Yearbook of the Poznań Linguistic Meeting 6 (2020), pp. 27–58
DOI: 10.2478/yplm-2020-0005

The younger, the better? Speech perception development in adolescent vs. adult L3 learners

Christina Nelson
University of Münster, Germany
c.nelson@uni-muenster.de

Abstract

Whereas the belief "the younger, the better" regarding foreign language learning seems to hold tenaciously, studies comparing learners of different starting ages attest that in instructed (as opposed to naturalistic) learning contexts, a younger age of onset does not automatically yield better results. However, little is known about how multilingual learners from different age groups develop in their non-native languages over time. The present study thus investigates the understudied domain of perceptual development with seven adolescents aged 12–13 and seven adults aged 19–39 (L1 German, L2 English) over the first year of L3 Polish instruction (tested after one, three, five, and ten months). The sound contrast under scrutiny was /v–w/, which only exists in the learners' L2 and L3. As expected, in the ABX task, adults performed better than adolescents in both languages at most testing times and generally showed a slight upward trend in both their L2 and L3 learning trajectories. For the adolescents, development was more non-linear. Further, a boosting 'novelty effect' was found for the younger learner group: After only a few hours of L3 instruction, they perceived the contrast more consistently and faster than in L2 and at any other testing time, performing within the adults' range.

Keywords: Multilingualism; age differences; perceptual development; L2/L3 phonology; instructed learning.

1. Introduction

One of the important prerequisites for the successful development of communicative skills is accurate speech perception (Cristia et al. 2012). Naturally, the question of how speech perception develops in second language (henceforth L2) learners has thus been a core issue in L2 research. To that end, different explanatory models have been brought forward, most of which share the assumption that the learner's first or native language (henceforth L1) and its sound inventory will, especially at the beginning of L2 learning, heavily influence and therefore guide the perception of L2 sounds (e. g. Flege's (2002) *Speech Learning Model*, SLM; Best and Tyler's (2007) *Perceptual Assimilation Model*, PAM). According to these models, different predictions can be made about which non-native sounds might be more challenging to perceive than others, depending on the degree of distance between L2 and L1 sounds.

This hypothesis also implies that developmental trajectories may look different for younger and older learners. Studies investigating L1 perception development provide evidence that native speech perception develops slowly throughout childhood and adolescence, up until the age of 14 and beyond (see e.g. Bent 2015; Hazan & Barrett 2000; Idemaru & Holt 2013; Johnson 2000). Since children's L1 sound systems appear to be more flexible and malleable, it is sensible to assume that they are quicker to form new categories for unknown L2 sounds if necessary, given a consistently high amount and quality of input (Flege 1995). However, this may not always be the case in classroom learning contexts (Cenoz 2002; Muñoz 2006), so that it is unclear whether age effects can be expected to play out in the same way as in naturalistic learning environments, i.e. in favour of younger learners. Indeed, studies conducted within formal learning contexts in various language domains such as phonology, grammar and vocabulary learning have pointed, rather, towards an advantage for older learners (Kopečková, Dimroth & Gut 2019; Pfenninger & Singleton 2017).

Another important question the majority of the current speech perception models do not address is the case of a speaker who has already acquired (or is currently in the process of learning) a non-native language, and then starts learning their – chronologically-speaking – third language (henceforth L3). Instead of being exposed to unfamiliar sounds solely against the backdrop of their L1, they would possibly already have enlarged or adapted, at least to some extent, their repertoire to their L2. This would make their sound repertoire both quantitatively and qualitatively different from that of a monolingual speaker at the onset of L2 acquisition.

In fact, in today's multicultural and globalized world, this scenario applies to a multitude of language learners. Still, the popular models of L2 speech perception, such as SLM and PAM, fail to make precise predictions for L3 learners (naturally, since these models were not geared towards L3 learning). However, there have been attempts to build on the Perceptual Assimilation

Model-L2 (Best & Tyler 2007) and generate predictions for multilingual/L3 learners in line with the general reasonings of the model (Wrembel et al. 2019). Wrembel et al. (2019: 522) suggest that L3 learners assimilate new L3 sounds to both L1 and L2 categories perceptually, and that they furthermore have a 'head start' in perceiving novel contrasts from the very beginning simply due to prior experience of learning different languages, which they refer to as "facilitative effect of multilingualism". They found initial evidence for their hypotheses in a study conducted with 14 L3 learners completing a cross-linguistic similarity task and a perceptual discrimination task, but admit that more empirical research, especially of longitudinal nature, is needed to corroborate their claims and "ultimately arrive at a comprehensive model of L3 speech learning" (Wrembel et al. 2019: 531).

The present longitudinal study thus aims to fill these research gaps by investigating the non-native perceptual development of seven adolescent[1] and seven adult L1 German speakers in both of their non-native languages (L2 English and L3 Polish). For four times over the course of their first year of L3 learning, the learners completed a perception task testing their ability to discriminate between the /v/–/w/ sound contrast, which exists in their L2 and L3, but not in their L1. Eventually, these data can contribute to advancing models of speech perception to include multilingual learners of different ages and learning contexts and develop a more thorough understanding of multilingual speech processing.

2. Age differences in naturalistic and instructed language learning contexts

With regard to language learning, there is the widespread belief that children are generally guaranteed to succeed at this task, picking up a new language fast and seemingly effortlessly (Pfenninger & Singleton 2017: 7). In turn, the common observation of generally increased difficulties of post-puberty L2 learners has led researchers to hypothesize that there is a critical period of language learning, after which a learner's ability to acquire a new language is inhibited due to maturational constraints (Lenneberg 1967). Numerous studies from the 1960s onwards have indeed found a declining linear correlation

[1] Note that throughout this paper the younger group of learners is referred to as either *children* or *adolescents*. For sake of brevity, *children* is used in all tables and figures.

between age of acquisition and ultimate attainment from early childhood until puberty: the younger a person at first exposure to a new language, the higher the chance that they will develop target-like abilities (see, e.g., Johnson & Newport 1989; DeKeyser 2000; Flege et al. 1999 for morphosyntax; Flege et al. 1995; Flege et al. 1999; Flege & MacKay 2004, 2011 for phonology).

However, the majority of age research that may be largely responsible for the common assumptions about child-adult differences has been conducted in naturalistic settings (Larson-Hall 2008: 36), and there are several reasons to assume that these findings cannot be extrapolated to instructed learning contexts. For instance, formal language instruction often requires a certain level of cognitive development. (Younger) children are generally thought to learn in an implicit way (DeKeyser 2003), but the consistent quantity and quality of input prerequisite for implicit learning may not be provided in language classrooms due to constraints of time and resources. While older learners may profit from more explicit methods of instruction and already possess the necessary skills to comprehend and discuss language in a more abstract way, younger learners may not be cognitively advanced enough in this regard (Muñoz 2001: 35).

Indeed, studies comparing learner groups of different ages of onset in classroom language learning settings have yielded mixed results depending on which language skill was investigated. Within the domains of grammar and vocabulary learning, older learners turned out to be better than, or at least just as good as, younger ones (de Bot 2014; Muñoz 2006; Pfenninger & Singleton 2017). What about phonology, though, the one language area where age effects seem to be especially visible (Flege et al. 1999; Scovel 1988)? Some studies have also reported advantages for later starters here (see Fullana 2006 for phonemic discrimination; García Lecumberri & Gallardo 2003 for accentedness and intelligibility). Yet, others have suggested the opposite. Cenoz (2002) and Larson-Hall (2008), for instance, both tested earlier and later learners after several years of L2 instruction. The former employed various measures of oral proficiency in a storytelling task, while the latter administered an oral grammaticality judgment test, a phonemic discrimination test, and an aptitude test. It was only in the phonology-related tests in both studies (pronunciation measure of oral proficiency in Cenoz's paper, phonemic discrimination test in Larson-Hall's) that the earlier-onset learners outperformed the older ones. However, in Cenoz's (2002) study, the mean difference was statistically significant but rather minute in effect, while Larson-Hall (2008: 50–51) explicitly warns of a direct age-to-ability inference in the specific context of the two age groups of her Japanese participants, pointing to other confounding factors such as motivation. Still, the findings of these two studies point to a special standing of phonology in the investigation of age differences, since they demonstrate that other aspects of language were not affected in the same way. However, much like the other studies, they only compare learners within close proximity agewise (always under 14 years old).

