



HAL
open science

The Role of Delayed Output on Second/Foreign Language Pronunciation in Children

Hana Ehbara, Martha Young-Scholten, Jalal Al-Tamimi

► **To cite this version:**

Hana Ehbara, Martha Young-Scholten, Jalal Al-Tamimi. The Role of Delayed Output on Second/Foreign Language Pronunciation in Children. *Advances in Early Childhood and K-12 Education*, 2021, Teaching Practices and Equitable Learning in Children's Language Education, pp.23 - 44. 10.4018/978-1-7998-6487-5.ch002 . hal-03729034

HAL Id: hal-03729034

<https://hal.science/hal-03729034>

Submitted on 1 Aug 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Role of Delayed Output on Second/Foreign Language Pronunciation in Children

Hana Ehbara

Newcastle University, UK

Martha Young-Scholten

Newcastle University, UK

Jalal Al-Tamimi

<https://orcid.org/0000-0002-0926-6660>

Newcastle University, UK

INTRODUCTION

Understanding L2 speech has been a pressing issue for researchers and educationalists. Influences include age of initial exposure, the learner's first language (L1) and various universals (Colantoni & Steele, 2008).

The literature on L2 phonology has focused to a great extent on the age of acquisition factor in phonological acquisition considering it the most crucial factor after L1 influence (Ioup, 2008); see Lenneberg (1967) on the critical period. Findings indicate that an earlier age of acquisition correlates with more successful second/ foreign language acquisition. However recent findings (e.g. Pfenninger, 2016), suggest that age is less important than thought, and that starting an L2 before puberty in a foreign language classroom does not warrant better L2 phonology. The age factor – in populations where learners reside in the L2 country – is confounded by the fact that younger learners often have better opportunities for exposure to the L2 from native speakers (Mack, 2003; Young-Scholten, 1995).

In the Output Hypotheses, Swain (1985, 1995, 2005) and Mackey (2007) argue that production is the tool for prompting the cognitive processes that enable learners to notice differences between theirs and others' production to create novel linguistic knowledge. Feedback through speaking practice is argued to allow the learner to reflect on his/her performance. This then encourages the learner to revise and correct his/her production in a continuous process to match the target language. This process is ongoing until the learner feels that his/her production matches the model input. Colantoni and Steele (2008) specifically propose a model for L2 phonology under which learners continue to modify their phonological system in the L2 comparing their output to a model input. Krashen (1985) argues that comprehensible input alone, without production is sufficient. For him "when acquirers are forced to produce language that they have not yet acquired, known as "forced speech" [...] "makes no direct contribution to language acquisition" (Krashen, 2018, para 1). There has been little research on delayed oral production.

Moreover, oral production training has rarely been investigated in L2 experimentation (e.g. Kartushina et al., 2016; Kartushina & Martin, 2019) and even more rarely in children or by comparing classroom instruction with online instruction. This chapter accordingly reports on a study of children learning English as a foreign language which examined input type (including computer-assisted instruction) and delayed production.

BACKGROUND

Children's acquisition of phonemes shows some similarities across languages among other things.¹ For example /ð/ and /ɹ/ are late acquired in some languages and are considered 'typologically marked' (McCleod & Crowe, 2018; Ohala, 2008). That universals are at play has also been taken up by in L2 phonological acquisition for example by Eckman ((Eckman, 1977) in his Markedness Differential Hypothesis. L2 sounds and sound structures absent from the learner's L1 and which are typologically marked are predicted to be more difficult, that is, a "phenomenon A in some language is more marked than B if the presence of A in a language implies the presence of B; but the presence of B does not imply the presence of A." (Eckman, 1977, p. 320).

Eckman (2008) argues for universal tendencies as an explanation for the observation that L2 development can resemble L1 learners' development. For syllables, Kløve and Young-Scholten (2008) conclude that L2 learners simplify consonant clusters not only to conform with their L1 syllable structures, but also with language universals. There are age differences: children favour deletion of consonants when fronted with clusters (Pater & Barlow, 2003) but adults favour insertion of vowels to break up clusters (Broselow & Finer, 1991; Fantazi, 2003; Hancin-Bhatt & Bhatt, 1997; Karimi, 1987; Kwon, 2006; Major, 1994).² Here markedness also plays a role in the form of the Sonority Hierarchy which governs how consonants in clusters arrange themselves (Broselow & Finer, 1991).

When it comes to training in pronunciation, Piske (2012) discerns several factors which influence acquisition and along with L1-L2 differences and quality and quantity of the input (see also Young-Scholten 1995), they include training tasks in perception and production which target structural differences between the learner's L1 and L2.

Few studies have explored the effect of training on production of consonants (Bradlow et al., 1997) or vowels (Thomson, 2011). While studies have examined the effect of various types of training, none have compared the effect of delayed production. Studies have indeed found that training in perception and production has positive effects and a number of experimental studies have used computer-aided pronunciation training software using native speaker input after which learners were asked to imitate the input (Kissling, 2013; Lord, 2005; Pearson et al., 2011; Weinberg & Knoerr, 2003). In such studies imitations were recorded to allow learners to compare their productions against native speaker models. This technique was also used to teach suprasegmentals (Chen et al., 2019; Chun et al., 2013; Hardison, 2004) and global speech aspects (Tanner & Landon, 2009). In other computer-assisted pronunciation training studies, training involved automatic speech recognition feedback (e.g. Neri et al., 2008).

THE INFLUENCE OF PRODUCTION

To output or not to output

Colantoni and Steele (2008, pp. 522–523) incorporated aspects from principles of L2 speech learning, aspects of perception models, and principles from phonetic production research into hybrid schemata of segmental acquisition. They argue that even if the learner has target-like representations and accurate motor planning, target-like productions cannot be guaranteed because articulation is shaped by L1 articulatory patterns as well as universal articulatory constraints. They put forward that a learner's output is either more or less accented and feeds back into his/her perception alongside other native input. This allows the learner to continue to finetune his/her productions and the mental category continues to evolve accordingly. If feedback from the learner's production is perceived as being no different from the model input, the categories become fossilised and thus will no longer evolve. In this case, it is said to have reached an end state. Krashen (1982) argues that the pressure to perform results in premature use of L2. Even though Krashen did not specify this for L2 pronunciation, it seems to be compatible with the role of the learner's output in his/her perception and phonological learning. If a learner does not practise speaking, there will essentially be no output and, thus, he/she will not have the opportunity to modify his/her productions to match the input.

To evaluate the role of production training and that of input, the study was designed in the following steps:

Step 1: The questions this study sought to answer were whether and, if so, participants' pronunciation varies as a function of 1) presentation method, that is perception-only practice vs. perception-and-production practice, and 2) input type, that is native English vs. Arabic-accented.

