**UNIVERSITAT DE BARCELONA**

# Estudios de Fonética Experimental
**Journal of Experimental Phonetics**

**LF**
*Laboratori de Fonètica*

# Effects of attention on the production of Spanish consonants by SFL learners

Zhao Liu[a], María Jesús Machuca Ayuso[b]

[a] East China Normal University (China), liuz_liuzhao@163.com
[b] Universitat Autònoma de Barcelona (Spain), mariajesus.machuca@uab.cat

| ARTICLE INFO | ABSTRACT |
|---|---|
| | The aim of this paper is to analyse the production of Spanish stops /b, d, g, p, t, k/ and liquids /l, ɾ/ by Mandarin Chinese learners of Spanish (*N*=10). Two different tasks were used to elicit data: one induced a focus on stops and the other a focus on liquids. Realizations were categorized based on spectral manifestation, and acoustic parameter values between the two groups were compared. The results showed that attention exerts a significant influence on voiced stops and tap production: more inaccurate productions were observed when the learners did not focus on these phonemes. The data also pointed to considerable intra-speaker variation. Learners tended to produce realizations with longer durations in new phonemes: /b, d, g, ɾ/. |

## 1. Introduction

Pronunciation, one of the most difficult components to teach, is not prioritized in Spanish as a foreign language (SFL) class or foreign language classes in general. On the one hand, consideration should be given to teaching it best and what knowledge is required by a teacher to teach it. It is worth noting that the communicative approach has relegated the role of pronunciation in SFL even further since the priority is to ensure that communication is effective. However, concepts relating to "good pronunciation", including intelligibility, comprehensibility, foreign accent, and fluency (Santamaría, 2015), are still addressed, and, thus, the characteristics of the learner's interlanguage are compared to the ideal pronunciation expected of a native speaker. To that end, teachers should be familiar with the phonic system of the language they are teaching, albeit they do not need to be phonetic experts, as some scholars have suggested (see, for example, Gil, 2007).

On the other hand, acoustic analysis of the phonetic categories produced by foreign language learners highlights the fact that they create new categories for both the sounds with which they are already familiar from their native language and new sounds (Flege, 1995).

The goal of this study was to look into two things: the importance of focusing learners' attention in the

SFL classroom to help them learn new sounds, and the analysis of sounds shared by the two languages (the language they are learning and the native language) to establish if they have the same acoustic characteristics.

## 2. Theoretical framework

Foreign language learning should, according to some authors (Schmidt, 2001, 2010; Ishihara & Cohen, 2010), be based on focused instruction. Although these works mostly focus on pragmatic learning, this same idea could be extended to improve the realization of phonemes, especially those not found in the learner's native language, considering how difficult it is to pronounce these phonemes like a native speaker. When learning a native language, this does not happen since the people who are in contact with native speakers help them develop these skills naturally from birth. Barquero (2020), in line with Schmidt (2001, 2010), distinguished between the concepts of attention, noticing, and understanding. Attention, according to Barquero, is the mechanism by which a speaker is aware of something, whereas noticing is the process by which a learner recognizes the existence of a linguistic phenomenon without regard for any form of rule or systematization. The following phase, which is to comprehend this phenomenon, is integrating this component inside the linguistic system of the target language. Likewise, Ruiz de Zarobe & Ruiz de Zarobe (2013) pointed out that in order for learners to acquire a certain element of a foreign language, they must be aware of that aspect; therefore, someone must inform them about it so that this awareness can take place. Thus, even at advanced levels of language acquisition, teachers must help learners become aware of the distinctions between the sounds of their native language and emphasize the importance of certain articulations to prevent their foreign accent from being noticeable. When it comes to teaching foreign language pronunciation, there is no clear consensus on which method is most appropriate. The literature discusses two approaches: one involves teaching pronunciation in communicative contexts (communicative language teaching), and the other

addresses the importance of focusing on phones and teaches the articulation mechanisms of a given sound (see Kang et al., 2017). Based on these two methodologies, Lan & Wu (2013) conducted an experiment on the realization of rhotics in consonant clusters in English by native speakers of Mandarin Chinese. Their findings revealed that pronunciation teaching is more effective focusing on forms, namely on phones, rather than communicative context. This type of instruction in foreign language teaching is closely linked to what Shintani (2016) called incidental acquisition. Essentially, it is understood as learning that is acquired by focusing on anything other than what is being taught. It is considered a facilitative learning process since two different linguistic characteristics are inadvertently learned at the same time. This can happen when learners acquire pronunciation by learning the lexicon; they assimilate words and thus unconsciously acquire the pronunciation and phonotactic traits of a specific phoneme without associating them with specific articulatory characteristics (Saito, 2014). According to Loewen (2015), however, learners should be able to distinguish between the new phonemes of the language they are learning and the similar phonemes in their native language. Moreover, distinctions should also be made between two similar phonemes in the foreign language. Thus, instruction plays a key role in achieving this goal, both in terms of production and perception. In isolated words or reading tasks, instruction on a specific phoneme together with knowledge of the L1's phonological system can yield extremely effective outcomes, but not in activities involving communicative contexts. For this reason, we included a test requiring participants to produce isolated words that contained stop and liquid consonants, phonological categories in which native Chinese speakers who are learning Spanish have the most difficulty (Cortés Moreno, 2014). All Chinese stops are voiceless, although aspiration is an important feature in the categorization of these phonemes; according to this parameter, stops in Mandarin Chinese can be divided into voiceless unaspirated stops /p, t, k/ and voiceless aspirated stops /$p^h$, $t^h$, $k^h$/. In Spanish, stop phonemes can be