True child–adult studies are much scarcer. The present author is only aware of two small-scale studies comparing the phonological abilities of children and adults in formal learning contexts. Kopečková, Dimroth & Gut (2019) tested ten children (aged 9–11) and 19 adults (aged 19–27), all L1 German speakers, in an experimental classroom setting with highly controlled input. For ten days, both groups received two hours of communication-based Polish instruction followed by one to two hours of language testing, of which this particular study only reports results of sibilant and vowel perception and production as measured in the first and the second week of the course. The adults were more successful than the children in perception, while there was no significant group difference in production. As to learning trajectories, their findings illustrate that "phonological learning at this very initial stage cannot be modelled as a straight path of progress towards higher accuracy" (Kopečková et al. 2019: 22) in either of the groups, further confirming conceptualizations of language learning as a complex and dynamic process. Similarly, Kopečková, Gut & Golin (2019) report non-linear learning trajectories in their longitudinal investigation of nine children and seven adults. They tested the learners' production of the /v/–/w/ contrast (non-existent in their L1) in both of their non-native languages for three times across the first year of L3 learning. In terms of production, the adults outperformed the children in their L2 for both sounds at all three testing times. In their L3, however, a language that was completely novel to both groups, the children produced both sounds more accurately at the first two testing times, i.e. five and ten weeks into L3 learning. At the final testing time, after ten months of L3 instruction, the picture flipped, and the adults suddenly performed better than the children. While the confusion rate (producing /v/ as [w] and vice versa) in both languages dropped for the adults over the course of the year, for the children it actually rose. Furthermore, an acoustic analysis revealed that the children's L2 phonological system appeared to be affected by their L3 learning more so than that of the adults', indicating a difference in stability. It remains to be seen to what extent their findings also hold true for perceptual development.

To sum up, a younger starting age does not automatically yield better results in instructed learning contexts, and findings may vary according to which language aspect was tested. Previous research from the past two decades has thus contributed significantly to our understanding of age as a predictor in foreign language learning and how it might differ in naturalistic vs. instructed settings. Yet, a number of relevant questions remain unexplored. First of all, most previous studies in formal learning contexts mentioned above tested learners at a later stage of acquisition and did not collect longitudinal data from the initial learning stages onwards. Also, participants in these studies have usually consisted of speakers within a small age range; child-adult comparisons are rare. As a result, little is known about paths of development of early vs. late learners in instructed learning contexts. Finally, most of the previous studies have only tested one language (usually English as L2), even though at many European schools nowadays, two or three languages are taught concurrently (see e.g. European Commission 1995). This is of great relevance considering that current language learning models theorize that languages share a common space in a multilingual speaker's mind (de Bot 2012). This assumption has found much support in studies on cross-linguistic influence, which have demonstrated that a speaker's languages interact and influence each other in all possible directions (Cenoz et al. 2001). Thus, investigating only one of the non-native languages in a multilingual speaker may not tell the whole story, as demonstrated in the next section, which introduces theoretical models of perceptual development.

3. Development of non-native speech perception

As stated earlier, L2 perception in the initial stages of learning is thought to heavily depend on the learner's L1 sound system. Hence, contingent on the degree of distance between L2 and L1 sounds, different predictions can be made as to which non-native sounds might be more challenging to perceive than others. To that end, the Perceptual Assimilation Model (PAM), which focuses on adult naïve listeners (functional monolinguals), proposes that they are likely "to perceptually assimilate the nonnative phone to the most articulatorily-similar native phoneme" when presented with a non-native phonetic segment for the first time (Best 1995). Any specific non-native phone can be identified as a good or poor equivalent of L1 phonemes, resulting in at least four different assimilation patterns for each non-native sound contrast (Best

1995: 194–98): They can be assimilated into the L1 phonological space as *good*, *acceptable* or *deviant* exemplars of an L1 category, or they can fall in between L1 phoneme categories.

In a later adaption of the model for L2 learners (as opposed to naïve listeners) dubbed PAM-L2, Best and Tyler (2007) predict a continuous refinement of non-native speech perception over time as a function of specific experience with using their L1 and L2. The present study investigates the non-native sound contrast /v/–/w/, and based on PAM and PAM-L2, some predictions can be put forward. The acquisition of the contrast between the voiced labiodental fricative /v/ and the labiovelar approximant /w/, as it exists in English and Polish, has previously been shown to be challenging for L1 speakers of German in both perception (Ankerstein & Morschett 2013; Iverson et al. 2008) and production (Kopečková, Gut & Golin 2019; Pascoe 1987). This can be explained by the absence of /w/ in the German sound inventory as well as by the slightly different phonetic realizations of /v/ in comparison to English and Polish. In German, /v/ is commonly realized as weak labiodental approximant [ʋ], and is therefore phonetically speaking almost in between English and Polish [w] and [v]. Learners may thus initially map both of these two L2/L3 phonemes onto the L1 category of /v/ (*single category assimilation*), evidenced by a high confusion rate of /v/ and /w/. Another possible scenario would be that one sound of the non-native contrast, arguably [w], is perceived as a much poorer member of the native language category /v/ than is the other. In that case, the discrimination between these two non-native sounds would be predicted to range from moderate to very good due to their different standings as deviant and good exemplars of the L1 category /v/.

Either way, over time and with growing exposure to the target language, PAM-L2 predicts a continuous refinement of L2 categories. As mentioned above, Wrembel et al. (2019) suggest building on PAM-L2 and try to extend its general predictions to L3 learners to account for their enlarged phonological repertoire as well as their greater language learning experience. However, at this point, it is unclear what their *developmental* predictions for L3 learning would be in this line of argumentation. Furthermore, their proposal does not provide any hypotheses regarding the learning trajectories of the other languages in the multilingual's system (i.e. L1, L2 and other L3s). Such predictions would particularly be of interest to learners who are still in the process of L2 acquisition at the time of L3 onset.

Generally, there are only a few small-scale studies investigating speech perception in formally-instructed multilinguals, and most of them are either not developmental (Cabrelli 2017; Onishi 2016) or their investigation does not extend beyond the target language (Kopečková, Dimroth & Gut 2019; Wrembel et al. 2019). Still, findings of perception studies comparing multilingual and mono- or bilingual speakers corroborate the notion of a 'multilingual advantage,' in the sense that multilingual learners are more sensitive to novel contrasts and hence learn to distinguish them earlier (see Kopečková 2016 for a critical review of previous research). This hypothesis of facilitative effects thanks to prior language exposure is also confirmed in Onishi's (2016) study with adult L1 Korean, L2 English, L3 Japanese speakers, in which she reports a positive correlation between L2 and L3 accuracy scores on several contrasts. Although there have been no studies to date that evaluate both L2 and L3 perception in children, evidence for a multilingual advantage in L3 perception has been reported for young and adult multilinguals alike (see e.g. Antoniou et al. 2015; Enomoto 1994; Onishi 2016; Tremblay & Sabourin 2012).

Looping back to the age question in L2 learning, for which neither PAM nor PAM-L2 offer any hypotheses, Flege's (1995) Speech Learning Model (SLM) makes some predictions based on L1–L2 sound system interactions. One of its central postulates is that "[t]he mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span and can be applied to L2 learning" (Flege 1995: 239). However, despite the assumption that learning *abilities* remain stable, there are some predictions as to how learning processes differ with age. According to SLM, a speaker's languages co-exist in a common phonetic space, which consists of sound categories established in the process of language acquisition. SLM posits that the formation of new (L2) sound categories becomes increasingly challenging with age, as the L1 phonetic subsystem grows in its assimilative power. This especially applies to L2 sounds that are somewhat similar yet not identical to an L1 sound category. While younger learners might be more flexible in adapting and fine-tuning their still-developing phonetic system to the new language and form new categories where needed, L1 categories in older learners are stronger "attractors" for unfamiliar sounds (Flege & MacKay 2011: 76). Older learners are therefore more likely to integrate similar L2 sounds into a corresponding L1 category. The lack of appropriate L2 categories, in turn, is thought to decrease the chances of accurate perception and production of the respective sound. Naturally, a prerequisite of the formation of appropriate categories is sufficient quantity and quality of L2 input (Baker Smemoe et al. 2002; Flege & MacKay 2004; MacKay et al. 2001), which, as discussed earlier, is likely not the case in formal learning contexts. Thus, a prediction that can be inferred from SLM regarding age differences in classroom language learning would not necessarily be higher accuracy in perceiving and producing non-native sounds for children or adolescents per se, but rather more variability in their performance. Although younger learners' phonological systems are assumed to be more malleable and therefore more likely to create appropriate categories corresponding to novel L2/L3 sounds, they may not receive enough input to inform category formation. The lack of information may then result in non-linear learning trajectories to an even greater extent than it would be expected in adults, whose phonological categories are, in turn, more stable at the time a new language enters the picture (for some empirical support for this hypothesis see, e.g., Kopečková 2012).