Step 2: Arabic was the L1 spoken by the participants in the study. The many differences between the phonology of the variety spoken in Misrata, Libya, and the Received Pronunciation English to which learners are exposed provide an ideal ground for investigating language learning. Table 1 shows the L1 inventory of 25 consonants excluding the voiced labial-velar approximant /w/ (Ehbara, forthcoming) and the target language (TL) additional sounds (boldfaced) excluding affricates. Consonants found in RP English but not Misrata-Libyan Arabic are the voiceless bilabial plosive /p/, affricates /tʃ, dʒ/, dental fricatives /θ, ð/,³ the voiced labiodental fricative /v/, the velar nasal /ŋ/ and the rhotic approximant /ɹ/.

Table 1. Consonant inventories of Misrata Libyan Arabic (Ehbara, forthcoming) and English

	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Uvular	Pharyngeal	glottal
Plosive	p b			t d			k g			ʔ
(Emphatic)				t ^ɕ d ^ɕ						
Nasal	m			n			ŋ			
Tap/trill				r						
Fricative		f v	θ ð	s z	ʃ ʒ			x ɣ	ħ ʕ	h
(Emphatic)				s ^ɕ						
Approximant				ɹ		j				
Lateral approximant				l						

Additionally, the r-sound in the L1 is a tap everywhere except when geminated. In RP English, the r-sound is a rhotic approximant word-initially and medially but is not realised post-vocally.

There are two L1 diphthongs: word-final [aw] and [aj] compared to RP English's /ɪə, eə, uə, eɪ, aɪ, oɪ, əʊ, aʊ/. The L1 also has a less complex syllable structure C(C)V(CC) than English: (CCC)V(CCCC). There are also phonotactic restrictions on consonant allowed in clusters (e.g. Roach, 2009).

Step 3: It is hypothesized that participants' pronunciation will vary as a function of 1) presentation method, that is perception-only-practice vs. perception-and-production practice, and 2) input type, that is native vs. Arabic-accented.

Step 4: Experiment:

- **Participants:**

The choice of children – and not adults – is twofold. First, seven-year-old children are prepubescent, which minimises the potential effects of the age factor. Second, the chances of finding child participants with no prior instruction in English are higher than those for adults, whose aspects of L2 language can potentially be fossilised.

The participants were 58 Libyan-Arabic children aged between 6;11 and 8;0 (+2 months in the delayed post-test). All were reported to have had no prior instruction in English before the training. Second language researchers (e.g. Flege, 1987) emphasise the significance of having homogeneous participant groups owing to the inherent variability of speech in general and L2 speech learning in particular and all were born and had been raised in Libya. Because gender could be a potential variable, both boys and girls were included as equally as possible in each condition. The participants were then divided into the three condition groups who were randomly placed into these, shown in Table 2. The Traditional Teaching

participants were drawn from the school where the Traditional teacher worked. They were all pupils from the same class attending year two in that school.

Table 2. Number of participants per condition, time of test, and gender

Post-test	Males	Females	Delayed post-test	Males	Females
Listen and Speak	9	11	Listen and Speak	7	7
Listen Only	10	8	Listen Only	9	8
Traditional	11	9	Traditional	9	9

- **Procedure:**

Deciding on amount of L2 input was informed by studies, although amount varies drastically. Duration of training in computer-assisted pronunciation training studies is relatively shorter than other interventions (Thomson & Derwing, 2014): one computer-assisted pronunciation study lasted 20 minutes (Guilloteau, 1997) and some classroom studies last many hours, e.g. 70 (Parlak, 2010). Studies do not always use control groups to separate the effect of input amount from input type.

It was decided to deliver the training for three weeks, five hours per week. For Listen and Speak, each word was allocated a three-minute listening practice and one minute of speaking practice.⁴ For Listen Only, each word was allocated an additional listening minute instead of the minute speaking practice. This was to eliminate the impact of differences in time allocated for each word between the Listen and Speak and Listen Only conditions. This yielded a total of four minutes of listening practice for each item, but children were not required to produce the words. The Listen and Speak and Listen Only used native speaker input provided by a programme called the Digital Literacy Instructor (DigLin) (Overall, 2014). For the Traditional condition, teaching was carried out by a typical ‘traditional’ Libyan teacher of English. The teacher spoke Libyan-accented English, which is common in Libya and was instructed to provide a three-week course to the children and include the test words and the same pictures as the experimental conditions within her teaching. The teacher was free to include, for example, grammar tasks or alphabet drills, as are typical of traditional classroom teaching but not to use any electronic devices as traditional classrooms never use them.

- **Design:**

The productions of the target words were compared using the same tasks immediately after the training (post-test) and 10 weeks later (delayed post-test) and across the three conditions. Because the children had no instruction in English before the training a pre-test was not possible (English is introduced in the national curriculum when pupils are between nine and ten years old). To eliminate the possibility of exposure to English, it was confirmed that none had had any by asking their parents.

- **Testing:**

Three tasks were used to elicit data: picture-naming, read aloud and delayed repetition. Stimuli were presented individually on PowerPoint slides and randomised to elicit two tokens from both tasks yielding two tokens per test word, with a total of 150 expected tokens. The stimuli comprised monosyllabic words containing sounds problematic for Libyan-Arabic speakers (see above) (Ehbara, forthcoming). Items were grouped on the basis that the target English and the L2 English counterpart as produced by Libyan adult speakers are not the same phoneme. In the L1, English /ɪ/ is typically produced [ɪ], /θ/ as [t], /tʃ/ and /dʒ/ as /f/ and /z/ respectively. As with diphthongs, they are monophthongal and coda clusters usually exhibit epenthesis. Thus, impressionistic transcription should be able to reveal differences in production. However, for phonemes transcribed the same across the two languages, such as plosives, transcription was not sufficient and further acoustic analyses of VOT was deemed necessary. Arabic data (see Ehbara (forthcoming) for stimuli items) were also elicited from the participants to allow comparison with English

production using a direct comparison of the results across languages. Thus, the stimuli were subdivided into Arabic and English, composed of monosyllabic words consisting of a bilabial or coronal plosive syllable-initially followed by each of the high front vowel /i:/, high back vowel /u:/ and low front vowel /a/.

Five seconds were allowed for each slide. Visual stimuli (photographs) for the delayed post-test were different from those for the post-test and the training visual materials to avoid task familiarity.

In the delayed repetition task, an intervening phrase was used to prompt production. Flege et al. (1995a) note that the intervening phrase reduces reliance on short-term memory and merely imitating. In this task, participants heard an audio clip of each item followed by an intervening phrase in Arabic /*taw:a ʕawədha mar:a ta:nja/* ‘now repeat it one more time’. This task yielded the vast majority of tokens. Because some children in this sample had not yet acquired literacy in Arabic to read the words, a similar delayed repetition task was used to elicit the Arabic data. The data for each speaker were collected in a single session lasting approximately between 20–30 minutes. The procedure carried out to elicit data during the delayed post-test was the same as that in the post-test.