voiceless, as in Chinese /p, t, k/, or voiced, as in /b, d, g/. This means that three of the phonemes analysed were common to the phonic systems of both languages, while the remaining three, the voiced stops, were new for learners of Spanish (Liu, 2019). Concerning liquid phonemes, only two were addressed in this study: the alveolar lateral /l/, which is also found in the Chinese system (Duanmu, 2007; Lin, 2007), and the tap, which is a new phoneme for learners of Spanish. Therefore, each of these series presented phonemes that were common to both the foreign and the native language (/p, t, k, l/) and new phonemes (/b, d, g, ɾ/).

Numerous studies have explored Spanish voiceless stops, but this article only refers to those that employed word reading corpora. As a result, it is possible to conduct a comparative study with the data presented in this work. Two phases can be distinguished in the temporal analysis of stops: first, the silent interval or hold phase, which corresponds to the moment the articulators come into contact in the labial region (/p/ or /b/), dental region (/t/ or /d/) or velar region (/k/ or /g/). Martínez Celdrán (1993) found that silent intervals varied from 70 to 140 ms for voiceless stops and 26 to 11 ms for voiced stops. Second, the average values reported in the various studies consulted for the VOT (voice onset time), the phase that corresponds to the moment the articulators separate and the air is suddenly released and the onset of vocal-fold vibration (Keating, 1984), ranged from 14 to 17.5 ms for voiceless labial stops; 19 to 20.7 ms for voiceless dental stops; and 32.2 to 35.4 for voiceless velar stops (Castañeda, 1986; Asensi et al., 1997; Villamizar, 2002; Martínez Celdrán & Fernández Planas, 2007). Setting aside the controversy concerning the features that establish the phonological voicing contrast, it is worth noting that no values for this phase can be found in Spanish voiced stops since voicing begins before the stop is released, and these values refer to the total duration of the consonant, which includes the two abovementioned phases. Thus, for example, Castañeda (1986) -69 ms for labial stops, -77.7 ms for dental stops, and -58 ms for velar stops. When considering only the release phase, some studies have reported values from 2 to

8 ms for voiced labial stops, 6 to 11 ms for dental stops, and 11 to 19 ms for velar stops (Lemus Sandoval, 2002). Another parameter that can be studied in voiced stops is the formant found in low-frequency areas. However, no studies have reported frequency values of the first formant for voiced stop consonants in Spanish since this parameter has not traditionally been considered important for the voicing contrast, as indicated by classical authors such as Liberman et al. (1958). A study by Machuca (1997) reported values for this parameter, albeit in relation to a spontaneous speaking style: 295 Hz for labial stops, 332 Hz for dental stops, and 334 Hz for velar stops.

With respect to liquid phonemes, numerous studies have reported varying durations for alveolar lateral consonants, depending on whether they are in onset or coda position, or in simple or complex onset position; however, the mean duration of this consonant range from 60 ms (Quilis, 1981) to 64 ms (Almeida & Dorta, 1993). The authors also reported different frequency values of the first three formants. For Quilis, et al. (1979), the values were 333, 1554 and 2564 Hz for the first (F1), second (F2) and third formant (F3), respectively. For Almeida & Dorta (1993), the values were 400, 1514, and 2571 Hz, if we take the average of the mean value of unstressed and stressed syllables. The values were slightly lower in Martínez Celdrán & Fernández Planas (2007): 301, 1474, and 2353 Hz.

In terms of rhotics, our work focused solely on tap (just one brief interruption of airflow). In Spanish, the tap has different phonetic categories, all of which are correct. The articulation can take the form of a flap, which corresponds to a brief silent interval (when the articulators are in contact) followed by a release phase (when the articulators relax and the air is suddenly released), or an approximant, in which the articulators come close together but do not touch, as described in this paper. Perceptually, these realizations are difficult to distinguish; however, when they are viewed on a spectrogram, it is easier to distinguish them. For instance, Figure 1 illustrates two productions by a

native Spanish speaker; the image on the left illustrates a flap, and the image on the right shows an approximant. According to Quilis (1993), the silent phase of the rhotic has an average duration of 20 ms, with values ranging from 16 ms to 36 ms. For a more detailed description of all rhotic realizations in Spanish, see Blecua (2001).