4. The present study

To further explore the multilingual perceptual development of adolescents vs. adults, the present study followed 14 speakers with the same language combination of L1 German and L2 English (but belonging to two different age groups) over the course of their first year of learning L3 Polish. The /v/–/w/ sound contrast was chosen as a particularly interesting feature to investigate in these speakers, because it exists in both their foreign languages, but not in their L1. As mentioned above, previous studies have demonstrated that this contrast is challenging for L1 German learners of English and Polish.

4.1. Predictions

Regarding the multilingual perceptual development of the adult and child language learners in this study, the following predictions were made:

(1) **Relationship between the non-native languages.** If the /w/–/v/ contrast can be distinguished reliably in L2 English, it will also be discerned in L3 Polish and vice versa. The individuals' L2 accuracy scores will therefore emerge as a predictor for their L3 performance.

(2) **Development over time.** With increased (L3) learning experience, both learner groups may start noticing an abstract structural proximity for this

contrast between their L2 and L3. As a result, they will perform more accurately and more consistently in both of their foreign languages over time (continual refinement of perception as predicted by PAM-L2).

(3) **Child-adult comparison.** As more experienced and cognitively more advanced learners with quantitatively greater exposure to the contrast through their L2 English, the adults will outperform the younger learners in both accuracy and response times.

4.2. Methods

4.2.1. Participants

Fourteen L3 learners of Polish took part in the longitudinal study. They were divided into two groups according to their age. Seven adolescents (aged 12–13) and seven adults (aged 21–39), all of whom spoke German as L1 and English as L2, completed perception tasks in both of their foreign languages. A summary of the participants' profiles is given in Table 1.

Table 1. Participants' language learning profiles, AOL = onset age of learning.

Group	N	Age (mean)	L1	L2 (mean AOL; self-assessed proficiency)	L3
Children	7	12–13 (12.2)	German	English (6.5; lower intermediate)	Polish
Adults	7	21–39 (26.4)	German	English (9.4; upper intermediate to advanced)	Polish

The groups were matched regarding their L3 Polish input, which amounted to three hours per week. For all participants, Polish was a completely new language, i.e. all heritage speakers were excluded from the final dataset, as were all early bilinguals and speakers of additional languages not tested in this study.[2] As Table 1 shows, the groups differed in their mean starting age of L2

[2] Five of the seven adults also took some French and Spanish lessons in secondary school, and were thus tested in these languages as well, but including all additional languages in the present analysis would exceed the scope of this paper.

learning. This is because educational policies have changed over the years regarding the school year in which English is introduced as a mandatory subject. The two groups also differed in their self-assessed L2 proficiency, which ranged from an upper-intermediate to an advanced level for the adult group, while the younger learners can all be located within a lower intermediate proficiency level. No objective measure of language proficiency was employed.

The adult participants were recruited in community college and university language classes and received a small financial compensation for participating in the study. Their reasons for learning Polish ranged from having a Polish partner to plans of taking part in a Polish–German Erasmus exchange. Initially, a greater number of speakers were recorded, out of which only seven participants matching the profile (no heritage speakers and no prior L3 knowledge) were present at all four testing times and thus remained in the dataset for final analyses.

The younger L3 learners in the present study were recruited at a school located in Germany within close proximity to the Polish border. They all chose Polish as a new subject (over French) and were total beginners. All of them had received instruction in their L2 English for several years. They also had four 45-minute English lessons at school (a total of three hours per week) alongside Polish lessons throughout the time span of the research project.

4.2.2 Design, materials, and procedure

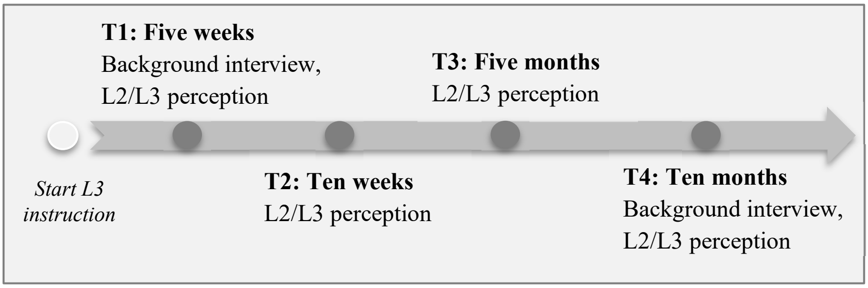

Figure 1. Data collection timeline.

Figure 1 visualizes the longitudinal data collection with four testing times. It started a month after the participants' start of L3 Polish lessons and stretched over the first year of learning. This longitudinal design is extremely informative when tackling research questions regarding learning trajectories, as every learner can act as their own control (Cabrelli 2013: 103). A timed ABX categorization task (see Figure 2) was employed to assess the learners' ability to discriminate between /v/ and /w/. This is a typical test of language-specific contrasts often used in perception studies (Strange & Shafer 2008), in which the participant listens to two minimal pair stimuli (AB) recorded by one speaker. Next, another (third) stimulus (X) is produced by a different speaker, and is either the same as the first (A) or the second one (B). The participant is then asked to decide whether the final stimulus (X) was more like the first (A) or the second one (B).

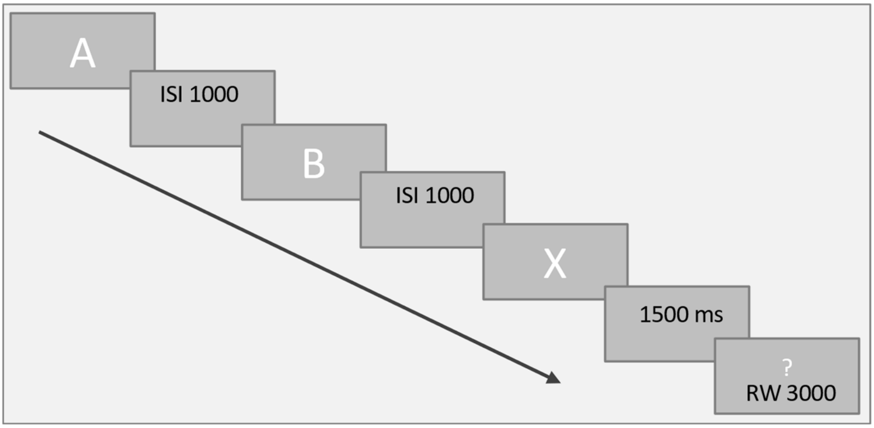

Fig. 2. ABX task procedure (ISI = inter-stimulus interval, RW = response window).

All participants were tested individually in a quiet room (located in the university or school buildings) and on separate occasions for each language, instructed by a native speaker of the respective language. This was done to help create the appropriate language mode. The task itself was administered through E-Prime 2.0 measuring accuracy and response time (henceforth RT), and the participants indicated their judgment by pressing a button on a button box. After a short practice session of four trials, they were given the opportunity to ask questions of clarification and then proceeded with the real experiment.

The five minimal pairs for L2 English and six minimal pairs for L3 Polish (see Table 2) were randomized and appeared in all possible combinations (ABA, ABB, BAB, BAA), resulting in 20 and 24 trials, respectively. If no response was logged within 3000 ms after the last stimulus, the trial was coded as incorrect, and the experiment automatically proceeded with the next trial. In the same session, other contrasts were tested as well but are not reported here. Thus, there was a total of 40 randomized trials in the English session and 44 in the Polish session, each with a short break halfway through, so that it took about five minutes to complete the task.

Table 2. Tokens used in ABX task.