- **Instrumentation**

To answer the research questions, every token was assigned a label ‘target-like’ if its transcription holistically matched the IPA transcription for the item, or a ‘non-target-like’ if it did not, irrespective of the type of input received by the treatment condition with a few exceptions. Mismatches in final voicing were not considered errors because final devoicing in L1 is well-documented in children’s speech (Jakobson, 1968; Stoel-Gammon, 1987; Templin, 1957). Additionally, /l/ vocalisation is not considered an error either because /l/ vocalisation is very common in children learning languages that employ velarised post-vocalic /l/ (Grunwell, 1985; Hodson & Paden, 1981; Stoel-Gammon & Dunn, 1985). When a speaker had an overall lateralisation and/or a lisp in all words including Arabic, this feature was not considered in the decision whether the production was ‘non-target-like’. These points aside, for a token to be rated target-like, the segments must match for place and manner of articulation as well as voicing in word-initial and word-medial positions.

Step 5: For each task, each participant was presented with 75 test items, as written words in the read aloud task, pictures in the picture-naming task, and audio clip of the word (from DigLin (Overall, 2014)) in a carrier phrase. This culminates in 225 English stimuli per participant per test; 13050 target words in the post-test (58 participants) and 10800 in the delayed post-test (48 participants).

In the delayed repetition learners could *generally* produce what they had heard from the audio prompt. This differed in memory recall tasks – read aloud and picture-naming – because learners could only produce the words they learned, and not all had learned all the test words, as shown in Table 3; 15 hours were insufficient for seven-year-olds to learn all 75 words.

Table 3. Mean number of words produced stratified by test, group and task.

Test	Training	Task	Mean out of 75
Post-test	Listen and Speak	Picture-naming	3.85
		Read aloud	2.00
		Delayed repetition	53.60
	Listen Only	Picture-naming	1.83
		Read aloud	1.00
		Delayed repetition	44.30
	Traditional	Picture-naming	8.35
		Read aloud	5.00
		Delayed repetition	24.30
Delayed post-test	Listen and Speak	Picture-naming	2.00
		Read aloud	5.00

	Listen Only	Delayed repetition	54.50
		Picture-naming	1.62
		Read aloud	1.00
	Traditional	Delayed repetition	48.10
		Picture-naming	7.13
		Read aloud	1.00
		Delayed repetition	24.90

The following patterns can be observed here: there is a considerable difference in words produced during the delayed repetition task and those of the memory recall tasks and the effect of the test interacts with both task and training condition and this diverges more in the delayed post-test. This also varies with training condition, increasing slightly for the Listen and Speak condition, remaining the same for Listen Only and decreasing for the Traditional condition. Traditional teaching, however, shows the highest rate in the picture-naming task whereas Listen Only the lowest. In the delayed repetition task, Traditional teaching had the lowest rate of fully produced words. For the two memory recall tasks, picture-naming task yielded consistently a marginally higher rate of words than the reading aloud.

The following sections show results for the two phases of the analyses; target-likeness rating scores and voice onset time durations.

Phase One: Target-likeness results

Turning to the research question, whether training condition had an effect on the phonology of what the children produced, Table 4 shows the percentages of average target rating scores by problematic sound class and by task for each training condition in both tests.

Table 4. Average target rating score per test, training condition, problematic sound class and task

	Post-test			Delayed post-test		
	Listen and Speak					
	Picture-naming	Read aloud	Delayed repetition	Picture-naming	Read aloud	Delayed repetition
Affricates	43%	0%	80%	0%		75%
Plosives	70%	56%	62%	29%	67%	55%
Diphthongs	50%		41%	100%	0%	43%
Clusters	22%	40%	30%			33%
Fricatives	12%		28%	0%		28%
Rhotics	0%		20%	0%	100%	19%
	Listen Only					
	Picture-naming	Read aloud	Delayed repetition	Picture-naming	Read aloud	Delayed repetition
Affricates	33%		62%	0%		58%
Plosives	56%	50%	54%	20%	33%	44%
Diphthongs	0%	0%	36%	0%		42%
Clusters	0%		23%	0%		15%
Fricatives	0%		22%	0%		17%
Rhotics	0%		12%	25%		9%
	Traditional					
	Picture-naming	Read aloud	Delayed repetition	Picture-naming	Read aloud	Delayed repetition
Affricates	24%	0%	46%	19%		46%
Plosives	21%	0%	39%	27%	50%	34%
Diphthongs	0%	0%	12%	0%		19%
Clusters	14%	33%	18%	29%		19%

Fricatives	0%	0%	18%	0%		22%
Rhotics	0%	0%	02%	0%		3%

The figures show that target rating scores varied between training conditions in terms of averages and rank. Cross-problematic sounds averages varied by training condition, test and task. For the delayed repetition task, Listen and Speak exhibited the highest average target rating scores during the post-test, followed by Listen Only. Traditional training condition had the lowest.

In the delayed test, the Traditional condition target ratings for coda clusters (19%) and dental fricatives (22%) exceeded those for the Listen Only training condition (15%, 17% respectively). Additionally, the rank of difficulty for the Listen Only and Traditional training condition changed. For the Traditional training condition, in the post-test, the highest score was evident in affricates followed by plosives, clusters and dental fricatives (same score), diphthongs and finally rhotic approximants had the lowest score. In the delayed test, this changed for clusters. Similarly, the ranking of clusters and dental fricatives for the Listen Only training condition varied by test.

For the picture-naming task, scores varied considerably. The highest average target rating score for Listen and Speak during the post-test was plosives (70%). In the delayed test it was diphthongs (100%). As for Listen Only training condition, the highest score was also evident in the class of plosives (56%) during the post-test. In the delayed test, it was rhotic approximants (25%). The highest average target rating score for Traditional training condition during the post-test was affricates (24%). In the delayed test, it was clusters (29%).

Materials from the read aloud task were scarce and figures did not represent scores very accurately as in the other elicitation materials. The Listen and Speak condition yielded tokens containing affricates, plosives and clusters in the post-test and plosives, diphthongs, and rhotic approximants in the delayed test. Listen Only training condition yielded tokens containing plosives and diphthongs in the post-test and only plosives in the delayed test. Traditional training condition on the other hand yielded tokens containing all of the sound classes in the post-test and only plosives in the delayed post-test.

For Listen and Speak training condition, plosives received the highest average target rating score in the read aloud task during the post-test (56%). In the delayed test, the class of rhotic approximants received the highest average target rating score (100%) within that group. Listen Only training condition exhibited a score of 50% for plosives in the post-test, whereas in the delayed test, this score declined to 33%. Traditional training condition exhibited the score of 33% for plosives in the post-test, which increased during the delayed test to 50%. It is clear from the above the interaction of the effects of training condition, test, sound class and task.

The data showed an interaction between gender and literacy – measured based on the ability to read the Arabic stimuli – that was consistent regardless of the training condition. Within literates, males performed better than their female counterparts. Within illiterates, females outperformed their male peers. For the Traditional Teaching condition, such comparisons were not possible as it did not have illiterate females.