The objective of this study was to analyse the production of shared and new phonemes in native Chinese speakers when they learn Spanish. It considered the learners' production of errors and their categorization according to the focus of attention when pronouncing the words under study, which contained both stop and liquid phonemes. It was based on the hypothesis that production errors in a given phoneme will decrease when learners focus on that phoneme.
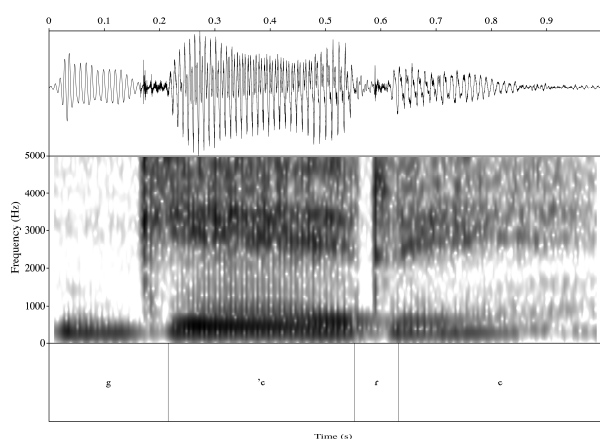


Oscillogram and spectrogram of the segment ['gere]
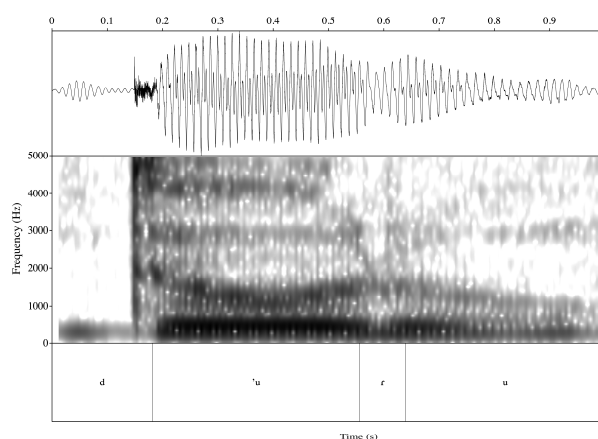


Oscillogram and spectrogram of the segment ['duru]

**Figure 1.** Different realizations of /ɾ/ produced by a native Spanish speaker.

## 3. Methodology

### 3.1 Participants

A total of 13 speakers were chosen to participate in this investigation. Because it was difficult to find male volunteers who matched all of the selection requirements, all participants were female. In order to select the participants, a Google Forms questionnaire containing questions about linguistic profiles was designed and distributed. Finally, two groups were created based on the linguistic profile of the participants.

The first group consisted of 10 native Mandarin Chinese speakers[1], hereinafter referred to as the CH group. All were born in China and had studied Spanish at a Chinese university. At the time they were recorded, they were enrolled in the Master's Degree in Spanish Language, Hispanic Literature and Spanish as a Foreign Language at the Universitat Autònoma de Barcelona (UAB). Their Spanish level was B2/C1 (DELE), and they speak three languages: Chinese, English, and Spanish.

The second group was made up of three native Spanish/Catalan speakers (the ES group). All participants were born in Catalonia, Spain, and in

[1] From here on, we will use the term "Chinese" to refer to Mandarin Chinese.

their final year of study at the UAB for a degree in translation and interpretation (Spanish and Chinese). Their Chinese proficiency level was B1. None of the speakers had ever lived in a foreign country, and all spoke Spanish, Catalan, Chinese, and English. In this case, both Spanish and Catalan are their mother tongue.

It should be noted that the speakers in both groups had learned English as a second language (L2) and were at CEFR level B2 (Common European Framework of Reference, 2001).

## 3.2 Task description

Two reading tasks were completed for data elicitation. These were constructed using a list of bisyllabic pseudowords [2] with the same CVCV structure. The onset consonant in the first syllable was one of the six Spanish stops /b, d, g, p, t, k/, while the consonant in the second syllable was either the lateral or the tap. Each pseudoword had the same vowels in the two syllables (i.e. one of the five vowels in Spanish /a, e, i, o, u/). The list contained a total of 60 words. Thus, there were 1560 tokens for liquids and 1560 tokens for stops. However, realizations in which the speaker produced a pause between the two syllables were eliminated since this may influence the duration of the consonant. This resulted in a total of 3102 tokens. The number of cases analysed for each phoneme is presented in the results section.

In both tasks, the participants were asked to focus their attention on the red-highlighted portion of the word. Figure 2 illustrates the two tasks: the first task (on the left) required participants to focus their attention on the stops, while the second task (on the right) required them to focus on the liquids.

## 3.3 Procedure

The recording was carried out on the same day. Prior to the data collection, a brief conversation was held in Spanish with the non-native participants to acclimate them to the language's context.

The recording took place in a soundproofed room in the phonetics laboratory of the UAB's Servei de Tractament de la Parla i del So. The microphone was a Rode NT1-A. The frequency response was 20-20,000 Hz and the dynamic range 132 dB. The mixing console was an Alesis Multimix and the sound card was a Roland Cakewalk UA-25EX. Adobe Audition CS5.5 was used as the recording tool. The voice files were digitized in 24-bit WAV PCM format with a sampling frequency of 44.1 KHz.



