Journal of Memory and Language 141 (2025) 104600

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Contribution of prior linguistic knowledge to L3 phonological perception and production

Tal Norman, Anat Prior [1], Tamar Degani [1,*]

University of Haifa, Haifa, Israel

ARTICLE INFO

Keywords:
Cross-language influence
Trilingual
Phonology
Oddity task

ABSTRACT

Adult phonological processing may be affected by previous linguistic knowledge. Here, we examine how phonological perception and production in a third-language (L3) are affected by multilinguals' first- (L1) and second-languages (L2). To this end, Arabic-Hebrew-English trilinguals ($n = 41$) completed an oddity (perception) task and a word repetition (production) task in English (the L3). Critically, word pairs ($n = 96$) targeted phonological contrasts that overlap between English and Arabic (L1), English and Hebrew (L2), English and both Arabic and Hebrew (Both) or exist uniquely in English (None). Results showed that words including phonological contrasts that exist in L1 Arabic (L1 & Both conditions) were perceived more accurately than those that do not exist in the L1 (L2 & None conditions). This pattern cannot be the mere result of item characteristics, because using the same items, a control group of Hebrew-English bilinguals ($n = 39$) responded more accurately when phonological contrasts overlapped with Hebrew (their L1). We further verified that the L2 contrasts had at least partially been acquired in the L2, by testing an additional group of trilinguals ($n = 27$), who performed above chance on these contrasts when embedded in an L2 task. Judgments collected from monolingual English evaluators revealed that trilingual productions exhibited the same pattern as that observed in perception, with more intelligible productions of contrasts which overlap with the L1, but not with the L2. Thus, multilinguals appear to draw on their L1 knowledge, but not on their L2 knowledge, while processing phonological information in the L3. The findings further underscore the relation between phonological perception and production in the L3.

Introduction

Processing of auditory verbal information is determined in part by the phonological categories the listeners impose. In the case of multilingual listeners, these categories may be shaped not only by the individual's experience with the target language, but also by broader prior linguistic knowledge. Ample evidence documents cross-language influences in phonological processing of bilingual individuals showing that perception and production of second-language (L2) phonological information is shaped by perceptual categories of the first-language (L1) (e.g., Best & Tyler, 2007; Evans & Alshangiti, 2018; Flege, Bohn, & Jang, 1997). Of note, it is difficult to dissociate whether such influences stem from entrenchment of early phonological experiences during infancy (in the native language), or whether they are indicative of ongoing interactions in the phonological system (among all acquired languages, regardless of age-of-acquisition). To better dissociate these mechanisms, the present study relies on evidence from trilingual individuals, examining phonological processing in the third-language (L3). In such trilingual individuals, L3 phonological processing may be affected by the L1, the L2, or both, revealing important principles in the operation of the adult phonological system. Further, whereas previous research often focused on self-selected trilinguals, namely individuals who have chosen to learn additional languages which has lead them to become trilinguals (e.g., Onishi, 2016; Sypiańska, 2022; Wrembel et al., 2020; Zhu & Mok, 2023), and on those who use multiple related languages (e.g., Garcia, 2013; Lloyd-Smith, 2023; Wrembel, 2010; though see Stoehr et al., 2024), the current study tests Arabic-Hebrew-English trilinguals who are sampled from a large community of multilingual speakers. Further, we examine phonological processing across both perception and production tasks, allowing characterization of the interplay between these modalities in trilingual phonological processing.

* Corresponding author at: Department of Communication Sciences & Disorders, University of Haifa, Haifa, Israel.
E-mail address: tdegani@research.haifa.ac.il (T. Degani).

[1] Equal contribution

https://doi.org/10.1016/j.jml.2024.104600
Received 4 March 2024; Received in revised form 6 December 2024; Accepted 11 December 2024
Available online 21 December 2024

L3 phonological processing – Source of language influences

Cross-language phonological influences have been extensively studied in the case of bilingual speakers. Specifically, L2 phonological processing is determined in part by phonological perceptual categories of the L1 (Best & Tyler, 2007). However, such influences may stem from two different reasons. First, it is possible that perceptual phonological categories are shaped early in development, during infancy, limiting the influence of subsequent exposure. Under such an account, perceptual categories are entrenched early on, and thus processing of new information in adulthood is heavily influenced by the organization of the phonological system as determined early in childhood, possibly within the first year of life (Kuhl, 2004). However, the mere presence of L1 cross-language influences during L2 phonological processing may also be explained within an interactive phonological system. Under this framework, all prior linguistic knowledge is called upon during a linguistic processing task. Thus, L1 influences on L2 processing should be accompanied, albeit to a lesser extent, with L2 influences on L1 processing (for review see e.g., McDonald & Kaushanskaya, 2020; Pavlenko, 2020 Table 1).

A complementary approach by which to reveal the operation of these mechanisms is to examine processing of trilingual individuals. In the case of trilinguals, processing of the L3 may be shaped by the L1, the L2, both the L1 and the L2, or neither of them. Theoretical work on cross-language influences in L3 has mainly focused on the domain of morphosyntax (for review see Puig-Mayenco et al., 2020), but its essence may be applicable across domains. Specifically, some models propose that L3 processing is influenced more strongly by the native language (the L1) (Hermas, 2010; 2015). In the case of phonological processing, such *L1 privilege* models align with the claim that the phonological system is determined by early exposure to the native language (Cabrelli-Amaro & Rothman, 2010; Kuhl, 2004). Exposure to additional languages during this critical period may lead to a greater variety of perceptual categories (e.g., Garcia-Sierra et al., 2016), but in the case of trilinguals who acquire their languages sequentially, as is the case here, the L1 should be the sole, or stronger, source of influence during L3 processing. In support of such accounts in the phonological domain, Llama and Cardoso (2018) investigated trilingual speakers of English, French and Spanish. Participants were advanced learners of Spanish as an L3, and there were two mirror groups regarding the L1 and the L2: French-English and English-French. Results demonstrated influence only from the L1 on VOT pronunciation in L3.

Other models, however, suggest that L3 processing is shaped by the L2. These so called *L2 Status Factor* accounts (Bardel & Falk, 2007; 2012) propose that acquisition and representation of both later-acquired languages, the L2 and the L3, are subserved by similar neuro-cognitive systems (Paradis, 2009; Ullman, 2001; 2020), making transfer across these languages more likely than from L1 to L3. Thus, the similarity in the acquisition circumstances of the L2 and the L3 makes the L2 the sole source of transfer during L3 processing. Supporting this account, Geiss et al. (2022) report that Italian-German-English trilinguals (heritage speakers of Italian) did not differ from German-English bilinguals in VOT values for stop consonants in English the L3. This finding demonstrates cross-language influence from the L2, German, but not from their heritage language, Italian. However, these trilinguals were more proficient in the L2 than in the L1, they had acquired German at an early age, and German is typologically closer to English (the L3) than is Italian (the L1). Thus, it is difficult to disentangle the possible influences of proficiency and typological similarity from those of order of acquisition.

Finally, a third class of more recent models suggest that L3 processing is influenced by all prior linguistic information, with some variants of this approach suggesting that the dominant source of influence is determined early in the process of L3 acquisition (Typological Primacy Model, TPM, Rothman, 2011; 2015), whereas others allow for continuous influences from both languages on a property by property basis (the Linguistic Proximity Model, LPM, Westergaard et al., 2017; Westergaard, 2021; and the Scalpel Model, Slabakova, 2017).

Evidence supporting such interactive influences is the most common in studies of perception and production in L3 phonology (for a recent review see Wang & Nance, 2023). Of relevance, these studies have mostly focused on specific phonological, or phonetic, contrasts in L3 and examined how their degree of overlap with learners' L1 and L2 might impact their processing. For example, Zhu and Mok (2023) examined how speakers of Cantonese (L1) and English (L2) produce consonant clusters in German (L3). Results showed that overlap with both previously learned languages influenced production accuracy in L3. Specifically, participants were more likely to produce an L3 cluster correctly if it included phonemes that exist in the L1, and if the cluster itself existed in the L2. Notably, participants in that study had acquired English, their L2, at an early age (before age 3) and were highly proficient in it. Further, the words for production were presented in writing, and German shares the Latin alphabet with English. These two factors might have increased the observed influence of L2 on L3 phonological production.

Parrish (2022) reported that bilinguals used categories from both their L1 and their L2 when categorizing sounds in a novel L3, suggesting that both previously learned languages brought to bear on the task (see also Cabrelli & Pichan, 2021; Patience, 2018). Kopečková and colleagues (2023) report influence from both L1 and L2 on L3 production, though influence was stronger from the L1. Notably, the balance of cross-language influences from the previously learned languages differed by the specific phonetic contrast examined (see also Archibald, 2023), lending support to the notion of property-by-property cross-language influence (Slabakova, 2017; Westergaard et al., 2017).

Finally, there are also cases when no cross-language influence was identified in the production of L3 phonemes. Amengual (2021) examined production of stop consonants in L3 Japanese of L1-English L2-Spanish speakers, and documented distinct productions for each of the languages, concluding that trilinguals were able to overcome influence from both their previously learned languages. In a small-scale study, Gut (2010) as well, does not identify clear evidence for cross-language influence from either L1 or L2 on the production of vowels in the L3 (English or German).

As evident in the literature presented above there is substantial variability across studies, but the most common pattern emerging is that of influences from both the L1 and the L2 during L3 phonological processing. This is corroborated in a recent comprehensive review of the literature (Wang & Nance, 2023). Of note, this review further highlights important limitations in the extant literature. First, about 16 studies focused on L3 phonological *production* and only six studies examined L3 phonological *perception* (see Tables 1 and 2 in Wang & Nance, 2023). Moreover, the generalizability of the findings may be constrained by the

Table 1
Background characteristics of the Trilingual Participants in Experiments 1a & 2; Mean (SD).

General Background	Trilingual Participants		
N	41		
Age (years)	20 (0.8)		
Gender (number of males/females)	5/36		
Education (years)	12.96 (0.36)		
Maternal Education (years)	13.25 (3.24)		
Language Specific Background	**Arabic**	**Hebrew**	**English**
Age Of Acquisition	0	7.73 (1.12)	8.22 (0.96)
Mean Subjective Proficiency (0–10)	9.67 (0.53)	7.55 (1.07)	6.27 (1.88)
Mean Subjective Current Use (%)	41.84 (15.86)	29.48 (13.32)	17.19 (13.05)
Semantic Fluency (sum of valid words in both categories)	17.02 (3.68)	12.90 (4.64)	9.76 (4.67)

Note: Of the 41 participants, 9 self-reported higher proficiency in English (L3) compared to Hebrew (L2).

Table 2
The 8 phonological contrasts in the 4 phonological overlap conditions, with example word-pairs.

Phonological Overlap	Contrast	Word-Pair
Both	ɪ / ʌ	fin / fun
	d / t	dry / try
Arabic	iː / ɪ	feel-fill
	t /θ	mat / math
Hebrew	stress location (penultimate/ ultimate)	present / present
	b / p	big / pig
None	ɑ / ʌ	dog / dug
	± size of consonant cluster (2/3)	cream / scream

Note: ± Neither Arabic nor Hebrew include words with clusters of 3 consonants.

fact that many of these studies were conducted with trilingual speakers of Indo-European languages (see Table 3 in Wang & Nance, 2023), and all but two relied on relatively small sample sizes. Therefore, the current study aims to improve on these shortcomings by testing a larger sample ($n = 41$) of trilingual speakers of two Semitic languages (Arabic L1 and Hebrew L2), with English as the L3. Finally, we tested the same participants across both perception and production, allowing examination of the patterns of cross-language influences in both aspects of L3 phonological processing.

Cross language influences in phonological perception and production

Different theoretical accounts of the relation between perception and production in non-native language learning have been proposed. According to the theoretical approach of PAM (Best, 1995) and PAM-L2 (Best & Tyler, 2007), perception is based on articulatory gestures. Thus, although not directly discussed in these models, perception and production should be tightly coupled. In support of this view, Evans and Alshangiti (2018) tested native Arabic speakers in their phonological perception and production of English vowels and consonants. They observed that individuals with better vowel identification were also more intelligible in their vowel productions, suggesting a link between production and perception. The SLM (Flege, 1995; see also SLM-r, Flege & Bohn, 2021) directly addresses the perception-production link, and claims that accurate L2 production is only possible after stable perceptual categories have been established (e.g., Casillas, 2020; Sakai & Moorman, 2018). Casillas (2020), for instance, tracked the perception and production performance of a group of adult L2 learners over a 10-week intensive immersion program. Results showed that shifts in L2 perception preceded those in production. In a *meta*-analysis of the literature, Sakai and Moorman (2018) show that L2 perceptual training can result in gains in L2 production, again highlighting the link between the two modalities (see also Bradlow et al., 1999). It may also be the case that perception and production are decoupled when particular phonemes are salient in perception, but their motor-articulatory execution is challenging (Nagle, 2018; see discussion in Wrembel et al., 2022).

Table 3
Mean stimuli characteristics as a function of phonological overlap condition.

Measure	Both	Arabic	Hebrew	None
Subtlex Word Frequency	3.23 (0.59)$_a$	3.09 (0.64)$_a$	2.90 (0.64)$_a$	3.02 (0.56)$_a$
Phonological Distance	0.09 (0.05)$_a$	0.08 (0.04)$_a$	0.02 (0.02)$_a$	0.51 (0.45)$_b$
Mean Length in Letters	3.56 (0.63)$_{a,c}$	4.33 (0.60)$_{a,c,b}$	5.12 (1.70)$_b$	4.29 (0.96)$_c$
Mean Length in Phonemes	3.29 (0.46)$_a$	3.42 (0.58)$_a$	5.62 (2.50)$_b$	3.83 (0.99)$_a$

Note: Standard deviations (SDs) are shown in parentheses. Averages in the same row that do not share alphabetic subscript differ at the $p < .05$ level, based on a one-way Anova with Bonferroni corrections for multiple comparisons.

Nagle and Baese-Berk (2022) review the extant literature and current theoretical accounts and highlight the great variability in methods and tasks which precludes strong conclusions from the available literature regarding perception-production links in L2. Furthermore, knowledge gained from L2 phonological processing does not necessarily generalize to L3 processing. Specifically, Wrembel et al., (2022) report stronger perception-production links in L2 than in L3, which they attribute to differences in language proficiency, as their participants were in the initial stages of L3 acquisition (for related evidence see Hanulíková et al., 2012).