Based on these observations, a generalised linear mixed effects model was built to capture these interactions in order to yield more accurate predictions. To this end, the following model was built:

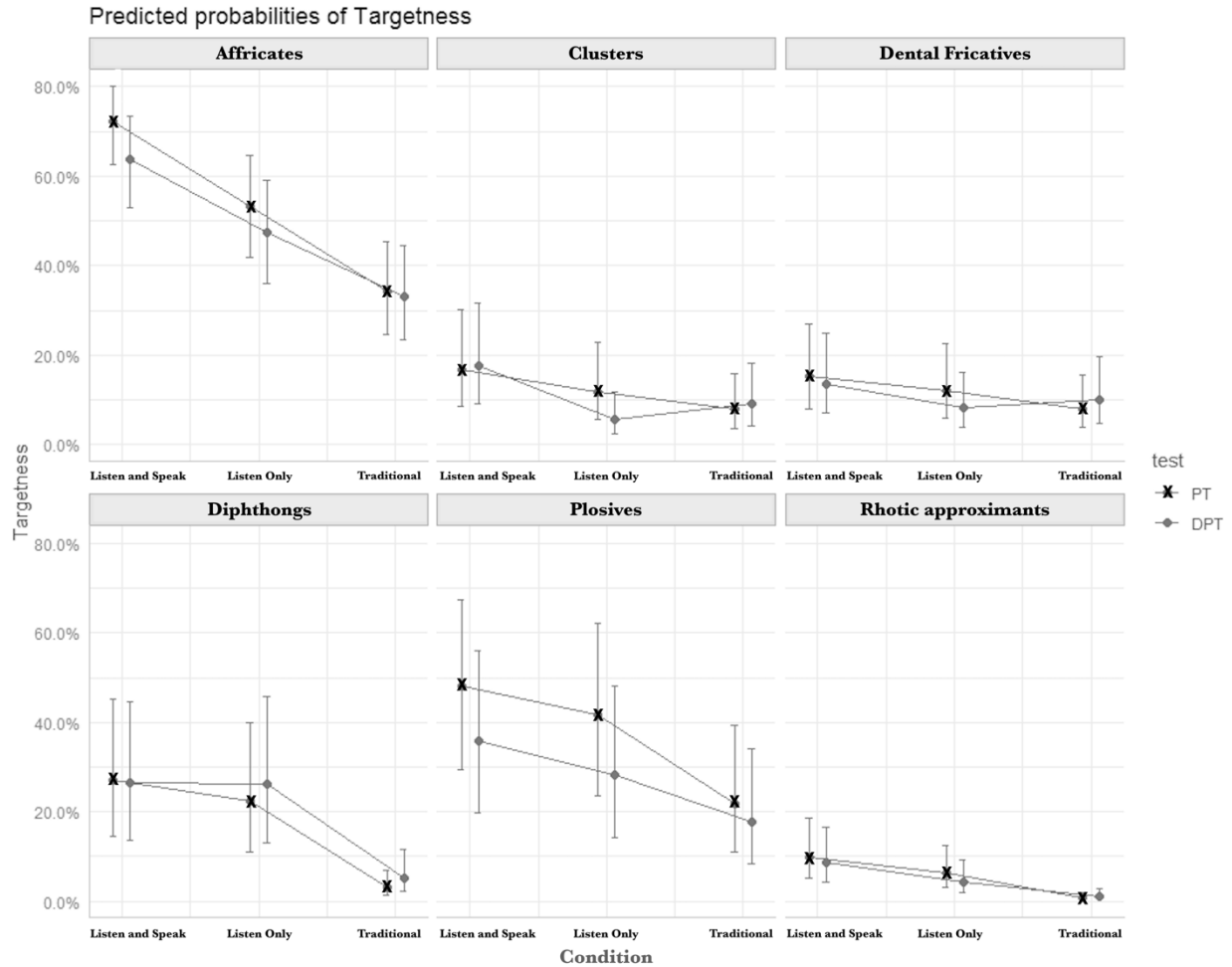
```
Target ~ Condition*Test*Sound + Task + Literacy*Gender + Age + (1 + Test | speaker) + (1 | item)
```

The output of the model includes coefficients for the simple effects and two-by-two and three-by-three interaction terms. Therefore, differences between these levels are examined using post-hoc tests to extract model-predicted probabilities of target-likeness rating for each training condition, in each test by sound class. The function `emmeans` (Lenth, 2019) in R was applied and the following code was used:

```
emmeans mdl, pairwise ~ Condition * Test | Sound, type = 'response')
```


Figure 1 shows model-predicted probabilities of target-likeness grouped by sound class, training condition, and time of test. The red line depicts predicted percentage rating during the post-test and the blue line predicts that in the delayed post-test. Affricates were the least challenging sound class, followed by plosives, and then diphthongs. The most challenging sound class was rhotic approximants especially for the Traditional Teaching condition. The figure also shows that overall, the Listen and Speak condition outperformed the other two training conditions across almost all the sound classes in both times of testing (an exception being in the class of diphthongs in the delayed post-test).

Figure 1. Model-predicted probabilities of target-likeness rating



The post-hoc within-test pairwise comparisons by training condition show that in the post-test, although the Listen and Speak condition outperformed its Listen Only counterpart, this difference was only statistically significant for the group of affricates ($p < .03$). The Listen and Speak condition also outperformed its Traditional counterpart and this difference was statistically significant for the group of affricates ($p < .0001$), coda clusters ($p = 0.0157$), diphthongs ($p < .0001$), plosives ($p < .0001$), and rhotic approximants ($p < .0001$), but not for dental fricatives. Finally, the Listen Only condition outperformed its Traditional counterpart during the post-test and the difference in the predicted probability of target-like rating between them was statistically significant for the group of affricates ($p = 0.015$), diphthongs ($p < .0001$), plosives ($p = 0.003$), and rhotic approximants ($p > .001$), but not for coda clusters, or dental fricatives.

Within the delayed post-test, the Listen and Speak condition once again outperformed its Listen Only counterpart – an exception was the class of diphthongs where their predicted probabilities of target-likeness rating were matched. However, this difference was statistically significant for the group of coda clusters only ($p > .001$). The Listen and Speak condition also outperformed its Traditional counterpart and the difference was statistically significant for the group of affricates ($p = 0.0001$), diphthongs ($p < .0001$), plosives ($p = 0.0083$), and rhotic approximants ($p = 0.0001$), but not for coda clusters, or dental fricatives. Finally, the Listen Only condition was outperformed by the Traditional condition in the delayed post-test within the groups of coda clusters and dental fricatives. However, this decline in performance did not illustrate a significant difference compared to the Traditional condition. The difference in the predicted probability of target-likeness rating was statistically significant between those latter conditions for the group of diphthongs only ($p < .0001$), where the Listen Only condition outperformed its Traditional counterpart.

