Article

Cross-Linguistic Interactions in Third Language Acquisition: Evidence from Multi-Feature Analysis of Speech Perception

Magdalena Wrembel [1,*] **, Ulrike Gut** [2] **, Romana Kopečková** [2] **and Anna Balas** [1]

[1] Faculty of English, Adam Mickiewicz University, Wieniawskiego 1, 61-712 Poznań, Poland; abalas@wa.amu.edu.pl

[2] English Department, University of Münster, Schlossplatz 2, 48149 Münster, Germany; gut@uni-muenster.de (U.G.); kopeckor@tcd.ie (R.K.)

[*] Correspondence: magdala@wa.amu.edu.pl

check for updates

Abstract: Research on third language (L3) phonological acquisition has shown that Cross-Linguistic Influence (CLI) plays a role not only in forming the newly acquired language but also in reshaping the previously established ones. Only a few studies to date have examined cross-linguistic effects in the speech perception of multilingual learners. The aim of this study is to explore the development of speech perception in young multilinguals' non-native languages (L2 and L3) and to trace the patterns of CLI between their phonological subsystems over time. The participants were 13 L1 Polish speakers (aged 12–13), learning English as L2 and German as L3. They performed a forced-choice goodness task in L2 and L3 to test their perception of rhotics and final obstruent (de)voicing. Response accuracy and reaction times were recorded for analyses at two testing times. The results indicate that CLI in perceptual development is feature-dependent with relative stability evidenced for L2 rhotics, reverse trends for L3 rhotics, and no significant development for L2/L3 (de)voicing. We also found that the source of CLI differed across the speakers' languages: the perception accuracy of rhotics differed significantly with respect to stimulus properties, that is, whether they were L1-, L2-, or L3-accented.

1. Introduction-Bilingual vs. Multilingual Perspective

In this contribution, we explore cross-linguistic interactions between phonological subsystems in third language acquisition, based on evidence from multifeature analysis of speech perception. Research on third language (L3) phonological acquisition has shown that Cross-Linguistic Influence (CLI) plays a role not only in forming the newly acquired language but also in reshaping the previously established ones (cf. Wrembel and Cabrelli 2018). There is scarcity of evidence from perceptual studies, however, which seems unfortunate considering that speech perception has been seen as driving the process of non-native phonological acquisition. The most influential second language (L2) phonology models use (cross-language phonetic) perception to explain the outcomes of L2 speech learning e.g., Speech Learning Model, (Flege 1995; Flege and Bohn 2020) and Perceptual Assimilation Model (Best 1995; Tyler 2019). It would, therefore, seem beneficial and necessary that L3 phonology research complements its findings by examining cross-linguistic interactions in multilingual perceivers in order to be ultimately in a more favorable position to both explicate their production and gain a more complete picture of multilingual phonological acquisition.

Third language acquisition (TLA) has recently gained recognition as an independent field of enquiry from second language acquisition (SLA). Scholars working on this new perspective maintain

that the former is inherently more complex than the latter, as it involves a quality change in the language learning and processing (e.g., Cenoz et al. 2001; De Angelis 2007). They imply that the process of learning the first foreign language (L2) is fundamentally different from the process of learning a third or additional language (L3/Ln), mainly because of enhanced language awareness, language learning strategies, and increased potential for cross-linguistic interactions between L1, L2, L3, or Ln that occur in additional language acquisition. A number of linguistic and psycholinguistic studies support these claims by providing evidence for the existence of qualitative and quantitative differences in processing the third language as compared to the first or second language (Cenoz and Jessner 2000; Cenoz et al. 2001; Hufeisen and Lindemann 1997). From a theoretical linguistic perspective, Flynn et al. (2004) argue that the study of L3 acquisition can offer new insights into the process of language learning that exceed those offered by investigations of the first or the second language.

One of the major differences between the acquisition of a second and a third language is that L3 learners have already acquired their first foreign language, and, thus, they can resort to some conscious linguistic knowledge as well as language-learning experience and strategies (cf. De Angelis 2007). Multilingual learners, thus, have at their disposal a broadened phonetic repertoire, a raised level of metalinguistic awareness, and potentially enhanced perceptual sensitivity, which may facilitate the learning of a subsequent phonological system (cf. Gut 2010; Wrembel 2015). In a dedicated volume on "Universal or diverse paths to English phonology", Gut et al. (2015) attempt a comprehensive comparison between the acquisition of phonology from a SLA vs. TLA perspective, showing that L3 learners' development of perception and production differs sharply from that of L2 learners' in being more differentiated and constrained by a greater number of factors.

The extant findings from L3 phonology research suggest that any of the previously or currently acquired languages can serve as a source for CLI in the perception and production of target segments and suprasegmentals, and that this phenomenon is multidirectional (cf. Cabrelli and Wrembel 2016). We have a growing understanding of the combination of factors conditioning the different types of phonological CLI in L3 learning, such as proficiency in the respective languages, (psycho)typology as well as the type of phonological task performed (for an overview, see Wunder 2014). However, so far the investigations have been mostly limited to a single feature and/or one testing time, thus exploring this question with more phonetic features and longitudinally seems paramount for our understanding of the relative effect of cross-linguistic processes in non-native speech learning, and speech perception in particular.

In the present paper, we examine L2 and L3 speech perception of two phonetic features, which have a different standing in the phonological repertoire of the multilinguals of this study, over the course of the first year of their instructed L3 learning. We seek to investigate how and to what extent phonological CLI may change over time in multilingual perceivers.

2. Non-Native Speech Perception

Considering that only a few studies to date have examined speech perception of multilingual learners (cf. Balas et al. 2019; Wrembel et al. 2019; Nelson 2020), models of L2 speech perception may serve as an informative starting point for the formulation of predictions for L3 learners, taking into consideration the learners' enlarged phonological repertoire as well as greater language learning experience. Most L2 speech perception models have predicted accuracy of perception on the basis of similarities and differences between L1 and L2 sounds. Starting with Lado (1957) Contrastive Analysis Hypothesis, L2 phonemes that are similar to L1 phonemes were considered easy to perceive and L2 sounds that are different from the L1 sounds difficult. Eckman (1977) Markedness Differential Hypothesis proposed that target language structures that are both different and more marked should prove difficult for learners, whereas structures that are different but less marked should not pose difficulties. The Speech Learning Model (SLM; Flege 1995) predicts that it is the fairly similar L2 sounds (to their L1 counterparts) that are most challenging for L2 learners to acquire, as they are subject to equivalence classification, i.e., they are perceptually equated with existing L1 categories.

Conversely, the sounds that do not resemble any of the L1 categories may enhance the process of category formation, and, hence, be perceived accurately. Similarly, the Perceptual Assimilation Model (PAM; Best 1995; Best and Tyler 2007) presupposes that not all target language sounds are equally challenging for learners, but it focuses on non-native contrasts rather than on individual phonemes. Discrimination of non-native sounds varies depending on how a non-native contrast is assimilated and goodness-rated to native language phonological categories, resulting in at least four different assimilation patterns for each non-native sound contrast (Best 1995, pp. 194–98).

Most relevantly for the present study, PAM predicts a continuous refinement of L2 learners' speech perception as a function of their extended experience with learning the L2 (PAM-L2; Best and Tyler 2007). With time, learners are likely to enjoy not only more L2 input but also to gain greater experience in producing the target contrasts and to increase their knowledge of L2 (minimal pair) vocabulary (Bundgaard-Nielsen et al. 2011). According to the model, L2 learners are, thus, expected to start perceiving within-category differences and develop new categories for the non-native sounds and contrasts. The way this category refinement may reshape in the context of L3 learning, particularly when L2 continues to develop too is still to be examined (for the first attempt, see Wrembel et al. 2019).

As non-native speech perception is characterized by considerable inter-listener variation, the Second Language Linguistic Perception Model (L2LPM; Escudero and Boersma 2004; Escudero 2005, 2009) concentrates on individual developmental paths on the basis of a detailed acoustic comparison of the production of L1 and L2 sounds. Two main learning scenarios are present for L2 learners, according to this model: When two L2 sounds are categorized to the same native language category, the learner needs to create a new category for one of the L2 sounds or split the existing category. When two L2 sounds are heard as separate L1 categories, the learner's task is to shift category boundaries to accommodate the L2 sounds. The latter scenario in which an L2 sound is perceived as more than one native category may be challenging as it may lead to overdifferentiation in the L2. The speed of perceptual learning in this model is, thus, predicted to depend on the particular learning scenario and richness of both L1 and L2 input that an individual learner enjoys in their learning environment.