**Figure 2.** Sample of the first task (above) and the second task (below).

## 3.4 Data analysis

Using Praat (Boersma & Weenink, 2017), the recordings were manually labelled and annotated. The spectrogram's acoustic features were used to categorize the various realizations. Categorization has been revised by both researchers, and they are detailed in the section corresponding to each phoneme.

Once the sounds had been categorized and further determined as either correct or incorrect, values on acoustical and temporal parameters were extracted

---

[2] In this study, pseudowords were employed to prevent foreign language learners from accessing the lexicon.

using a Praat script created by the authors. In the findings section, the acoustic parameters that were analysed are listed for each category.

The statistical analysis of the data was carried out with the RStudio program (R CoreTeam, 2019). Different tests were used depending on the variable type. When both variables were categorical, we employed the chi-squared test. The variables were the focus of attention (stops or liquids) and different realizations produced by the speakers. Duration and frequency values were transformed to a logarithmic scale (base 10) to avoid possible inter-speaker variation. Nevertheless, in the results section, we include the numeric data of the acoustic parameters for comparison with data from the literature.

When the response variable was numerical, the non-parametric Kruskal-Wallis test was used. A detailed examination of the data indicated that they lacked a normal distribution or equality of variance, indicating that they did not meet the prerequisites for performing a parametric ANOVA test. Thus, the test used in this case was the non-parametric Kruskal-Wallis test. The independent variables were group and focus of attention, while the dependent variables were each of the acoustic parameters.

## 4. Results

This section details the results of each realization. We first present different realizations made by both groups of speakers. Then, we extracted values for a set of acoustical parameters for those realizations that have been classified as correct production, and we compare them for both groups of speakers.

### 4.1. Tap

#### 4.1.1. Error categorization

In total, 765 realizations of this phoneme were analysed; 179 corresponded to the native Spanish speaker and 586 to the language learners. All realizations made by native speakers were as

expected for this phoneme. Two different phonetic categories were observed:

a) Approximant (121 cases, 67.6%). In this category, clear formant structures are observed. This was also the most common category among native speakers.
b) Realizations with a brief hold (58 cases, 32.4%). In this category, we can observe a small blank portion, corresponding to the phase when the articulatory organs come into touch. After then comes the release phase, where a higher concentration of energy can be observed. This is the moment when the organs suddenly separate.

The learners, besides the two canonical realizations (i.e. approximants: 170 cases, 29.01%; and stops: 302 cases, 51.54%), also presented other types that were considered non-canonical (errors):

a) Lateral (86 cases, 14.68%). A greater segmental duration is observed compared to the approximant realization of the rhotic.
b) Trill (13 cases, 2.22%). These realizations present silent intervals and opening phases, which represent, respectively, the moment the articulatory organs come into contact and the moment they relax.
c) Combination of an approximant and a stop. In this realization, characteristics of an approximant and a stop can be observed at the same time (15 cases, 2.56%). A formant structure can be seen in the closing phase and at the end of the consonant segment, before the following vowel, and a release phase related to the sudden relaxation of the articulatory organs is observed.

Figure 3 shows the percentage of each realization in both groups. The x-axis shows the realizations broken down by group, and the y-axis shows the percentage of each realization. Each realization was classified as an approximant (labelled 4), a lateral (4_l), a flap (4_ocl), a combination of approximant and release phase (4_ocl_aprox), or a trill (4_r). The graph shows that native Spanish speakers (ES group) presented a higher percentage of

approximants (67.6%), while learners (CH group) had a higher percentage of stops (51.54%).
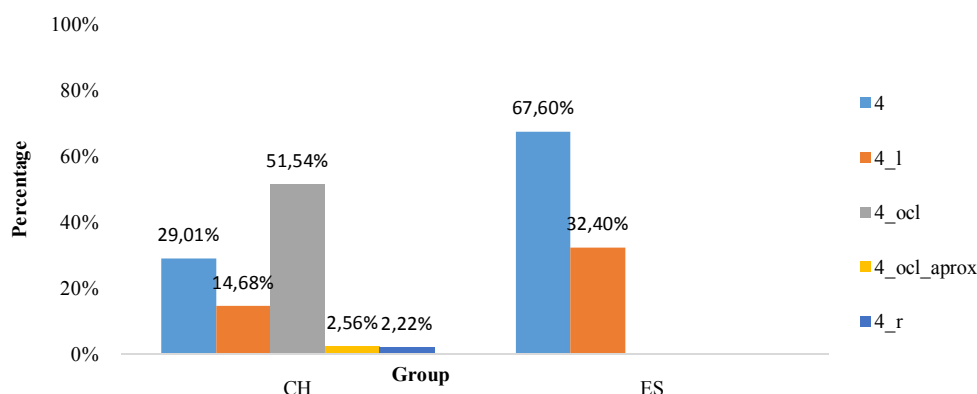


**Figure 3.** Realizations of /ɾ/ by both groups of speakers.

### 4.1.2. The effect of attention

As mentioned above, the main purpose of this study was to determine if the correct realization of these phonemes depends on whether learners pay attention to the phonological category they are producing.