Some common phonological processes observed in the data were mainly substitution, including deaffrication in affricates and th-fronting in dental fricatives, and cluster reduction. In plosives, the main process observed was voicing. In diphthongs, the most common process was monophthonging mainly with /əʊ, eɪ/ for all groups and /ɪə/ for the Traditional condition only. The case for the latter group reflects their traditional teacher’s pronunciation.

Phase Two: VOT results

VOT measurements for L2 English /p, b, t, d/ were also extracted using Praat (Boersma & Weeninck, 2016) and compared to L1 Arabic /b, t, d/ as well as English /p, b, t, d/ from each training conditions’ respective input.

Table 5 shows mean VOT durations by plosive for the target input for the experimental groups and the Traditional condition. The traditional teacher clearly shows non-English VOT durations.

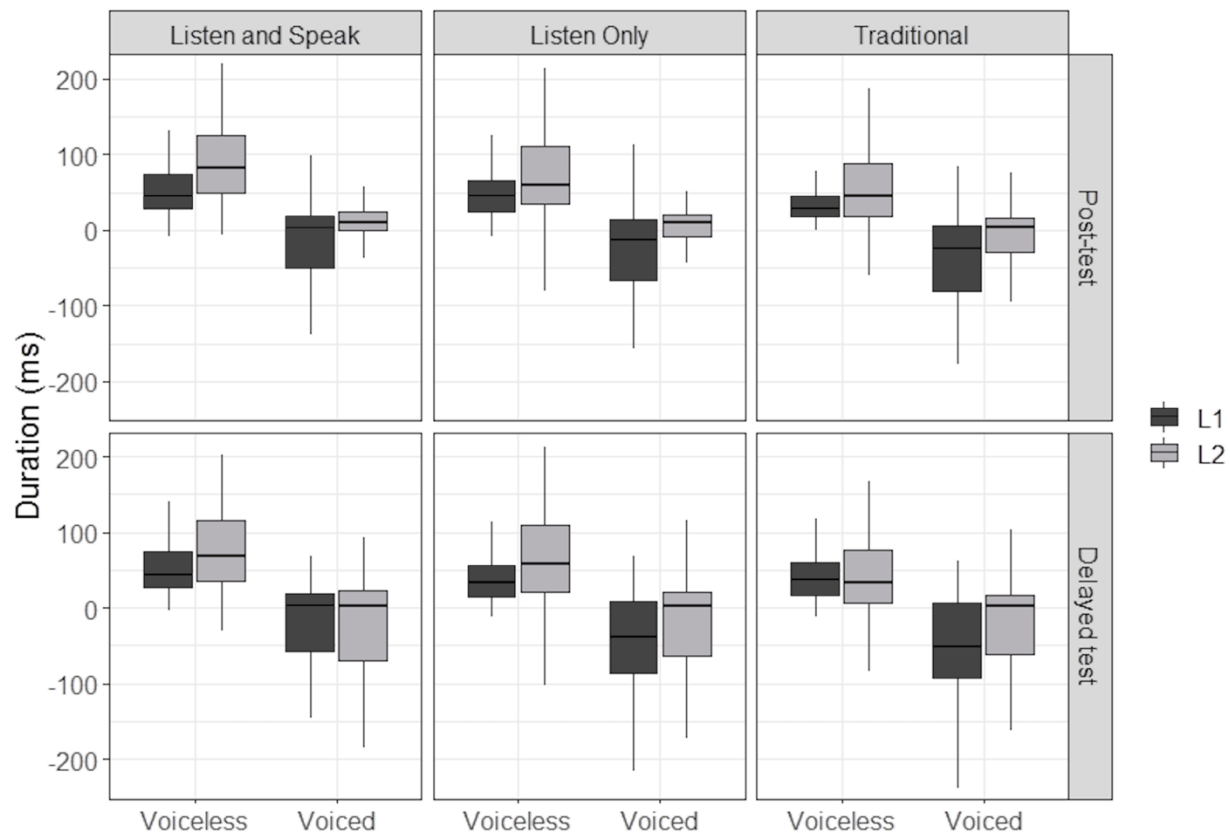
		DigLin	Teacher
/p/	(+)	103	27
/b/	(+)	6	3
	(-)		-76
/t/	(+)	125	35
/d/	(+)	19	
	(-)		-65

Figure 2 shows VOT durations for L1 and L2 as produced by the participants. Although overall L2 VOT values are higher than those for L1 (exceptions are voiced categories in Listen and Speak voiceless categories in the Traditional both during the delayed post-test), they do not seem to vary considerably from one another. It is also noticed that the L1 VOT durations for the experimental groups, and more notably for Listen and Speak, are longer on the positive side and shorter on the negative side.

Table 5. Mean voice onset times (ms)

		DigLin	Teacher
/p/	(+)	103	27
/b/	(+)	6	3
	(-)		-76
/t/	(+)	125	35
/d/	(+)	19	
	(-)		-65

Figure 2. VOT durations (ms). The line inside the boxplot represents the mean



Linear mixed effects models were used to compare the measurements in the three language groups; L1, L2 and target language TL to examine whether the learners in each training condition established L2 VOT categories *statistically* independent of their L1. It was not possible to include the three training conditions in a single model given the Traditional training received Arabic-accented English compared to the experimental conditions. To control for potential speaker-related effects such as gender and literacy, as well as VOT-related effects such as place of articulation, vowel context, and voice (voiceless plosives were pooled together and so were voiced ones), these were incorporated in the linear mixed effects model. The model added by-speaker and by-item random intercepts and by-speaker random slopes for voice, place, and vowel context. Adding the effect of test was not possible for the TL data. Therefore, data were examined for each test individually. The following model was built:

```
Duration ~ Language*Voice + Gender*Literacy + Place + Vowel + (1 + Voice + Vowel + Place | speaker) + (1 | item)
```

The outputs of the models include coefficients for simple effects and interaction terms. Therefore, differences between these levels are examined using the post-hoc tests `emmeans` using the following code:

```
emmeans mdl, pairwise ~ Language * Voice)
```

The results show that overall, none of the training conditions managed to establish L2 categories that are statistically independent of either their L1 counterparts or the TL categories for either within-voiceless or within-voiced plosives. An exception is the Listen Only condition whose L1 VOT category for voiceless plosives (/t, d/ values pooled together) was statistically ($p = 0.045$) 75 ms shorter than the TL counterpart

in the post-test but not in the delayed post-test. This does not mean that learning did not take place as the L2 categories were mostly indistinguishable from the TL categories. Moreover, they have all successfully managed to statistically distinguish their voiced and voiceless categories including bilabial plosives. This is surprising given /p/ is absent in Libyan Arabic and native English speakers find it difficult to identify it when produced by Arabic speakers (Flege & Port, 1981).

Step 6: Findings

Hypothesis One: The hypothesis was only partly confirmed for target-likeness rating results. Whilst the Listen and Speak condition outperformed the two other training conditions, this difference did not reach statistical significance consistently across all the problematic sounds in both tests. This is because not all sounds are equal in their degree of difficulty in production (Diehl & Lindblom, 2004) even within the same training type.

