive in the womb. We hypothesized that orthographic input in the familiar Latin-alphabet-based script would draw the participants' attention to segments, at the cost of suprasegmental (that is prosodic) information, and thus impede subsequent prosody-based language recognition with the low-pass filtered stimuli. We predicted larger detrimental effects of deep (that is, less transparent) orthography compared to shallow orthography, and negligible or no effects of transcriptions using an unfamiliar alphabet.

An analysis of the post-test language recognition scores revealed that after 5 minutes of exposure to the novel language auditorily, the adult participants' accuracy in recognizing the exposed language in the competition of another very similar language in low-pass filtered stimuli was above chance, but it was also affected by orthography. The correct recognition of the exposed language and rejection of the competitor language in the post-test was significantly attenuated by exposure to deep orthography as well as by exposure to unfamiliar segmental orthography.

Our prediction that orthography would hamper the ability to perceptually attune to novel language prosody was borne out. There was a detrimental effect of deep orthography, as presenting listeners with non-transparent transcriptions of the exposure speech in the novel language attenuated their ability to recognize the novel-language prosody at the post-test. Our data do not permit us to determine with confidence whether presenting listeners with shallow-orthography transcription during exposure led to attenuated recognition as well. Numerically, the performance was intermediate and it was not reliably worse than the performance in the audio-only condition (given the non-significance of the auditory-only versus shallow orthography contrast); at the same time though, it was not better than the performance in the deep orthography condition (given the overlap of the confidence intervals for the shallow and deep orthography effects). Thus, our second prediction that shallow orthography would be less detrimental than deep orthography was not confidently confirmed, and neither was our third prediction that transcriptions using an unfamiliar script would be the least, or not at all, harmful to prosodic sensitization, making the performance in this condition similar to that of the audio-only condition. Instead, there was a clear reduction in the unfamiliar-orthography condition, just like in the deep-orthography condition, indicating that even attempting to read an unfamiliar alphabetic script while listening to the exposure speech (which had a natural tempo) had a distracting effect on the implicit learning of prosody.

4.1. Tuning in to a novel prosody in adulthood

Our findings demonstrate that after exposure to 5 minutes of natural speech in an unfamiliar language, adult listeners are able to recognize this language when hearing low-pass filtered utterances spoken by a new speaker. The ability to identify the exposed language, and to reject another unfamiliar but closely related language, was likely driven by the listeners' sensitivity to the prosodic cues available (that is, intonation and rhythm) which were well preserved in the low-pass filtered signal, unlike segmental cues which were mostly removed (possibly with some exceptions such as cues to vowel height). Previous studies have shown that in low-pass filtered speech stimuli, adults can recognize their native language variety and distinguish it from a non-native language or variety (Vicenik & Sundara, 2013), as well as recognize foreign accents in their native language (Kolly et al. 2014), or that both child and adult listeners can discriminate speaker gender and accent (Weatherhead et al. 2019, Bozkurt & Soley 2022, Jacewicz et al. 2023). The present findings show that adults are able to achieve prosody-based language recognition with non-native and unfamiliar languages, after only very brief exposure to one of them. It thus appears that the tracking of and sensitizing to the prosody of a novel language – which is one of the first mechanisms that humans use to tune in to their native language very early in life – is available also in adulthood.

Correctly recognizing that a short low-pass filtered utterance comes from a just-encountered language (language A), and that a short low-pass filtered utterance from a similar language (B) is not from language A is, in fact, a rather striking ability, perhaps even surpassing infants' prosody-based language discrimination abilities that have been reported so far (especially if one assumes that adults

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

would be less adept at perceptual sensitization to a new language, Long 1990, Munro & Mann 2005). Developmental literature shows that newborns and 2-month-old infants can discriminate low-pass filtered languages only when they are very different rhythmically (e.g., English versus Italian) and/or when one of the languages is the infants' native language (Mehler et al. 1988, Nazzi et al. 1998). Low-pass filtered speech from languages that are rhythmically rather close such as English and Dutch seems to be indiscriminate across the lifespan, namely, by newborns (Nazzi et al. 1998), older infants (Johnson et al. 2003, holds also for Basque versus Spanish, Molnar et al. 2014), as well as by adults (Ramus et al., 2003).