2.1. Development of Non-Native Speech Perception

Previous research on the role of experience in the perception of non-native sounds and contrasts has yielded mixed results. Flege (1991); Baker et al. (2002); Kopečková (2012), and Rallo Fabra and Romero (2012) reported (immersion) experience effects on the discrimination and identification of at least some L2 English vowels and consonants of speakers of diverse L1 backgrounds, Cebrian (2006) found no significant differences between experienced and inexperienced Catalan-Spanish bilinguals in categorizing English /iː/ and /ɪ/ vowels. The former group of English learners had resided in Canada for an average of 25 years, while the latter group consisted of undergraduate students of English philology living in Barcelona. Cebrian (2006) reported both learner groups to rely on duration rather than spectral cues in the perception of the target contrast. In a similar vein, Broesma (2005) showed that highly experienced Dutch learners of English can accurately categorize word-final lenis-fortis contrasts, but do not use native-like weighing of cues for voicedness for this familiar contrast (present in Dutch) in an unfamiliar coda position.

Mixed findings have also been reported in perception training studies. For instance, Bradlow et al. (1999) found a long-term increase in identification accuracy of English liquids by L1 Japanese speakers. Anderson (2011) showed in a study with American English learners of Spanish that after about three weeks of identification training, some of the learners perceived the Spanish tap-trill contrast highly variably first, but then it perceptually stabilized with time; that some perceived the acoustic differences rather well in the beginning, but also revealed little change and no bifurcation of the existing phoneme category, and finally that there were also "non-learners" who showed no progress in the perception of this novel contrast. The question of refinement of non-native categories for diverse phoneme types and most crucially, under what type of learning experience it happens, thus remains at present unanswered.

2.2. Previous L3 Speech Perception Studies

As argued in previous sections, one type of learning experience that may offer important insights into the process of phonological learning in general and cross-linguistic interaction in particular is that of additional/L3 phonological learning. In one of the first studies examining phonological CLI in L3 acquisition, Wrembel et al. (2019) showed that beginner L3 Polish learners perceptually assimilate L3 sibilants to both their L1 German and L2 English categories, with preference for the latter. They can perceive subtle differences between highly similar vowel sounds across the three languages and seem to develop separate L3 categories for them. Beginner L3 learners were, thus, theorized in this study to behave similarly to experienced L2 learners thanks to their extended prior linguistic and learning experience. These are important initial insights, yet longitudinal studies examining the development of speech perception beyond only the L3 are needed to gain a more holistic picture of cross-linguistic mapping processes in multilingual learners, and possible changes thereof over time.

Some first attempts for this methodologically challenging endeavor appeared in Balas et al. (2019) and Nelson (2020). Although an examination of category formation in multilingual speech perception was the main aim of neither of these longitudinal studies, the reported findings into the development of L2 and L3 perception jointly shed at least some light on the process. In a study that stems from the same research project as the present paper, Balas et al. (2019) examined the perception of L2 and L3 rhotic sounds in two groups of young multilinguals five and nine months into their first year of L3 learning. Both L1 Polish and L1 German speakers were found to perceive L2 English rhotics highly accurately and consistently after about five years of learning the language, suggesting fairly stable phonetic categories for this novel sound (in relation to their L1) and no perceptual change as a result of the one year of additional language learning. L1 German speakers were further found to perceive the novel L3 Polish alveolar trills and taps highly accurately, and significantly better and more consistently than L1 Polish speakers did in perceiving L3 German uvular fricatives; the accuracy in perceiving the novel sound further dropped significantly between the two testing times for the latter learner group. The findings were interpreted as suggesting a joint effect of the learner's L1, but not L2, markedness and L2/L3 proficiency in the perception of rhotic sounds by multilingual learners. The present contribution expands on and refocuses this study.

Nelson (2020) examined young and adult L3 learners' perception of the /v-w/ contrast, present in their L2 but not L1, reporting more accurate and faster discrimination ability in the L3 than in the L2 after only a few hours of L3 input. The author hypothesized a positive 'novelty effect' for the L3 learners, maintaining that very initial learners may not automatically assimilate novel sounds to their pre-existing categories (whether those of L1 or L2) but rather resource acoustic cues available to them and tap possible yet different processing and phonological skills at that stage of L3 phonological learning. With respect to their L2 perception development, the young learners evidenced a drop in accuracy after around 10 weeks of their L3 learning, which was interpreted as suggesting a reverse cross-linguistic effect in the form of a temporary 'perceptual confusion'. However, after ten months of learning the L3, the novelty effect as well as the negative cross-linguistic effect disappeared for the young L3 learners, who perceived the contrast in their L2 and L3 similarly (67% and 74% accuracy levels).

To sum up, a common denominator for the existing L3 perception studies is that the phonological space of multilinguals seems to be reshaped relatively early in the course of learning the new L3, and that category boundaries can be expanded to accommodate L1, L2, and L3 categories of similar phonetic types, while new L3 categories for novel phonetic types may be formed. Initial sensitivity to phonetic contrasts may also deteriorate with time as a result of language interactions and be modulated by the status of various contrasts in L3 acquisition, including that of markedness. In the present paper, we attempt to contribute to these emerging findings by examining the perception of novel rhotic sounds (both in the L2 and L3 of the multilinguals, and more marked in their L3) and the perception of final obstruent (de)voicing (more marked in their L2) in the first months of L3 learning.

3. The Present Study

The aim of this study is to explore the development of speech perception in young multilinguals' non-native languages (L2 and L3) and to trace the patterns of cross-linguistic mappings over the first year of L3 learning. This study forms a part of the international MULTI-PHON research project, in which speech perception and production was investigated with a battery of tests in two parallel groups of young adolescents in Polish and German schools.

3.1. Participants

The participants were 13 L1 Polish speakers (aged 12–13) who had been learning English as their L2 at school for five years (pre-intermediate level) and who had just started to learn German as their L3 in an instructed setting. They were observed over the first year of L3 learning. Our strict inclusion criteria featured no prior command of German, only Polish as an L1, no additional languages, and data availability at all testing times, thus, for the sake of the present analysis the number of participants was reduced from a larger participant pool (initially 24) to 13 speakers with a homogeneous profile (see Table 1).

Table 1. Participant profiles.

	Mean	*SD*
Age (years)	12.25	0.41
AOL2 (age of onset of L2 English)	5.50	0.80
AOL3 (age of onset of L3 German)	12.25	0.62
Hrs of L2 instruction per week	3	--
Hrs of L3 instruction per week	5	--
Self-evaluation in L2 *	3.65	0.51
Self-evaluation in L3 *	3.33	0.58
Female/male ratio	8/5	--

* Self-evaluation of proficiency was assessed on a 5-point scale (1 = very poor, 5 = very good).

An informed consent was obtained from all the subjects who participated in the study, their parents, and the school authorities where the data was collected. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ministry of Education in Brandenburg on 17/07/2017 (ref. number 51/2017).

Language background interviews were conducted in the participants' L1 Polish at the very onset of the project in order to collect information about the individual learner's language backgrounds, including information about their language learning history (i.e., age of learning, length and intensity of instruction), language use (declared percentage in varied situations/contexts), self-evaluation of proficiency (at the onset of instructed L3 learning), and attitudes towards foreign language learning.

3.2. Features under Investigation

Two phonetic features were selected for investigation, rhotics and final obstruent devoicing, since they have a relatively different standing in the phonological repertoire of the L3 learners in this study (see Table 2). The former sounds are realized differently in each of the speakers' three languages, whereas the latter process is productive in their L1 and L3 but not L2.

Table 2. Selected features under analysis in the present study.

Language	Rhotics	Obstruent Devoicing in Syllable-Final Position
Polish	/r/	Yes
English	/ɹ/	No
German	/ʁ/ and /ɐ̯/	Yes

3.2.1. Rhotics

In spite of belonging to a phonological natural class, for which there are more phonological than phonetic arguments (cf. Ladefoged and Maddieson 1996), rhotics exhibit large interlanguage variability. In the three languages under investigation in this paper, the distribution of rhotics is as follows: Polish has the alveolar trill, which may be produced as a tap intervocalically or in fast speech (Jassem 2003). In standard German, the conservative uvular trill /ʀ/, occurring in word-initial or in stressed positions, is usually produced as the uvular fricative /ʁ/ (Kohler 1999). English rhotics include British English postalveolar approximant /ɹ/ and prevocalically [/ɹ/], and an American English retroflex approximant (Ladefoged and Maddieson 1996) articulated either with tongue retroflexion or bunching (Ladefoged 2001). The English rhotic is generally voiced except when adjacent to a voiceless obstruent. It occurs in syllable-initial (e.g., run/ɹʌn/), and syllable-final position (e.g., poor/pɔɹ/) (not in British English), both as singletons and in clusters (e.g., tree/tɹi/; heart /haɹt/). Both English rhotic sounds are continuants, as opposed to the 'interrupted' variants such as taps or trills in Polish. Worthy of note is the fact that in all three languages, the rhotic sounds are represented orthographically using the <r> letter. This suggests that orthography may promote multiple and multidirectional phonological transfer (cf. Rafat 2011).