The difference between the Listen and Speak and Listen Only condition was only statistically meaningful in affricates in the post-test and coda clusters in the delayed post-test. The Listen Only did not exceed the Listen and Speak in any sound class despite the two training conditions receiving the same type of input. These results reveal that overall, delayed speaking (Listen Only) does not have an advantage over *pressure* to speak (Listen and Speak) when it comes to target-like realisations for this amount of training. Although in many cases the experimental conditions were not statistically different, overall oral production practice does have an advantage over delayed production either. Furthermore, the Listen Only condition, despite receiving native input, was outperformed by the Traditional condition, who received Arabic-accented input, in coda clusters and dental fricatives over time. It is suggested that this might also be due to an interaction between input type (native vs. accented) and training type (output vs. no output practice). The only sound class whereby no statistical difference was observed between any pair of conditions in either time of testing was dental fricatives.

Studies of child language acquisition indicate that rhotics and dental fricatives are among the *late* acquired sounds, that is acquired after the age of five (McLeod & Crowe, 2018). This is due to their relative markedness based on cross-linguistic typology compared with the other sound classes involved in this chapter. According to Shriberg (1993) native English children do not fully master the production of the rhotic approximants until the age of eight and seem to be one of the most problematic sounds for them (Shriberg & Kwiatkowski, 1994; Smit et al., 1990). For British-English children, in particular, the rhotic approximant /ɹ/ is mastered by the age of 6;0 to 6;5 and the dental fricatives are mastered *after* the age of 7;0 (Dodd et al., 2003, 2013). McLeod and Crowe (2018) further reanalysed the data of English-speaking children from various studies (including Dodd et al., 2003, 2013; Shriberg, 1993) and applied 75-85% percentage of consonant correct (PCC) criteria from nine studies and 90-100% PCC from eight studies. Their findings indicate that these sounds are acquired between the age of 5;0 and 7;0 with the voiceless dental fricative being, on average, the last-acquired consonant. The th-fronting process experimental conditions exhibited for this sound class is typical of English-speaking children. The Traditional teaching learners exhibited stopping instead, a process also found in Arabic adult L2 learners of English. This is due to the accented input adults typically receive as is the case for the Traditional teacher's production. In Educated Spoken Arabic, (inter)dental fricatives were also among the late sounds to be acquired (Amayreh, 2003). Amayreh (2003) provides two explanations for this: one is the lack of input from Educated Spoken Arabic given that this sound is introduced in school when children are six years of age. However, he does not rule out the potential of Arabic children being exposed to the sound from sources other than school. In schools, it is highly likely that teachers also use colloquial variants [t] and [d] for /θ/ and /ð/ respectively instead of Modern Standard Arabic reflecting substitution patterns found in the children's data. The other reason he suggests is the relative markedness of (inter)dental fricatives based on cross-linguistic typology.

The Traditional teaching learners' pronunciation showed a general improvement in the delayed post-test. This seems to result from the input they received from the native audio clips during the delayed repetition task during both times of testing.

The merger VOT categories are also an indicator of learning even for the Traditional condition whose members received very minimal native input during both times of testing and further supports backward transfer (Kartushina et al., 2016).

Finally, the findings on word learning implicate that a 15 hour course for learning 75 novel words was insufficient. Children in the three conditions failed to retrieve the majority of words especially during the memory recall tasks.

SOLUTIONS AND RECOMMENDATIONS

The findings implicate that a 15 hour course for learning 75 novel words was insufficient. It is well-known that hours dedicated to L2 teaching in school classrooms is insufficient for intensive language training. The use of computer-assisted pronunciation training software can help with this by allowing learners to work independently at their own pace. Moreover, second language teachers should focus on the importance of oral practice and supplying lessons with native speaker input. Interlanguage phonological processes are to be expected. They are evident in English-speaking children and thus, do not necessarily reflect a lack of learning.

FUTURE RESEARCH DIRECTIONS

Whilst the examination of oral production was appropriate, it is worthwhile to examine perception in future studies. This is to ascertain whether the participants' performance in perception tasks would match that of their production. Further research is also required to examine traditional foreign-accented teachers' productions during the classroom-based instruction to capture the actual input learners receive during training. Further research is also needed to examine the long-term training effect on perception and production. Moreover, further research is required to determine the amount of training child learners require to develop literacy decoding and to learn test words efficiently. Moreover, research on the acquisition of L2 phonology by children are relatively rare compared to other domains of L2 acquisition. The ones that do exist rarely cover lexical learning in any depth. There are few studies of children's early learning of vocabulary in the classroom (an exception is (an exception is Heimbach, 1994 for example). This should not be the case, so it needs to be pursued.

CONCLUSION

Overall, results for the segmental and lexical learning domain show a disadvantage for the delayed production practice (Listen Only) in comparison to the Listen and Speak training. This indicates the importance of output practice in second language speech learning contrary to Krashen's claim that output does not contribute to L2 learning. Results also show an advantage for native input. Younger learners do not automatically develop accent-free pronunciation. Results from phonetic learning (VOT) show that neither of the training conditions created statistically independent L2 phonetic categories. Instead, all of the three conditions created merged L2 categories that reflect L1 and TL values. Despite the lack of data from learners prior to the training and the relatively short period of training, it is argued that the experimental learners' L1 are altered as a result of the merger category supporting the findings by Flege (1995) and MacKay et al. (2001) on Italian-English bilinguals. They argue that the effect of one language on the other is bidirectional. The effect of L1 on L2 is known as forward transfer whereas that from L2 on L1 is known as backward transfer (Kartushina et al., 2016) and it is reported in adults as well. The learners have successfully distinguished their voiced and voiceless categories despite the lack of a voiceless bilabial plosive in their L1, an area which has been reported as problematic among adult Arabic L2 learners.

The results also vary as a function of problematic sound class. Colantoni and Steele (2008) assert that not all sounds are equally difficult because articulatory constraints and typological markedness also play a role in L2 acquisition. Affricates are the least marked sounds in the study set and this is where statistical differences were observed, whereas dental fricatives are one of the most marked sounds and there was no statistical differences in target-likeness ratings between any of the conditions.⁵ In terms of the lack of statistical differences in VOT patterns between the training conditions and the Traditional condition's patterns resembling more of those from DigLin than their teacher, it is argued that some foreign TL sounds are readily perceivable to naive listeners without any training necessary (Best, 1994, 1995; Best & Tyler, 2007).