4.2. Tuning in to novel prosody can be hampered by orthographic input

Adult tracking of novel prosody, however, is undermined if listeners are concurrently presented with a complex orthographic form of the novel language. This supports the idea that learning to read an alphabetic system steers speech perception not only towards the segmental content of speech but also away from the suprasegmental features, that is, prosody. The reason why learners attune less to prosody is probably because the segmentally-based visual cues promote segmentally-based listening. One could also argue that the limited tracking of prosody was observed here because the listeners' attention was split into two modalities. Although our experimental design does not allow us to eliminate this possibility, we do not consider it a likely explanation, given that we found reduced accuracy not for all but only some of the orthography conditions: only for the deep-orthography and unfamiliar-script conditions did we observe reduced accuracy as compared with the no-orthography condition, not for the shallow-orthography condition. What the present results clearly show is that the availability of orthographic input *does not improve* sensitization to novel language prosody. This contrasts with the literature on other domains of L2 suggesting that, if accompanying audio with orthography (such as captioning and textual enhancement) has any effect at all, it boosts word-form recognition and vocabulary and grammar learning in a second language (Markham 1999, Winke et al. 2010, Montero Perez et al. 2013). The literature has proposed that textual input helps learners become fluent listeners because it involves top-down conceptual and lexical knowledge as well as helps learners parse and decode the auditory stream (Montero Perez 2020). Thus, although little to no data is available for visual spelling effects on the learning of prosody, evidence from word learning and speech parsing studies speaks in favor of visual cues rather than otherwise.

To what extent and under which conditions alphabetic visual input *hampers* prosodic attunement needs to be researched further across various language populations and learning scenarios. Based on the present findings, however, we conclude that captioning may not always serve as an *aiding* input modality, because especially in the beginning stages of language learning, the segmental basis of the orthographic input guides the learner away from auditorily sensitizing to the global suprasegmental patterns. And such orthography-induced segment-based listening that literate learners may experience from the earliest stages of L2 learning, demotes their auditory tracking of the suprasegmental features, consciously or subconsciously. This might be precisely the reason why mastering prosody in an L2 is a challenge even in advanced L2 users.

With regard to the degree of transparency in sound-to-grapheme mapping, our results do not allow us to determine with confidence whether and how it modulated the orthographic interference with prosody tracking. While performance in the deep-orthography condition was reliably worse than in the audio-only condition, performance in the shallow-orthography condition was in between the audio-only and deep-orthography conditions and not reliably different from either. Therefore, further research is needed in this direction. A possible avenue would embrace the developmental perspective and measure prosody sensitization – with and without orthography – in preliterate and literate children. The prediction here is that prosody sensitization would be least affected by any type of orthography in preliterate children, and the hampering effects of orthography would become stronger with the children's growing knowledge of spelling. This would be an effect parallel to the finding that knowledge of orthography affects parsing and learning words, or more specifically, that compared to their literate peers matched on IQ, preliterate children are better at learning determiner+noun combinations than isolated nouns (Havron et al. 2018).

The reduction of prosody tracking was found, contrary to our predictions, also after exposure to captions written in an unfamiliar script. Despite being unfamiliar to the participants, they had had ample experience with the alphabetic orthographies of their L1 Czech and L2 English, possibly creating the expectation that the novel language we exposed them to also used an alphabetic, segmental, orthography even if the characters were strange. What is more, the transcriptions using Hebrew characters were strictly segment-based: the number of speech sounds per word corresponded to the number of characters displayed. Therefore, one possible explanation of our findings is that the unfamiliar segment-based orthography drove similar segment-focused listening just like the Latin-alphabet-based script. If that is the case, a number of predictions ensue: first, even the unfamiliar orthography could promote segment-based parsing and thus also facilitate segmental acquisition of the novel language at the cost of prosody; second, the harmful effects of orthography on prosody tracking could diminish with exposure to a non-segmental script, such as logographic or syllabic writing if the learners are aware of its non-segmental basis; and third, the effects of segmentally-based orthography may perhaps be smaller in learners whose L1 writing system is not alphabetic, that is, not segment-based. Alternatively, our finding of reduced prosody recognition after exposure to the unfamiliar script transcriptions may also be explained as due to increased cognitive demands of trying to read the unknown symbols, which resulted in overall reduced attention to the auditory signal as such. That would predict not only reduced tracking of prosody but also reduced learning of the segmental details. A future study could help decide between the alternative explanations if it compares the effects of shallow orthography using familiar versus unfamiliar characters on the tracking both of the prosodic properties (reduced learning predicted by both possible explanations) and of segmental properties of speech in a novel language (reduced learning predicted by only the latter explanation).