3.2.2. Final Obstruent Devoicing

The three languages under investigation differ in the realization of coda obstruents. While English retains a voicing contrast in a syllable-final position, this opposition is neutralized in German and Polish (Gonet 2001; Smith et al. 2009). Although both German and Polish manifest final obstruent devoicing, Polish additionally applies the rule of regressive voicing assimilation (Rubach 1984). Both languages have also been associated with less than a total neutralisation of the underlying voicing contrast, in that small differences in one or more acoustic properties, such as the length of the preceding vowel, have been reported when compared to underlying voiceless counterparts (Slowiaczek and Dinnsen 1985). English, in contrast to German and Polish, typically manifests the marked voiced/voiceless contrast among word final obstruents, even though individual variation has also been reported, as well as the effect of phonological environment on the production of specific word-final obstruents (Gonet 2001; Smith et al. 2009). Finally, English voiced word-final obstruents have primarily been characterized by longer duration of the preceding vowel and not necessarily by glottal pulsing (Krause 1982).

3.3. Research Questions and Hypotheses

In order to investigate cross-linguistic interactions in multilinguals' speech perception, the following research questions were posed in the study:

1. Is there evidence of CLI in the perception of L2 English and L3 German?

It is hypothesized that cross-linguistic interactions in the two foreign languages may differ and result in variable performance on the measures of perception accuracy and reaction time (RT), depending on the language status (L2 vs. L3) as well as the investigated feature (rhotics vs. final obstruent devoicing). Better performance on both measures and, thus, less CLI is expected for the more established L2 as compared to the newly acquired L3.

Hypothesis 1 (H1). *Both phonological feature and language determine perception accuracy and reaction times. There will be less CLI in the learners' L2 English than their L3 German.*

2. Is there a perceptual development over time caused by a change in CLI? Does the perceptual development in L3 parallel that in L2? In this study, cross-linguistic interactions were operationalized as

the respondents' preferences for L1-, L2-, L3-accented stimuli in the performed forced-choice goodness

task. We expect different patterns to hold for the two foreign languages acquired. We expect to observe a change in CLI patterns as a function of the testing time (T1 vs. T2).

Hypothesis 2 (H2). *There will be changes in CLI across time. The developmental patterns of CLI differ between the learners' L2 English and L3 German.*

3.4. Materials and Methods

The participants performed perceptual tasks in both their L2 English and L3 German, respectively, to test their perception of rhotics and final obstruent (de)voicing. Response accuracy and reaction times were recorded for analyses at two testing times (T1, after 5 months and T2, after 10 months of L3 learning). To create appropriate language modes, the data collection for each of the languages was carried out on two separate days with L1 speakers of the respective languages as instructors.

A forced-choice (FC) goodness task was selected for the present study as an alternative to more traditional perceptual paradigms such as discrimination or identification. Perception discrimination tasks, in which the listener decides whether two stimuli are the same or different, seemed to be of little use as the aim was to test the association of a given variant of a sound with a chosen language in the multilingual's repertoire. Identification tasks in turn are inherently notorious for specifying response alternatives (including difficulties concerning non-transparent orthography), the problem being magnified in the case of three phonological systems in interaction. Moreover, identification tasks are not useful for testing allophonic differences across languages. Overly complex perception tasks needed to be avoided, too: when task complexity increases, perceivers have been found to switch to a primarily phonological level of reasoning (Strange 2009). Therefore, a forced-choice goodness task was selected for the present research, which allowed for elicitation of an association of a given allophone across multiple languages while the complexity of stimulus identification was avoided.

More specifically, the participants in this study heard two renditions of the same phrase differing on the last stimulus items embedded in a carrier phase. By pressing one of two buttons (marked 1 and 2) on a button box, they had to decide which phrase sounds more natural (i.e., more target-like) to them. One rendition was a target realization and the other was an accented language realization, where only the investigated feature was manipulated. For example, for rhotics, in the English version of the task, the stimuli included the target-like phrase "You will hear the word ring /ɹɪŋ/" followed by the Polish-like realization of the rhotic sound "You will hear the word ring /riŋ/".

For rhotic sounds, this included two trials of pair items as the target item was positioned next to two other possible realizations, while for obstruent (de)voicing, it featured a single trial as the target was presented in opposition to voiced or devoiced/voiceless. The order of presentation of target and non-target stimuli was counterbalanced across trials.

Thus, in the English version, there were stimuli with English target rhotics as well as with Polish and German rhotics. Likewise, in the German version, the stimuli included German target rhotics embedded in a carrier phrase as well as Polish- and English-accented manipulated rhotics in the target words. In case of obstruent (de)voicing, the stimuli in the English version included the target-like phrase "You will hear the word have" /hæv/, followed by a manipulated realization of the final obstruent /hæf/. Similarly, in the German version, the target words embedded in a carrier phrase ("Du hörst das Wort Hand" /hant/) included final obstruents that were either voiceless (thus target like) or voiced (i.e., L2-accented).

The stimuli in each language version involved 10 pair items containing rhotics, 13 to 14 pair items featuring final (de)voicing, and three training pair items that preceded the testing blocks. In total, the FC task, thus, included 26 English and 27 German pair items for the participants to respond to.

The target rhotics occurred either in word-initial or medial position and included:

- For English: *ring, rabbit, red, round, giraffe* (with the manipulated items realized as having an L1-Polish-accented alveolar trill or an L3-German-accented uvular fricative).

- For German: *rot*, *Regen*, *Reise*, *Fahrrad*, *verloren* (with the manipulated items realized as having an L1-Polish-accented alveolar trill or an L2-English accented post-alveolar approximant).

The final obstruent (de)voicing stimuli were in coda positions and featured as follows:

- For English: *days*, *grab*, *leg*, *could*, *stab*, *big*, *skies*, *give*, *love*, *food*, *judge*, *have*, *rob* (with the manipulated items realized with voiceless final obstruents, which could be interpreted as either L1-Polish or L3-German-accented)
- For German: *Hand*, *Berg*, *Quiz*, *lieb*, *Kleid*, *Mund*, *Honig*, *Hund*, *Fahrrad*, *Kind*, *vierzig*, *brav*, *Korb*, *gelb* (with the manipulated items realized with voiced final obstruents, which could be interpreted as L2 English-accented).

The stimuli were randomized across trials in E-prime. The inter-stimulus interval was set at 500 ms and the participants had a 3000 ms response limit, thus, the task was timed. The participants' performance on the timed forced-choice goodness task was examined in terms of accuracy and reaction time (RT). The latter was included as a proxy for the perceptual difficulty of the tested stimuli.

The stimuli were recorded by three female native speakers of the respective languages, who were fluent advanced speakers of the other two languages in the triad of languages. The stimuli were produced naturalistically to avoid artificial concatenation. To ensure naturalness, several recordings of the same items were performed and validated by selecting the ones in which the performed accented manipulation sounded the most acceptable to the researchers. The process of stimulus validation was based on the perceptual assessment of each stimulus by native speakers of the respective languages. We adopted a perceptual 'category goodness' criterion, which was deemed to have the best ecological validity given the nature of the FC goodness task administered to the participants.

As far as the three speakers who produced the stimuli are concerned, their stay in a foreign country ranged from a few months to a few years. While we acknowledge the fact that their L1 production could be affected by a highly proficient knowledge of the L2/Ln, it is debatable if the prototypical monolingual rendition should be sought as the target production of the stimuli, in the light of the recent discussions on the native monolingual norm in research on multilingual acquisition (see e.g., Sorace 2020; Kroll 2020). Moreover, monolingual speakers of German, Polish, or English are increasingly impossible to find. Therefore, it was not our goal to search for a native monolingual rendition of the target items, but rather to allow for a potential variation represented by native speakers of particular languages who are multilingual speakers themselves.

4. Results

Due to violation of the assumption of normality and homogeneity of variance of the present dataset, nonparametric tests were used for between-subjects (Mann–Whitney *U*-test) and within-subjects (Wilcoxon signed-rank test) comparisons. The statistical tests were run using STATISTICA 10. The performed analyses included perceptual development over time, feature comparison, language comparison, individual variability, and CLI analysis, which will be presented in the following subsections.