The current study aims to contribute to this literature by testing more advanced L3 speakers, with a focus on the ways in which cross-language influences manifest in production and perception. As noted above, the research on cross-language influences has mostly been tested independently in L3 production and perception (Wang & Nance, 2023). An exception to this is a study by Liu and Lin (2021) which tested a group of L1 Mandarin-Chinese L2 English beginning learners of L3 Russian or Japanese. They focused on the stop system, on which the L1 and L2 differ in VOT from that used in the L3. In a perception task, there were cross-language influences in the perception of L3 voiceless stops. Specifically, learners relied on VOT length which is the relevant dimension in their L1 and L2. In the case of voiced stops, because word initial pre-voicing is a unique feature in the L3 that does not exist in the L1 or L2 of the speakers, learners appear to have been more accurate, with less interference from prior knowledge. Most critically, because the same participants also completed a production task using the same experimental items, the authors were able to examine the correlation across perception and production. They found a positive correlation between perception and production in the case of voiceless stops, where participants show cross-language influences from L1 and L2. However, there was no perception-production correlation in the case of pre-voiced stops, which are unique to the L3, and also pose an articulatory challenge. Continuing this single study, the current work examines whether L1 and L2 influence L3 processing in a perceptual oddity task as well as a word production task, to gauge the alignment of cross-language influences in perception and production. Importantly, we examine a relatively large number of phonological features and contrasts, to provide a broad description of the phenomena.

The current study

To examine the degree to which L3 phonological perception and production are influenced by prior linguistic knowledge, the current study tested a group of Arabic-Hebrew-English adult trilinguals in their L3 English. Participants performed an oddity discrimination task and also produced the same items, which were then submitted to comprehensibility ratings by native English-speaking evaluators. Critically, stimuli in both tasks targeted English phonological contrasts that either exist in L3 and L1, in L3 and L2, in L3 and both L1 and L2, or are unique to the L3. To boost our confidence that the trilinguals' performance can be ascribed to the overlap of these critical stimuli across languages, we also tested a control group of Hebrew-English bilinguals with no knowledge of Arabic on the perception task (a subtractive design, see Westergaard et al., 2023). By comparing the observed pattern in this group with that of the trilingual group we were able to discriminate between effects that stem from general phonological properties of English, as manifested in the selected item set, and those that result from cross-language influences. We further examined perceptual performance of comparable trilinguals on an L2 perception task, to examine the extent to which phonological contrasts that exist exclusively in L2 (and not in the L1) have been acquired in that language by such trilinguals. Importantly, trilinguals were moderately proficient in their L3 after having studied it for about 10 years, in contrast to most previous research which has investigated the early stages of L3 learning. Moreover, trilinguals were sampled from a large community who had not self-selected to become trilingual, providing the study ecological standing.

Experiment 1 presents findings from an oddity perception task, and Experiment 2 presents the production data, as well as analysis of correspondence across modalities.

Overall, two overarching questions guided the current study. One, to what extent are cross-language influences from L1 and L2 evident in L3 phonological processing. Two, are L3 phonological perception and production aligned with each other.

All stimuli, data, and analysis code are available on the OSF platform (https://osf.io/mdn34/?view_only = 65b37294d96e4020977eeb742fb2890d).

Experiment 1: Cross-language influences in L3 phonological perception

Experiment 1a: Arabic-Hebrew-English trilinguals in L3 English

Method

Participants. A total of 41 participants (ages 19–22)[2] with normal or corrected to normal vision and with no language, learning, or hearing disabilities, participated in Experiment 1a. All were Arabic-Hebrew-English trilingual undergraduate students living in Arabic speaking communities in Israel, who started learning Hebrew and English as their L2/L3 between the ages of 4–10 in a school setting in Israel. Although not explicitly stated in the curriculum, most instruction includes American English as the target variety. All participants were at least partially immersed in a Hebrew speaking environment at the time of testing, as they were all University students in an institution where Hebrew is the language of instruction. Participants were not fluent in any additional languages other than Arabic, Hebrew, and English. They signed an informed consent approving their participation. Their background characteristics as a function of language (L1-Arabic, L2-Hebrew, L3-English), based on their self-report ratings on a language history questionnaire (Abbas et al., 2024) and on an objective proficiency measure of semantic fluency (Kavé, 2005), are summarized in Table 1. In the semantic fluency task, participants provided as many exemplars as possible from two categories per language, over 60 s per category. All procedures associated with the reported experiments were approved by the IRB of the University of Haifa.

Stimuli. The task focused on 8 phonological contrasts in English, with different degrees of overlap with participants' L1 and L2, which are presented in Table 2. In particular, to reveal whether phonological overlap between the L1-Arabic or L2-Hebrew and the L3-English modulates L3 processing, these L3 phonological contrasts belonged to one of four conditions: (1) exist in both the L1-Arabic and the L2-Hebrew; (2) exist only in the L1-Arabic; (3) exist only in the L2-Hebrew; or (4) do not exist in either the L1-Arabic or the L2-Hebrew. Thus, each of the 4 phonological overlap conditions (i.e., Both, Arabic, Hebrew, None) included two different contrasts. For each contrast, we selected 12 word-pairs, resulting in 24 English word-pairs per condition, and a total of 96 word-pairs (192 words). The full list of stimuli is available on the OSF (https://osf.io/mdn34/?view_only = 65b37294d96e4020977eeb742fb2890d).

Across phonological overlap conditions, there were no significant differences in English Frequency based on log Subtlex Frequency (from the English Lexicon Project, Balota et al. 2007, $F(3, 92) = 1.27$, $MSE = 0.36$, $p = 0.288$, see Table 3). In addition, for each pair, we computed an objective phonological distance score between the two words of each pair by computing a Levenshtein's distance on their detailed phonological representations using the PanPhon 0.7 Python package (https://pypi.org/project/panphon/0.7/), which applies a set of rules for adding diacritics and modifiers to IPA segments based on 20 phonological features. On this phonological distance measure, there were significant differences among the conditions ($F(3, 92) = 23.85$, $MSE = 0.05$, $p < 0.001$), such that word pairs in the None condition had significantly higher phonological distance compared to the other conditions. Because phonological distance should make the differentiation between word pairs easier, if anything this should make this condition easier to process (whereas based on cross-language influences this condition should be most difficult). Nonetheless, phonological distance was included as a control variable in the analyses as were the measures of mean Length (for each pair) in Letters ($F(3, 92) = 8.56$, $MSE = 1.14$, $p < 0.001$) and in Phonemes ($F(3, 92) = 14.43$, $MSE = 1.94$, $p < 0.001$) which also differed across phonological conditions (see Table 3).

The stimuli set included audio recordings of 192 English words – 96 word-pairs (i.e., two words that differed on the critical phonological contrast, such as big/pig). Each of the 192 words was recorded by 3 different female native speakers of American English. For the oddity perception task, word triplets were created based on these 96 word pairs. In particular, we created two different triplets for each word pair. Across participants, different orders of the two words within the triplets were created. These included triplets in which one word was different than the other two, counterbalanced across 1st, 2nd and 3rd position within the triplet (e.g., big, big, pig; pig, big, pig), or triplets in which all 3 words were the same (e.g., pig, pig, pig). In each triplet, recordings were drawn from all three speakers, such that the first-, second-, and third- word in each triplet, were not produced by the same speaker. Recordings from the 3 speakers were counterbalanced across trials.

Procedure. In each trial of the lexical oddity task, participants listened to three-word recordings (e.g., big, big, pig) and were required to determine whether one of the three words was different from the other two or whether all three recordings were the same. Participants were instructed to press 1, 2, or 3 if the first-, second-, or third- word was different, respectively, or 0 if all words were the same, as quickly and accurately as possible. Before starting the task, participants read the instructions, were introduced to two examples of word triplets, and performed 6 practice trials, on which visual feedback was provided. At the start of each trial, a fixation cross was presented for 1000 ms. Then, three word-recordings were played consecutively over headphones while a blank white screen was displayed. At the offset of the third audio file, a question mark appeared on the screen for 4000 ms, or until a response was made. Response latencies were measured from the onset of the question mark.

Each participant was presented with 192 triplets in random order. There were 144 triplets with an odd-one out word – 48 trials for each of the three possible word positions (1st, 2nd or 3rd) and 48 trials on which all words were the same. Thus, participants had to listen to all three words in order to make a decision. The task was programmed and run via E-Prime 3.0 software (Psychology Software Tools, Pittsburgh, PA). The experiment was administered in a sound attenuated room and lasted approximately 30 min.

Experiment 1a was conducted as part of a larger longitudinal study, investigating the influence of L1/L2 knowledge on L3 processing. The overall experimental procedure of this project included 3 sessions administered on the same day. The first included English tasks, the second Arabic tasks, and the third Hebrew tasks. The phonological perception and production tasks were administered during the English session, directly following a semantic decision task in English. Of relevance, all experimental tasks in this first session were conducted in English (L3), and their instructions were presented in English. Nonetheless, consent forms were signed in Hebrew, and oral communication was carried out naturally in Arabic with the trilingual experimenters.

[2] These were taken from a larger group of participants that took part in a longitudinal study examining cross-language influences on L3 processing across various language domains. Only participants that produced at least 14 (out of 24) word recordings with an intact audio signal for each type of phonological contrast in the lexical production task (see Experiment 2 below), were included in the current analysis.

Thus, although the overall session was designated as an English session, a mixed language mode likely characterizes this experiment.

Results

Analysis approach

Results were analyzed using Linear Mixed-Effects (LME) models as implemented in the 'lme4′ (Baayen et al., 2008) in R (Version 4.1.2; R Core Team, 2020). To evaluate the influence of phonological overlap between the L1-Arabic/L2-Hebrew and L3-English on L3 phonological perception, we analyzed accuracy in the task (following a binomial distribution) as a function of Phonological Overlap condition. Random effect structure was determined by the *buildmer* function in the 'buildmer' package (v. 2.2, Voeten, 2021), which uses the *glmer* function for binomial distribution from the 'lme4′ package (v. 1.1.-21, Bates et al., 2015). The maximal model submitted to *buildmer* included random intercepts and random slopes justified by the design (Barr et al., 2013). Critically, the fixed effect of Phonological Overlap was forced to be included in the selected model. The selected model was refitted using the *glmer* function, and *p*-values for all fixed main effects and interactions were determined using the *anova* function from the 'stats' package. Further, simple effects and pairwise comparisons were employed using the *testInteraction* function from the 'phia' package (v. 0.2–1, Martinez, 2015), which computes chi-square test with Bonferroni adjustments for multiple comparisons.

Data analysis

The initial dataset of Experiment 1a included 7872 data points that resulted from 41 participants and 192 target words (96 word-pairs). To ensure that all analyzed data points reflected genuine decision making in the task, the dataset was inspected in terms of RT outliers. Thus, trials with RT shorter than 100 ms were excluded from analysis ($n = 504$), as these are unlikely to reflect a cognitive decision process, resulting in 7368 data points for analysis. Next, mean accuracy rates by participant were inspected, indicating that all participants performed above chance level (25 % as there were 4 possible responses) in the task (i.e., all had above 47 % accuracy).

Then, a maximal model of fixed and random effects was submitted to the *buildmer* function. This model included (1) the fixed effect of the variable of interest (dummy coded): Phonological Overlap ('None', 'Hebrew', 'Arabic', Both', with 'None' as the reference level); (2) the fixed effect of the control variables (continuous and normalized) to control for possible variation in the set of critical items: Phonological distance, length in phonemes, length in letters, and log Subtlex frequency; and (3) the random effects of Participant and Item, with by-Participant and by-Item intercepts, and by-Participant slope for Phonological Overlap.

The final model selected by *buildmer* included the fixed effect of the variable of interest – Phonological Overlap; the fixed effect of two control variables – Phonological Distance (the phonological distance between the two English words within each word-pair) and Length in Phonemes; and the random intercepts of Participant and Item. See Table 4 for model summary obtained from the *summary* function in R.

Due to this dummy coding scheme, the estimates in the summary table reflect simple effects of each factor relative to its reference level, when all other factors are at their reference. Thus, we further estimated main effects using the *anova* function which evaluates the significance of the main effects (across all levels of the other variables, rather than only at their reference level as in the summary tables). Values from the *anova* function, are reported in the text.

Within the selected model, the main effect of Phonological Overlap was significant ($F(3) = 40.91$, $p < .001$; See Fig. 1A). Pairwise comparisons of the 4 phonological overlap conditions revealed that only phonological overlap with L1-Arabic facilitated L3-English phonological perception. Specifically, accuracy in the 'Arabic' condition was higher than in the 'None' ($\chi^2(1) = 75.46$, $p < .001$) and 'Hebrew' ($\chi^2(1) = 79.72$, $p < .001$) conditions. Similarly, accuracy in the 'Both' condition was higher than in the 'None' ($\chi^2(1) = 65.91$, $p < .001$) and 'Hebrew' ($\chi^2(1) = 66.10$, $p < .001$) conditions. Lastly, the difference between the 'None' and 'Hebrew' conditions was not significant ($\chi^2(1) = 1.54$, $p = 1.00$), nor was the difference between the 'Both' and 'Arabic' conditions ($\chi^2(1) = 0.58$, $p = 1.00$), indicating that L3-English phonological perception was not further facilitated by phonological overlap with the L2-Hebrew.

Table 4
Summary of the selected LME model predicting *mean accuracy as a function of Phonological Overlap in Experiment 1a.*

Fixed Effects	***b***	***SE***	***z***	***p***
(intercept)	−0.202	0.204	−0.992	0.321
Phonological Overlap (Hebrew)	0.338	0.273	1.240	0.215
Phonological Overlap (Arabic)	2.318	0.267	8.687	< 0.001
Phonological Overlap (Both)	2.145	0.264	8.118	< 0.001
Control Variables				
Phonological Distance	0.856	0.105	8.138	< 0.001
Mean Length in Phonemes	−0.202	0.204	−0.992	0.321
Random Effects	***Variance***		***SD***	
Item (intercept)	0.497		0.705	
Participant (intercept)	0.199		0.447	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the 'None' condition in the Phonological Overlap variable), when other continuous factors are held constant at the mean (e.g., control variables), without a correction for multiple comparisons. For main effects see *F* values in the text.

To further examine if this pattern is consistent across the contrasts included in each overlap condition, we conducted an analysis of accuracy by Contrast Type (see Appendix A for the full analysis), which revealed substantial differences between the two contrasts within the 'None' (e.g., cream-scream; dog-dug) and 'Both' (e.g., dry-try; fin-fun) conditions. This suggests that the differences between the 4 phonological overlap conditions might be partially driven by idiosyncratic characteristics of the specific phonological contrasts used in each condition, and not exclusively by phonological cross-language influences (see Fig. 1B). This possibility was further examined in Experiment 1b.

Experiment 1b: Hebrew-English bilinguals in L2 English

Experiment 1b was conducted to further examine whether the observed differences between the 4 phonological overlap conditions in Experiment 1a can be ascribed to phonological cross-language influences on L3 phonological perception. This experiment was identical to Experiment 1a, except that it tested a different group of participants – Hebrew-English bilinguals who had either minimal ($n = 2$) or no ($n = 37$) reported knowledge of Arabic. Here, we examined the cross-language influences of the L1-Hebrew on the phonological perception of L2-English. Critically, because these Hebrew-English bilinguals did not have any significant knowledge of Arabic, facilitation effects were expected to be exhibited only for English phonological contrasts that exist in Hebrew, namely in the 'Both' and 'Hebrew' conditions. As such, a distinct pattern of performance among Hebrew-English bilinguals (Experiment 1b), in comparison to Arabic-Hebrew-English trilinguals (Experiment 1a), using a subtractive design approach (Westergaard et al., 2023), will increase our confidence that performance in this task is influenced by phonological overlap with participants' previously learned languages. In contrast, a similar pattern of performance across both groups will indicate that the observed effects arise from the idiosyncratic characteristics of the specific phonological contrasts used in the study.