Native Spanish speakers presented a correct production in both cases. In contrast, foreign language learners presented incorrect realizations 11.22% of the time (33 out of 294 cases) when they were told to focus on that phoneme while reading the word. The number of mistakes climbed to 27.74% when they were told to pay attention to stops (81 out of 292 cases). The results showed a significant difference due to the focus of attention (liquids and stops): $\chi^2(1) = 24.46$, $p = 7.597e\text{-}07$ *.

In the CH group, chi-squared tests were used to determine if the attention effect was attributable to intra-speaker variability, with the responses as the dependent variable and the speaker and focus of attention as the independent variable. Figure 4 shows the realizations based on this latter variable.

The chi-squared results indicate that this variable caused a significant difference: $\chi^2(19) = 498.8$, $p < 2e\text{-}16$ *. The post-hoc test indicated that learners made more errors when distracted by focusing on the phonemes they were not learning (see Figure 4).
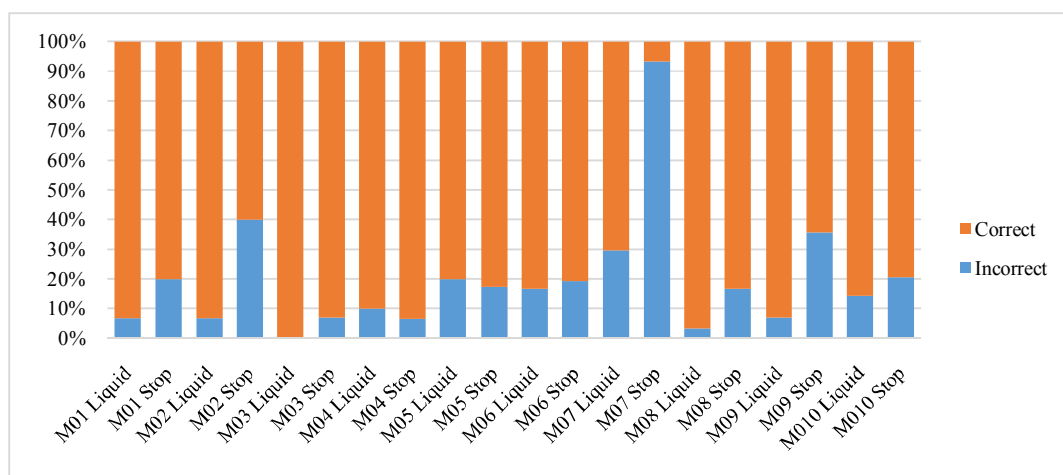


**Figure 4.** Errors (%) in the realization of /ɾ/ by the learners based on the variable "speaker_attention".

### 4.1.3. Acoustical analysis

#### 4.1.3.1. Approximant

For approximants, four acoustic parameters were taken into account: duration and frequencies of the first three formants (F1, F2, and F3). Note in Table 1 that there is a higher level of data dispersion in the F2 and F3 frequencies for both groups. The duration of segments produced by the learners (group CH) also presented greater variability compared to that of the native speakers (group ES).

| | | Approximant realization | | | | | Stop realization | |
|---|---|---|---|---|---|---|---|---|
| | N | F1 (Hz) | F2 (Hz) | F3 (Hz) | Duration (ms) | N | F1 (Hz) | Duration (ms) |
| CH group | 168 | 436.9 (177.7) | 1584.8 (391.3) | 2591.1 (467.4) | 34.1 (11.8) | 302 | 477.8 (195.7) | 39.3 (10.6) |
| ES group | 121 | 397.1 (140.5) | 1806.2 (399.7) | 2618.2 (461.0) | 19.6 (5.6) | 58 | 516.8 (243.5) | 30.7 (10.5) |

**Table 1.** F1, F2, F3 and duration values of the approximant and stop realization for the phoneme /ɾ/ produced by speakers of both groups.

According to the statistical results, the F1 values in the CH group were significantly higher than in the ES group: $F[1, 289] = 4$, $p = 0.04$ *. The F2 values were also significant, however the native group (ES) had greater F2 values than the learner group (CH): $F[1, 289] = 21$, $p = 5e\text{-}06$ *. The duration of the realization in the learner group (34.1 ms) was greater than that of the native group (19.6 ms): $F[1, 289] = 127$, $p < 2e\text{-}16$ *. By contrast, the F3 values did not show any significant differences: $F[1, 289] = 0.05$, $p = 0.08$.

#### 4.1.3.2. Silent interval and release (flap)

For this category, only the first formant can be observed; thus, two acoustic parameters were taken into account: duration and F1 frequency, which are reported in Table 1.

In accordance with the statistical analyses, there was no significant difference with respect to F1 frequency ($F[1, 359] = 0.9$, $p = 0.3$). By contrast, the learners produced this segment with a longer duration (39.3 ms) than the native speakers (30.7 ms): $F[1, 359] = 33$, $p = 1e\text{-}08$ *.