4.1. Nonparametric Tests of Perception Accuracy and RT

4.1.1. Perceptual Development over Time: Perception Accuracy at T1 and T2

The performed across-time comparison did not show much development in perception accuracy for the multilingual learners. The only statistical difference between the two testing times in the performance in L2 and L3 for the two features under investigation was found for L3 German rhotics (and not in the expected direction), in which case the perception accuracy was higher at T1 than at T2 ($Z = 4.5$, $p < 0.05$) (see Table 3).

Table 3. Perception accuracy for second language (L2) and third language (L3) at testing time one (T1) and testing time two (T2).

Language	Feature	Time	N	Perception Accuracy		Wilcoxon Matched-Pairs Test	
				Mean	SD	Z	p
L2 English	obstr_devoicing	T1	155	0.55	0.50	0.75	0.4509
		T2	155	0.51	0.50		
	rhotics	T1	95	0.92	0.27	0.24	0.8139
		T2	95	0.92	0.28		
L3 German	obstr_devoicing	T1	149	0.40	0.49	0.31	0.7603
		T2	149	0.42	0.50		
	rhotics	T1	89	0.76	0.39	4.50	* 0.0000
		T2	89	0.38	0.45		

* $p < 0.05$.

4.1.2. Perceptual Development over Time: RT at T1 and T2

The performed Wilcoxon matched pairs signed rank test for the comparison of reaction times (RT) at two testing times (T1 vs. T2) did not show much development over time either. The only statistically significant result was found for L3 German obstruent devoicing ($Z = 2.14$, $p < 0.05$), with the processing time being longer at T1 than at T2 (see Table 4).

Table 4. Reaction time (RT) for L2/L3 at T1 and T2.

Language	Feature	Time	N	RT		Wilcoxon Matched-Pairs Test	
				Mean	SD	Z	p
L2 English	obstr_devoicing	T1	155	644.0	437.7	1.59	0.1122
		T2	155	581.4	397.0		
	rhotics	T1	95	600.3	398.9	1.88	0.0596
		T2	95	519.3	308.6		
L3 German	obstr_devoicing	T1	149	802.4	444.0	2.14	* 0.0325
		T2	149	729.3	458.9		
	rhotics	T1	89	682.6	471.7	0.22	0.8222
		T2	89	647.7	418.2		

* $p < 0.05$.

On the whole, the results did not demonstrate much development over time in perception accuracy and processing speed as measured by means of a FC task. It appears, however, that the L2 English is the more established phonological system, while L3 German is more susceptible to changes over the two testing times (i.e., a significant change in the perception accuracy of rhotics and in processing speed for obstruent devoicing). There is no consistency though in the observed developmental changes (the decrease in RT for the perception of obstruent devoicing is as expected, whereas the decrease in accuracy of rhotics perception appears counterintuitive).

4.1.3. Feature Comparison: Perception Accuracy

In the performed feature comparison, the Mann–Whitney *U*-test for perception accuracy demonstrated statistical differences in three out of four conditions: L2 English rhotics were perceived with greater accuracy than obstruent devoicing both at T1 ($Z = -6.18$, $p < 0.05$) and T2 ($Z = -6.51$, $p < 0.05$), while for L3 German the same held true at T1 ($Z = -5.19$, $p < 0.05$) (see Table 5).

Table 5. Comparison of perception accuracy of features at T1 and T2.

Language	Time	Feature	N	Accuracy		Mann–Whitney *U*-Test	
				Mean	SD	Z	*p*
L2 English	T1	obstr_devoicing	159	0.55	0.50	−6.18	* 0.0000
		rhotics	97	0.92	0.26		
	T2	obstr_devoicing	165	0.50	0.50	−6.51	* 0.0000
		rhotics	101	0.89	0.31		
L3 German	T1	obstr_devoicing	166	0.40	0.49	−5.19	* 0.0000
		rhotics	97	0.73	0.41		
	T2	obstr_devoicing	159	0.44	0.50	0.52	0.6028
		rhotics	94	0.40	0.46		

* *p* < 0.05.

4.1.4. Feature Comparison: RT

When the two features were compared in terms of reaction time, only one statistical difference was attested for L3 German at T1, when RT were longer for final devoicing than for rhotics (Z = 2.98, *p* < 0.05). Otherwise, the processing speed did not differ across features (see Table 6).

Table 6. Comparison of RT to features at T1 and T2.

Language	Time	Feature	N	RT		Mann–Whitney *U*-Test	
				Mean	SD	Z	*p*
L2 English	T1	obstr_devoicing	159	657.4	445.6	1.02	0.3100
		rhotics	97	594.0	397.1		
	T2	obstr_devoicing	165	585.3	425.9	1.08	0.2791
		rhotics	101	512.9	305.2		
L3 German	T1	obstr_devoicing	166	794.9	457.9	2.98	* 0.0029
		rhotics	97	681.2	479.6		
	T2	obstr_devoicing	159	719.9	453.0	1.25	0.2130
		rhotics	94	657.8	430.5		

* *p* < 0.05.

4.1.5. Language Comparison: Perception Accuracy

To compare the perception performance across languages, a Mann–Whitney *U*-test was performed, which demonstrated statistically significant differences for three out of four conditions, i.e., the perception accuracy was higher for rhotics in L2 English than in L3 German at both T1 (Z = 4.0, *p* < 0.05) and T2 (Z = 7.63, *p* < 0.05), and for obstruent devoicing at T1 (Z = 2.7, *p* < 0.05). A higher proficiency in the more established L2 was reflected in better accuracy performance in perception (see Table 7).

Table 7. Perception accuracy comparison between L2 English and L3 German.

Feature	Time	Language	N	Accuracy		Mann-Whitney *U* Test	
				Mean	**SD**	**Z**	**p**
obstr_devoicing	T1	L2 English	159	0.55	0.50	2.70	* 0.0070
		L3 German	166	0.40	0.49		
	T2	L2 English	165	0.50	0.50	1.02	0.3075
		L3 German	159	0.44	0.50		
rhotics	T1	L2 English	97	0.92	0.26	4.00	* 0.0001
		L3 German	97	0.73	0.41		
	T2	L2 English	101	0.89	0.31	7.63	* 0.0000
		L3 German	94	0.40	0.46		

* $p < 0.05$.

4.1.6. Language Comparison: RT

A Mann–Whitney *U*-test for reaction time comparison between L2 English and L3 German demonstrated statistically significant differences for three out of four conditions, i.e., RTs were longer in L3 German than in L2 English for the perception of obstruent devoicing at both T1 and T2 and for the perception of rhotics at T2. On the whole, it took longer to process the perception task in the L3 than in the L2 (see Table 8).

Table 8. Reaction time comparison between L2 English and L3 German.

Feature	Time	Language	N	RT		Mann–Whitney *U*-Test	
				Mean	**SD**	**Z**	**p**
obstr_devoicing	T1	L2 English	159	657.4	445.6	−3.70	* 0.0002
		L3 German	166	794.9	457.9		
	T2	L2 English	165	585.3	425.9	−3.26	* 0.0011
		L3 German	159	719.9	453.0		
rhotics	T1	L2 English	97	594.0	397.1	−0.74	0.4606
		L3 German	97	681.2	479.6		
	T2	L2 English	101	512.9	305.2	−2.36	* 0.0184
		L3 German	94	657.8	430.5		

* $p < 0.05$.

4.1.7. Correlation: Perception Accuracy and RT

No statistically significant correlations were found between perception accuracy and reaction time for L2 English and L3 German performance in the perception of rhotics and final devoicing at either T1 or T2 (see Table 9).

Table 9. Correlation between perception accuracy and RT at T1 and T2.

Language	Time	Feature	n	r(X.Y)	t	p
L2 English	T1	obstr_devoicing	13	0.371	1.32	0.2121
		rhotics	8	−0.618	−1.92	0.1027
	T2	obstr_devoicing	13	0.288	1.00	0.3400
		rhotics	8	−0.577	−1.73	0.1345
L3 German	T1	obstr_devoicing	14	0.203	0.72	0.4859
		rhotics	8	0.287	0.73	0.4911
	T2	obstr_devoicing	14	0.062	0.22	0.8333
		rhotics	8	−0.099	−0.24	0.8160

4.2. GLM Modelling

We fitted our data to a generalized linear model (GLM), with the dependent variable being perception accuracy and independent variables including RT, testing time (T1 and T2) and feature (obstruent devoicing and rhotics). The analysis was performed separately for each language and based on the number of token items rather than participants.

The GLM analysis for L2 English revealed a significant effect of feature on the perceptual accuracy in L2 English [F(1,522) = 92.79, $p < 0.05$)], while the testing time and RT were not significant predictors (see Table 10).