Language	Contrast word-initial	Contrast word-medial
English	whale–veil, wine–vine, wet–vet, wheel–veal	wayward–wavered, leeward–levered
Polish	łódka–wódka, łata–wata	piła–piwa, stały–stawy, brała–brawa, lała–lawa

4.2.4. Data analysis

As a first step, *d'* sensitivity scores were computed from the accuracy scores for each participant. Building on insights from signal detection theory, this is commonly done in perception studies to check for response bias (see e.g. Verde & Macmillan 2006). Since there are only two response options in an ABX task (button 1 and 2), a participant could theoretically achieve a 50% accuracy score by only pressing button 1, for instance, which would not reflect their true ability to discern the contrast. Such a response bias is taken into account when calculating *d'* sensitivity scores, which would turn out lower than the raw accuracy scores in such a case. Correlation measures were run on aggregated L2 and L3 scores to investigate the relationship between the two. Both groups' accuracy and *d'* sensitivity scores are strongly and linearly correlated at all testing times and for both languages ($r = 0.99$).[3] Scatter plots confirmed that

[3] A Pearson correlation test was run on all participants' mean accuracy scores per language and testing time ($N = 112$, normally distributed) and corresponding *d'* sensitivity scores.

accuracy scores below 50% result in negative *d′* values, indicating insensitivity to the contrast. There were no outliers, as for example cases of moderate to high accuracy scores in connection with lower *d′* numbers, so that a possible response bias can be ruled out for all participants. Since accuracy scores are easier to interpret and allow for a straightforward comparison of L2 and L3 performance, they are reported in the results section.

L3 accuracy, as well as L3 response time (RT) data, were then modelled with (generalized) mixed-effects linear regression using the `glmer()` and `lme()` function, respectively, from the lme4 package (Bates et al. 2015) in R (R Development Core Team 2019). Generalized linear mixed modelling has been shown to be appropriate for binomially distributed data such as accuracy scores (Breslow and Clayton 1993; Jaeger 2008) as well as for repeated-measures designs (Dixon 2008), while response time data is continuous and was therefore modelled in a linear mixed-effects regression.

The following transformations of variables and model specifications were applied. First, the variable *Testing time* was transformed into a sliding contrast using the `contr.sdif` function from the MASS package. A sliding contrast is more appropriate than the default treatment contrast when assessing the participants' learning trajectory, as it compares adjacent testing times (T1 to T2, T2 to T3, and T3 to T4), instead of using T1 as a sole reference point for comparison (Schad et al. 2019). None of the variables were centered. In order to handle convergence issues, a bound optimization by quadratic approximation (BOBYQA; Powell 2009) with a set maximum of 200,000 iterations was applied (Miller 2018).

As commonly happens with such data, the response time values were positively skewed (according to visual inspection as well as the Shapiro-Wilk test for normality). They were therefore log-transformed (Levshina 2015: 66) with the `log1p` R base package function, which consequently improved the model fit. Furthermore, outliers on the upper end, i.e. individual items with very slow RT, were removed. The criterion of inclusion was based on the absolute three deviations around the median (MAD), which was calculated with the `outliers_mad` function from the Routliers package (Leys et al. 2013). Out of the 1058 correct L3 items for both groups combined, 35 of the adults' and 30 of the children's responses were thus discarded (6.1% in total, spread fairly evenly across testing times and task-internal conditions). On the lower end (very short RT), no responses were excluded due to the relatively long interval of 1500 ms between the last stimulus and the response window.

The crucial effects investigated in the models were *Testing time* (T1, T2, T3, T4), *Group* (children vs. adults), *Position* (word-initial vs. word-medial), and *Condition* (ABA vs. ABB).[4] The additional factor *Mean L2 accuracy*[5] was then added to assess to what extent the participants' performance on the same contrast in their L2 impacted their ability to perceive it in their L3. Models with and without *Mean L2 accuracy* were compared using the `anova()` function from the stats package (R Development Core Team 2019). As advised by Barr et al. (2013), first, the most complex model consistent with the experimental design was fitted, followed by removing only terms required to allow a non-singular fit[6], which, in this case, affected most random slopes. The resulting mixed effect structure is presented in greater detail in Table 3.

Table 3. Mixed effect structure in final models.

		Accuracy model (glmer)	RT model (lmer)
Dependent variable		L3 accuracy (binary)	L3 response time (in ms) for correct responses only
Independent variables	Random intercepts	Participant Minimal pair	Participant Minimal pair
	Random slopes	–	Testing time | Participant
	Fixed effects	Group Testing time Condition Position Mean L2 accuracy	
	Interactions	Group × Testing time Group × Condition Group × Position Group × Mean L2 accuracy Testing time × Mean L2 accuracy Group × Testing time × Mean L2 accuracy	

[4] The four kinds of trial types every minimal pair was tested in (ABA, ABB, BAB, BAA) were collapsed into two conditions: Either X was the same word as the first stimulus (ABA, BAB) or the second stimulus (ABB, BAA). The former was recoded as ABA and the latter as ABB.
[5] *Mean L2 accuracy* was calculated separately for each participant, testing time, position and condition.
[6] Almost all random slopes that were initially included either explained very little of the variance or had correlation parameters with values close to ±1. They were therefore discarded to avoid singular model fit.

The models were therefore suitable to test the three predictions made beforehand:

(1) Relationship between the non-native languages – indicated by the effect of *Mean L2 accuracy*

(2) Development over time – indicated the effect of *Testing time*

(3) Child-adult comparison – indicated by the effect of *Group*

Beyond these research questions, potential effects of the task-internal factors *Condition* and *Position* could indicate group differences in processing strategies and the use of possible phonetic cues. They are thus reported on as well.

The model outputs shown in the results section, including p-values (based on conditional F-tests with Kenward-Roger approximation for the degrees of freedom) and Nakagawa's marginal/adjusted R^2, were obtained with the `tab_model` function from the sjPlot package (Lüdecke 2020).

5. Results

5.1. Descriptive statistics: Accuracy in each language across testing times

Table 4. Mean (SD) accuracy scores.

		T1	T2	T3	T4
Adults	L2	0.79 (0.41)	0.83 (0.38)	0.89 (0.32)	0.89 (0.32)
	L3	0.81 (0.39)	0.89 (0.32)	0.82 (0.38)	0.92 (0.27)
Children	L2	0.59 (0.49)	0.58 (0.50)	0.78 (0.42)	0.67 (0.47)
	L3	0.78 (0.42)	0.67 (0.47)	0.67 (0.47)	0.74 (0.44)

Table 4 shows the mean accuracy scores and standard deviation of both groups at all testing times and in both languages. As Figure 3 visualizes, the adults' performance in the two languages appeared fairly aligned and considerably high from the very beginning, with a slight upward trend from T1 to T4. The children, on the other hand, started off very differently: While their L3 score was comparable to the adults' at 78% at the first testing time, their L2 performance was much lower then, with an accuracy mean score of only 59%. However, at T3 and T4, their ability to discriminate between the contrast in L2 improved. In L3, the children's performance plummeted marginally at T2 and T3, but almost recuperated back to where it was at T1 at T4 (74%). Hence, L2 and L3 scores largely overlapped at T4. Comparing children and adults in their performance descriptively, adults did better in both languages at all testing times but T1, where both groups performed equally well in their L3.

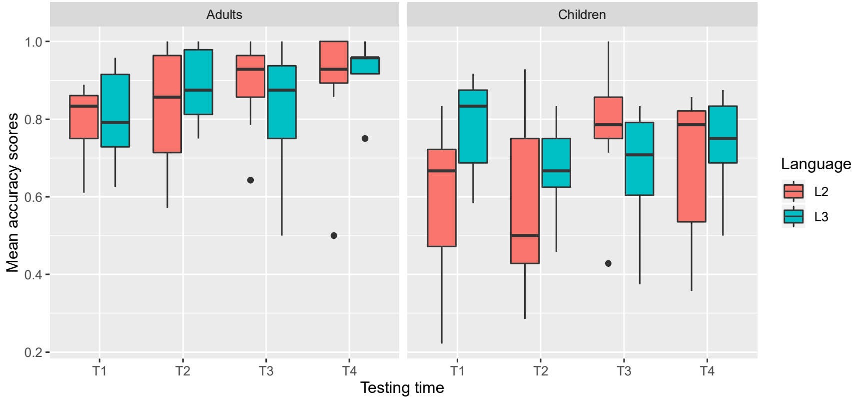

Figure 3. Development of L2/L3 accuracy scores per group.