At the level of the neural processing of speech, the brain's ability to track words and syllables (that is, chunks of speech larger than segments), indexed by the strength and phase coherence of the neural oscillatory activity in the delta and theta bands, has been shown to correlate with the listeners' familiarity with and proficiency in the target language (being most precise for those who had the target language as their L1, Ding et al. 2016, Lizarazu et al. 2021); hampered neural speech tracking particularly in the delta band has been reported as a marker of dyslexia (Hämäläinen et al. 2012). Experiments following up on the present finding that orthography limits the ability to attune to and recognize novel language prosody

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

should investigate how orthographic input affects neural speech tracking: for instance, whether orthography interferes with neural tracking of words or syllables in a novel language, or, how the interactions between orthographic input and neural speech tracking of the different-sized chunks develops with proficiency in a second language.

4.3. Implications for second language acquisition theory and teaching practice

The present finding that adults can sensitize to the prosody of a novel language suggests that the typical struggle with prosody in second-language learning might be due to exogenous, experiential factors rather than endogenous factors such as a closed sensitive period. In our view, our results therefore align with theories of L2 acquisition which propose that the effects of the learner's age on L2 acquisition success are not due to maturation *per se* but rather the external learning conditions (Bialystok 1997, Moyer 2004, Baumert et al. 2020, Singleton & Lesniewska 2021, and Flege & Bohn 2021). Prior research suggests that promoting the salience of prosody by manual gestures or by providing L2 learners with prosodically enhanced input can lead to prosodic bootstrapping in the L2, that is, facilitate the learning of the L2 even beyond the prosodic patterns themselves (Baills et al. 2019, Yuan et al. 2019, Campfield & Murphy 2014). Such prosodic bootstrapping of other linguistic knowledge is hypothesized to be an important characteristic of first-language development (e.g., Gervain et al. 2020).

Arguably, one of the experiential factors that block access to the prosody of the target language is premature exposure to its orthography. Orthographic representations of L2 words are typically omnipresent in the input from the earliest stages of second-language development in literate learners (and often even in illiterate ones, Sbertoli & Arnesen 2014). It has been proposed that the involvement of reading and writing in the early stages of L2 development supports the explicitness of learning and analytical thinking about language (see Discussion in Miterrer & Reinisch 2015). Unlike segments, prosody is generally not marked (in an alphabetic writing system) and needs to be acquired implicitly. Premature exposure to alphabetic spelling in the learning of a new language may therefore pose a problem for the acquisition of L2 prosody, since the learners' explicit attention is drawn to the segments, at the cost of prosody. Our finding of implicit auditory attunement to a novel language after even a brief exposure period is in line with recent research showing that adults are well able to learn by implicit exposure to a second or foreign language (Alexander et al. 2023).

It is possible that if learners miss the chance to sensitize to L2 prosody early in L2 development, due to high exposure to orthography in the input, they may be prevented from sensitizing to the L2 prosody in a similar fashion as they did in their L1. We do not mean to suggest that the acquisition of L2 segments or lexical items, for which segment-based orthography may be helpful, is not important in L2 acquisition. It is possible though that prosody is even more important than segments at the initial stages of the acquisition of any language (native or second): in the same way it bootstraps the development of L1, prosody might serve as a stepping stone into a successful acquisition of the second language. As outlined in the introduction, it is possible that the apparent age-related detriment in L2 acquisition is due to the lack of initial sensitization to L2 prosody, which reduces the potential bootstrapping or facilitation effects that prosody may have on the learning of other language levels and further language competences in the L2 such as word segmentation or morphosyntactic relations. This means that the lack of early prosody attunement may be much more hindering to L2 development than just making the learner sound non-native. This is why it is worth investigating the factors that may facilitate prosody sensitization at the initial stages of L2 learning, such as postponed exposure to L2 orthography. To what extent does presenting orthography adversely affect the learning of the L2 sound patterns overall or the learning of prosody in particular? This remains to be resolved in future research. In any case, presenting (unfamiliar or deep) orthography distracts the learner away from the sound of the L2 speech. This should be considered in foreign language classrooms where learners are routinely exposed to even deep and unfamiliar orthographies right from the first moment without the necessary awareness of any adverse effects on sound-pattern learning this may have.