Table 10. Results of a linear model for the dependent variable—Accuracy for L2 English.

Effect	SS	df	MS	F	*p*
Intercept	94.55	1.00	94.55	506.05	* 0.0000
RT	0.42	1.00	0.42	2.27	0.1326
Testing time	0.30	1.00	0.30	1.59	0.2075
Feature	17.34	1.00	17.34	92.79	* 0.0000
Time*Feature	0.02	1.00	0.02	0.10	0.7559
Error	96.59	517.00	0.19		

* $p < 0.05$.

The Bonferroni pairwise comparisons confirmed that there were statistically significant differences ($p < 0.001$) between perception accuracy for rhotics and obstruent devoicing, with the former feature generating higher accuracy rates (see Table 11).

Table 11. Mean Accuracy with respect to Feature for L2 English.

Feature	n	Mean	SD	SE
obstr_devoicing	324	0.52	0.50	0.03
rhotics	198	0.91	0.29	0.02

The GLM analysis for L3 German failed to find a significant effect of RT, however, the remaining variables proved to be significant predictors for perceptual accuracy in L3 German, namely testing time [(F(1, 516) = 11.85, $p = 0.000$)], feature [(F(1, 516) = 10.55, $p = 0.001$)], and the Time*Feature interaction [(F(1, 516) = 18.05, $p = 0.000$)] (see Table 12).

Table 12. Results of a linear model for the dependent variable—Accuracy for L3 German.

Effect	SS	df	MS	F	*p*
Intercept	38.07	1.00	38.07	168.58	* 0.0000
RT	0.06	1.00	0.06	0.28	0.5938
Testing Time	2.68	1.00	2.68	11.85	* 0.0006
Feature	2.38	1.00	2.38	10.55	* 0.0012
Time*Feature	4.08	1.00	4.08	18.05	* 0.0000
Error	115.40	511.00	0.23		

* $p < 0.05$.

The Bonferroni pairwise comparisons pointed to a statistically significant difference ($p = 0.017$) between the two testing times in L3 German, with higher perception accuracy observed at T1 (see Table 13).

Table 13. Mean Accuracy with respect to Testing Time for L3 German.

Testing Time	N	Mean	SD	SE
T1	263	0.52	0.49	0.03
T2	253	0.42	0.48	0.03

Bonferroni correction confirmed a statistically significant difference between perceptual accuracy of the two investigated features (p = 0.0008), with rhotics being perceived more accurately than obstruent devoicing in L3 German (Table 14).

Table 14. Mean Accuracy with respect to Feature for L3 German.

Feature	N	Mean	SD	SE
obstr_devoicing	325	0.42	0.49	0.03
rhotics	191	0.57	0.47	0.03

The Bonferroni pairwise comparisons confirmed that there were statistically significant differences for perceptual accuracy in L3 German between the following variables: (1) obstruent devoicing at T1 and rhotics at T1 (p < 0.0001); (2) obstruent devoicing at T2 and rhotics at T1 (p < 0.0001); (3) rhotics at T1 and rhotics at T2 (p < 0.0001) (see Table 15).

Table 15. Mean Accuracy with respect to the Time*Feature interaction for L3 German.

Testing Time	Feature	N	Mean	SD	SE
T1	obstr_devoicing	166	0.40	0.49	0.04
T1	rhotics	97	0.73	0.41	0.04
T2	obstr_devoicing	159	0.44	0.50	0.04
T2	rhotics	94	0.40	0.46	0.05

4.3. Individual Differences

Figures 1–8 show that, in general, more inter- and intraspeaker variability occurs in L3 German than in L2 English, in which individual perceptual performance seems more homogeneous across learners. This is especially true for the perception of the English rhotic where six learners show ceiling performance at both testing times. Pronounced changes in perception accuracy across time are, however, apparent for individual learners. In the case of Subject 20, for instance, their perception of both L2 English obstruent voicing and rhotics drops drastically from T1 to T2 and also shows a drop in perception accuracy in the L3 German rhotic from well above chance to well below it from T1 to T2 (see Figures 1–4, 7 and 8). Subject 12, in turn, performs consistently accurately in their perception of the L2 sounds under examination (Figures 1–4). Their perception of the L3 counterparts drops between the two testing times, most dramatically for rhotics (Figures 5–8). Some increase in L2 English perception of final obstruents (Figures 1 and 2) together with a dramatic improvement of L3 German perception of the same feature (Figures 5 and 6) was evidenced in Subject 6. See Figures 1–8, illustrating perception accuracy of individual subjects in L2 English and L3 German at T1 and T2 for obstruent devoicing and rhotics (with group means marked as horizontal black lines on the graphs).
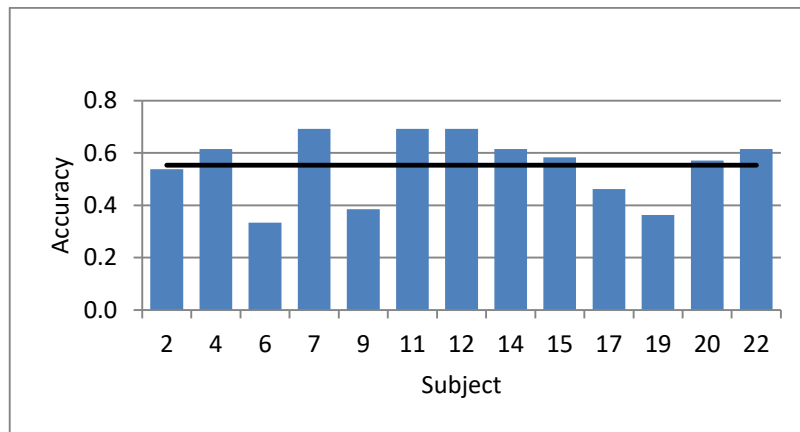
Figure 1. Perception accuracy in L2 English for obstruent devoicing at T1.
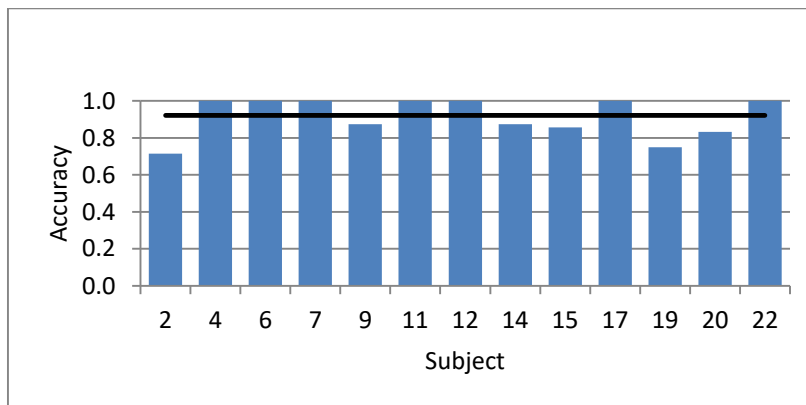

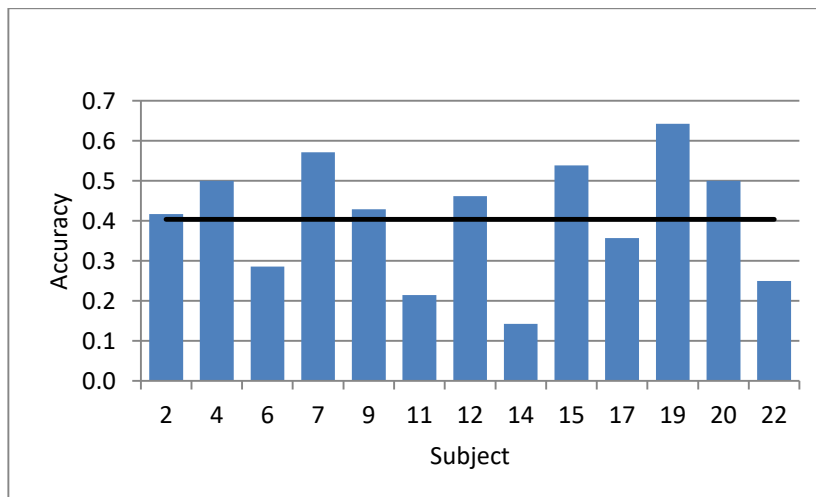**Figure 2.** Perception accuracy in L2 English for rhotics at T1.