5.2. Modelling L3 accuracy scores

After plotting and inspecting the dataset descriptively, the L3 accuracy scores were fitted to generalized linear mixed-effects models to assess the effect of group, testing time, and L2 accuracy on the ability to discriminate between /v/ and /w/ in L3. Table 5 displays the accuracy model outputs. Note that odds ratios below 1 indicate a lower estimated accuracy score than the reference value, while those above 1 suggest a higher one (reference values were selected according to alphabetical order, i.e. "Adults" for *Group*, "T1" for *Testing time*, "ABA" for *Condition*, and "initial" for *Position*). It was decided to present both models here, one including *Mean L2 accuracy* as a predictor (*Accuracy model 2*) and one without (*Accuracy model 1*), as they show partially different results. It should be pointed out, however, that including this factor did not improve the model fit significantly ($p = 0.2362$).

The major differences between the two models are the main effect of *Group* as well as significant developments according to testing time in the first model. According to both models, children had lower L3 accuracy scores than adults, but this difference only reaches statistical significance in the first model ($\beta = -1.19$, $SE = 0.48$, $z = -2.54$, $p = 0.011$). While there was no significant development over time in either of the groups when comparing adjacent testing times in model 2, in model 1 the adults improved significantly from T1 to T2 ($\beta = 0.68$, $SE = 0.33$, $z = 2.87$, $p = 0.036$) as well as from T3 to T4 ($\beta = 1.04$, $SE = 0.36$, $z = 2.09$, $p = 0.004$), and the children's accuracy score dropped significantly from T1 to T2 ($\beta = -1.29$, $SE = 0.42$, $z = -3.08$, $p = 0.002$). In terms of the task-internal factors examined here, everyone performed better on medial-position contrasts than on initial-position ones[7] according to both models, although the effect is only significant for the younger group of learners in model 2 ($\beta = 0.64$, $SE = 0.32$, $z = 2.00$, $p = 0.045$). However, both models indicate different effects for the groups in relation to trial condition. Whereas the adults' accuracy scores were significantly higher when the items are presented in the ABB order as opposed to the ABA order in both models ($\beta = 0.56$, $SE = 0.24$, $z = 2.35$, $p = 0.019$), the effect was the opposite for the children, although only significantly so in the first model ($\beta = -0.61$, $SE = 0.30$, $z = -2.04$, $p = 0.042$).

Interestingly, although adding *Mean L2 accuracy* in the second model impacted the significance of some of the detected effects, its effect on the L3 accuracy scores is not entirely straightforward. There is no main effect or interaction with *Group*; however, there is a significant three-way interaction of *Group*, *Mean L2 accuracy* and *Testing time* when comparing T3 to T2 ($\beta = 3.55$, $SE = 1.69$, $z = 2.10$, $p = 0.035$). The effect plot of this three-way interaction (created with the effects package in R, Fox 2003) in Figure 4 helps interpret this result. For testing time T2, the model suggests that L3 accuracy scores and mean L2 accuracy scores are inversely connected for the younger learners, which is neither the case for the adults at this testing time nor for the younger learners themselves at T3. The figure also visualizes why generally no main effect for *Mean L2 accuracy* on L3 accuracy scores was found – the regression lines either have only gentle slopes or are entirely flat at most other testing times.

---

[7] The same trend (medial>initial position) was found for L2 accuracy scores.

Table 5. L3 accuracy model output.

	Accuracy model 1 (without L2 accuracy)			Accuracy model 2 (with L2 accuracy)		
Predictors	Odds Ratios	CI	p	Odds Ratios	CI	p
(Intercept)	5.08	2.13 – 12.08	<0.001	2.91	0.88 – 9.59	0.08
groupChildren	0.31	0.12 – 0.76	0.011	0.48	0.13 – 1.76	0.267
T2-T1	1.98	1.04 – 3.76	0.036	0.54	0.04 – 7.43	0.647
T3-T2	0.55	0.29 – 1.06	0.073	2.03	0.18 – 22.44	0.563
T4-T3	2.84	1.39 – 5.79	0.004	0.97	0.10 – 9.54	0.976
positionMedial	1.53	0.68 – 3.44	0.305	1.42	0.62 – 3.26	0.41
conditionABB	1.75	1.10 – 2.79	0.019	1.63	1.00 – 2.66	0.049
groupChildren:T2-T1	0.28	0.12 – 0.63	0.002	2.64	0.16 – 43.94	0.499
groupChildren:T3-T2	1.81	0.81 – 4.05	0.15	0.13	0.01 – 2.09	0.151
groupChildren:T4-T3	0.5	0.21 – 1.20	0.121	2.86	0.20 – 42.01	0.443
groupChildren: positionMedial	1.71	0.94 – 3.10	0.078	1.89	1.01 – 3.54	0.045
groupChildren: conditionABB	0.54	0.30 – 0.98	0.042	0.56	0.31 – 1.04	0.066
L2_acc				2.16	0.70 – 6.70	0.181
groupChildren:L2_acc				0.53	0.14 – 2.03	0.353
T2-T1:L2_acc				4.99	0.21 – 117.13	0.318
T3-T2:L2_acc				0.2	0.01 – 3.29	0.259
T4-T3:L2_acc				3.86	0.28 – 53.97	0.315
groupChildren:T2-T1:L2_acc				0.04	0.00 – 1.26	0.067
groupChildren:T3-T2:L2_acc				34.81	1.27 – 950.77	0.035
groupChildren:T4-T3:L2_acc				0.11	0.00 – 2.77	0.18
Random Effects						
σ2	3.29			3.29		
τ00	0.49 Participant			0.45 Participant		
	0.15 Minimal_pair			0.15 Minimal_pair		
ICC	0.16			0.16		
N	14 Participant			14 Participant		
	6 Minimal_pair			6 Minimal_pair		
Observations	1343			1343		
Marginal R2 / Conditional R2	0.137 / 0.277			0.156 / 0.288		

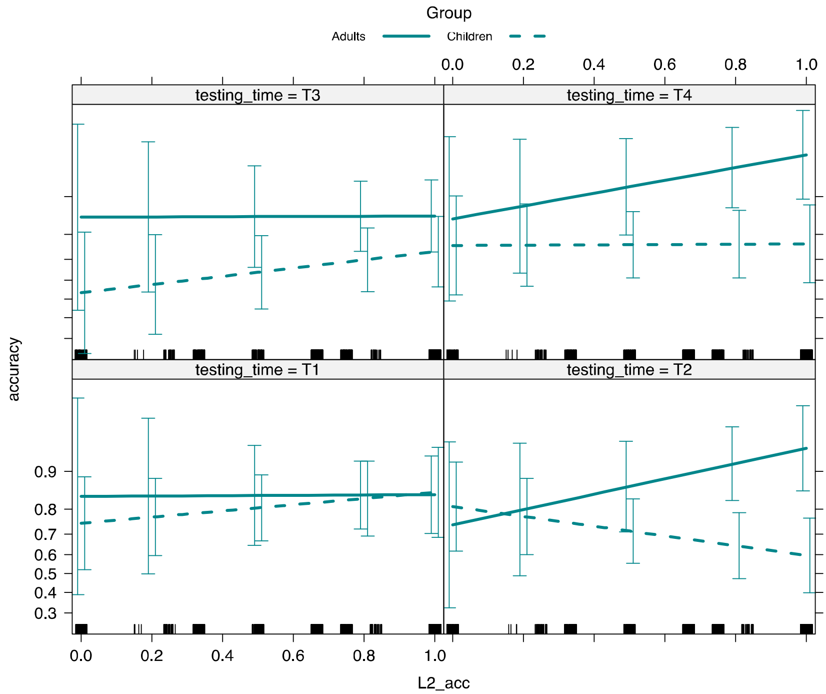

Figure 4. Effect plot for L3 accuracy: Group Testing time Mean L2 accuracy.

5.3. Descriptive statistics: Response time

Table 6. Mean (SD) of response times in milliseconds.