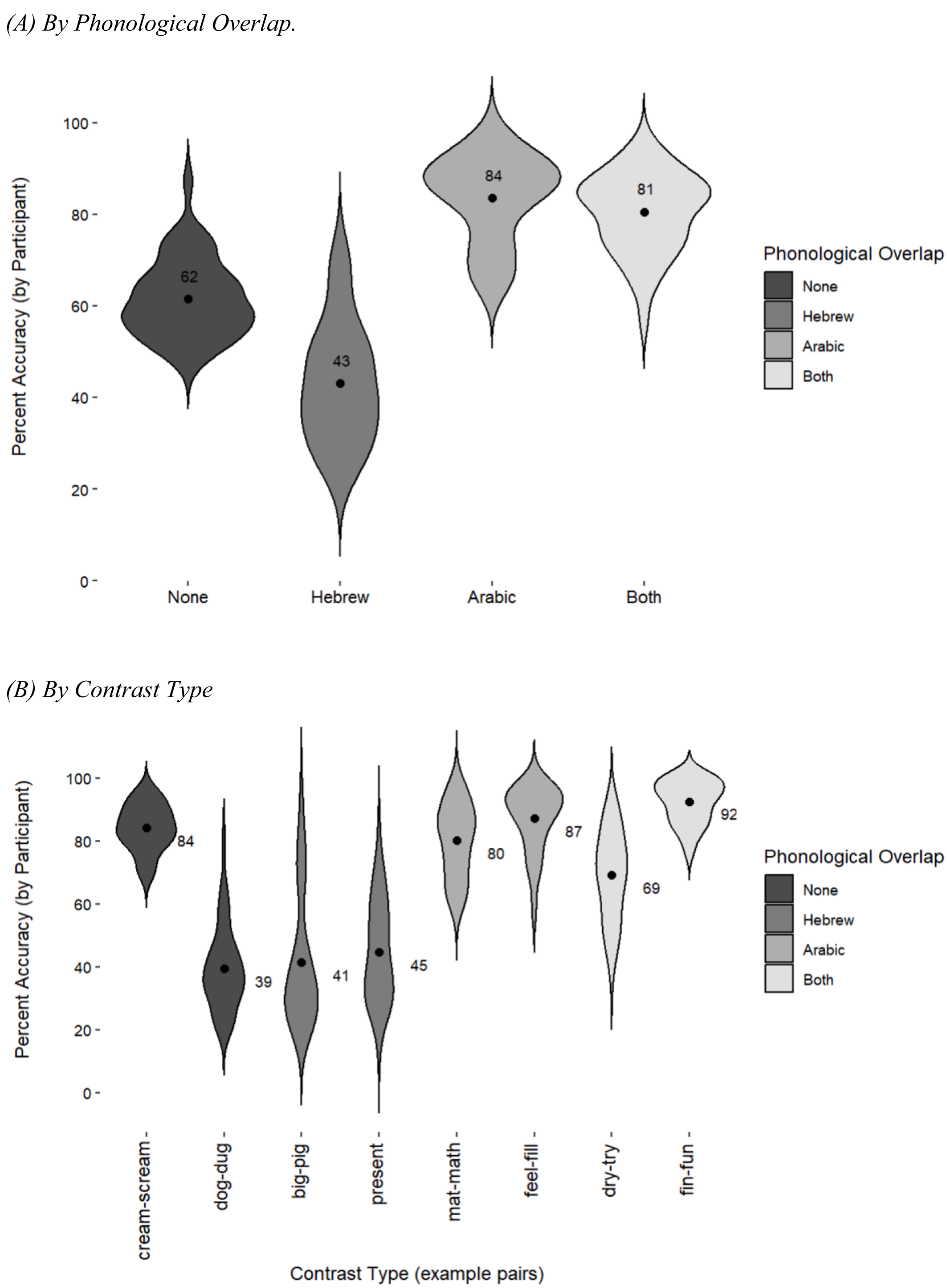

Fig. 1. Mean Perception Accuracy in Experiment 1a. A) By phonological overlap B) By contrast Type.

Method

Participants

A total of 39 participants (ages 20–32) with normal or corrected to normal vision and with no language, learning, or hearing disabilities participated in Experiment 1b. All were Hebrew-English bilingual university students living in Hebrew speaking communities in Israel, who started learning English as their L2 between the age of 4–10 in a school setting in Israel. Participants were not fluent in any additional languages other than Hebrew and English and did not report current use of any other language at the time of testing (see participant characteristics in Table 5). Note, however, that 2 participants reported learning Arabic in school but did not report any current proficiency or use of Arabic. Additionally, all participants may have been passively exposed to Arabic to different extents as a function of living in Israel, in which both Hebrew and Arabic are spoken, and learning at a university in which roughly 40 % of the student population are native Arabic speakers. Participants signed an informed consent approving their participation.

Stimuli and procedure

The set of stimuli was identical to that used in Experiment 1a (see Table 2 above). Participants performed the same lexical oddity task

Table 5
Background characteristics of the Bilinguals tested in Experiment 1b; Mean (SD).

General Background	Bilingual Participants	
N	39	
Age (years)	24.87 (3.05)	
Gender (number of males/females)	10 / 29	
Education (years)	14.51 (1.74)	
SES (mother education years)	15.15 (2.59)	
Language Specific Background	**Hebrew**	**English**
Age Of Acquisition	0 (0)	7.28 (1.54)
Mean Subjective Proficiency (0–10)	9.63 (0.79)	7.40 (1.41)
Mean Subjective Current Use (%)	59.08 (16.00)	44.90 (3.47)

described in Experiment 1a. However, because of limitations due to the COVID-19 pandemic, participants in this group performed the task in their own homes with headphones, using Eprime-Go (Psychology Software Tools, Pittsburgh, PA). Explanations and instructions were provided by the experimenter over a video meeting, prior to completing the task and the LHQ questionnaire.

Results

To reveal possible differences in the sensitivity to Phonological Overlap between the Hebrew-English bilinguals (Experiment 1b) and the Arabic-Hebrew-English trilinguals (Experiment 1a), we analyzed the results from the two experiments jointly. Following the same criteria described above in Experiment 1a, 931 data points were excluded, resulting in a combined data set that consisted of 14,379 data points.

The maximal model submitted to the *buildmer* function was identical to the one used in Experiment 1a, except for two changes. First, it included the fixed effect of Group (dummy coded; 'Bilinguals', 'Trilinguals', with 'Bilinguals' as the reference level) and the interaction between Group and Phonological Overlap. Second, the maximal random-effect structure also included the by-Item slope for Group. See Table 6 for the final model summary.

Table 6
Experiments 1a & 1b: Summary of the LME model predicting mean accuracy as a function of Phonological Overlap and Group.

Fixed Effects	*b*	*SE*	*z*	*p*
(intercept)	−0.347	0.168	−2.069	0.039
Phonological Overlap (Hebrew)	1.536	0.248	6.204	< 0.001
Phonological Overlap (Arabic)	1.057	0.214	4.943	< 0.001
Phonological Overlap (Both)	2.232	0.220	10.130	< 0.001
Group (Trilinguals)	0.169	0.178	0.951	0.341
Group (Trilinguals): Phonological Overlap (Hebrew)	−1.025	0.216	−4.754	< 0.001
Group (Trilinguals): Phonological Overlap (Arabic)	1.125	0.227	4.951	< 0.001
Group (Trilinguals): Phonological Overlap (Both)	−0.190	0.236	−0.804	0.421
Control Variables				
Phonological Distance	0.796	0.082	9.653	< 0.001
Mean Length in Phonemes	−0.245	0.081	−3.039	0.002
Random Effects	***Variance***		***SD***	
Item (intercept)	0.294		0.543	
Item (Group – Trilinguals)	0.389		0.624	
Participant (intercept)	0.163		0.404	
Participant (Phonological Overlap – Hebrew)	0.030		0.172	
Participant (Phonological Overlap – Arabic)	0.089		0.299	
Participant (Phonological Overlap – Both)	0.125		0.354	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the 'None' condition in the Phonological Overlap variable), when other categorical factors are at their reference level (i.e., the 'Bilinguals' in the Group variable), without correcting for multiple comparisons. For main effects see *F* values in the text.

Within the selected model, the main effect of Phonological Overlap was significant ($F(3) = 24.88$, $p < .001$), and the main effect of Group was marginally significant ($F(1) = 4.08$, $p = .043$). Critically, the two-way interaction between Phonological Overlap and Group was significant ($F(3) = 32.02$, $p < .01$), indicating that native Arabic Trilinguals and native Hebrew Bilinguals were differentially affected by the 4 phonological overlap conditions (see Fig. 2). Table 7 presents the pairwise comparisons between the two groups of participants, in each phonological overlap condition.

As shown in Table 7, the native Hebrew Bilinguals showed a substantially distinct pattern of results, in comparison to the native Arabic Trilinguals, in the 'Arabic' and 'Hebrew', but not in the 'Both' and 'None' overlap conditions. In the 'Arabic' overlap condition, native Arabic trilinguals were significantly more accurate than the native Hebrew Bilinguals. In striking contrast, in the 'Hebrew' overlap condition, the pattern of results was reversed. This pattern of results strongly supports the notion that the perception of English phonological contrasts was influenced by their phonological overlap with participants' native language, Hebrew and Arabic respectively.

We once again examined the consistency of the effects across Contrast Types within each overlap condition (see Appendix B for the full analysis). As evident in Fig. 2B, here again there was some variability across Contrast Types. Specifically, in two instances the group differences were not the same for the two contrasts included within one overlap condition. Thus, in the 'None' overlap condition, native Arabic trilinguals were more accurate than native Hebrew bilinguals in the *cream-scream* contrast, but the groups performed similarly in the *dog-dug* contrast type. In addition, in the 'Hebrew' overlap condition, the native Hebrew Bilinguals were significantly more accurate than the native Arabic Trilinguals in the *big-pig*, but not in the *present-present* contrast type. This suggests that although overall performance appears to be driven by cross-language overlap with the L1, there are still idiosyncratic effects linked to the particular contrast examined. We return to this issue in the General Discussion.

The results of Experiments 1a and 1b indicate that phonological overlap with Hebrew influenced perception in English for native Hebrew bilinguals but not for Arabic-Hebrew-English trilinguals. This outcome could either be due to the trilingual participants' prior L2 phonological knowledge not being activated during L3 word perception, or because the trilingual participants had not acquired the critical L2-Hebrew phonological contrasts that were embedded within the English word-pairs. Experiment 1c was conducted to examine this latter possibility.

Experiment 1c: Arabic-Hebrew-English trilinguals in L2 Hebrew

Experiment 1c was conducted in order to test whether Arabic-Hebrew-English trilinguals could reliably discriminate the L2-Hebrew phonological contrasts when these appear in L2-Hebrew and not in L3-English. If trilinguals are unable to perceive these contrasts in Hebrew, then we would not expect them to affect trilingual performance in L3 English. Experiment 1c therefore assessed trilinguals' ability to distinguish word-pairs in L2-Hebrew including the two phonological contrasts from Experiment 1a that exist in Hebrew but not in Arabic (i.e., p/b and stress location). If these trilinguals have not acquired the specific L2-Hebrew contrasts used in Experiment 1a, their accuracy on these L2-Hebrew contrasts would not differ from chance, and our claim regarding the lack of L2 influence on L3 word perception (in Experiment 1a) will need to be revisited.

Method

Participants

Following the recruitment criteria of Experiments 1a, a total of 27 Arabic-Hebrew-English trilingual university students (aged 19–25) were tested. The data of 8 additional participants were excluded due to technical difficulties (n = 2), due to participants having knowledge of

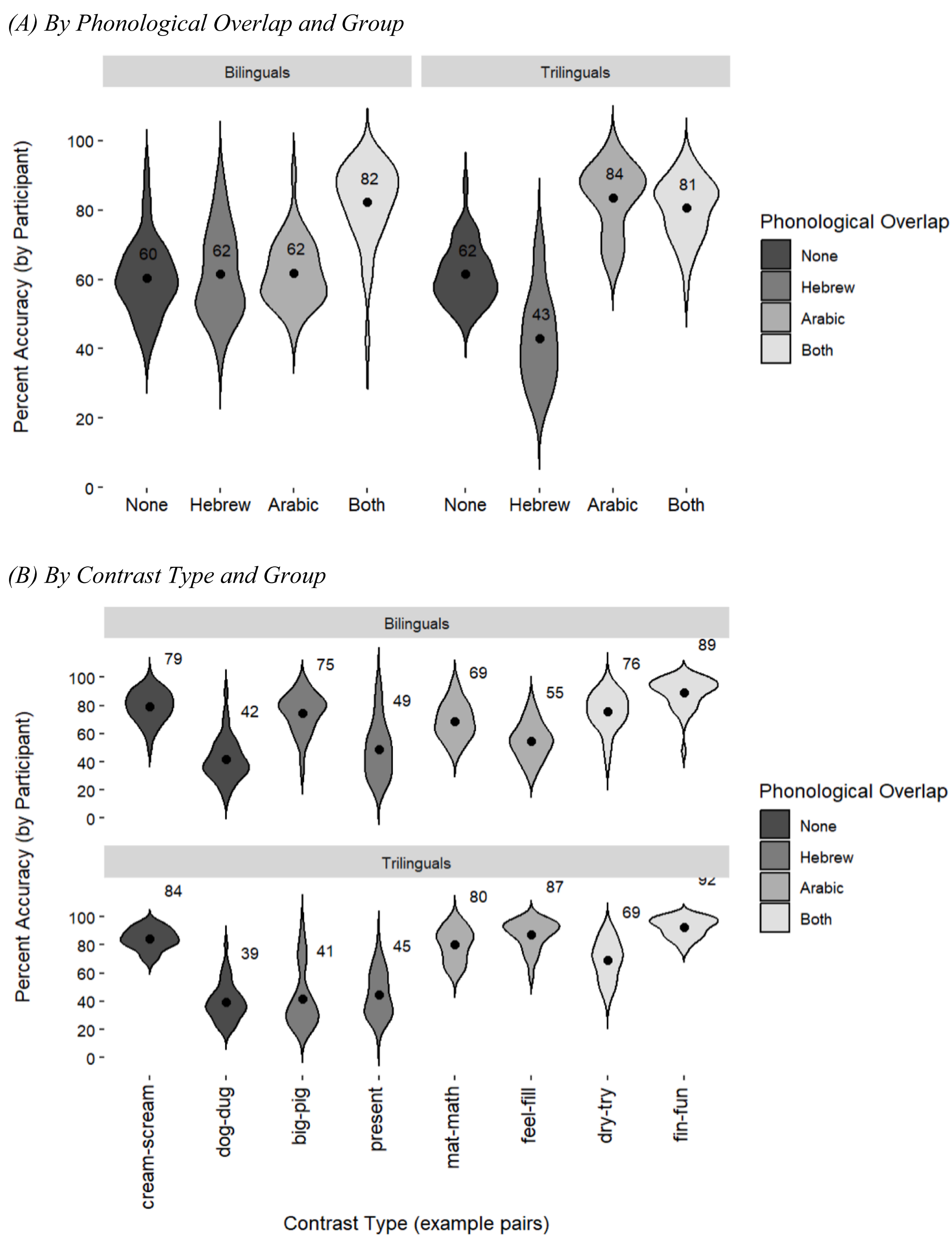

Fig. 2. Mean Perception Accuracy in Experiment 1a&b. A) By Phonological Overlap and Group; B) By By Contrast Type and Group.