### 4.2. Alveolar lateral

#### 4.2.1. Error categorization

In total, 779 realizations of this phoneme were analysed, with 180 belonging to native Spanish speakers and 599 to language learners.

There were 93 instances of errors (15.5%) and 506 cases of correct realizations (84.5%) in the learner group. All realizations of native speakers were categorized as correct production. The learners' errors were categorized as follows:

a) Tap /ɾ/. The rhotic exhibits an approximant realization, that is, a formant structure with a short duration of the consonant (64.5% of the errors).
b) Tap /ɾ/ with a small portion of silent interval. In this case, the realization presented a silent interval corresponding to the contact between the articulatory organs and a release phase in which the articulators were suddenly relaxed (21.5% of errors).
c) Dental approximant, an allophone of the phoneme /d/ in Spanish (11.8%).
d) Trill with a closing and an opening phase (2.2%).

Figure 5 shows each of the aforementioned realizations. Segments labelled "l" are correct realizations, whereas all other cases are incorrect

realizations of /l/: "l_4" and "l_4_ocl" correspond to different manifestations of [ɾ]; "l_D" to a dental approximant and "l_r" to a trill.
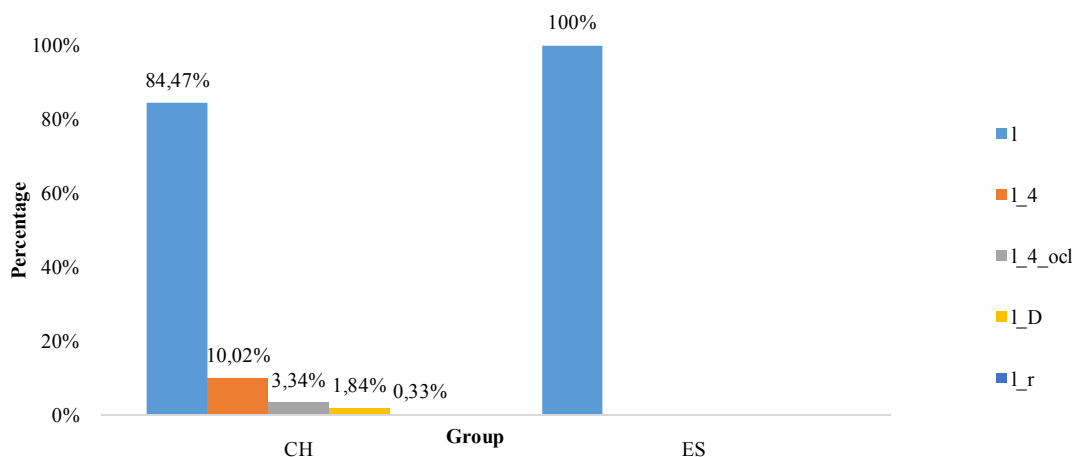


**Figure 5.** Realizations of /l/ by both groups of speakers.

### 4.2.2. The effect of attention

A chi-squared test was carried out to determine whether the errors were due to the phoneme on which the learners focused their attention. The results showed that there were no significant differences ($\chi^2(1)$ = 4.3405, $p$ = 0.03722 *). However, the errors increased when they were distracted by having to focus their attention on the other type of sound (18.8% of errors when they

focused on the stops versus 12.33% when they focused on the liquids). However, these results depended on the speaker ($\chi^2(19)$ = 327.92, $p$ < 2.2e-16 *). Figure 6 shows the error percentages based on the focus of attention and the learner. Some speakers did not make any errors in the realization of lateral consonants when they paid attention to these consonants, but all made errors when their attention shifted to the stops in the same word.
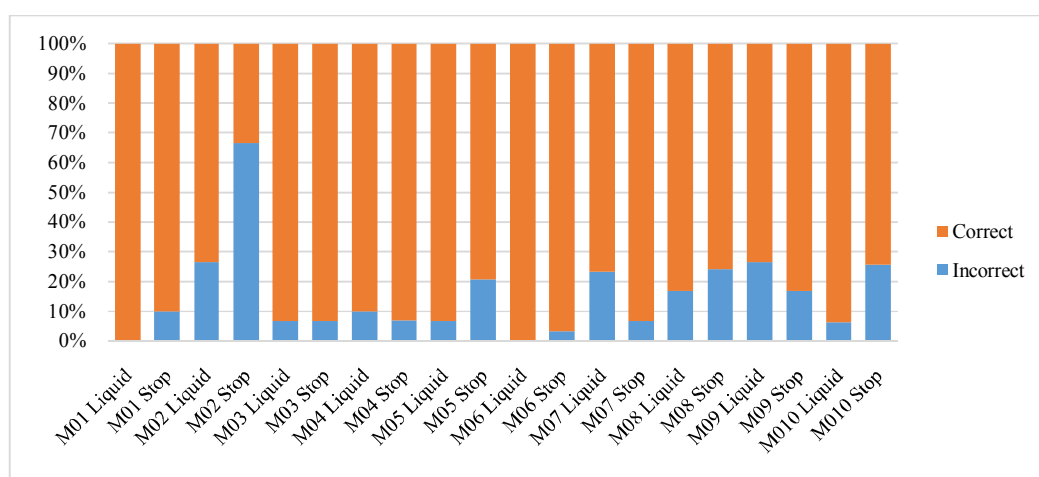


**Figure 6.** Errors (%) in the realization of /l/ by the learners based on the variable "speaker_attention".

### 4.2.3. Acoustical analysis

For this category, we analyzed the duration and the values of the first three formants. Recall that comparison has only been carried out on realizations that have been categorized as correct.