		T1	T2	T3	T4
Adults	L2	385 (352)	332 (221)	395 (273)	393 (252)
	L3	453 (300)	383 (288)	418 (385)	373 (237)
Children	L2	642 (579)	559 (443)	503 (433)	432 (376)
	L3	431 (264)	492 (418)	443 (314)	387 (231)

Table 6 details the mean and standard deviations of the untrimmed response times in milliseconds (for accurate responses only) for both groups and languages across all testing times. The groups showed somewhat distinct learning trajectories (also see Figure 5). For the adults, L2 and L3 response times were similar from the very beginning and seemed to align even more over time. Initially, they were marginally faster in their L2. At T3 and T4, the difference shrunk further. While there was a slight downward trend of L3 response time, the adults' L2 performance remained relatively stable. The children, on the other hand, started off fairly slow in their L2 but became faster over time, decreasing from T1 to T4 by more than 200 ms on average. Mirroring the accuracy score findings, the children performed much better in L3 than L2 at T1, but this language effect flattened afterwards. From T2 onwards, response times were increasingly aligned; the difference between L2 and L3 decreased steadily.

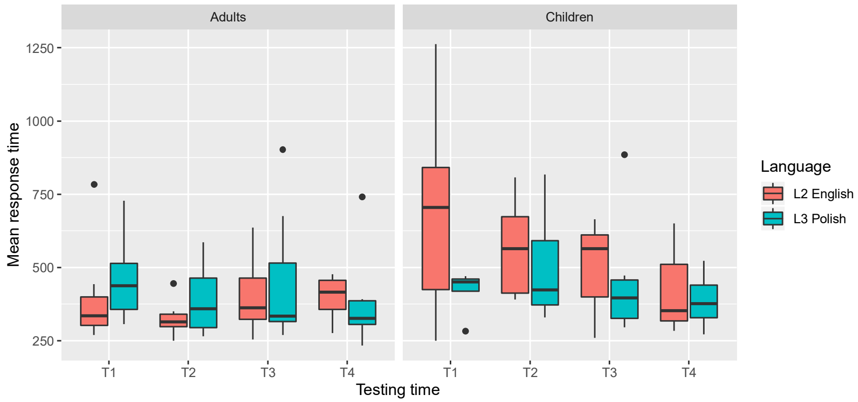

Figure 5. Mean response time per group and language.

5.4. Modelling response time trajectories in L3

Table 7 shows the model output of the log-transformed L3 response times (reference values were again selected according to alphabetical order, i.e. "Adults" for *Group*, "T1" for *Testing time*, "ABA" for *Condition*, and "initial" for *Position*). This time only the model including the factor *Mean L2 accuracy* is

Table 7. Response time model output.

	Response time model		
Predictors	Estimates	CI	p
(Intercept)	5.69	5.44 – 5.93	<0.001
groupChildren	0.05	−0.27 – 0.36	0.775
positionMedial	0.09	0.00 – 0.18	0.058
conditionABB	0.09	0.01 – 0.17	0.034
T2-T1	−0.73	−1.37 – −0.09	0.028
T3-T2	0.72	0.10 – 1.35	0.028
T4-T3	0.49	−0.16 – 1.13	0.142
L2_acc	−0.04	−0.31 – 0.22	0.749
groupChildren:positionMedial	−0.01	−0.14 – 0.12	0.903
groupChildren:conditionABB	−0.09	−0.21 – 0.03	0.122
groupChildren:T2-T1	0.53	−0.24 – 1.29	0.18
groupChildren:T3-T2	−0.67	−1.47 – 0.13	0.107
groupChildren:T4-T3	−0.78	−1.60 – 0.04	0.066
groupChildren:L2_acc	0.04	−0.29 – 0.38	0.798
T2-T1:L2_acc	0.64	−0.04 – 1.32	0.065
T3-T2:L2_acc	−0.87	−1.55 – −0.19	0.015
T4-T3:L2_acc	−0.43	−1.12 – 0.26	0.221
groupChildren:T2-T1:L2_acc	−0.5	−1.35 – 0.35	0.248
groupChildren:T3-T2:L2_acc	0.91	0.00 – 1.81	0.052
groupChildren:T4-T3:L2_acc	0.81	−0.11 – 1.72	0.086
Random Effects			
σ2	0.21		
$\tau 00_{\text{Participant}}$	0.03		
$\tau 00_{\text{Minimal\_pair}}$	0		
$\tau 11_{\text{Participant.1.Testing\_timeT2}}$	0.12		
$\tau 11_{\text{Participant.1.Testing\_timeT3}}$	0.09		
$\tau 11_{\text{Participant.1.Testing\_timeT4}}$	0.05		
$\rho 01_{\text{Participant.1.Testing\_timeT2}}$	−0.85		
$\rho 01_{\text{Participant.1.Testing\_timeT3}}$	−0.67		
$\rho 01_{\text{Participant.1.Testing\_timeT4}}$	−0.69		
ICC	0.14		
$N_{\text{Minimal\_pair}}$	6		
$N_{\text{Participant}}$	14		
Observations	989		
Marginal R2 / Conditional R2	0.054 / 0.184		

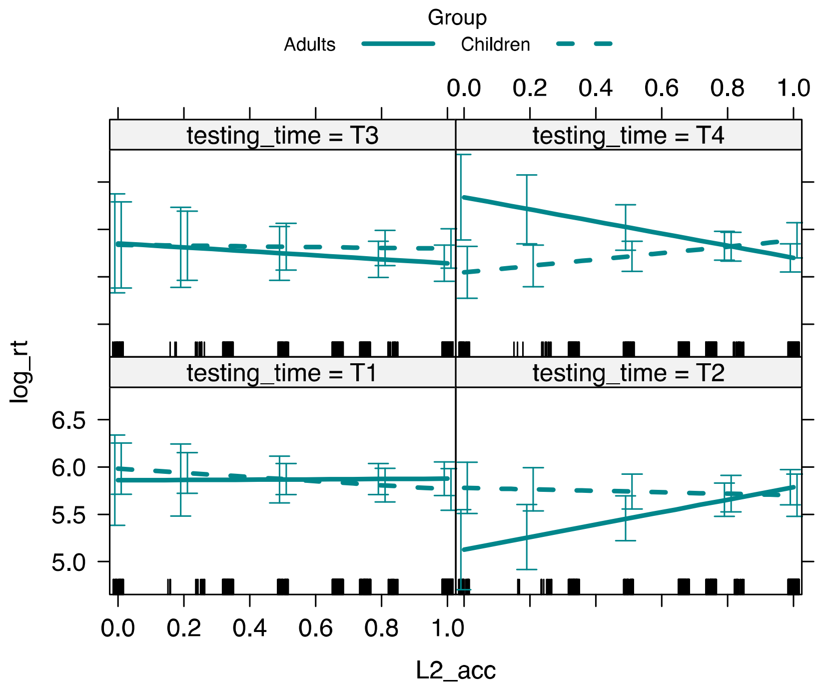

Figure 6. Effect plot for L3 RT: Group × Testing time × Mean L2 accuracy.

presented, because it was significantly better than the one without (p = 0.0128). There was no significant group difference in how fast the participants responded accurately (children were predicted to be marginally faster than the adults). While the adults responded faster from T1 to T2 (β = −0.73, SE = 0.39, t(112) = −2.28, p = 0.028), and then slower from T2 to T3 (β = 0.72, SE = 0.32, t(58) = 2.26, p = 0.028), there was no significant development over time for the children when comparing adjacent testing times. The position of the sound within the word did not have a significant impact on response times according to the model. However, trial condition did have an effect on the adults, who were faster in the ABA condition (β = 0.09, SE = 0.04, t(932) = 2.12, p = 0.034). The adolescents, in turn, responded faster to ABB items, although not significantly so. No main effect of *Mean L2 accuracy* was found, only a significant interaction with testing time comparing T2 and T3 for the adults (β = −0.87, SE = 0.35, t(71) = −2.50, p = 0.015). The effect plot (see

Figure 6) reveals that T2 was the only testing time where higher L2 accuracy scores corresponded to longer L3 response times for the adults. Finally, it should be pointed out that the model yields a low R^2 score, which indicates high inter- and perhaps even intra-learner variability that cannot be explained by the factors measured here.

6. Discussion

6.1. Relationship between the non-native languages

It was predicted that if the /w/–/v/ contrast can be distinguished consistently in L2 English, it will also be discerned in L3 Polish and vice versa, so that a participants' L2 mean score would emerge as predictor for their L3 performance in the accuracy and response time models. This hypothesis could not be upheld according to either of the models, as no main effect of *Mean L2 accuracy* on L3 performance was found. Some of the significant interactions with *Group* and *Testing time* identified in the models suggest that, especially for the children, the relationship between L2 and L3 was dynamic and evolved over time, sometimes indicating a positive relation and at other times perhaps even a competitive one.