Table 7
Pairwise comparisons for the effect of Group by Phonological Overlap condition.

Phonological Overlap	Group Comparison	*Df*	χ^2	*p*
None	Bilinguals = Trilinguals	1	0.91	1.000
Hebrew	Bilinguals > Trilinguals	1	21.20	< 0.001
Arabic	Bilinguals < Trilinguals	1	51.05	< 0.001
Both	Bilinguals = Trilinguals	1	0.01	1.000

languages other than Arabic, Hebrew and English (n = 2), or because participants reported extremely high subjective proficiency in Hebrew (above 9; n = 4). Although the current sample matched the one tested in Experiment 1a on the critical Hebrew proficiency and use measures, there were group differences in Age ($F(1) = 34.20$, $p < 0.001$), years of Education ($F(1) = 17.38$, $p < 0.001$), and Subjective Proficiency in English ($F(1) = 6.18$, $p < 0.05$), and thus these measures were taken into account in the statistical analyses (see Table 8).

Stimuli and procedure

Four phonological contrasts in Hebrew were included, of which two existed in Hebrew but not in Arabic (belonged to the Hebrew condition of Experiments 1a & 1b), and two exist in both Hebrew and Arabic (belonged to the Both condition of Experiments 1a & 1b). As in the previous experiments, for each phonological contrast 12 word pairs were selected, resulting in 24 Hebrew word pairs per condition, and a total of 48 word pairs (96 words). The four contrast types and examples of word pairs are presented in Table 9. The full list of stimuli is available on the OSF (https://osf.io/mdn34/?view_only =

Table 8
Background characteristics of Participants in Experiments 1c vs. 1a & 2; Mean (SD).

General Background Measures	Experiment 1c	Experiments 1a & 2
N	27	41
Gender (number of males/females)	3/24	5/36
Age (years)*	21.90 (0.75)	19.90 (0.75)
Education (years)*	14.10 (1.77)	12.96 (0.36)
Maternal Education (years)	13.20 (3.45)	13.25 (3.24)
Language Specific Background Measures		
Age Of Acquisition – Arabic	0 (0)	0 (0)
Age Of Acquisition – Hebrew	7.96 (0.71)	7.73 (1.12)
Age Of Acquisition – English	7.93 (1.07)	8.22 (0.96)
Mean Subjective Proficiency (0–10) – Arabic	9.70 (0.48)	9.67 (0.53)
Mean Subjective Proficiency (0–10) – Hebrew	7.93 (0.74)	7.55 (1.07)
Mean Subjective Proficiency (0–10) – English*	7.37 (1.63)	6.27 (1.88)
Mean Subjective Current Use (%) – Arabic	43.00 (20.40)	41.84 (15.86)
Mean Subjective Current Use (%) – Hebrew	23.90 (10.20)	29.48 (13.32)
Mean Subjective Current Use (%) – English	19.10 (15.90)	17.19 (13.05)

Note: Of the 27 participants, 12 self-reported higher proficiency in English (L3) compared to Hebrew (L2). * marks a significant group difference at the $p < .05$ level, based on independent samples t-tests.

Table 9
Stimuli example for Experiment 1c.

Condition	Contrast	Hebrew Word-Pair /IPA/	English Translation
Both (Exist in Arabic & Hebrew)	ɪ / ʌ	/ kɪr / – / kʌr/	wall – cold
	d / t	/mɪda/ – /mɪta/	size – bed
Hebrew (not realized in Arabic)	b / p	/banɪm/ – /panɪm/	boys – face
	stress location (penultimate/ ultimate)	/ˈnaʕal/ – /naˈʕal/	shoe – (he) locked

65b37294d96e4020977eeb742fb2890d).

Word pairs in the two conditions (Hebrew and Both) were well matched on Log Frequency (taken from Open Subtitles Corpus (sub-s2vec; https://github.com/jvparidon/subs2vec; Van Paridon & Thompson, 2021; $F(1, 46) = 0.01$, $MSE = 0.19$, $p = 0.935$), on mean Length in Letters ($F(1, 46) = 1.72$, $MSE = 0.59$, $p = 0.196$), and on Phonological Distance between the two words of each pair (which was calculated as in Experiment 1a; $F(1, 46) = 0.14$, $MSE = 0.00$, $p = 0.71$). However, there was a significant difference in mean Length in Phonemes between the two conditions ($F(1, 46) = 4.56$, $MSE = 0.89$, $p = 0.038$), and thus this measure was considered in the statistical analyses (Table 10).

Each of the 96 words was recorded by 3 different female native speakers of Hebrew. Then, word triplets were created based on these 48 word pairs using the same method employed in Experiment 1a.

Table 10
Stimuli characteristics for Experiment 1c.

Measure	Both Condition	Hebrew Condition
Mean Log Frequency	1.34 (0.47)	1.35 (0.39)
Mean Length in Letters	3.04 (0.83)	3.33 (0.70)
Mean Length in Phonemes*	3.83 (1.01)	4.42 (0.88)
Phonological Distance	0.03 (0.02)	0.03 (0.07)

Note: * marks a significant difference between conditions at the $p < .05$ level, based on independent samples t-tests.

Procedure

The procedure was identical to the one used in Experiment 1a. Each participant was presented with 96 triplets in a random order. There were 72 triplets including an odd-one out word – 24 trials for each of the three possible word positions (1st, 2nd or 3rd) and 24 trials in which all words were the same.

Results

Data analysis

The data analysis approach was identical to the one adopted in the previous experiments. The initial data set included 2592 trials that resulted from 27 participants and 47 word-pairs (one pair was removed due to an error in one of the recordings in the b/p contrast). Then, following the exclusion criteria described in Experiment 1a, an additional 161 trials were removed, resulting in 2377 trials.

The critical analysis for current purposes was to examine whether trilinguals were able to discriminate the two Hebrew phonological contrasts which did not overlap with Arabic. To this end a binomial test was conducted to compare accuracy on each contrast to chance level of 25 %, since participants had to choose 1 out of 4 possible responses on each trial. These analyses revealed that trilinguals were significantly better than chance at perceiving these contrasts. Specifically, for the b/p contrast mean accuracy was 0.62 ($SD = 0.49$; $z = 20.2$, $p < 0.001$), and for stress location, mean accuracy was 0.70 ($SD = 0.46$; $z = 25.3$, $p < 0.001$). These analyses suggest that trilingual participants had acquired, at least to some extent, the specific Hebrew contrasts tested here, although they do not exist in their L1 Arabic.

Beyond this critical demonstration of at least partial acquisition of the L2 Hebrew phonological contrasts, the current dataset further allows examination of the effect of L1 on L2 phonological perception. To this end, we compared performance on the Hebrew and Both conditions. Specifically, the maximal model submitted to the *buildmer* function included (1) the fixed effect of the variable of interest (dummy coded): Phonological Condition (Both and Hebrew, with Both as the reference level); (2) the fixed effect of the control variables (continuous and normalized) to control for Length in Phonemes, and participants' Age, Years of Education, and English Proficiency; and (3) the random effects of Participant and Item, with by-Participant and by-Item intercepts, and by-Participant slope for Phonological Condition. See Table 11 for the final model summary.

Results showed a significant effect of Phonological Condition ($F(1) = 116.20$, $p < 0.001$), indicating that trilinguals were more accurate in perceiving Hebrew words including contrasts that exist in both Hebrew and Arabic (Both: $M = 0.96$; $SD = 0.21$), relative to words including contrasts unique to Hebrew (Hebrew: $M = 0.66$; $SD = 0.48$). For the analysis by Contrast Type see Appendix C.

Summary of Experiment 1

The results of Experiment 1 demonstrate that L1 phonological

Table 11
Experiment 1c: Summary of the LME model predicting mean accuracy as a function of Phonological Overlap Condition.

Fixed Effects	*b*	*SE*	*z*	*p*
(intercept)	3.463	0.227	15.274	< 0.001
Phonological Overlap (Hebrew)	−2.673	0.246	−10.873	< 0.001
Control Variables				
Education in Years	0.493	0.126	3.90	< 0.001
Random Effects	***Variance***		***SD***	
Item (intercept)	0.386		0.621	
Participant (intercept)	0.293		0.541	

Note: *p*-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the 'Both' condition in the Phonological Condition variable). For main effects see *F* values in the text.

representations influence phonological perception in both the L2 (Experiment 1b and 1c) and the L3 (Experiment 1a). However, we did not find evidence for influences from L2 phonology on L3 phonological perception (Experiment 1a), despite evidence that the L2 phonological contrasts in question were at least partially acquired by the trilinguals (Experiment 1c). Experiment 2 tested the extent to which similar effects are present in trilinguals' L3 phonological production.

Experiment 2: Cross language influences in L3 production

In order to assess cross-language influences on trilinguals' L3 production, we opted to test whether native speakers of English (the trilinguals' L3) would be able to correctly identify the phonemes produced by the trilinguals. Thus, we recorded trilinguals' productions and then submitted these to be evaluated by native English evaluators.

Method

The participants and stimuli in Experiment 2 were the same as those tested in Experiment 1a.

Procedure

Experiment 2 investigated L1/L2 phonological influences on L3 phonological production and was performed by participants immediately after Experiment 1a. In this experiment, participants performed a word repetition task in their L3-English, during which they listened to 192 English words and were instructed to repeat each word they heard while being recorded directly by the program (E-Prime 3.0 software using Chronos response box and associated Electret microphone, Psychology Software Tools, Pittsburgh, PA).

Participants' L3 productions were evaluated by native English speakers (henceforth "evaluators"), in a 2-alternative forced choice (2-AFC) task (Akahane-Yamada et al., 1996; see also the minimal pair identification task, Bradlow et al., 1999). Specifically, native English speaking evaluators listened to the productions of the Arabic-Hebrew-English trilinguals and were asked to choose, for each word (e.g., pig), one of two words presented on the screen. The options presented were the actual word that the native Arabic trilingual heard and repeated (e.g., pig) and its counterpart word (e.g., big). The task was programmed in PsychoPy3 Experiment Builder (v2021.2.3; Peirce et al., 2019) and was run online via Pavlovia (https://pavlovia.org).

Native English Evaluators. Evaluators were 196 native English speakers (ages 18–35), living in the US, with no knowledge of Arabic or Hebrew, with no language, learning, or hearing disabilities, and with normal or corrected to normal vision. All were undergraduate students recruited via Prolific, an online crowdsourcing platform (https://www.prolific.co). All signed an informed consent approving their participation.

Trilingual Word Recordings. Recordings of only 84 out of the initial 96 word pairs from each trilingual participant were included in the evaluation phase. The recordings of 12 word pairs that differed in stress location (see Table 2 above), were excluded from the current study due to the complexity in visually displaying the phonological distinction between the two words in this type of contrast. Additionally, we excluded 684 out of 6888 possible recordings (9.9 %), in which either the file edges were truncated due to technical problems, there was excessive background noise, or the signal was too weak to allow for processing. These 684 word-recordings were replaced by designated intact filler recordings of the same words, whose evaluations were not included in the final analysis. Thus, the 2-AFC task included 6204 intact word-recordings that were edited using the Audacity software.

First, all audio files were normalized using the *normalize* function (peak amplitude = -1.0). Second, files recorded only in one channel were converted to stereo. Third, those files that had significant background noise were edited using the Noise Reduction and Noise Gate functions in Audacity. These noisy files belonged to various participants, and some participants had more noisy files than others. However, noisy files were distributed randomly across various word pairs and the different contrasts. Finally, all audio files were trimmed at the edges, leaving approximately 200 ms of silence at the beginning and end of each recording. Note that for each of the 41 Arabic-Hebrew-English trilinguals there was a minimum of 14 and a maximum of 24 word recordings per contrast.

Experimental Lists. The final 84 word pairs were divided into two lists (A and B) of 84 English words, each consisting of only one word from each pair (e.g., either 'pig' or 'big'). For each list (A and B), the final word-recordings from the 41 trilinguals were counterbalanced across 41 versions, resulting in a total of 82 versions. Each native English evaluator was exposed to only one word from each pair and to word recordings of multiple trilingual speakers. Finally, each version was presented to 2–3 different evaluators, resulting in between 28 (14 intact word-recordings per trilingual participant X 2 evaluations per recording) to 72 (24 intact word-recordings per trilingual participant X 3 evaluations per recording) evaluations per contrast type per trilingual participant.

2-AFC Task Procedure: The task was administered remotely and lasted approximately 10 min. Before starting the experiment, evaluators performed an audio check, read the instructions, and were introduced to 8 practice trials. The evaluators were told that that they will listen to various English words, and for each word, they will have to decide as quickly and accurately as possible which of two words presented on the screen is the word they heard, by key press. The instructions made no explicit reference to the fact that they were about to hear accented English speech. At the start of each trial, a fixation cross was presented for 500 ms and then a word recording was presented. At the offset of the audio file, two written words were displayed on the screen to the right and left sides of the fixation cross for 5 s or until a response was made. The location of each word within the pair (i.e., right / left) was determined randomly in each trial. Then, a blank screen was presented for 300 ms. Each native English evaluator performed the task on a total of 84 critical word recordings and 6 "catch" recordings (e.g., project), after which one of the two visually presented words was the target word (i.e., project) and the other substantially differed from the target word, in both form and meaning (e.g., smile), so the response on these "catch" trials was straightforward. This was done to allow us to exclude evaluators that were not attentive during the task.

Results

The initial dataset of word production evaluations in Experiment 2 included 14,816 data points (i.e., individual evaluations) that resulted from 6204 intact word-recordings of 41 trilingual participants producing 168 target words (84 word-pairs), which were evaluated each by 2/3 native English speakers (196 evaluators in total). The same exclusion criteria used in Experiments 1a and 1b were employed to ensure that all analyzed data points indeed reflected genuine decision making in the 2-AFC task. Thus, trials with RT shorter than 100 ms were excluded (n = 9), resulting in 14,807 data points for analysis. Then, a maximal model of fixed and random effects was submitted to the *buildmer* function. This model was identical to the one submitted in the analysis of Experiment 1a, except for the maximal random-effect structure, which also included by-Evaluator intercept and by-Evaluator slope for Phonological Overlap, to consider the variance due to the random selection of evaluators in the 2-AFC task.

The final model selected by *buildmer* included the fixed effect of the variable of interest – Phonological Overlap; the fixed effect of the control variables – Phonological Distance and Length in Phonemes and the random intercepts of Participant, Item, and Evaluator. See Table 12 for model summary obtained from the *summary* function in R.