One implication of our findings is thus that using orthographic representations in second or foreign-language classrooms almost all of the time should be discouraged if the acquisition of prosody is to be promoted. The idea that early exposure to L2 orthography interferes with the attunement to L2 prosody needs to be tested further. If confirmed, this will lend support to teaching approaches which see L2 and L1 learning as parallel processes (Morgan-Short et al. 2012) and encourage listening before speaking and definitely before writing. Until then, we cannot eliminate the risk that early exposure to orthography and insistence on correct spelling (often common even with child L2 learners in schools) has harmful effects on the acquisition of L2 prosody which is essential for the successful acquisition of sound patterns, and possibly helpful for the acquisition of patterning on other levels, of the target language (Anderson-Hsieh et al. 1992, Warren et al. 2009). Unlike in L1 acquisition and in immersive non-formal L2 acquisition, formalized instructional L2 learning routinely uses immediate exposure to L2 orthography. Understanding the potential unintended effects of early exposure to orthography on the acquisition of L2 sound patterns is essential both from the perspective of SLA theory and practice. Our study provides an indication of the potential disadvantages of early L2 orthography use and warrants further investigation.

5. Conclusion

After passively listening to 5 minutes of speech from a previously encountered natural language, adults recognize this language above chance in low-pass filtered recordings of new persons speaking the exposed or a different related language. The accuracy in recognizing the exposed language is lower if the exposure includes orthographic, alphabetic, representation – especially if the writing does not transparently reflect the auditory realization of speech sounds (that is, deep orthography) or if it is in an unfamiliar script. These findings demonstrate that adults are able to attune to the prosodic patterns of a novel language during a brief passive exposure and that presenting novice learners with the written form of the language may interfere with the prosody tracking ability. That has implications for linguistic theories of sensitive periods, as well as for applied language teaching research. In follow-up research, we aim to investigate the developmental trajectory and neural underpinnings of the interactions between prosodic sensitization and orthography. Future studies should also test how orthography may affect novel language prosody tracking for different combinations between the listeners' L1 and the particular new language.

Data availability statement. The stimuli, materials, the experiment package, data, and analysis scripts are available at https://osf.io/p4b5m/.

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press

Acknowledgments. We are grateful to Norbert Vanek for helping us purchase the Māori audiobook and for shipping it to us from Auckland, New Zealand, and to Barbora Dvořáková for assistance in data collection.

Funding statement. The research was funded by the Czech Science Foundation, grant no. 21-09797S and by the European Regional Development Fund, project "Beyond Security: Role of Conflict in Resilience-Building," reg. no.: CZ.02.01.01/00/22_008/0004595.

Competing interest. The authors declare no competing interest.

References

Alexander, E., **Van Hedger, S. C.**, & **Batterink, L. J.** (2023). Learning words without trying: Daily second language podcasts support word-form learning in adults. *Psychonomic Bulletin & Review*, **30**(2), 751–762.

Anderson-Hsieh, J., **Johnson, R.**, & **Koehler, K.** (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, **42**(4), 529–555.

Baills, F., **Suarez-Gonzalez, N.**, **Gonzalez-Fuente, S.**, & **Prieto, P.** (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, **41**(1), 33–58.

Bassetti, B. (2017). Orthography affects second language speech: Double letters and geminate production in English. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **43**(11), 1835–1842.

Bates, D., **Mächler M**, **Bolker B**, **Walker S** (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**(1), 1–48.

Baumert, J., **Fleckenstein, J.**, **Leucht, M.**, **Köller, O.**, & **Möller, J.** (2020). The long-term proficiency of early, middle, and late starters learning English as a foreign language at school: A narrative review and empirical study. *Language Learning*, **70**(4), 1091–1135.