However, it should also be noted that both groups generally started off with fairly high L3 accuracy scores even at T1, after only a few weeks of exposure to the novel language. This can tentatively be interpreted as evidence for the multilinguals' reliance on their existing linguistic repertoire when learning a new language, like other studies such as Onishi (2016) have found. The learners' familiarity with the contrast through their L2 English might have given them a head start in learning this contrast in their L3, as Wrembel et al. (2019) have suggested in their adaption of PAM-L2 for multilingual learners. Of course, in order to substantiate this claim, it would be necessary to test a control group to see if they perform differently on this contrast.

Yet, it cannot be dismissed that a task such as the one administered here may also tap other cognitive abilities including, for example, the phonological short-term memory (PSTM), which refers to the retention of verbal information over short periods of time (Dewaele 2013). This route of argumentation entails that PSTM would be a powerful predictor of L3 perceptual accuracy scores at the initial stages, independently of the learners' L2 abilities. Some support for this lies in the observation that adults perceived the contrast more faithfully when the last two words that they heard where the same (ABB condition) than when the first and the last words were identical (ABA condition), whereas there was less of an effect for the children. It would be difficult, though, to disentangle these factors even with the help of partial correlations, since PSTM has frequently been identified as a salient predictor of foreign language pronunciation aptitude (Dewaele 2013). However, it should be pointed out that the reasonably long inter-stimulus intervals used here are thought to target phonological categorization and processing (Strange and Shafer 2008), and that the results are thus still likely to reflect the learners' categorical representation of the /v/–/w/ contrast. It would nonetheless be interesting for future research to measure PSTM and other cognitive functions as a correlate or confound of L2/L3 performance at the initial stages.

6.2. Development over time

Based on PAM-L2's premise of a continual refinement of perception with increasing exposure, it was hypothesized that both learner groups may start noticing an abstract structural proximity for the /v/–/w/ contrast between their L2 and L3 with increased (L3) learning experience, resulting in a more accurate and consistent performance in both of their foreign languages over time. This prediction can only be partially confirmed. For the adults, an upward trend was visible for both languages, and it is reasonable to assume that the participants processed the contrast similarly in both languages from the very beginning of L3 learning. The children's performance, on the other hand, was more variable across testing times in both languages. Especially their T1 performance was noteworthy regarding both accuracy scores and response time: They were more accurate and faster in discriminating between /v/ and /w/ in their L3 than in their L2, despite the fact that they had only received a few hours of L3 input at that point. This positive "novelty effect" of the L3 gives reason to assume that the younger learners did not automatically assimilate the sounds in the very initial stages of language learning. The unfamiliarity with the tested items may have helped them focus on the acoustic cues presented to them, and perhaps different processing or phonological skills were tapped as a result. Another interesting observation about their perceptual development was that at T2, 10 weeks into L3 learning, their L2 score dropped down to chance level. A possible interpretation would be that their L2 subsystem was affected by the exposure to their new language, and they were temporarily "confused". However, with more time and input, the novelty effect as well as the confusion waned, and by the end of the schoolyear, they performed fairly similarly on this contrast in their two non-native languages.

Generally, such variable trajectories of L2 and L3 phonological learning are not surprising. Longitudinal studies such as the one conducted by Kopečková, Gut & Golin (2019) have provided evidence that the development of (multilingual) speech production is non-linear, even more so for children than for adults, and that (non-native) languages are likely to interact over time. Interactions can be complex and dynamic, and the absence of aligned L2 and L3 scores does not necessarily mean that languages do not influence each other.

6.3. Child-adult comparison

The previous subsection already mentioned child-adult differences regarding distinct learning trajectories. But what about a direct comparison of their ability to perceive the non-native /v/–/w/ contrast? It was predicted that the adults would outperform the younger participants, since they were more experienced and cognitively more advanced learners with quantitatively greater exposure to the contrast through their L2. This hypothesis can be largely confirmed with regard to accuracy scores, which were higher at all testing times in both languages. A significant main effect of group was only found in the model without *Mean L2 accuracy*, though. L2 performance presumably correlated with *Age group*, so when it was added to the analysis, it took some of the explained variance from the *Age group* factor, therefore watering down or concealing its effect.

However, no group difference in favour of the older learners could be identified with regard to response times. Lending further support to the existence of some kind of positive novelty effect that only exists for the children, they were even marginally faster than the adults in discriminating between the contrast in the very initial stages of L3 learning. It is thus reasonable to assume some age-related processing differences in perceptual learning of adults and adolescents, also considering the different effects of trial type and contrast position that were found for the groups.

6.4. Limitations

It goes without saying that there are some limitations to this study that ought to be considered carefully when interpreting the findings. First of all, the factors included in the models only explained a small part of the variance and thus pointed to a great overall variability in the performance of both groups. This illustrates that phonological learning is a highly variable and complex process which is probably guided by a multitude of confounding factors that can also be age-related, such as learning context, motivation, type of input, previous language knowledge, etc. What exactly these factors are remains to be established in future studies. Therefore, any strong interpretations concerning the adults' more consistent performance as merely a function of age or maturation should be avoided. After all, the adults in this study were (in addition to most likely being more cognitively advanced) more experienced language learners with higher L2 language proficiency. Furthermore, it is unclear what happens after the first year of L3 learning. Some studies in formal learning contexts suggest that age effects in favour of young learners "may not emerge until a substantial amount of input has been gained" (Larson-Hall 2008: 35).

It should also be stressed that the sample of learners was very small with only seven participants in each group, which evades making any sweeping generalizations from the individual participants in this study to a larger population of adolescent and adult learners. Another inherent limitation concerns the lack of an objective measure of (global) L2 proficiency and exposure in this study. The two age groups were not matched in that regard and, as stated above, this weakens any claims of performance differences existing due to age-related processing differences, for instance, as prior language experience might well be responsible here.

7. Conclusion

This work investigated age differences in multilingual perceptual development. The findings are largely in agreement with previous age research conducted in formal learning contexts, pointing to the advantage of older learners. However, thanks to the longitudinal design of the study that included two non-native languages, several novel observations have been made, such as, e.g., a positive 'novelty effect' for the young L3 learners at the very initial stages and a non-linear path of development for both languages. Adults, on the other hand, were more aligned on this shared non-native contrast from the very beginning and improved more consistently over time. These findings highlight the need for more (long-term) longitudinal studies with multiple participant examinations in between.

Finally, gaining a better understanding of age differences in combination with language proficiency regarding processing of a novel language can inform teaching practices and practitioners' expectations. What is more, it can raise their awareness for the fact that many students in their classroom are actually multilingual and can thus draw on their other languages for support.

8. Acknowledgments

The research reported here was supported by the German Research Foundation (DFG). I would like to thank the adults and adolescents who participated in this study and my colleagues for their kind help with collecting the data (especially Kristy Sigmeth). Also, my thanks go out to two anonymous reviewers who provided valuable input regarding the statistical analysis and the manuscript in general. Finally, I am especially grateful to Romana Kopečková for her guidance throughout the entire process of writing this paper. All remaining errors are mine.

The dataset and R script are available online at: https://osf.io/x7nsr/

References

Ankerstein, C.A. & R. Morschett. 2013. Do you hear what I hear?: A comparison of phoneme perception in native and Saarlandian German nonnative speakers of English. *Saarland Working Papers in Linguistics* 4. 1–8.

Antoniou, M., E. Liang, M. Ettlinger & Patrick C. M. Wong. 2015. The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition* 18(4). 683–95.

Baayen, R.H., D.J. Davidson & D.M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.

Baker Smemoe, W., P. Trofimovich, M. Mack & J. Flege. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. *Proceedings of the Annual Boston University Conference on Language Development* 26. 36–47.

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

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

Bent, T. 2015. Development of perceptual flexibility. In *Proceedings of the XVIII International Congress of Phonetics Sciences*. Glasgow, Scotland.

Best, C.T. 1995. A direct realist view of cross-language speech perception. In W. Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171–204. Baltimore, MD: York Press.