Within the selected model, the main effect of Phonological Overlap ($F(3) = 18.20$, $p < .001$) was significant (see Fig. 3A). Pairwise comparisons of the four Phonological Overlap conditions revealed that once

Table 12
Summary of the selected LME model predicting mean production accuracy as a function of Phonological Overlap.

Fixed Effects	*b*	*SE*	*z*	*p*
(intercept)	0.74	0.16	4.54	< 0.001
Phonological Overlap (Hebrew)	0.27	0.26	1.05	0.292
Phonological Overlap (Arabic)	1.57	0.23	6.93	< 0.001
Phonological Overlap (Both)	1.72	0.22	7.68	< 0.001
Control Variables				
Phonological Distance	0.97	0.12	7.90	<0.001
Mean Length in Phonemes	−0.22	0.10	−2.20	0.028
Random Effects	***Variance***		***SD***	
Item (intercept)	0.36		0.60	
Participant (intercept)	0.07		0.27	
Evaluator (intercept)	0.06		0.25	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the 'None' condition in the Phonological Overlap variable), when other continues factors are held constant at the mean (e.g., control variables), without correcting for multiple comparisons. For main effects see F values in the text.

again, only phonological overlap with L1-Arabic facilitated L3-English phonological production. Specifically, accuracy was higher in the 'Arabic' condition than in the 'None' ($\chi^2(1) = 48.04$, $p < .001$) and 'Hebrew' ($\chi^2(1) = 34.25, p < .001$) conditions. Similarly, accuracy in the 'Both' condition was higher than in the 'None' ($\chi^2(1) = 58.92, p < .001$) and 'Hebrew' ($\chi^2(1) = 41.43, p < .001$) conditions. Lastly, the difference between the 'None' and 'Hebrew' conditions was not significant ($\chi^2(1) = 1.11, p = 1.00$), nor was the difference between the 'Both' and 'Arabic' conditions ($\chi^2(1) = 0.58$, $p = 1.00$), indicating that L3-English phonological production was not further facilitated by phonological overlap with the L2-Hebrew.

An additional analysis of accuracy by Contrast Type (see Appendix D for the full analysis) revealed the same pattern of differences between the 7 contrast types as was observed in Experiment 1a (Fig. 3B). Recall that the stress contrast, namely *present-present* was not included in the production evaluations.

Perception production links in L3 phonology

Given the similar influence of cross-language overlap observed across the phonological perception and production tasks, we further examined the association between performance across modalities. To this end, we computed mean accuracy scores for each participant, in each contrast type, within each task, resulting in 287 data points for analysis (41 participants X 7 contrasts) in each task. Then, using the *lm* function in R, we fitted a linear regression model predicting Mean Production Accuracy (continuous) by Mean Perception Accuracy (continuous). Perception Accuracy positively and significantly predicted Production Accuracy ($b = 0.50$, $SE = .02$, $t = 24.50$, $p < .001$; See Table 13 for the model goodness-of-fit statistics). This indicates that trilinguals' ability to correctly produce L3 phonemes embedded within L3 words is highly linked to their ability to perceive them (see Fig. 4).

General discussion

The current study explored cross-language influences on L3 phonological processing, across perception and production. Performance of Arabic-Hebrew-English trilinguals was compared to that of Hebrew-English bilinguals in a phonological oddity perception task, and the trilinguals' L3 phonological productions were judged by native English speakers. In both L3 perception and production, we observed cross-language influences from Arabic, the trilinguals' L1, but not from Hebrew, their L2, despite evidence that L2 Hebrew contrasts had been acquired. Further, we found close correspondence across participants' performance in L3 perception and production.

Consistent with previous research showing that phonological representations in the L1 influence perception of additional languages (for L2 e.g. Evans & Alshangiti, 2018; for L3, e.g. Llama & Cardoso, 2018), here as well, English phonological contrasts that are also represented in L1 Arabic were processed more accurately by trilinguals. These same contrasts were not facilitated for a comparison group of Hebrew-English bilinguals, who do not speak Arabic. This finding supports the interpretation that it is not the inherent ease of processing or salience of the selected phonological contrasts, but rather the overlap with L1 which is driving performance.

Three pieces of evidence further support the influence of L1 on additional language processing in the current study. First, Hebrew-English bilinguals outperformed Arabic-Hebrew-English trilinguals on phonological contrasts in English that overlap with Hebrew (but not with Arabic, see Experiment 1a and 1b; Table 7). This difference likely reflects the fact that these contrasts were part of bilinguals' L1 but of trilinguals' L2. Second, trilinguals exhibited more accurate performance in their L2 as well when perceiving phonological contrasts that also exist in their L1 (Experiment 1c). Finally, in the production task, once again, trilinguals were more successful in producing L3 phonemes when these overlapped with their L1, but similar facilitation was not evident for phonemes overlapping exclusively with the L2.

This pattern of results extends existing literature (Wang & Nance, 2023). Specifically, whereas the majority of studies testing L3 phonological processing focused on production tasks (Geiss et al, 2022; Kopečková et al., 2023; Zhu & Mok, 2023), here we provide important evidence regarding cross-language influences in the perception of L3 phonology. Further, the task we employed directly tested participants' ability to discriminate between words including L3 phonological contrasts, and did not require trilinguals to make explicit, *meta*-linguistic judgments (e.g., Wrembel et al., 2020), or to map phonology to letters (e.g. Evans & Alshangiti, 2018; Liu & Lin, 2021; Onishi, 2016; Stoehr & Martin, 2022). Importantly, in the current task all materials were presented exclusively in the L3, such that any observed influence from prior linguistic knowledge stems from participants activating internal representations and not from bottom-up activation through task stimuli (in contrast with e.g. Parrish, 2022). Thus, the current results broaden our understanding of processing sounds in L3 in a more naturalistic manner.

Such influences of L1 on L3 phonological processing may be an example of cross-language interactions, whereby processing of any given language is influenced by all linguistic knowledge available to the individual. If this is the case, then an L2 should similarly exert an influence when trilinguals process their L3. However, in the current study we do not find evidence for influences from trilinguals' L2 on their L3 processing, in either perception or production. Specifically, the processing of L3 phonological contrasts that overlap only with trilinguals' L2 (Hebrew), did not differ from that of contrasts that are unique to the L3. Critically, these same contrasts shared across English and Hebrew were processed more accurately by Hebrew-English bilinguals than by trilinguals, once again suggesting that trilinguals' low accuracy does not reflect only the difficulty of the contrasts in English. Further, we also compared processing of L3 phonological contrasts that overlap exclusively with trilinguals' L1 with that of contrasts that overlap with trilinguals' L1 and L2 (the 'both' condition). Across production and perception, the addition of overlap with trilinguals' L2 did not further facilitate processing beyond that of contrasts overlapping exclusively between L3 and L1.

Such an absence of L2 influences on trilingual phonological processing may be expected in cases where L2 phonological contrasts were not sufficiently acquired. Clearly, L2 phonological knowledge cannot impact L3 phonological processing if it has not been learned. To rule out this possibility we tested a new group of trilinguals sampled from the same population, and examined their L2 phonological perception (Experiment 1c). The findings demonstrated that the trilinguals were able to accurately perceive these contrasts in their L2 (i.e., substantially above chance performance). Thus, these particular L2 phonological contrasts were at least partially acquired, and could have theoretically

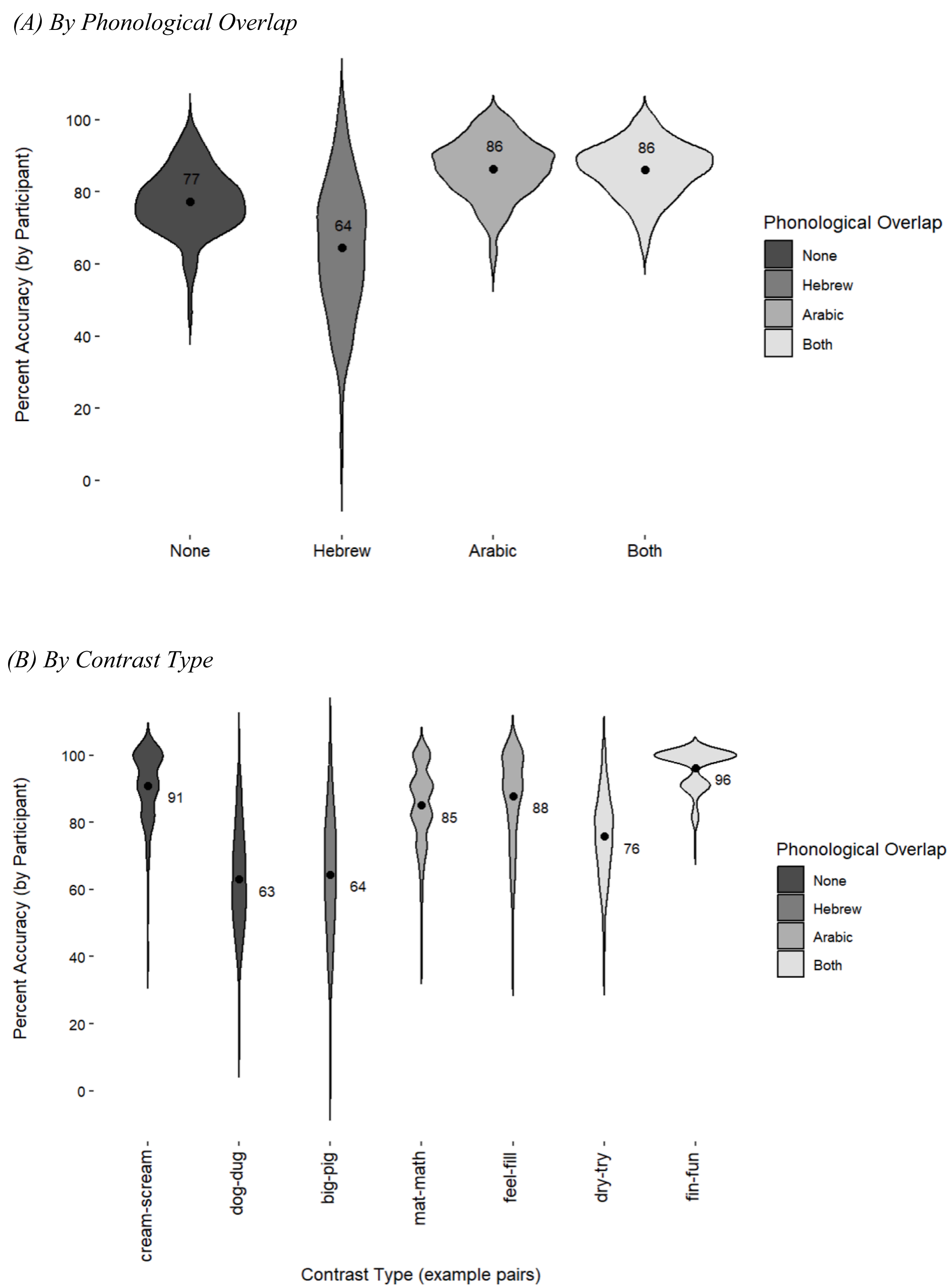

Fig. 3. Mean Production Accuracy. (A) By Phonological Overlap. (B) By Contrast Type.

Table 13
Goodness-of-fit for the linear regression model predicting Mean Production Accuracy by Mean Perception Accuracy.

Statistic	*Value*
Multiple R-squared	0.678
Adjusted R-squared	0.677
Akaike Information Criterion (AIC)	−608.496
Bayesian Information Criterion (BIC)	−597.518

been activated during L3 processing to affect performance.

Nonetheless, we did not observe any influence from L2 on L3 phonological processing. This observed pattern is consistent with two alternative explanations, discussed in the introduction in the context of L1 phonological influences on L2 processing in bilinguals. One possibility is that phonological processing is exclusively determined by knowledge acquired in infancy, with perceptual categories being entrenched early on (Kuhl, 2004). A second option is that the absence of L2 influences in the current study stems from reduced accessibility of L2

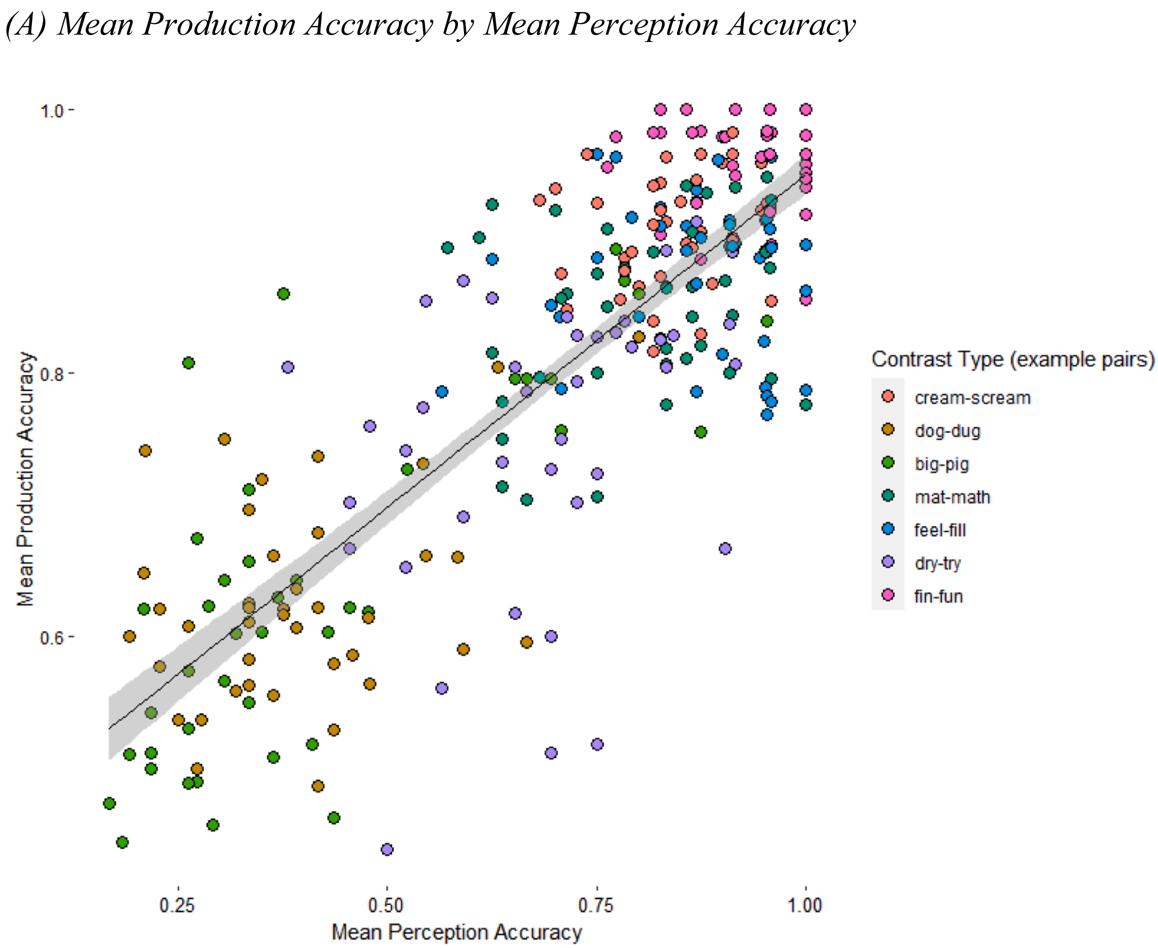

Fig. 4. (A) Mean Production Accuracy by Mean Perception Accuracy. (B) Mean Production Accuracy by Mean Perception Accuracy and Contrast Type in each Phonological Overlap condition.

phonological representations for trilinguals. Here, the overall assumption would be of an interactive system in which all available knowledge is called upon, but the degree of influence is proportional to the strength of activation of these representations in the phonological system. The results of Experiment 1c in fact demonstrate that, although performance was well above chance, when presented with L2 word pairs, the trilinguals processed phonological contrasts that were unique to the L2 less efficiently than contrasts that were shared with the L1. This may suggest that L2 unique phonological representations were less accessible for these trilinguals, thus constraining possible cross-language influences from these L2 representations to L3 processing. A test case for this possibility is to examine trilinguals with stronger L2 phonological representations, for whom we would predict cross-language influences from both L1 and L2 during L3 processing. Although the trilinguals in the current study were relatively proficient in their L3, and were partially immersed in their L2, testing trilinguals with even higher L2 proficiency might reveal influences from L2 on L3 phonological processing, and tease apart these two theoretical explanations (though see e.g., Samuel & Larraza, 2015; Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2009 for possible constraints on acquisition of L2 phonological representations after infancy).