Table 2 presents the mean values and the standard deviation for these parameters in both groups. As we can see, the main difference lies in the frequency values of F2, which were higher among the native speakers than the learners.

| | F1 (Hz) | F2 (Hz) | F3 (Hz) | Duration (ms) |
|---|---|---|---|---|
| **CH group** (**N = 506**) | 401.7 (107.9) | 1502.2 (352.9) | 2661.4 (489.6) | 85.9 (29.4) |
| **ES group** (**N = 180**) | 376.6 (103.5) | 1818.7 (468.5) | 2748.6 (490.0) | 84.1 (22.6) |

**Table 2.** F1, F2, F3 and duration values of the lateral realization of /l/ produced by speakers in both groups.

The statistical test indicated that these parameters differed significantly among two groups: F1 frequency ($F[1, 685] = 9.2754$, $p = 0.002323$ *), F2 frequency ($F[1, 685] = 68.303$, $p < 2.2e-16$ *), and F3 frequency ($F[1, 685] = 6.3077$, $p = 0.01202$ *). The duration did not yield a significant difference ($F[1, 685] = 0.20425$, $p = 0.6513$).

### 4.3. Voiced stops

In total, 779 realizations of voiced stop were analysed, 180 of which corresponded to the native Spanish speakers and 599 to the language learners.

### 4.3.1. Error categorization

277 tokens produced by foreign language learners were categorized as errors (47.8%); 100 labial stops (36.1%), 75 dental stops (27.1%), and 102 velar stops (36.8%). All these mistakes were related to the voicing; that is to say, the participants devoiced them. However, M02 also presented other kinds of errors, such as nasalization (five cases) and fricativization (one case); these accounted for 6% of all labial errors. These six cases were excluded from the statistical analysis. Similarly, native speakers produced 10 cases of devoiced realizations (5.5%), which were also excluded from the statistical analysis.

### 4.3.2. The effect of attention

The effect of focusing on a certain phoneme influenced the production of errors in the realization of voiced stops. Thus, at the labial place of articulation, the fact that the speakers focused their attention on the liquids when producing the stops caused an increase in errors from 13.71% (focus on stops) to 59.23% (focus on liquids): $\chi^2(1) = 56.18$, $p = 6.613e-14$ *.

To determine whether this attention effect was due to inter-speaker variability, chi-squared tests were carried out within the learner group. The results revealed that, for /b/, there was a speaker-related effect on the production of errors when the focus of attention was modified ($\chi^2(19) = 1096$, $p < 2.2e-16$ *). Figure 7(a) shows correct and incorrect realizations broken down by the speaker.
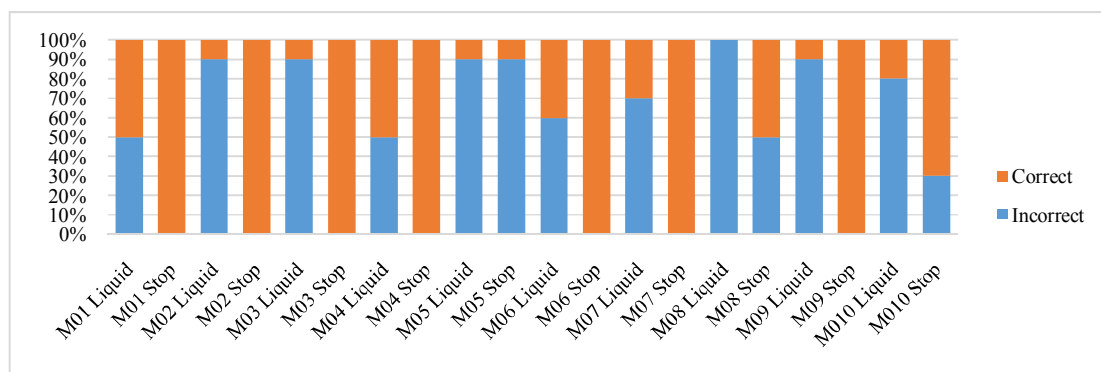
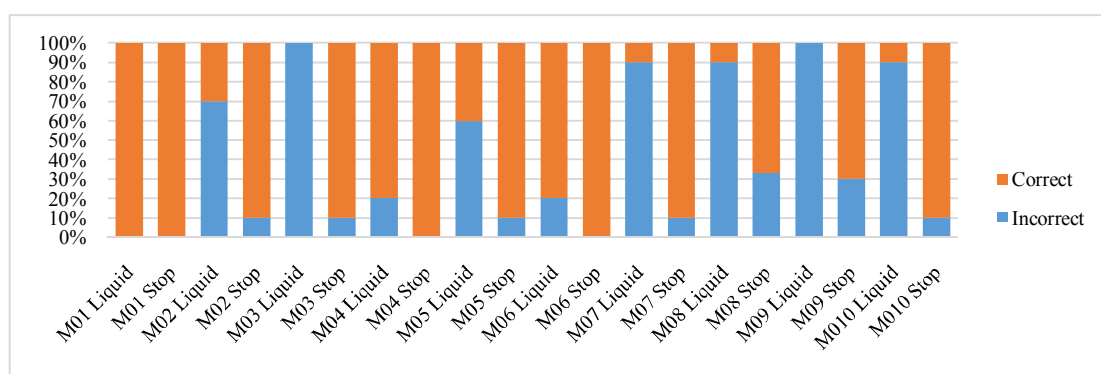**Figure 7(a).** Errors (%) in the realization of /b/ by the learners based on the variable "speaker_attention".