Indeed, as proponents of output practice such as Swain (1995) and Swain (2005) for general L2 learning and Colantoni and Steele (2008) for L2 speech learning postulate, output practice allows learners to reflect on their productions, comparing them to the TL in a continuous refining process. Output also allows learners to develop articulatory muscle memory for the target sounds and structures with more practice comes better tuning of articulatory-motor skills. Differences between child adult L2 learning cannot be explained merely by age as previous studies mostly emphasise but rather on type of input. It was demonstrated how learners of the same age (prepubescent) can have different performances resulting from differing types of input. Not all L2 phonology is governed by backward transfer. Some aspects are explained by universal language acquisition tendencies regardless of the age group. Universal tendencies also relate to acoustic salience and articulatory settings. It was demonstrated that in the delayed test, whereby the Listen and Speak condition maintained its rank in target-likeness and the Listen Only condition's performance levels out with that of the Traditional condition who received accented input.

ACKNOWLEDGMENT

This research received funding from the Libyan Ministry of Higher Education. It was supported by The National Centre for Language Training, Droob Al-Elm Language School, and Shuhada Qaser Ahmed Primary School by providing their training facilities.

REFERENCES

- Alqattan, S. (2015). *Early phonological acquisition by Kuwaiti Arabic children*. [Unpublished doctoral dissertation]. Newcastle University, Newcastle upon Tyne, United Kingdom.
- Altaha, F. (1995). Pronunciation errors made by Saudi university students learning English: Analysis and remedy. *International Journal of Applied Linguistics*, 109(1), 110–123.
- Amayreh, M. M. (2003). Completion of the consonant inventory of Arabic. *Journal of Speech, Language, and Hearing Research*, 46, 517–529.
- Ayyad, H. S. (2011). *Phonological development of typically developing Kuwaiti Arabic-speaking preschoolers*. [Unpublished doctoral dissertation]. University of British Columbia, Vancouver, Canada.
- Barros, A. M. (2003). *Pronunciation difficulties in the consonant system experienced by Arabic speakers when learning English after the age of puberty*. [Unpublished masters dissertation]. West Virginia University, Morgantown, United States.
- Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, 167, 167–224.
- Best, C. T. (1995). A direct realist view of cross-language speech perception: New directions in

research and theory. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research* (pp. 171–204). York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. Munro J (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (Vol. 17). John Benjamins Publishing Company.

Boersma, P., & Weeninck, D. (2016). *Praat: doing phonetics by computer (version 6.0.092)*. <http://www.praat.org>

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299–2310.

Broselow, E., & Finer, D. (1991). Parameter setting in second language phonology and syntax. *Interlanguage Studies Bulletin (Utrecht)*, 7(1), 35–59.

Chen, Y., Khattab, G., & Al-Tamimi, J. (2019). The effect of explicit instruction and auditory/audio-visual training on Chinese learners' acquisition of English intonation. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2286-2290). Australasian Speech Science and Technology Association Inc.

Chun, D. M., Jiang, Y., & Ávila, N. (2013). Visualization of tone for learning Mandarin Chinese. *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, 77–89.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, 29(3), 489–534.

Diehl, R. L., & Lindblom, B. (2004). Explaining the structure of feature and phoneme inventories: The role of auditory distinctiveness. In A. Greenberg & F. Popper (Eds.), *Speech processing in the auditory system* (pp. 101–162). Springer.

Dodd, B., Holm, A., Crosbie, S. C., & Zhu, H. (2013). Children's acquisition of phonology. In B. Dodd (Ed.), *Differential diagnosis and treatment of children with speech disorder* (2nd ed.). John Wiley & Sons, Incorporated.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics & Phonetics*, 17(8), 617–643.

Dyson, A., & Amayreh, M. (2000). Phonological errors and sound changes in Arabic-speaking children. *Clinical Linguistics & Phonetics*, 14(2), 79–109. <https://doi.org/10.1080/026992000298850>

Eckman, F. (1977). Markedness and the Contrastive Analysis Hypothesis. *Language Learning*, 27(2), 315–330.

Eckman, F. (2008). Typological markedness and second language phonology. In J. G. Edwards Hansen & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 95–115). John Benjamins.

Ehbara, H. (forthcoming). *Impact of computer-assisted pronunciation training on Libyan child learners of English*. [Unpublished doctoral dissertation]. Newcastle University, Newcastle upon

Tyne, United Kingdom.

Fantazi, G. (2003). *Perception and production of syllable structure and stress by adult Libyan Arabic speaker acquiring English in the UK*. [Unpublished doctoral dissertation]. Durham University, Durham, United Kingdom.

Flege, J. E. (1987). The instrumental study of L2 speech production: Some methodological considerations. *Language Learning*, 37(2), 285–296.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (Vol. 92, pp. 233–277). York Press.

Flege, J. E., Munro, M. J., & MacKay, I. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America*, 97(5), 3125–3134.

Flege, J. E., & Port, R. (1981). Cross-language phonetic interference: Arabic to English. *Language and Speech*, 24(2), 125–146.

Grunwell, P. (1985). *Phonological assessment of child speech*. College Hill Press.

Guilloteau, N. C. (1997). *Modification of phonetic categories in French as a second language: Experimental studies with conventional and computer-based intervention methods*. The University of Texas at Austin.

Hancin-Bhatt, B., & Bhatt, R. M. (1997). Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition*, 19(3), 331–378.

Hardison, D. M. (2004). Generalization of computer assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8(1), 34–52.

Heimbach, R. L. (1994). *The effect of interaction on the comprehension and acquisition of new lexical items by Kindergarten ESL learners*. Temple University.

Hodson, B. W., & Paden, E. P. (1981). Phonological processes which characterize unintelligible and intelligible speech in early childhood. *Journal of Speech and Hearing Disorders*, 46(4), 369–373.

Ingram, D., Christensen, L., Veach, S., & Webster, B. (1980). The acquisition of word-initial fricatives and affricates in English by children between 2 and 6 years. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (pp. 169–192). Elsevier.

Ioup, G. (2008). Exploring the role of age in the acquisition of a second language phonology. In J. G. Edwards Hansen & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 41–62). John Benjamins.

Jakobson, R. (1968). *Child language, aphasia and phonological universals*. Mouton.

Karimi, S. (1987). Farsi speakers and the initial consonant cluster in English. *Interlanguage Phonology: The Acquisition of a Second Language Sound System*, 305–318.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics*, 57, 21–39.