The absence of observed cross-language influences from L2 on L3 processing contrasts with many previous studies, which report influences from both L1 and L2 on L3 phonological processing (e.g., Llama et al., 2010; Lloyd-Smith, 2023; Onishi, 2016; Zhu & Mok, 2023), or even stronger influence from L2 than from L1 (Geiss et al, 2022; Luo et al, 2020). Several factors may explain this discrepancy. First, as described above, the current task did not explicitly present materials in participants' L2 (as in e.g., Parrish, 2022), which might have led to reduced accessibility of L2 representations, and therefore reduced L2 influence. Relatedly, trilinguals in the current study were part of a large population whose language repertoire is determined by their sociolinguistic context, and not by individuals' choices to learn additional languages, as in some previous studies (e.g., Onishi, 2016; Wrembel et al., 2020; Zhu & Mok, 2023). These non-self-selected trilinguals might have somewhat lower *meta*-linguistic inclinations than trilinguals choosing to engage in language learning, thus possibly attenuating the influence of their explicit L2 learning experience, as suggested by the L2 status hypothesis (Bardel & Falk, 2007). Moreover, many previous studies sampled trilinguals in the first stages of acquiring the L3 (e.g. Cabrelli & Pichan, 2021; Kopečková et al., 2023; Sypiańska, 2022) whereas the current participants had studied the L3 for over 10 years. Thus, it may be the case that L2 influences are more prominent in the early stages of acquisition (Cabrelli-Amaro & Rothman, 2010), when L3 processing is more explicit and less automatic. Perhaps as proficiency in the L3 increases, phonological processing of L3 representations becomes less explicit, thus attenuating the influence of the similarity in learning conditions between the L2 and the L3 (Cal & Sypiańska, 2020).

Second, in the current study trilingual participants' L1 and L2 were both typologically distant from the L3, whereas in several previous studies the L2 came from the same linguistic family as the L3 (Geiss et al., 2022; Lipińska, 2015; Zhu & Mok, 2023). In the morpho-syntactic domain, the TPM (Rothman, 2015) suggests that the source of cross-language influences in L3 is determined early in acquisition based on

(B) Mean Production Accuracy by Mean Perception Accuracy and Contrast Type in each Phonological Overlap condition

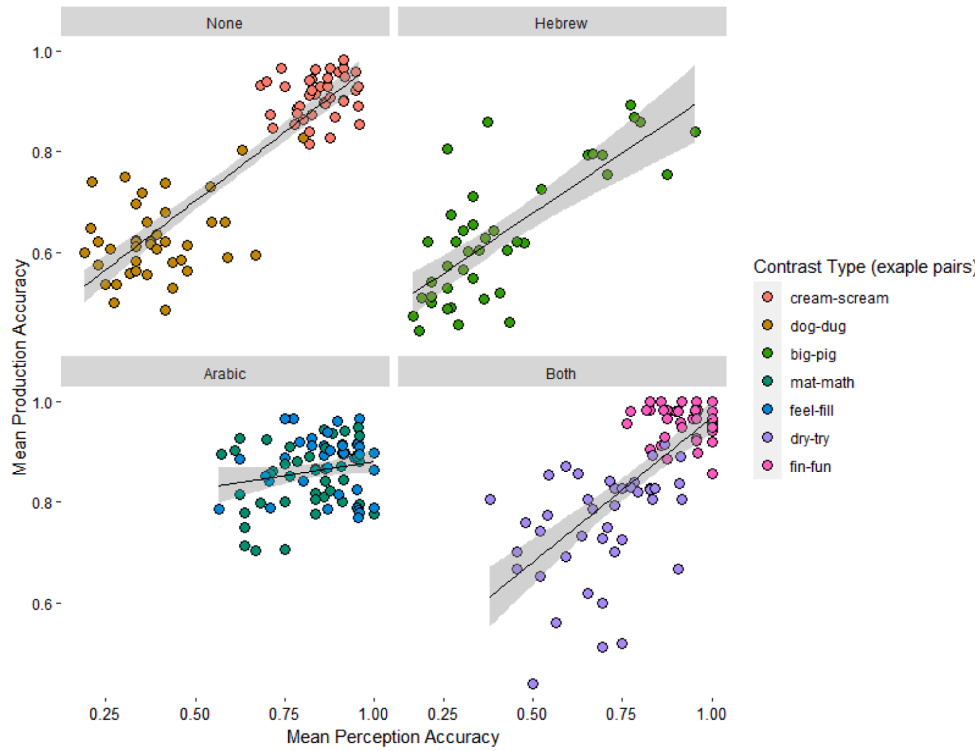

Fig. 4. (*continued*).

overall typological similarity. It is possible that typology, or in this case overlap in the phonemic inventory of each previous language with the L3 (Archibald, 2023; Cabrelli & Pichan, 2021), also contributes to determining cross-language influences in L3 phonology. In the current study, where there is no clear advantage for the L2 in terms of overall similarity, trilinguals appear to resort to L1 phonological representations when processing the L3.

Third, trilingual participants in the current study had clearly acquired their languages in a sequential manner. Thus, they were all native speakers of L1, Arabic, who had acquired L2, Hebrew, through formal instruction and later immersion, and were incontrovertibly L1 dominant in their proficiency. This is in contrast to studies which focused on other trilingual populations where the L1 and L2 are less easily distinguished from each other in terms of dominance (e.g. Geiss et al., 2022; Zhu & Mok, 2023). Thus, it is possible that in the case of a formally later-acquired L2, as targeted here, L2 exerts a weaker influence on L3 processing.

Of note, the current findings do align with some previous studies which also report strong L1 influences on L3 production (e.g., Kopečková et al., 2023; Llama & Cardoso, 2018). For instance, Zhang and Levis (2021) studied production of a phonological contrast in English L3, which does not exist in learners' L1 (Southwestern Mandarin) but does exist in their L2 (Standard Mandarin), and report significant difficulty, demonstrating cross-language interference from the L1 without facilitative cross-language influence from the L2. Interestingly, both L1 and L2 were typologically distant from the L3 in that study, which is also the case in our current investigation.

Nonetheless, before reaching strong conclusions in favor of exclusive L1 cross-language influences in the phonological domain, it is important to keep in mind that there was significant variability across the contrasts examined within the current study (see also e.g., Kopečková et al., 2023; Onishi, 2016). In particular, the approach we adopted in the current study was one in which we selected two phonological contrasts for each overlap condition, in an effort to provide a comprehensive view on the patterns of cross-language influences. When examining the data, some contrast specific variability is clearly evident, in that performance on contrasts belonging to the same overlap condition sometimes differed significantly. For instance, in both perception and production we found that trilinguals were highly accurate in perceiving and producing clusters of 3 consonants, even though these do not overlap with either their L1 or their L2. This diverges from their less successful processing of the o/u vowel contrast, which is also unique to the L3. The higher accuracy for the consonant clusters, in both the trilingual and bilingual groups, can be linked to the fact that this contrast differs in the number of segments, which may be more salient than the vowel contrast. Similarly, within the Hebrew phonological contrasts, performance on the b/p distinction contrast was higher than performance on the stress location contrast for the Hebrew-English bilinguals. As a result, Hebrew-English bilinguals did not show overall facilitation in the Hebrew overlap condition (including the difficult stress location contrast) relative to the no overlap condition (including the easy consonant cluster contrast). A fine-grained analysis of these phonological features is beyond the scope of the current discussion. Nonetheless, this pattern suggests that cross-language phonological overlap is only one of the factors determining successful L2 and L3 phonological processing.

Interestingly, the pattern we observed here of exclusive L1 cross-language influences on L3 phonological processing diverges from the emerging pattern in other language domains examined in the same trilingual population. In the lexical domain, Elias et al. (in press) documented equivalent cross-language influences from L1 and L2 when processing L3 cognate words in a semantic decision task. In the syntactic domain, Abbas et al., (2021) observed influences from both the L1 and the L2 during L3 sentence reading, and Silawi et al., (under review) found stronger L2 influences during L3 sentence production, under high cognitive load. Taken together, this line of research emphasizes how cross-language influences might manifest differently in different language domains, even within a given trilingual population. Therefore, theoretical work on cross-language influences should acknowledge that models put forth for one language domain do not necessarily generalize to other language domains (Degani et al., 2022).

In terms of the correspondence across perception and production, the current work shows significant correlations between trilinguals' perception and production of phonological contrasts. Thus, phonological contrasts that were easily distinguished in the oddity task were more accurately produced by trilinguals. This pattern is consistent with the PAM-L2 (Best & Tyler, 2007) and SLM (Flege, 1995; see also SLM-r, Flege & Bohn, 2021) for L2 processing, extending them to L3 processing as well. As noted by Nagle and Baese-Berk (2022) there is great variability in the tasks used to probe for possible perception-production correspondences, and Hanulíková and colleagues (2012) show that not all perceptual tasks correlate with production to the same extent. In the current study, the observed correlation between perception and production may be partly explained by the fact that the production task we used (word repetition) in fact included a perceptual component. Future studies should further explore the overlap across tasks as a modulating dimension of the correspondence across perception and production.

The current findings, of strong perception-production links in L3, align with some but not all previous studies. Specifically, Wrembel et al., (2022) found stronger perception-production links in L2 than in L3, but they ascribed this difference to participants' proficiency in the two languages. In that study, participants were in the initial stages of L3 acquisition, whereas participants in the current study had been studying the language for several years. Thus, the current results support the notion of stronger perception-production links in more proficient languages. However, the current results diverge from those of Liu and Lin (2021). In that study, alignment of perception and production was only evident for an L3 contrast where cross-language influence from both the L1 and the L2 was observed, but not in the case of a unique L3 phonological contrasts. In our study, in contrast, we find significant correspondence across perception and production for the different overlap conditions that we examined (see Fig. 4). One possible explanation here is that the participants in Liu and Lin (2021) were once again at the initial stages of L3 acquisition, as opposed to the experienced learners tested here. Taken together, these patterns suggest that with growing proficiency and stabilization of L3, perception-production links might become a general characterization of the phonological processing system. We offer this conclusion tentatively in light of the limited empirical data available, and highlight this as an area ripe for further investigation.

To conclude, the current work provides a comprehensive exploration of L3 phonological processing. Trilingual speakers whose L1 and L2 differ typologically from their L3 showed influence only from the L1 on L3 processing. Although the bilingual control group we tested allows us to confidently ascribe the findings to phonological overlap across languages, the presence of contrast-specific effects highlights the need for caution when drawing conclusions from examinations of a single phonological contrast. Further, as described above, several characteristics of the current population may have contributed to the observed results, stressing the importance of sampling across the wide variability of trilingual speakers. Specifically, many individuals in the world become multilingual as a corollary of their socio-linguistic context, but much of academic L3 research focuses on individuals choosing to study additional languages. In addition, much of L3 research has examined speakers of typologically similar Indo-European languages (Wang & Nance, 2023). Thus, the current study advances the field on both these dimensions.

CRediT authorship contribution statement

Tal Norman: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Anat Prior:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization. **Tamar Degani:** Writing – review & editing, Writing – original draft, Supervision, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A: Experiment 1a analysis of accuracy by Contrast Type

Analysis approach

The analysis approach was identical to the one employed in the analysis by Phonological Overlap of Experiment 1a. Importantly, to evaluate the influence of phonological overlap between the L1-Hebrew/L2-Arabic and L3-English on L3 phoneme perception, at the contrast level, here we analyzed the accuracy measure as a function of Contrast Type.

Data analysis

The final dataset from Experiment 1a, including 7368 data points, was analyzed. Similarly, a maximal model of fixed and random effects was submitted to the *buildmer* function. This model was identical to the one submitted in the analysis by Phonological Overlap, except that instead of the fixed effect of Phonological Overlap, it included the fixed effect of Contrast Type (dummy coded; '2/3 consonants', 'ɑ/ʌ', 'b/p', 'stress location', 't/θ', 'i:/ɪ', 'd/t', 'ɪ/ʌ', with '2/3 consonants' as the reference level).

The final model selected by *buildmer* included the fixed effect of the variable of interest – Contrast Type; the fixed effect of the control variable Log Subtlex Frequency; and the random effects of Participant and Item, with by-Participant and by-Item intercepts. See Table A1 for model summary obtained from the *summary* function in R.

Table A1
Experiment 1a: Summary of the LME model predicting mean accuracy as a function of Contrast Type; Effect Size (b), Standard Errors (SE), z-value (z), and p-value (p) of the fixed effects; Variance and Standard Deviation (SD) of the random effects.

Fixed Effects	***b***	***SE***	***z***	***p***
(intercept)	2.001	0.20	9.80	< 0.001
Contrast Type (ɑ/ʌ)	−2.49	0.26	−9.57	< 0.001
Contrast Type (b/p)	−2.39	0.26	−9.19	< 0.001
Contrast Type (stress location)	−2.15	0.26	−8.28	< 0.001
Contrast Type (t/θ)	−0.42	0.26	−1.58	0.114
Contrast Type (i:/ɪ)	−0.03	0.27	0.11	0.907
Contrast Type (d/t)	−1.20	0.26	−4.54	< 0.001
Contrast Type (ɪ/ʌ)	0.63	0.28	2.23	0.026
Control Variables				
Mean Log Subtlex Frequency	0.16	0.07	2.38	0.017
Random Effects	***Variance***		***SD***	
Word-Pair (intercept)	0.30		0.55	
Participant (intercept)	0.20		0.45	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the '2/3 consonant' condition in the Contrast Type variable), without correcting for multiple comparisons. For main effects see *F* values in the text.