Best, C.T. & M.D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M.J. Munro (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins.

Bot, K. de. 2012. Rethinking multilingual processing: From a static to a dynamic approach. In J. Cabrelli Amaro, S. Flynn & J. Rothman (eds.), *Third language acquisition in adulthood*, 79–94. Amsterdam: John Benjamins.

Bot, K. de. 2014. The effectiveness of early foreign language learning in the Netherlands. *Studies in Second Language Learning and Teaching* 3. 409–418.

Breslow, N.E. & D.G. Clayton. 1993. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88(421). 9–25.

Cabrelli, J. 2017. Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism* 21(6). 698–717.

Cabrelli, J. 2013. Methodological issues in L3 phonology. *Studies in Hispanic and Lusophone Linguistics* 6(1).

Cenoz, J. 2002. Age differences in foreign language learning. *International Journal of Applied Linguistics* 135–136. 125–142.

Cenoz, J., B. Hufeisen & U. Jessner (eds.). 2001. *Cross-linguistic influence in third language aquisition: Psycholinguistic perspectives*. Clevedon, UK & Buffalo: Multilingual Matters.

Cristia, A., A. Seidl, C. Vaughn, R. Schmale, A. Bradlow & C. Floccia. 2012. Linguistic processing of accented speech across the lifespan. *Frontiers in Psychology* 3.

DeKeyser, R. 2000. The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition* 22. 499–534.

DeKeyser, R. 2003. Implicit and explicit learning. In C.J. Doughty and M.H. Long (eds.), *The handbook of second language acquisition*, 312–348. Oxford: Blackwell.

Dewaele, J. 2013. Learner-internal psychological factors. In J. Herschensohn & M. Young-Scholten (eds.), *The Cambridge handbook of second language acquisition*, 159–179. Cambridge: Cambridge University Press.

Dixon, P. 2008. Models of accuracy in repeated-measures designs. *Journal of Memory and Language* 59(4). 447–456.

Enomoto, K. 1994. L2 perceptual acquisition: The effect of multilingual linguistic experience on the perception of a ‘less novel’ contrast. *Edinburgh Working Papers in Applied Linguistics* 5. 15–29.

European Commission. 1995. *White paper on education and training. Teaching and learning: Towards the learning society*.

Flege, J.E. 1995. Second language speech learning: Theory, findings and problems. In W. Strange (ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, 233–277. Timonium: York Press.

Flege, J.E. & I.R.A. MacKay. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26(01).

Flege, J.E. & I.R.A. MacKay. 2011. What accounts for 'age' effects on overall degree of foreign accent? In M. Wrembel, M. Kul & K. Dziubalska-Kołaczyk (eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010*. (Vol. 2.) Frankfurt am Main: Peter Lang.

Flege, J.E., M.J. Munro & I.R.A. MacKay. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication* 16. 1–16.

Flege, J.E., G.H. Yeni-Komshian & S. Liu. 1999. Age constraints on second language learning. *Journal of Memory and Language* 41. 78–104.

Fox, J. 2003. Effect displays in R for generalised linear models. *Journal of Statistical Software* 8(15). 1–27.

Fullana, N. 2006. The development of English (FL) perception and production skills: Starting age and exposure effects. In C. Muñoz (ed.), *Age and the rate of foreign language learning*, *S. language acquisition*, 41–64. Bristol: Multilingual Matters.

García Lecumberri, M. & F. Gallardo. 2003. English FL sounds in school learners of different ages. In M. del Pilar Garcia Mayo & M. Garcia Lecumberri (eds.), *Age and the acquisition of English as a foreign language*, 115–135. Clevedon: Multilingual Matters.

Hazan, V. and S. Barrett. 2000. The development of phonemic categorization in children aged 6–12. *Journal of Phonetics* 28(4). 377–396.

Idemaru, K. and L.L. Holt. 2013. The developmental trajectory of children's perception and production of English /r/–/l/. *The Journal of the Acoustical Society of America* 133(6). 4232–4246.

Iverson, P., D. Ekanayake, S. Hamann, A. Sennema & B.G. Evans. 2008. Category and perceptual interference in second-language phoneme learning: An examination of English /w/–/v/ learning by Sinhala, German and Dutch speakers. *Journal of Experimental Psychology: Human Perception and Performance* 34(5). 1305–1316.

Jaeger, T.F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59(4). 434–446.

Johnson, C.E. 2000. Childrens' phoneme identification in reverberation and noise. *Journal of Speech, Language, and Hearing Research* 43(1). 144–157.

Johnson, J.S. & E.L. Newport. 1989. Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology* 21(1). 60–99.

Kopečková, R. 2016. The bilingual advantage in L3 learning: A developmental study of rhotic sounds. *International Journal of Multilingualism* 13(4). 410–425.

Kopečková, R. 2012. Differences in L2 segmental perception: The effects of age and L2 experience. In C. Muñoz (ed.), *Intensive exposure experiences in second language learning*, 234–255. Clevedon: Multilingual Matters.
Kopečková, R., C. Dimroth & U. Gut. 2019. Children's and adults' initial phonological acquisition of a foreign language. *Journal of Second Language Pronunciation*. 374–401.
Kopečková, R., U. Gut & C. Golin. 2019. Acquisition of the /v/–/w/ contrast by L1 German children and adults. In S. Calhoun, P. Escudero, M. Tabain & P. Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*, 3745–3749. Canberra: Australasian Speech Science and Technology Association.
Larson-Hall, J. 2008. Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research* 24(1). 35–63.
Lenneberg, E.H. 1967. *Biological foundations of language*. New York: Wiley.
Levshina, N. 2015. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins.
Leys, C., C. Ley, O. Klein, P. Bernard & L. Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49(4). 764–766.
Lüdecke, D. 2020. *SjPlot - Data visualization for statistics in social science.* Zenodo.
MacKay, I.R.A., J.E. Flege, T. Piske & C. Schirru. 2001. Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America* 110(1). 516–528.
Miller, S.V. 2018. Mixed effects modeling tips: Use a fast optimizer, but perform optimizer checks. <http://svmiller.com/blog/2018/06/mixed-effects-models-optimizer-checks/>. Retrieved 3 Sep 2019.
Muñoz, C. 2001. Factores escolares e individuales en el aprendizaje formal de un idioma extranjero. *ELUA. Estudios de Lingüística Universidad de Alicante* (Anexo 1). 249–270.
Muñoz, C. (ed.). 2006. *Age and the rate of foreign language learning*. Bristol: Multilingual Matters.
Onishi, H. 2016. The effects of L2 experience on L3 perception. *International Journal of Multilingualism* 13(4). 459–475.
Pascoe, G. 1987. *Die Aussprache des Englischen an bayerischen Schulen: Untersuchungen an Gymnasien, Realschulen und Hauptschulen*. München: Profil.
Pfenninger, S.E. & D.M. Singleton. 2017. *Beyond age effects in instructional L2 learning: Revisiting the age factor*. Bristol & Blue Ridge Summit, PA: Multilingual Matters.
Powell, M.J.D. 2009. *The BOBYQA algorithm for bound constrained optimization without derivatives*. (DAMTP 2009/NA06.) Cambridge: University of Cambridge.
R Development Core Team. 2019. *R: A language and environment for statistical computing. R foundation for statistical computing*. Vienna. R Foundation for Statistical Computing.
Schad, D.J., S. Vasishth, S. Hohenstein & R. Kliegl. 2019. How to capitalize on a priori contrasts in linear (mixed) models: A tutorial. *ArXiv:1807.10451 [Stat]*.

Scovel, T. 1988. *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Cambridge & New York: Newbury House.

Strange, W. and V.L. Shafer. 2008. Speech perception in second language learners: The re-education of selective perception. In J.G. Hansen Edwards & M.L. Zampini (eds.), *Phonology and second language acquisition*, 153–191. Amsterdam: John Benjamins.

Tremblay, M.C. & L. Sabourin. 2012. Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals, and multilinguals. *Journal of the Acoustical Society of America* 132. 3465–3474.

Wrembel, M., M. Marecka & R. Kopečková. 2019. Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism* 16(4). 513–533.

Corresponding author:
Christina Nelson
Westfälische Wilhelms-Universität Münster
Englisches Seminar
Johannisstr. 12-20
D-48143 Münster
Germany
c.nelson@uni-muenster.de