Figure 3. Perception accuracy in L3 German for obstruent devoicing at T1.
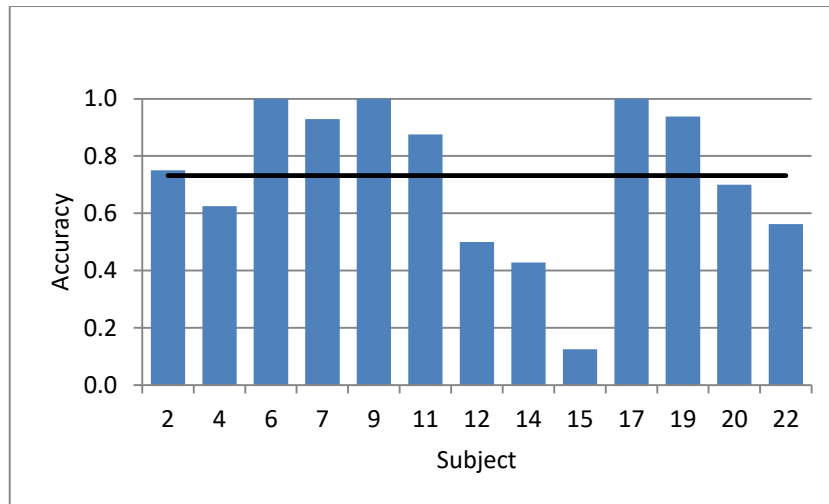
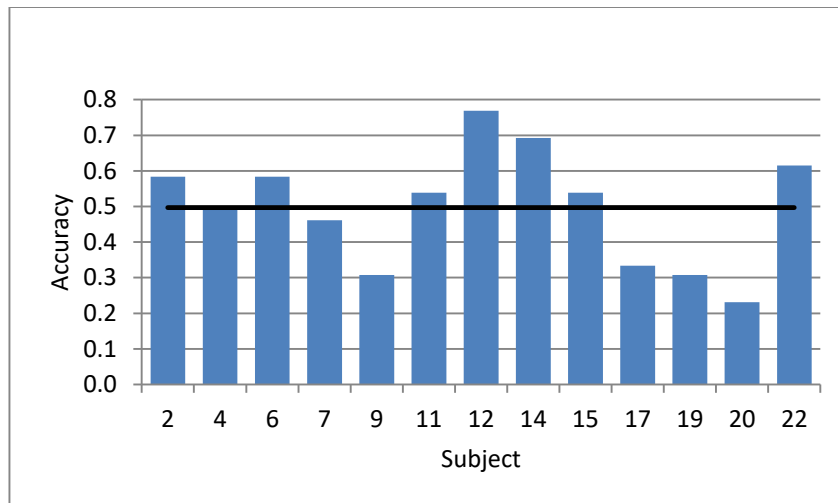**Figure 4.** Perception accuracy in L3 German for rhotics at T1.

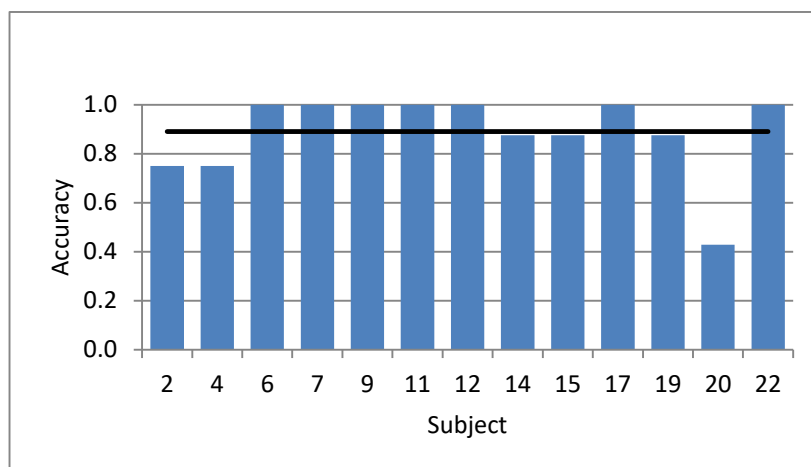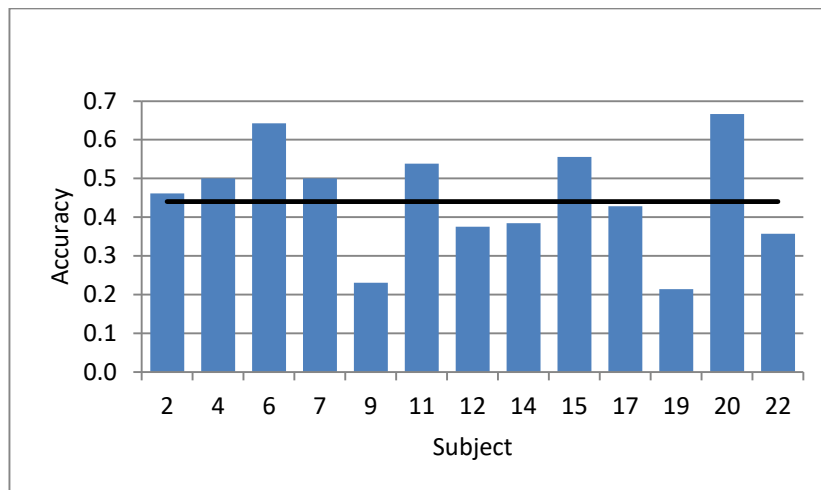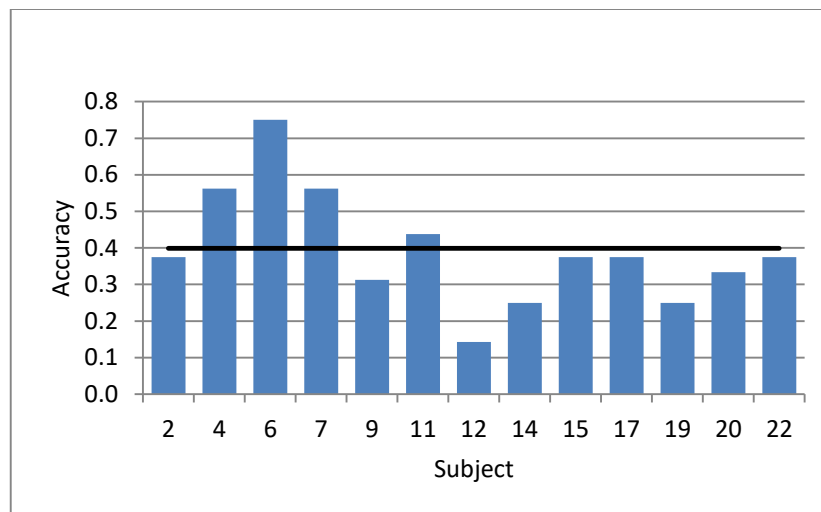**Figure 5.** Perception accuracy in L2 English for obstruent devoicing at T2.

Figure 6. Perception accuracy in L2 English for rhotics at T2.

Figure 7. Perception accuracy in L3 German for obstruent devoicing at T2.

Figure 8. Perception accuracy in L3 German for rhotics at T2.

4.4. CLI

In order to explore cross-linguistic mappings in the perception of the multilingual learners of this study, we further explored their perception accuracy (as the dependent variable) with respect to the different stimulus properties of the perception task employed (i.e., L1-accented, L2-accented, L3-accented) as independent variables.

For rhotics, the performed ANOVA (with L2 and L3 treated jointly) demonstrated that there was a statistically significant difference in perception accuracy between these three conditions ($F_{(2;24)} = 46.38$, $p < 0.05$). The post-hoc Scheffé test for multiple comparisons showed that the differences between all pairs of differently accented stimuli were significant ($p < 0.05$). The accuracy of perceiving the correct rhotic stimuli in L2 English was the highest when the other manipulated stimulus was L3-(German) accented, while it was the least accurate when the unnatural stimulus was L2-(English) accented in L3 German (see Figure 9). Interestingly, however, when we compared the latencies of responses in all these conditions, there were no statistical differences found in RT for rhotics irrespective of the source of accent in the manipulated stimuli.
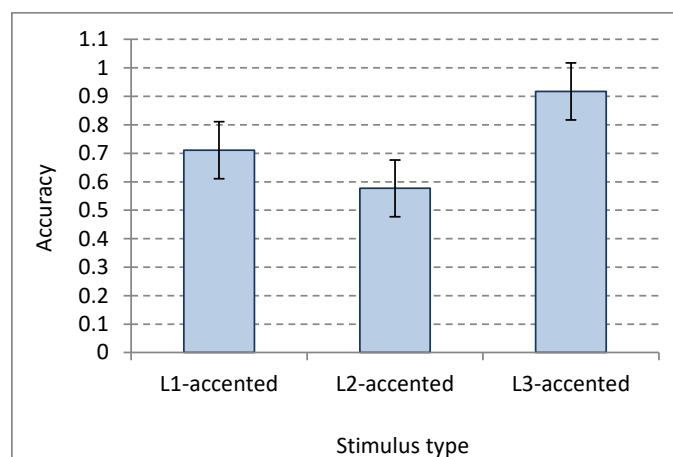
Figure 9. Perception accuracy of rhotics according to stimuli types (L1-accented, L2-accented, L3-accented), with 0.95 confidence intervals as whisker bars.