Within the selected model, the main effect of Contrast Type ($F(3) = 44.74, p < .001$) was significant. Table A2 presents the pairwise comparisons of the 7 contrast types and Fig. 1B illustrates the mean accuracy by Contrast Type (within each phonological overlap condition).

Table A2
Experiment 1a: Pairwise comparisons of the 7 contrast types; Degree of Freedom (df); Chi- square value (χ^2); and p-value (p) for each comparison.

Contrast Comparison	**Phonological Overlap**	***df***	χ^2	***p***
2/3 consonants – ɑ/ʌ ±	None – None ±	1	91.579	< 0.001
2/3 consonants – b/p	None – Hebrew	1	84.534	< 0.001
2/3 consonants – stress location	None- Hebrew	1	68.579	< 0.001
2/3 consonants – t/θ	None – Arabic	1	2.492	1.000
2/3 consonants – i:/ɪ	None – Arabic	1	0.014	1.000
2/3 consonants – d/t	None – Both	1	20.596	< 0.001
2/3 consonants – ɪ/ʌ	None – Both	1	4.989	0.714
ɑ/ʌ – b/p	None – Hebrew	1	0.166	1.000
ɑ/ʌ – stress location	None – Hebrew	1	1.774	1.000
ɑ/ʌ – t/θ	None – Arabic	1	67.089	< 0.001
ɑ/ʌ – i:/ɪ	None – Arabic	1	95.229	< 0.001
ɑ/ʌ – d/t	None – Both	1	27.079	< 0.001
ɑ/ʌ – ɪ/ʌ	None – Both	1	134.616	< 0.001
b/p – stress location ±	Hebrew – Hebrew ±	1	0.870	1.000
b/p – t/θ	Hebrew – Arabic	1	60.886	< 0.001
b/p – i:/ɪ	Hebrew – Arabic	1	87.975	< 0.001
b/p – d/t	Hebrew – Both	1	23.089	< 0.001
b/p – ɪ/ʌ	Hebrew – Both	1	126.335	< 0.001
stress location – t/θ	Hebrew – Arabic	1	46.879	< 0.001
stress location – i:/ɪ	Hebrew – Arabic	1	69.090	< 0.001
stress location – d/t	Hebrew – Both	1	14.015	0.005
stress location – ɪ/ʌ	Hebrew – Both	1	104.110	< 0.001
t/θ – i:/ɪ ±	Arabic – Arabic ±	1	2.878	1.000
t/θ – d/t	Arabic – Both	1	9.241	0.066
t/θ – ɪ/ʌ	Arabic – Both	1	14.486	0.004
i:/ɪ – d/t	Arabic – Both	1	22.393	< 0.001
i:/ɪ – ɪ/ʌ	Arabic – Both	1	4.553	0.920
d/t – ɪ/ʌ ±	Both – Both ±	1	45.621	< 0.001

Note: ±comparisons of contrasts within the same phonological overlap condition.

Appendix B: Joint analysis of Experiments 1a & 1b – Accuracy by Contrast Type and Group

Analysis approach

The analysis approach was identical to the one employed in the joint analysis of experiments 1a and 1b of accuracy by Phonological Overlap and Group. Importantly, to reveal differences in the sensitivity to Contrast Type between the group of Hebrew-English bilinguals (Experiment 1b) and the group of Arabic-Hebrew-English trilinguals (Experiment 1a), here we analyzed the accuracy measure as a function of Contrast Type and Group.

Data analysis

The combined dataset from both experiments (1a & 1b), including 14,379 data points, was analyzed. A maximal model of fixed and random effects was submitted to the *buildmer* function. This model was identical to the one submitted in the analysis by Phonological Overlap and Group, except that instead of the fixed effect of Phonological Overlap, it included the fixed effect of Contrast Type (dummy coded; '2/3 consonants', 'ɑ/ʌ', 'b/p', 'stress location', 't/θ', 'i:/ɪ', 'd/t', 'ɪ/ʌ', with '2/3 consonants' as the reference level) and the interaction between Group and Contrast Type.

The final model selected by *buildmer* included the fixed effect of the variables and interaction of interest – Contrast Type, Group, and the interaction between them; and the random effects of Participant and Item, with by-Item intercept and by-Item slop for Group. See Table B1 for model summary obtained from the *summary* function in R.

Table B1
Experiments 1a & 1b: Summary of the LME model predicting mean accuracy as a function of Contrast Type and Group; Effect Size (b), Standard Errors (SE), z-value (z), and p-value (p) of the fixed effects; Variance and Standard Deviation (SD) of the random effects.

Fixed Effects	***b***	***SE***	***z***	***p***
(intercept)	1.45	0.15	9.50	< 0.001
Contrast Type (ɑ/ʌ)	−1.80	0.21	−8.64	< 0.001
Contrast Type (b/p)	−0.34	0.21	−1.63	0.104
Contrast Type (stress location)	−1.52	0.21	−7.29	< 0.001
Contrast Type (t/θ)	−0.63	0.21	−2.99	0.003
Contrast Type (I/i:)	−1.25	0.21	−6.03	< 0.001
Contrast Type (d/t)	−0.30	0.21	−1.40	0.161
Contrast Type (ɪ/ʌ)	0.68	0.23	3.03	0.002
Group (Trilinguals)	0.48	0.17	2.84	0.005
Group (Trilinguals): Contrast Type (ɑ/ʌ)	−0.57	0.21	−2.66	0.008
Group (Trilinguals): Contrast Type (b/p)	−1.93	0.22	−8.85	< 0.001
Group (Trilinguals): Contrast Type (stress location)	−0.64	0.22	−2.95	0.003
Group (Trilinguals): Contrast Type (t/θ)	0.20	0.22	0.89	0.374
Group (Trilinguals): Contrast Type (i:/ɪ)	1.31	0.23	5.77	< 0.001
Group (Trilinguals): Contrast Type (d/t)	−0.78	0.22	−3.54	< 0.001
Group (Trilinguals): Contrast Type (ʌ/I)	−0.03	0.25	−0.13	0.896
Random Effects	***Variance***		***SD***	
Item (intercept)	0.18		0.43	
Item (Group – Trilinguals)	0.11		0.33	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the '2/3 consonant' condition in the Contrast Type variable), when other categorical factors are at their reference level (i.e., the 'Bilinguals' in the Group variable), without correcting for multiple comparisons. For main effects see *F* values in the text.

Within the selected model, the main effect of Contrast Type was significant ($F(3) = 34.80$, $p < .001$), however, the main effect of Group was not significant ($F(1) = 2.66$, $p = .103$). Critically, the two-way interaction between Contrast Type and Group was significant ($F(3) = 39.98$, $p < .001$). Therefore, we further performed pairwise comparisons of the 7 contrast types, separately within each Group. Table B2 presents the pairwise comparisons of the 7 contrast types within each Group and Fig. 2B above, illustrates the estimated mean accuracy by Contrast Type and Group.

Table B2
Experiments 1a & 1b: Pairwise comparisons of the 7 contrast types within each group; Degree of Freedom (df), Chi- square value (χ^2), and p-value (p) for each comparison.

Contrast Comparison	**Phonological Overlap**	***df***	**Trilinguals (1a)**		**Bilinguals (1b)**	
			χ^2	*p*	χ^2	*p*
2/3 consonants – ɑ/ʌ ±	None – None ±	1	84.43	< 0.001	74.56	< 0.001
2/3 consonants – b/p	None – Hebrew	1	77.94	< 0.001	2.65	1.000
2/3 consonants – stress location	None – Hebrew	1	69.53	< 0.001	53.19	< 0.001
2/3 consonants – t/θ	None – Arabic	1	2.64	1.000	8.91	0.159
2/3 consonants – i:/ɪ	None – Arabic	1	0.05	1.000	36.33	< 0.001
2/3 consonants – d/t	None – Both	1	17.21	0.002	1.97	1.000
2/3 consonants – ɪ/ʌ	None – Both	1	5.41	1.000	9.19	0.136
ɑ/ʌ – b/p	None – Hebrew	1	0.15	1.000	50.96	< 0.001
ɑ/ʌ – stress location	None – Hebrew	1	0.80	1.000	1.97	1.000
ɑ/ʌ – t/θ	None – Arabic	1	60.07	< 0.001	33.59	< 0.001
ɑ/ʌ – i:/ɪ	None – Arabic	1	90.13	< 0.001	7.43	0.359
ɑ/ʌ – d/t	None – Both	1	27.79	< 0.001	54.04	< 0.001
ɑ/ʌ – ɪ/ʌ	None – Both	1	128.02	< 0.001	129.56	< 0.001
b/p – stress location ±	Hebrew – Hebrew ±	1	0.26	1.000	33.24	< 0.001
b/p – t/θ	Hebrew – Arabic	1	54.45	< 0.001	1.89	1.000
b/p – i:/ɪ	Hebrew – Arabic	1	83.39	< 0.001	19.98	< 0.001
b/p – d/t	Hebrew – Both	1	23.92	< 0.001	0.05	1.000
b/p – ɪ/ʌ	Hebrew – Both	1	120.29	< 0.001	21.57	< 0.001
stress location – t/θ	Hebrew – Arabic	1	47.28	< 0.001	19.45	0.001
stress location – i:/ɪ	Hebrew – Arabic	1	74.61	< 0.001	1.75	1.000
stress location – d/t	Hebrew – Both	1	19.19	0.001	35.78	< 0.001
stress location – ɪ/ʌ	Hebrew – Both	1	110.08	< 0.001	102.06	< 0.001
t/θ – i:/ɪ ±	Arabic – Arabic ±	1	3.50	1.000	9.64	0.107
t/θ – d/t	Arabic – Both	1	6.60	0.573	2.56	1.000
t/θ – ɪ/ʌ	Arabic – Both	1	15.64	0.004	35.53	< 0.001
i:/ɪ – d/t	Arabic – Both	1	19.49	0.001	21.99	< 0.001
i:/ɪ – ɪ/ʌ	Arabic – Both	1	4.48	1.000	79.12	< 0.001
d/t – ɪ/ʌ ±	Both – Both ±	1	41.35	< 0.001	19.58	0.001

Note: ±comparisons of contrasts within the same phonological overlap condition.

Appendix C: Experiment 1c analysis of accuracy by Contrast Type

Analysis approach

For the analysis by *Contrast Type,* the maximal model submitted to the *buildmer* function included (1) the fixed effect of the variable of interest (dummy coded): Contrast Type ('d/t', 'ɪ/ʌ', 'b/p', and 'stress location', with 'd/t' as the reference level); (2) the fixed effect of the control variables (continuous and normalized) to control for items' Length in Phonemes, and participants' Age, Years of Education, and English Proficiency; and (3) the random effects of Participant and Item, with by-Participant and by-Item intercepts, and by-Participant slope for Contrast Type. See Table C1 for the final model summary.

Data analysis

Within the selected model, the main effect of Contrast Type was significant ($F(3) = 28.20$, $p < 0.001$), indicating that trilinguals' accuracy rates significantly differed as a function of Contrast Type. Table C2 presents pairwise comparisons between the four contrast types. As shown in the table, only contrasts across phonological conditions significantly differed from each other. Thus, the two Hebrew contrasts – 'b/p' and 'stress location' – significantly differed from the two Both contrasts – 'd/t' and 'ɪ/ʌ'. Participants were more accurate in distinguishing between Hebrew word-pairs with the contrasts 'd/t' ($M = 0.95$; $SD = 0.21$) and ɪ/ʌ ($M = 0.96$; $SD = 0.20$), than between pairs with the contrasts 'b/p' ($M = 0.62$; $SD = 0.49$) and 'stress location' ($M = 0.70$; $SD = 0.46$).

Table C1
Experiment 3: Summary of the LME model predicting mean accuracy as a function of Contrast Type.

Fixed Effects	***b***	***SE***	***z***	***p***
(intercept)	3.241	0.308	10.520	< 0.001
Contrast Type (i/a)	0.287	0.443	0.648	0.517
Contrast Type (b/p)	−2.506	0.398	−6.297	< 0.001
Contrast Type (stress location)	−2.209	0.386	−5.721	< 0.001
Control Variables				
Education in Years	0.366	0.149	2.457	< 0.05
Mean Length in Phonemes	−0.141	0.124	−1.142	0.253
Mean Subjective Proficiency of English	−0.012	0.141	−0.088	0.930
Random Effects	***Variance***		***SD***	
Item (intercept)	0.381		0.617	
Participant (intercept)	0.229		0.479	
Contrast (i/a)	0.107		0.328	
Contrast (b/p)	0.578		0.760	
Contrast (stress location)	0.466		0.682	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (e.g., the 'd/t' condition in the Contrast Type variable) without correcting for multiple comparisons. For main effects see *F* values in the text.

Table C2
Pairwise comparisons of the four contrasts.

Contrast Comparison	***Df***	χ^2	***p***
d/t – ɪ/ʌ	1	0.42	1.000
d/t – b/p	1	39.65	< 0.001
d/t – stress location	1	32.73	< 0.001
ɪ/ʌ – b/p	1	39.48	< 0.001
ɪ/ʌ – stress location	1	37.10	< 0.001
b/p – stress location	1	0.71	1.000

Appendix D: Experiment 2 analysis of accuracy by Contrast Type

Analysis approach

The analysis approach was identical to the one employed in the analysis by Phonological Overlap of Experiment 2. Importantly, to evaluate the influence of phonological overlap between the L1-Hebrew/L2-Arabic and L3-English on L3 phoneme production, at the contrast level, here we analyzed the accuracy measure as a function of Contrast Type.

Data analysis

The final dataset from Experiment 2, including 14,807 data points, was analyzed. A maximal model of fixed and random effects was submitted to the *buildmer* function. This model was identical to the one submitted in the analysis by Contrast Type of Experiment 1a. The final model selected by *buildmer* included the fixed effect of the variable of interest Contrast Type; and the random effects of Participant, Item, and Evaluator with by-Participant, by Evaluator, and by-Word-Pair intercepts. See Table D1 for model summary obtained from the *summary* function in R.

Table D1
Experiment 2: Summary of the LME model predicting mean accuracy as a function of Contrast Type; Effect Size (b), Standard Errors (SE), z-value (z), and p-value (p) of the fixed effects; Variance and Standard Deviation (SD) of the random effects.

Fixed Effects	***b***	***SE***	***z***	***p***
(intercept)	2.52	0.14	17.42	< 0.001
Contrast Type (ɑ/ʌ)	−1.96	0.18	−11.10	< 0.001
Contrast Type (b/p)	−1.92	0.18	−10.81	< 0.001
Contrast Type (t/θ)	−0.71	0.18	−3.91	< 0.001
Contrast Type (i:/ɪ)	−0.47	0.18	−2.58	0.001
Contrast Type (d/t)	−1.32	0.18	−7.38	< 0.001
Contrast Type (ɪ/ʌ)	0.83	0.21	4.06	< 0.001
Random Effects	***Variance***		***SD***	
Item (intercept)	0.13		0.36	
Participant (intercept)	0.07		0.26	
Evaluator (intercept)	0.06		0.25	

Note: p-values are based on Wald z-scores. Fixed effects reflect simple effects relative to the reference level of the factor (the '2/3 consonant' condition in the Contrast Type variable), without correcting for multiple comparisons. For main effects see *F* values in the text.