Bialystok, E. (1997). The structure of age: In search of barriers to second language acquisition. *Second language research*, **13**(2), 116–137.

Boersma, P. & **Weenink, D.**(1992–2023). *Praat: doing phonetics by computer* [Computer program]. Version 6.4.01, retrieved 30th November 2023 from http://www.praat.org/.

Bozkurt, C., & **Soley, G.** (2022). Adult listeners can extract age-related cues from child-directed speech. *Quarterly Journal of Experimental Psychology*, **75**(12), 2244–2255.

Campfield, D. E., & **Murphy, V. A.** (2014). Elicited imitation in search of the influence of linguistic rhythm on child L2 acquisition. *System*, **42**, 207–219.

Chun, D.M, & **Levis, J.M.** (2020). Prosody in Second Language Teaching: Methodologies and effectiveness. In Gussenhhoven, C. & Chen, A. (Eds.), *The Oxford handbook of language prosody* (pp. 619–630). Oxford: Oxford University Press.

Choi, S., & **Kang, O.** (2023). The roles of suprasegmental features in assessing paired speaking tasks in high-stakes language assessment. *System*, **119**, 103183.

Clynes, A., & **Deterding, D.** (2011). Standard Malay (Brunei). *Journal of the International Phonetic Association*, **41**(2), 259–268.

DeCasper, A. J., **Lecanuet, J. P.**, **Busnel, M. C.**, **Granier-Deferre, C.**, & **Maugeais, R.** (1994). Fetal reactions to recurrent maternal speech. *Infant Behavior and Development*, **17**(2), 159–164.

Dehaene, S., **Pegado, F.**, **Braga, L. W.**, **Ventura, P.**, **Nunes Filho, G.**, **Jobert, A.**, … **Cohen, L.** (2010). How learning to read changes the cortical networks for vision and language. *Science*, **330**(6009), 1359–1364.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient for delta. In Karnowski, P. C. & Szigeti, I. (Eds.) *Language and language-processing (pp.* 231–241*).* Frankfurt am Main: Peter Lang

Ding, N., **Melloni, L.**, **Zhang, H.**, **Tian, X.**, & **Poeppel, D.** (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, **19**(1), 158–164.

Escudero, P., & **Boersma, P.** (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, **26**(4), 551–585.

Escudero, P., **Simon, E.**, & **Mulak, K. E.** (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, **17**(2), 384–395.

Flege, J. E., & **Bohn, O. S.** (2021). The revised speech learning model (SLM-r). In Wayland, Ratree (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge: Cambridge University Press.

Gerken, L. (1994). Young children' s representation of prosodic phonology: Evidence from English-speakers' weak syllable productions. *Journal of Memory and Language*, **33**(1), 19–38.

Gervain J., **Christophe A.**, **Mazuka R.** (2020). Prosodic bootstrapping. In Gussenhoven, C., & Chen, A. (Eds.), *The Oxford handbook of language prosody (pp.* 563–573*).* Oxford: Oxford University Press.

Gleitman, L. & **Wanner, E.** (1982). L anguage acquisition: The state of the art. In E. Wanner & L. Gleitman (Eds.), *The state of the art* (pp. 3–48). Cambridge: Cambridge University Press.

Grabe, E. & **Low, E.** (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Ed.), *Laboratory phonology* **7** (pp. 515–546). Berlin, New York: De Gruyter Mouton.

Granier-Deferre, C., **Ribeiro, A.**, **Jacquet, A. Y.**, & **Bassereau, S.** (2011). Near-term fetuses process temporal features of speech. *Developmental science*, **14**(2), 336–352.

Goetry, V., **Urbain, S.**, **Morais, J.**, & **Kolinsky, R.** (2005). Paths to phonemic awareness in Japanese: Evidence from a training study. *Applied Psycholinguistics*, **26**(2), 285–309.

Gordon, R. L., **Jacobs, M. S.**, **Schuele, C. M.**, & **McAuley, J. D.** (2015). Perspectives on the rhythm–grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences*, **1337**(1), 16–25.

Hallé, P. A., **De Boysson-Bardies, B.**, & **Vihman, M. M.** (1991). Beginnings of prosodic organization: Intonation and duration patterns of disyllables produced by Japanese and French infants. *Language and Speech*, **34**(4), 299–318.