For final devoicing, given the binary response option as well as difficulty in strictly disentangling L1-based source of CLI in the perception of this feature from arguably the lack of it (L3-target stimuli), the results evidence CLI primarily from L1 and/or L3 in the case of perceiving L2 final obstruent voicing (accuracy levels at chance levels, with acceptance of L1/L3-based and L2-based stimuli to comparable levels). However, in the case of L3 final obstruent devoicing, L1-based CLI prevailed (L1-accented/L3-based stimuli were generally perceived as being more natural than L2-accented stimuli) (t = 4.12, $p < 0.05$).

As far as the reaction time is concerned, none of the independent variables (i.e., feature, stimulus type) entered into the GLM analysis proved to be significant, nor did the interaction between feature and stimulus type ($p > 0.05$). It follows that no statistical differences were found in RT, irrespective of the source of accent in the manipulated stimuli, in the perception of both of the investigated features, although there was a visible trend for the L2-accented stimuli in the perception of L3 obstruent devoicing taking longer to process that the L2-accented stimuli in L3 rhotics.

5. Discussion

Our results show that the effects of CLI on multilinguals' perception differ across both their two languages and the two features under investigation, thus confirming Hypothesis 1. Overall, perception accuracy is higher in their L2 English than in their L3 German and processing speed is faster, as predicted by Hypothesis 1. Moreover, perception accuracy in the L2 English, which they had been learning for 5–6 years, is more stable across time than for the L3 German, confirming Hypothesis 2. Our results, thus, suggest that CLI is lowest for the perception of the L2 English, especially for rhotics, where most of the investigated learners seem to have established stable perceptual categories. However, we did not test learners' perception in their L2 English after a few weeks of learning the new language German, and, thus, might have missed the short-term effect of influence from the new L3, the 'perceptual confusion' found by Nelson (2020). In fact, one individual learner did show a drop in L2 perception accuracy even after ten months of learning the L3, which might have the same underlying cause.

Our results further show that overall perception accuracy is higher for rhotics than for final obstruent in both languages. Perception of final obstruent devoicing in both the learners' L2 and L3 is at chance level at both testing times, evidencing no improvement for any of the learners, while perception of the rhotics is significantly higher in both languages, with individual speakers reaching ceiling performance. Contrary to the predictions of our Hypothesis 1, this suggests a high level of CLI for the former feature, even in the L2 English, for which learners had been attending school lessons for 5–6 years. One explanation might be the lower perceptual saliency of final obstruent (de)voicing

compared to the different articulations of the rhotics in the three languages under investigation. Moreover, the phonological process of obstruent voicing in coda position is characterized by a complex interaction of phonetic cues beyond that of glottal pulsing (Krause 1982). As shown in Broesma (2005), even highly proficient learners of English do not use native-like weighing of cues for the perception of voicedness in an unfamiliar position. Our learners may, thus, have had a hard time to attend to the relevant phonetic cues, longer duration for the preceding vowel in particular, to distinguish between the pairs of tested stimuli.

By the same token, evidence for CLI was found in the learners' perception in their L3 German: their accuracy of perceiving the German rhotic /R/ was higher after 5 months of learning than after 10 months. It appears that some restructuring of perceptual categories is still under way in the first ten months of exposure to a new language, thus echoing findings by Balas et al. (2019). However, again, this restructuring seems to be feature-dependent rather than a general mechanism as these changes were found only for the perception of the rhotics but not for the perception of final obstruent devoicing. Our findings thus appear to partially contradict the predictions of PAM-L2 (Best and Tyler 2007), which would expect a continuous refinement of the learners' speech perception as a function of their extended experience with learning the language. Possibly, this refinement only takes place after more input than our learners had enjoyed in their L3 after 10 months of learning. Not incompatible with this line of reasoning, it might be that the L3 learners in this study had been increasingly exposed to foreign-accented realizations of the German rhotic sound in their classroom environment, whether from their peers or their Polish teacher of German, thus, developing a nontarget representation of naturalness for it. Their own experience with producing the articulatorily challenging sound in the first year of learning German may have also contributed to the process of their category formation for the sound (cf. Bundgaard-Nielsen et al. 2011). As an alternative explanation for the drop in perceptual performance, one could point to a possible decreased attention to the task at the second testing point as compared with the novelty of the first testing time that triggered more focused interest and auditory processing in the participants. This finding would be in line with Nelson (2020) observations concerning the initial 'novelty effect' in perceptual performance of her child and adult L3 learners.

The source of cross-linguistic influence on the perception accuracy of the multilingual learners was found to vary: the accuracy of perceiving the L2 English rhotic /ɹ/ was higher when it was contrasted with an L3-German accented stimulus than with an L1-Polish accented stimulus in the FC task. This would point towards a stronger influence of the L1 than the L3 in the perception of rhotics in the L2, although recall that overall L2 rhotics were perceived highly accurately by the learners. On the other hand, in L3 German, the accuracy of perceiving the rhotic was lowest when it was contrasted with an L2 English accented stimulus rather than with an L1-Polish stimulus, which leads to the conclusion that the L2 rather than the L1 was a stronger source of CLI for the L3 perception of this feature. This would seem to suggest initially a greater influence of the L2 than the L1 on the perception of L3 rhotics, a finding that was also reported in Wrembel et al. (2019). Indeed, initial L3 learners appear to map new non-native phones to both their L1 and L2, which may be interpreted as aligning with the general reasoning of most L2 speech perception models: non-native phones are perceived in relation to previously established (or currently being established) categories depending on the degree of perceived cross-linguistic similarity between the phones concerned. The way in which such perceived cross-linguistic mappings are to be most effectively elicited in multilingual perceivers presents one of the greatest methodological challenges in future L3 speech research.

Regarding obstruent devoicing, it was not possible to disentangle the sources of CLI for L2 perception (due to the identical nature of L1-accented and L3-based stimuli). However, if we assume the existence of CLI at this stage of L3 learning, L3 perception of the devoicing feature was arguably influenced more significantly by the L1 than the L2, considering the more marked status of obstruent voicing in L2 as well as the similar standing of this feature in the L1 and L3.

Our results further showed that factors other than CLI might influence speech perception. Higher accuracy in the L2 than in the L3 and the fact that L2 is processed faster than the L3 are

viewed as evidence that what also matters in non-native speech perception is experience. Our results corroborate the effects of language learning experience on non-native consonant perception, similarly to some previous studies (e.g., Bradlow et al. 1999; Rose 2010; Anderson 2011), which reported some improvement for more experienced participants or after perception training, but also considerable variation across subjects and the phones tested, as predicted by L2LPM (Escudero and Boersma 2004; Escudero 2005, 2009).

No correlation was found between the learners' perception accuracy and reaction time in the perception of rhotics and final devoicing in either language and at either observation point. This suggests that processing speed is quite independent of the degree of establishment of perceptual categories and may not be the most informative proxy for evaluations of the learnability of different sounds, at least for L3 learning contexts.

As for the role of markedness, in the present study we tested one feature which was more marked in the L2 than in the L1 and L3 (i.e., final obstruent devoicing) and one feature which was more marked in the L3 than in the L2 (i.e., German uvular vs. English postalveolar rhotics). L2 English rhotics were more accurately chosen when contrasted with L3 German stimuli, possibly suggesting a stronger influence of the less marked L1 rhotic than of the most marked L3 rhotic on the L2 perception of a relatively unmarked rhotic variant. Contrastively, in L3 German, the less marked L2 rhotic influenced perception to a greater extent than the more marked L1 rhotic. In final obstruent devoicing, the accuracy was around or below the chance level, and it seems the more marked L2 variant has not been internalized by the learners at all. Therefore, in order to further disentangle the influence of language status from markedness of the tested feature, more studies that would use various combinations of markedness and language status are needed.

6. Conclusions

The overall results indicate that CLI in perceptual development is feature-dependent with relative stability evidenced for L2 rhotics, reverse trends for L3 rhotics, and no significant development for L2/L3 (de)voicing. We also found that perception accuracy of rhotics differed significantly with respect to stimulus properties, (i.e., whether they were L1-accented, L2-accented, or L3-accented) and that it took longer to process the perception task in the L3 than L2. On the whole, major findings include a nonlinear development of foreign language phonology, diverse CLI patterns that are feature-dependent, and differential learnability of phonetic features. We hope the present findings will be an incentive to extend current theoretical frameworks beyond L2 speech perception models to account for these phenomena in multilingual speech perception.