Within the selected model, the main effect of Contrast Type ($F(6) = 59.62, p < .001$) was significant. Table D2 presents the pairwise comparisons of the 7 contrast types and Fig. 3B illustrates the estimated mean accuracy by Contrast Type (within each phonological overlap condition).

Table D2
Experiment 2: Pairwise comparisons of the 7 contrast types. Degree of Freedom (df), Chi- square value (χ^2), and p-value (p) for each comparison.

Contrast Comparison	**Phonological Overlap**	***df***	χ^2	***p***
2/3 consonants – ɑ/ʌ ±	None – None ±	1	125.07	< 0.001
2/3 consonants – b/p	None – Hebrew	1	118.48	< 0.001
2/3 consonants – t/θ	None – Arabic	1	15.88	0.001
2/3 consonants – i:/ɪ	None – Arabic	1	6.95	0.176
2/3 consonants – d/t	None – Both	1	55.24	< 0.001
2/3 consonants – ɪ/ʌ	None – Both	1	15.24	0.002
ɑ/ʌ – b/p	None – Hebrew	1	0.09	1.000
ɑ/ʌ – t/θ	None – Arabic	1	56.19	< 0.001
ɑ/ʌ – i:/ɪ	None – Arabic	1	77.60	< 0.001
ɑ/ʌ – d/t	None – Both	1	15.89	0.001
ɑ/ʌ – ɪ/ʌ	None – Both	1	208.76	< 0.001
b/p – t/θ	Hebrew – Arabic	1	51.70	< 0.001
b/p – i:/ɪ	Hebrew – Arabic	1	72.31	< 0.001
b/p – d/t	Hebrew – Both	1	13.58	0.005
b/p – ɪ/ʌ	Hebrew – Both	1	200.73	< 0.001
t/θ – i:/ɪ ±	Arabic – Arabic ±	1	1.88	1.000
t/θ – d/t	Arabic – Both	1	12.67	0.008
t/θ – ɪ/ʌ	Arabic – Both	1	59.79	< 0.001
i:/ɪ – d/t	Arabic – Both	1	24.12	< 0.001
i:/ɪ – ɪ/ʌ	Arabic – Both	1	41.61	< 0.001
d/t – ɪ/ʌ ±	Both – Both ±	1	120.23	< 0.001

Note: ±Comparisons of contrasts within the same phonological overlap condition.

Data availability

All stimuli, data and analysis code are available: (https://osf.io/mdn34/?view_only=65b37294d96e4020977eeb742fb2890d).

References

Abbas, N., Degani, T., & Prior, A. (2021). Equal opportunity interference: Both L1 and L2 influence L3 morpho-syntactic processing. *Frontiers in Psychology, 12*, Article 673535. https://doi.org/10.3389/fpsyg.2021.673535

Akahane-Yamada, R., Tohkura, Y., Bradlow, A. R., & Pisoni, D. B. (1996). Does training in speech perception modify speech production?. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*. https://doi.org/10.1109/ICSLP.1996.607434

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *The Journal of the Acoustical Society of America, 149*(2), 1271–1284. https://doi.org/10.1121/10.0003559

Archibald, J. (2023). Using a contrastive hierarchy to formalize structural similarity as I-proximity in L3 phonology. *Linguistic Approaches to Bilingualism, 13*(5), 614–637. https://doi.org/10.1075/lab.22051.arc

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods, 39*, 445–459. https://doi.org/10.3758/BF03193014

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research, 23*(4), 459–484. https://doi.org/10.1177/0267658307080557

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). John Benjamins.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*(1–7), 2014. http://cran.r-project.org/package=lme4.

Best, C. T. (1995). Learning to perceive the sound pattern of English. *Advances in Infancy Research, 9*, 217.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception – Commonalities and complementarities. In *Language Experience in Second Language Speech Learning: in honor of James Emil Flege*, O.S. Bohn and M.J. Munro (Eds.), pp. 13–34. Doi: 10.1075/lllt.17.07bes.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /1/: Long-term retention of learning in perception and production. *Perception and Psychophysics, 61*(5), 977–985. https://doi.org/10.3758/BF03206911

Cabrelli, J., & Pichan, C. (2021). Initial phonological transfer in L3 Brazilian Portuguese and Italian. *Linguistic Approaches to Bilingualism, 11*(2), 131–167. https://doi.org/10.1075/lab.18048.cab

Cabrelli-Amaro, J., & Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of non-native phonological systems. *IRAL - International Review of Applied Linguistics in Language Teaching, 48*(2–3), 275–296. https://doi.org/10.1515/iral.2010.012

Cal, Z., & Sypiańska, J. (2020). The interaction of L2 and L3 levels of proficiency in third language acquisition. *Poznan Studies in Contemporary Linguistics, 56*(4), 577–603. https://doi.org/10.1515/psicl-2020-0019

Casillas, J. V. (2020). Phonetic category formation is perceptually driven during the early stages of adult L2 development. *Language and Speech, 63*(3), 550–581. https://doi.org/10.1177/0023830919866225

Degani, T., Prior, A., & Wodniecka, Z. (2022). Editorial: Modulators of cross-language influences in learning and processing. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.898793

Elias, M., Van Hell, J., Prior, A., & Degani, T. (in press). Cognate facilitation in different-script trilinguals as a function of task demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*.

Evans, B. G., & Alshangiti, W. (2018). The perception and production of British English vowels and consonants by Arabic learners of English. *Journal of Phonetics, 68*, 15–31. https://doi.org/10.1016/j.wocn.2018.01.002

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange, & J. J. Jenkins (Eds.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research, 92* (pp. 233–277). York PR.

Flege, J. E., & Bohn, O. S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). Cambridge University Press.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052

Garcia, A. (2013). The effects of L2 proficiency on L3 phonological acquisition: A preliminary test of the L2 proficiency hypothesis. In E. Voss (Ed.), *Selected Proceedings of the 2011 Second Language Research Forum* (pp. 173–186). Somerville, MA: Cascadilla Proceedings Projects.

Garcia-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2016). Relationships between quantity of language input and brain responses in bilingual and monolingual infants. *International Journal of Psychophysiology, 110*, 1–17. https://doi.org/10.1016/j.ijpsycho.2016.10.004

Geiss, M., Gumbsheimer, S., Lloyd-Smith, A., Schmid, S., & Kupisch, T. (2022). Voice onset time in multilingual speakers: Italian heritage speakers in Germany with L3 English. *Studies in Second Language Acquisition, 44*(2), 435–459. https://doi.org/10.1017/S0272263121000280

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism, 7*(1), 19–38. https://doi.org/10.1080/14790710902972248

Hanulíková, A., Dediu, D., Fang, Z., Bašnaková, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language Learning, 62*, 79–109. https://doi.org/10.1111/j.1467-9922.2012.00707.x

Hermas, A. (2010). Language acquisition as computational resetting: Verb movement in L3 initial state. *International Journal of Multilingualism, 7*(4), 343–362. https://doi.org/10.1080/14790718.2010.487941

Hermas, A. (2015). The categorization of the relative complementizer phrase in third-language English: A feature re-assembly account. *International Journal of Bilingualism, 19*, 587–607. https://doi.org/10.1177/1367006914527019

Kavé, G. (2005). Phonemic fluency, semantic fluency, and difference scores: Normative data for adult Hebrew speakers. *Journal of Clinical and Experimental Neuropsychology, 27*(6), 690–699. https://doi.org/10.1080/13803390490918499

Kopečková, R., Gut, U., Wrembel, M., & Balas, A. (2023). Phonological cross-linguistic influence at the initial stages of L3 acquisition. *Second Language Research, 39*(4), 1107–1131. https://doi.org/10.1177/02676583221123994

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831–843. https://doi.org/10.1038/nrn1533

Lipińska, D. (2015). Production of L3 vowels: Is it possible to separate them from L1 and L2 sounds? *Research in Language, 13*(1), 77–92. https://doi.org/10.1515/rela-2015-0011

Liu, J., & Lin, J. (2021). A cross-linguistic study of L3 phonological acquisition of stop contrasts. *SAGE Open, 11*(1). https://doi.org/10.1177/2158244020985510

Llama, R., & Cardoso, W. (2018). Revisiting (Non-)native influence in VOT production: Insights from advanced l3 Spanish. *Languages, 3*(3). https://doi.org/10.3390/languages3030030

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism, 7*(1), 39–57. https://doi.org/10.1080/14790710902972255

Lloyd-Smith, A. (2023). Perceived foreign accent in L3 English: The effects of heritage language use. *International Journal of Multilingualism, 20*(3), 753–767. https://doi.org/10.1080/14790718.2021.1957899

Luo, J., Li, V. G., & Mok, P. P. K. (2020). The perception of Cantonese vowel length contrast by Mandarin speakers. *Language and Speech, 63*(3), 635–659. https://doi.org/10.1177/0023830919879471

Martinez, H. D. R. (2015). Analysing interactions of fitted models. *CRAN-R Project.*

McDonald, M., & Kaushanskaya, M. (2020). Factors modulating cross-linguistic co-activation in bilinguals. *Journal of Phonetics, 81*. https://doi.org/10.1016/j.wocn.2020.100981

Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning, 68*(1), 234–270. https://doi.org/10.1111/lang.12275

Nagle, C. L., & Baese-Berk, M. M. (2022). Advancing the state of the art in L2 speech perception-production research: Revising theoretical assumptions and methodological practices. *Studies in Second Language Acquisition, 44*(2), 580–605. https://doi.org/10.1017/S0272263121000371

Onishi, H. (2016). The effects of L2 experience on L3 perception. *International Journal of Multilingualism, 13*(4), 459–475. https://doi.org/10.1080/14790718.2016.1217604

Paradis, M. (2009). *Declarative and procedural determinant of second languages*. Amsterdam: John Benjamins.

Parrish, K. (2022). The categorization of L3 vowels near first exposure by Spanish-English bilinguals. *Languages, 7*(3). https://doi.org/10.3390/languages7030226

Patience, M. (2018). Acquisition of the tap-trill contrast by L1 Mandarin-L2 English-L3 Spanish speakers. *Languages, 3*(4). https://doi.org/10.3390/languages3040042

Pavlenko, A. (2020). L2 influence on L1 in late bilingualism. *Issues in Applied Linguistics, 11*(2), 175–205. https://doi.org/10.5070/L4112005033

Pavlovia Surveys (Open Science Tools, Nottingham, UK). Retrieved from https://pavlovia.org.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from https://support.pstnet.com/.

Psychology Software Tools, Inc. [E-Prime Go]. (2020). Retrieved from https://support.pstnet.com/.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research, 36*(1), 31–64. https://doi.org/10.1177/0267658318809147

R Core Team. (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research, 27*(1), 107–127. https://doi.org/10.1177/0267658310386439

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism: Language and Cognition, 18*(2), 179–190. https://doi.org/10.1017/S136672891300059X

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics, 39*(1), 187–224. https://doi.org/10.1017/S0142716417000418

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language, 81*, 51–71. https://doi.org/10.1016/j.jml.2015.01.003

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*, 240–255. https://doi.org/10.1016/j.jml.2004.11.001

Sebastián-Gallés, N., Vera-Constán, F., Larsson, J. P., Costa, A., & Deco, G. (2009). Lexical plasticity in early bilinguals does not alter phoneme categories: II. Experimental evidence. *Journal of Cognitive Neuroscience, 21*, 2343–2357. https://doi.org/10.1162/jocn.2008.21152

Silawi, R., Prior, A. & Degani, T. (under review). Task demands affect patterns of cross-language influence during L3 speech production. (~pp. 1-40).

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism, 21*(6), 651–665. https://doi.org/10.1177/1367006916655413

Stoehr, A., & Martin, C. D. (2022). The impact of orthographic forms on speech production and perception: An artificial vowel-learning study. *Journal of Phonetics, 94*. https://doi.org/10.1016/j.wocn.2022.101180

Stoehr, A., Jevtović, M., de Bruin, A., & Martin, C. D. (2024). Phonetic and lexical crosslinguistic influence in early Spanish–Basque–English trilinguals. *Language Learning, 74*(2), 332–364. https://doi.org/10.1111/lang.12598

Sypiańska, J. (2022). The L3 Polish lateral in unbalanced bilinguals: The roles of L3 proficiency and background languages. *Languages, 7*(2). https://doi.org/10.3390/languages7020102

Ullman, M. T. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research, 30*(1), 37–69. https://doi.org/10.1023/A:1005204207369

Ullman, M. T. (2020). The declarative/procedural model: A neurobiologically motivated theory of first and second language. In B. VanPatten, & J. Williams (Eds.), *Theories in Second Language Acquisition* (pp. 128–161). Routledge.

Van Paridon, J., & Thompson, B. (2021). subs2vec: Word embeddings from subtitles in 55 languages. *Behavior Research Methods, 53*(2), 629–655. https://doi.org/10.3758/s13428-020-01406-3

Voeten, C. (2019). buildmer: Stepwise elimination and term reordering for mixed–effects regression. R package version 1.3. Retrieved from https://CRAN.Rproject.org/package=buildmer.

Wang, D., & Nance, C. (2023). Third language phonological acquisition: Understanding sound structure in a multilingual world. *Language and Linguistics Compass*. John Wiley and Sons Inc., 10.1111/lnc3.12497.

Westergaard, M. (2021). Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research, 37*(3), 379–407. https://doi.org/10.1177/0267658319884116

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The linguistic proximity model. *International Journal of Bilingualism, 21*(6), 666–682. https://doi.org/10.1177/1367006916648859

Westergaard, M., Mitrofanova, N., Rodina, Y., & Slabakova, R. (2023). Full transfer potential in L3/Ln acquisition: Cross-linguistic influence as a Property-by-Property process. In J. Cabrelli, A. Chaouch-Orozco, J. González Alonso, S. M. Pereira Soares, E. Puig-Mayenco, & J. Rothman (Eds.), *The Cambridge Handbook of Third Language Acquisition* (pp. 219–242). Cambridge Handbooks in Language and Linguistics. Cambridge University Press.

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism, 7*(1), 75–90. https://doi.org/10.1080/14790710902972263

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (2020). Cross-linguistic interactions in third language acquisition: Evidence from multi-feature analysis of speech perception. *Languages, 5*(4), 1–21. https://doi.org/10.3390/languages5040052

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (2022). The relationship between the perception and production of L2 and L3 rhotics in young multilinguals; an exploratory cross-linguistic study. *International Journal of Multilingualism*. https://doi.org/10.1080/14790718.2022.2036158

Zhang, W., & Levis, J. M. (2021). The southwestern mandarin /n/-/l/ merger: Effects on production in standard mandarin and english. *Frontiers in Communication, 6*. https://doi.org/10.3389/fcomm.2021.639390

Zhu, Y., & Mok, P. (2023). Cross-linguistic influences on the production of third language consonant clusters by L1 Cantonese–L2 English–L3 German trilinguals. *International Journal of Multilingualism*. https://doi.org/10.1080/14790718.2023.2208866