Hämäläinen, J. A., **Rupp, A.**, **Soltész, F.**, **Szücs, D.**, & **Goswami, U.** (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: an MEG study. *Neuroimage*, **59**(3), 2952–2961.

Harlow, R., **Keegan, P.**, **King, J.**, **Maclagan, M.**, & **Watson, C.** (2009). The changing sound of the Māori language. In Stanford, J. & Preston D. (Eds.), *Variation in indigenous minority languages* (pp. 129–152).Amsterdam: John Benjamins.

Havron, N., **Raviv, L.**, & **Arnon, I.** (2018). Literate and preliterate children show different learning patterns in an artificial language learning task. *Journal of Cultural Cognitive Science*, **2**(1), 21–33.

Jacewicz, E., **Fox, R. A.**, & **Holt, C. E.** (2023). Dialect and gender perception in relation to the intelligibility of low-pass and high-pass filtered spontaneous speech. *The Journal of the Acoustical Society of America*, **154**(3), 1667–1683.

Johnson, E. K., & **Jusczyk, P. W.** (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of memory and language*, **44**(4), 548–567.

Johnson, E. K., **Jusczyk, P. W.**, **Cutler, A.**, & **Norris, D.** (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology*, **46**(1), 65–97.

Kolly, M. J., **Leemann, A.**, & **Dellwo, V.** (2014). Foreign accent recognition based on temporal information contained in lowpass-filtered speech. In *Proceedings of INTERSPEECH 2014* (pp. 2175–2179). Singapore.

Levitt, A. G., & **Wang, Q.** (1991). Evidence for language-specific rhythmic influences in the reduplicative babbling of French-and English-learning infants. *Language and Speech*, **34**(3), 235–249.

Lizarazu, M., **Carreiras, M.**, **Bourguignon, M.**, **Zarraga, A.**, & **Molinaro, N.** (2021). Language proficiency entails tuning cortical activity to second language speech. *Cerebral Cortex*, **31**(8), 3820–3831.

Long, M. H. (1990). Maturational constraints on language development. *Studies in second language acquisition*, **12**(3), 251–285.

Lüdecke, D., **Aust, F.**, **Crawley, S.**, & **Ben-Shachar, M.** (2020). *Package 'ggeffects'. Create tidy data frames of marginal effects for "ggplot" from model outputs, 23.*

Maclagan, M., **Watson, C. I.**, **King, J.**, **Harlow, R.**, **Thompson, L.**, & **Keegan, P.** (2009). Investigating changes in the rhythm of Maori over time. In *Tenth annual conference of the international speech communication association* (pp. 1535–1538). Brighton.

Mampe, B., **Friederici, A. D.**, **Christophe, A.**, & **Wermke, K.** (2009). Newborns' cry melody is shaped by their native language. *Current biology*, **19**(23), 1994–1997.

Mandel, D. R., **Nelson, D. G. K.**, & **Jusczyk, P. W.** (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, **11**(2), 181–196.

Markham, P. (1999). Captioned videotapes and second language listening word recognition. *Foreign Language Annals*, **32**, 321–328.

Mehler, J., **Jusczyk, P.**, **Lambertz, G.**, **Halsted, N.**, **Bertoncini, J.**, & **Amiel-Tison, C.** (1988). A precursor of language acquisition in young infants. *Cognition*, **29**(2), 143–178.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, **32**(4), 543-563.

Mitterer, H., & **McQueen, J. M.** (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PLoS One*, **4**(11), e7785.

Mitterer, H., & **Reinisch, E.** (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language*, **85**, 116–134.

Mok, P. P. K., **Lee, A.**, **Li, J. J.**, & **Xu, R. B.** (2018). Orthographic effects on the perception and production of L2 mandarin tones. *Speech Communication*, **101**, 1–10.

Molnar, M., **Gervain, J.**, & **Carreiras, M.** (2014). Within-rhythm class native language discrimination abilities of Basque-Spanish monolingual and bilingual infants at 3.5 months of age. *Infancy*, **19**(3), 326–337.

Montero Perez, M., **Van Den Noortgate, W.**, & **Desmet, P.** (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System*, **41**, 720–739.