- Kartushina, N., & Martin, C. D. (2019). Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning*, 69(1), 71–105.
- Kharma, N., & Hajjaj, A. (1989). *Errors in English among Arabic speakers: Analysis and remedy*. Longman.
- Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal*, 97(3), 720–744.
- Kløve, M. H., & Young-Scholten, M. (2008). Repair of L2 syllables through metathesis. *International Review of Applied Linguistics in Language Teaching*, 39(2), 103–133.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.
- Krashen, S. D. (2018). Down with forced speech. In Y. N. Leung, J. Katchen, S. Y. Hwang, & Y. Chen (Eds.), *Reconceptualizing English language teaching in the 21st century* (pp. 99–106). Crane Publishing Company.
- Kwon, B.-Y. (2006). *Korean speakers' production of English consonant clusters: Articulatory and perceptual accounts* (Vol. 1). ProQuest.
- Lenth, R. (2019). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. <https://cran.r-project.org/package=emmeans>
- Lord, G. (2005). (How) can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 557–567.
- Mack, M. (2003). The phonetic systems of bilinguals. *Mind, Brain, and Language: Multidisciplinary Perspectives*, 309–349.
- MacKay, I. R. A., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America*, 110(1), 516–528.
- Mackey, A. (2007). *Conversational interaction in second language acquisition: A collection of empirical studies*. Oxford University Press.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge University Press.
- Major, R. C. (1994). Chronological and stylistic aspects of second language acquisition of consonant clusters. *Language Learning*, 44(4), 655–680.
- McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology*, 27(4), 1546–1571.
- Neri, A., Mich, O., Gerosa, M., & Giuliani, D. (2008). The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5), 393–408.
- Ohala, D. K. (2008). Phonological acquisition in a first language1. In J. G. Edwards Hansen & M. L. Zampini (Eds.), *Phonology in second language acquisition* (pp. 19–39). John Benjamins.
- Overall, J. W. (2014). *DigLin*. <http://diglin.eu/>
- Parlak, Ö. (2010). Does pronunciation instruction promote intelligibility and comprehensibility? *Annual Meeting for TESOL*, 1–5.

- Pater, J., & Barlow, J. A. (2003). Constraint conflict in cluster reduction. *Journal of Child Language*, 30(3), 487–526.
- Pearson, P., Pickering, L., & Da Silva, R. (2011). The impact of computer assisted pronunciation training on the improvement of Vietnamese learner production of English syllable margins. *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, 169–180.
- Pfenninger, S. (2016). All good things come in threes: Early English learning, CLIL and motivation in Switzerland. *Cahiers de IILSL*, 48, 119–147.
- Piske, T. (2012). Factors affecting the perception and production of L2 prosody: Research results and their implications for the teaching of foreign languages. In J. Romero-Trillo (Ed.), *Pragmatics, prosody and English language teaching* (pp. 41–59). Springer.
- Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.
- Shriberg, L. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech, Language, and Hearing Research*, 36(1), 105–140.
- Shriberg, L., & Kwiatkowski, J. (1994). Developmental phonological disorders 1: A clinical profile. *Journal of Speech and Hearing Research*, 37(5), 1100–1126.
- Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa articulation norms project and its Nebraska replication. *Journal of Speech and Hearing Disorders*, 55(4), 779–798.
- Stoel-Gammon, C. (1987). Phonological skills of 2-year-olds. *Language, Speech, and Hearing Services in Schools*, 18(4), 323–329.
- Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. University Park Press.
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Swain, M. (1995). The functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honor of HG Widdowson* (pp. 125–144). Oxford University Press.
- Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 495–508). Routledge.
- Tanner, M., & Landon, M. (2009). The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility. *Language Learning & Technology*, 13(3), 51–65.
- Templin, M. (1957). *Certain language skills in children; their development and interrelationships*. University of Minnesota Press.
- Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, 28(3), 744–765.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3), 326–344.

Weinberg, A., & Knoerr, H. (2003). Learning French pronunciation: Audiocassettes or multimedia? *Calico Journal*, 20(2), 315–336.

Young-Scholten, M. (1995). The negative effects of “positive” evidence on L2 phonology. In L. Eubank, L. Selinker, & M. Sharwood Smith (Eds.), *The current state of interlanguage* (pp. 107–121). John Benjamins Publishing Company.

ADDITIONAL READING

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge University Press.

Maddieson, I. (1984). *Patterns of sounds: Cambridge studies in speech science and communication*. Cambridge University Press.

Piske, T. (2017). The earlier, the better? Some critical remarks on current EFL teaching to young learners and their implications for foreign language teacher education. In R. Wilden & R. Porsch (Eds.), *The professional development of primary EFL teachers. National and international research* (pp. 45–57). Waxmann.

Piske, T., & Young-Scholten, M. (2009). *Input matters in SLA* (Vol. 35). Multilingual Matters.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191–215.

Settinieri, J. (2008). Teaching pronunciation. In S. Gramley & V. Gramley (Eds.), *Bielefeld introduction to applied linguistics: A course book* (pp. 1–12). Aisthesis Verlag.

Tessier, A. -M. (2015). *Phonological acquisition: Child language and constraint-based grammar*. Macmillan Education UK.

Trofimovich, P., Lightbown, P. M., Halter, R. H., & Song, H. (2009) Comprehension-based practice: The development of L2 pronunciation in a listening and reading program. *Studies in Second Language Acquisition*, 31(4), 609–639.

Wenk, B. J. (1979). Articulatory setting and de-fossilization. *Interlanguage Studies Bulletin*, 4, 202–220.

Zampini, M. (2008). L2 speech production research: Findings, issues, and advances. In J. G. Edwards Hansen & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 219–250). John Benjamins.

KEY TERMS AND DEFINITIONS

Backward Transfer: an influence from L2 on L1.

Deaffrication: a change of a feature in the affricate.

Final Devoicing: a phonological process by which voiced obstruents occurring word-finally become (partially) voiceless.

Forward Transfer: an influence from L1 on L2.

/l/ Vocalisation: a phonological process involving the change of velarised /l/ into a (semi)vowel.

Markedness: a relationship between elements of phonological class based on their degree of complexity.

Th-Fronting: a phonological process involving the substitution of [f] or [v] for /θ/ and /ð/ respectively.

Voice Onset Time: the duration between the start of the release of a plosive and the onset of glottal pulsing.

ENDNOTES

¹ Differences across languages can be explained on the basis of frequency effects (Zamuner, 2003) among other things. Role of input is important, and some children may still acquire this sound earlier if it is used often. When the acquisition of /v/ for example was compared between English, Bulgarian, Swedish and Estonian children, it was found that it was acquired earlier in all languages but English. English children were relatively late to acquire it. Ingram (1999) concluded that the reason for this was that the occurrence of /v/ in English was relatively less frequent compared with the other languages from the study. However, the role of frequency will not be further explored here.

² A study by Young-Scholten, Akita, and Cross (1999) compared two groups of adult L2 learners, one presented with orthographic input simultaneously with auditory input, and the other presented with auditory input only. In examining their production of L2 complex consonant clusters, participants in the former condition (auditory and orthographic input) exhibited epenthesis whereas participants from the second condition (auditory input only) exhibited productions reflecting those found in L1 child acquisition, that is cluster simplifications.

³ Although the voiced affricate and the interdental fricatives are part of Modern Standard Arabic, /d͡ʒ/, is realized as [ʒ] and /θ, ð/ are realized as [t, d] respectively.

⁴ The speaking practice inherently involved the participants listening to their own output. This allowed the participants of this group to reflect on their output and potentially modify their speech production throughout the treatment period as suggested by the Hybrid model in (Colantoni & Steele, 2008).

⁵ See Maddieson (1984) for the distribution of affricates and dental fricatives in the World's languages.