References

Anderson, Gregory. 2011. *Non-Linear Dynamics of Adult Non-Native Phoneme Acquisition Perception & Production*. Ph.D. thesis, George Mason University, Fairfax, VA, USA.

Baker, Wendy, Pavel Trofimowich, Molly Mack, and James E. Flege. 2002. The effect of perceived phonetic similarity on non-native sound learning by children and adults. In *Proceedings of the 26th Annual Boston University Conference on Language Development*. Edited by Barbora Skarabela, Sarah Fish and Anna H. -J. Do. Somerville: Cascadilla, pp. 36–47.

Balas, Anna, Romana Kopečková, and Magdalena Wrembel. 2019. Perception of rhotics by multilingual children. In *Proceedings of the 19th International Congress of Phonetic Sciences*. Edited by Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren. Melbourne: Australasian Speech Science and Technology Association, pp. 3725–29.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 171–204.

Best, Catherine T., and Michael D. Tyler. 2007. Non-native and second language speech perception: Commonalities and complementaries. In *Second Language Speech Learning*. Edited by Murray J. Munro and Ocke-Schwen Bohn. Amsterdam: John Benjamins Publishing, pp. 13–34.

Bradlow, Ann. R., Reiko Akahane-Yamada, David B. Pisoni, and Yoh'ichi Tohkura. 1999. Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception and Psychophysics* 61: 977–85. [CrossRef] [PubMed]

Broesma, Mirjam. 2005. Perception of familiar contrasts in unfamiliar positions. *Journal of the Acoustical Society of America* 117: 3890–901. [CrossRef] [PubMed]

Bundgaard-Nielsen, Rikke. L., Catherine T. Best, and Michael D. Tyler. 2011. Vocabulary size matters. The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32: 51–67. [CrossRef]

Cabrelli, Amaro Jennifer, and Magdalena Wrembel. 2016. Investigating the acquisition of phonology in a third language—A state of the science and an outlook for the future. *International Journal of Multilingualism* 13: 395–409. [CrossRef]

Cebrian, Juli. 2006. Experience and the use of no-native duration in L2 vowel categorization. *Journal of Phonetics* 34: 372–87. [CrossRef]

Cenoz, Jasone, and Ulrike Jessner. 2000. *English in Europe: The Acquisition of a Third Language*. Clevedon: Multilingual Matters.

Cenoz, Jasone, Britta Hufeisen, and Ulrike Jessner. 2001. *Cross-Linguistic Influence in Third Language Acquisition: Psycholinguistic Perspectives*. Clevedon: Multilingual Matters.

De Angelis, Gessica. 2007. *Third or Additional Language Acquisition*. Clevedon: Multilingual Matters.

Eckman, Fred. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27: 315–30. [CrossRef]

Escudero, Paola. 2005. Linguistic Perception and Second Language Acquisition. LOT Dissertation Series 113. Ph.D. thesis, Utrecht University, Utrecht, The Netherlands.

Escudero, Paola. 2009. Linguistic perception of 'similar' L2 sounds. In *Phonology in Perception*. Edited by Paul Boersma and Silke Hamann. Berlin: Mouton de Gruyter, pp. 151–90.

Escudero, Paola, and Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26: 551–85. [CrossRef]

Flege, James E. 1991. Orthographic evidence for the perceptual identification of vowels in Spanish and English. *The Quarterly Journal of Experimental Psychology* 43A: 701–31. [CrossRef] [PubMed]

Flege, James E. 1995. Second language speech learning: Theory, problems, findings. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Baltimore: York Press, pp. 233–77.

Flege, James E., and Ocke-Schwen Bohn. 2020. The revised Speech Learning Model. Available online: https://www.researchgate.net/publication/342923320_The_revised_Speech_Learning_Model (accessed on 19 August 2020).

Flynn, Suzanne, Clair Foley, and Inna Vinnitskaya. 2004. The Cumulative-Enhancement Model for language acquisition: Comparing adults' and children's patterns of development in L1, L2 and L3 acquisition of relative clauses. *The International Journal of Multilingualism* 1: 3–16. [CrossRef]

Gonet, Wiktor. 2001. Voicing control in English and Polish: A pedagogical perspective. *International Journal of English Studies* 1: 73–92.

Gut, Ulrike. 2010. Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism* 7: 19–38. [CrossRef]

Gut, Ulrike, Robert Fuchs, and Eva-Maria Wunder. 2015. *Universal or Diverse Paths to English Phonology*. Berlin: De Gruyter Mouton.

Hufeisen, Britta, and Beate Lindemann. 1997. *Tertiärsprachen, Theorien, Modelle, Methoden*. Tübingen: Staufenburg.

Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33: 103–7. [CrossRef]

Kohler, Klaus. 1999. German. In *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press, pp. 86–89.

Kopečková, Romana. 2012. Differences in L2 segmental perception: The effects of age and L2 experience. In *Intensive Exposure Experiences in Second Language Learning*. Edited by Carmen Muñoz. Clevedon: Multilingual Matters, pp. 234–55.

Krause, Sue Ellen. 1982. Developmental use of vowel duration as a cue to postvocalic stop consonant voicing. *Journal of Speech, Language and Hearing Research* 25: 388–93. [CrossRef] [PubMed]

Kroll, Judith. 2020. Learning and Using Two Languages May Change Your Mind. Available online: http://education.uci.edu/uploads/7/2/7/6/72769947/bb_kroll_20200420.pdf (accessed on 3 November 2020).

Ladefoged, Peter. 2001. *Vowels and Consonants: An Introduction to the Sounds of Languages*. Oxford: Blackwells.

Ladefoged, Peter, and Ian Maddieson. 1996. *The Sounds of the Worlds' Languages*. Oxford: Blackwell.

Lado, Robert. 1957. *Linguistics across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor: University of Michigan Press.

Nelson, Christina. 2020. The younger, the better? Speech perception development in adolescent vs. adult L3 learners. In *Young Linguists' Insights: From Exploration to Explanation in Psycho- and Sociolinguistic Eesearch*. Edited by Katarzyna Jankowiak and Joanna Pawelczyk. Special Issue, Yearbook of the Poznań Linguistic Meeting. vol. 6, pp. 27–58.

Rafat, Yasaman. 2011. Orthography-Induced Transfer in the Production of Novice Adult English-speaking Learners of Spanish. Ph.D. thesis, University of Toronto, Toronto, ON, Canada.

Rallo Fabra, L., and Gallero J. Romero. 2012. Native Catalan learners' perception and production of English vowels. *Journal of Phonetics* 40: 491–508. [CrossRef]

Rose, Marda. 2010. Differences in discriminating L2 consonants: A comparison of Spanish taps and trills. In *Selected Proceedings of the 2008 Second Language Research Forum*. Edited by Matthew T. Prior, Yukiko Watanabe and Sang-Ki Lee. Somerville: Cascadilla Proceedings Project, pp. 181–96.

Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: The Structure of Polish*. Dordrecht: Foris.

Slowiaczek, Louisa M., and Daniel A. Dinnsen. 1985. On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics* 13: 325–41. [CrossRef]

Smith, Bruce L., Rachel Hayes-Harb, Michael Bruss, and Amy Harker. 2009. Production and perception of voicing and devoicing in similar German and English word pairs by native speakers of German. *Journal of Phonetics* 37: 257–75. [CrossRef]

Sorace, Antonella. 2020. L1 attrition in a wider perspective. *Second Language Research*. [CrossRef]

Strange, Winifred. 2009. Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics* 39: 456–66. [CrossRef]

Tyler, Michael. 2019. PAM-L2 and phonological category acquisition in the foreign language classroom. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*. Edited by Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen and Mette Hjortshøj Sørensen. Denmark: Department of English, School of Communication & Culture, Aarhus University, pp. 607–30.

Wrembel, Magdalena. 2015. *In Search of a New Perspective: Cross-Linguistic Influence in the Acquisition of Third Language Phonology*. Poznań: Wydawnictwo Naukowe UAM.

Wrembel, Magdalena, and Jennifer Amaro Cabrelli, eds. 2018. *Advances in the Investigation of L3 Phonological Acquisition*. London: Routledge.

Wrembel, Magdalena, Marta Marecka, and Romana Kopečková. 2019. Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism.* 16: 513–33. [CrossRef]

Wunder, Eva-Maria. 2014. On the Hunt for Lateral Phonological Crosslinguistic Influence in Third or Additional Language Acquisition. Ph.D. thesis, University of Münster, Münster, Germany.