Montero Perez, M. (2020). Multimodal input in SLA research. *Studies in Second Language Acquisition*, **42**(3), 653–663.

Morais, J., **Cary, L.**, **Alegria, J.**, & **Bertelson, P.** (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, **7**(4), 323–331.

Morgan-Short, K., **Steinhauer, K.**, **Sanz, C.**, & **Ullman, M. T.** (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, **24**(4), 933–947.

Moyer, A. (2004). *Age, accent and experience in second language acquisition*. Clevedon, England: Multilingual Matters, 192.

Munro, M., & **Mann, V.** (2005). Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics*, **26**(3), 311–341.

Nazzi, T., **Bertoncini, J.**, & **Mehler, J.** (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, **24**(3), 756.

R Core Team (2021). R: A language and environment for statistical computing. *R foundation for statistical computing*, Vienna, Austria. https://www.R-project.org/.

Ramus, F., **Dupoux, E.**, & **Mehler, J.** (2003). The psychological reality of rhythm classes: Perceptual studies. In *Proceedings of the 15th international congress of phonetic sciences* (Vol. **3**, pp. 337–342).

Richards, D.S., **Frentzen, B.**, **Gerhardt, K.J.**, **McCann, M.E.** & **Abrams, R.M.** (1992). Sound levels in the human uterus. *Obstetrics & Gynecology*, **80**, 186–190.

Rowling, J.K. (2022). *Hare Pota me te Whatu Manapou [Harry Potter and the Philosopher's Stone]*. Translated by Nā Leon Heketū Blake. Auckland: Auckland University Press, 332.

Saito, K., **Trofimovich, P.**, & **Isaacs, T.** (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, **37**(2), 217–240.

Sbertoli, G., & **Arnesen, H.** (2014). Littératie et apprentissage de la langue dans l'intégration des immigrés en Norvège. *Les Politiques Sociales*, **12**(1), 46–61.

Singleton, D., & **Leśniewska, J.** (2021). The critical period hypothesis for L2 acquisition: An unfalsifiable embarrassment? *Languages*, **6**(3), 149.

Suppanen, E., **Huotilainen, M.**, & **Ylinen, S.** (2019). Rhythmic structure facilitates learning from auditory input in newborn infants. *Infant Behavior and Development*, **57**, 101346.

Trofimovich, P., & **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), 1–30.

Trofimovich, P., & **Baker, W.** (2007). Learning prosody and fluency characteristics of second language speech: The effect of experience on child learners' acquisition of five suprasegmentals. *Applied Psycholinguistics*, **28**(2), 251–276.

Vicenik, C., & **Sundara, M.** (2013). The role of intonation in language and dialect discrimination by adults. *Journal of Phonetics*, **41**(5), 297–306.

Wan, A. (2012). *Instrumental phonetic study of the rhythm of Malay*. [PhD thesis, Newcastle University]. [http://hdl.handle.net/10443/1682]

Warren, P., **Elgort, I.**, & **Crabbe, D.** (2009). Comprehensibility and prosody ratings for pronunciation software development. *Language Learning & Technology*, **13**(3), 87–102.

Weatherhead, D., **Friedman, O.**, & **White, K. S.** (2019). Preschoolers are sensitive to accent distance. *Journal of Child Language*, **46**(6), 1058–1072.

Wermke, K., **Teiser, J.**, **Yovsi, E.**, **Kohlenberg, P. J.**, **Wermke, P.**, **Robb, M.**, … & **Lamm, B.** (2016). Fundamental frequency variation within neonatal crying: Does ambient language matter?. *Speech, Language and Hearing*, **19**(4), 211–217.

Winke, P., **Gass, S.**, & **Sydorenko, T.** (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning and Technology*, **14**, 65–86.

Yuan, C., **Gonzalez-Fuente, S.**, **Baills, F.**, & **Prieto, P.** (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, **41**(1), 5–32.

Zhou, C. & **Hamann, S.** (2020). Cross-linguistic interaction between phonological categorization and orthography predicts prosodic effects in the acquisition of Portuguese liquids by L1-Mandarin learners. In *Proceedings of INTERSPEECH 2020* (pp. 4486–4490).

https://doi.org/10.1017/S1366728925000082 Published online by Cambridge University Press