**Figure 7(b).** Errors (%) in the realization of /d/ by the learners based on the variable "speaker_attention".
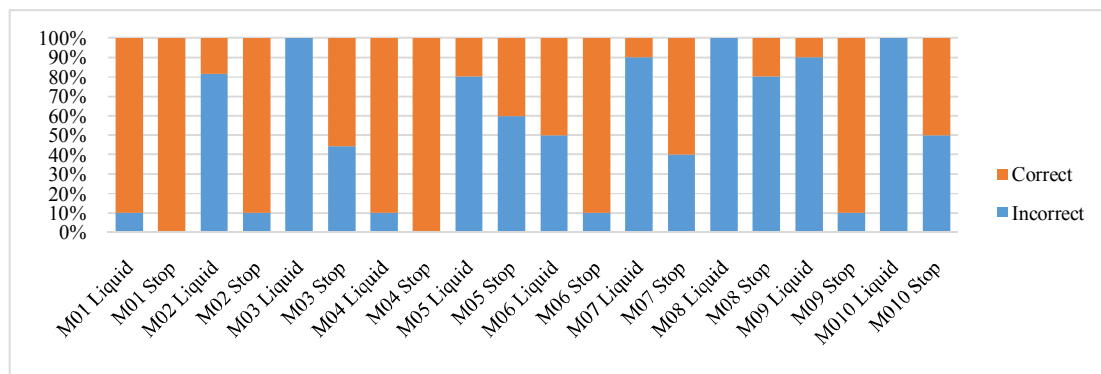


**Figure 7(c).** Errors (%) in the realization of /g/ by the learners based on the variable "speaker_attention".

In the case of dental stops, an increase in errors was also observed when the participants produced a stop while their focus of attention was on liquids. Errors rose from 10.08%, when the speakers were instructed to pay attention to the stops, to 50% when they were told to pay attention to liquids ($\chi^2(1)$ = 97.896, $p$ < 2.2e-16 *). As with labial consonants, this effect depended on the speaker ($\chi^2(19)$ = 1181.3, $p$ < 2.2e-16 *). Figure 7(b) shows the error

percentages broken down by focus of attention and speaker.

Finally, velars also presented an increase in errors when the speakers were instructed to focus on the liquid consonants while producing the stop consonants. The percentage increased significantly, from 24.03% (focus on stops) to 59.54% (focus on liquids): $\chi^2(1)$ = 31.984, $p$ = 1.554e-08 *. The chi-

squared test showed that these results depended on the speaker ($\chi^2(19) = 1069.5$, $p < 2.2\text{e-}16$ *). The percentages are shown in Figure 7(c).

### 4.3.3. Acoustical analysis

The acoustic parameters analysed in this category were the frequency values of the first formant, the duration values of the hold phase of the consonant and the values of the release phase (see Table 3). The two phases were not added together to obtain the VOT duration of voiced stops so that a comparison could be established for each phase between the learners and the native speakers. It is worth noting that the number of cases of the learner speakers depended on the errors they produced and,

therefore, varied depending on the consonant. As can be observed in both groups, the frequency values of F1 were very similar in labial and dental consonants and increased in velar consonants. However, for both labial consonants ($F[1, 159] = 48.247$, $p = 3.758\text{e-}12$ *) and dental consonants ($F[1, 180] = 32.412$, $p = 1.247\text{e-}08$ *), the values differed between the native group and the learner group, and the values for the native group were always lower. It is important to highlight the variation observed in the F1 values of /g/ in the case of the native speakers; however, the results between the two groups were still significant ($F[1, 150] = 6.1482$, $p = 0.013$ *).

|  | CH group | | | ES group | | |
|---|---|---|---|---|---|---|
|  | /b/ | /d/ | /g/ | /b/ | /d/ | /g/ |
| *N* | 100 | 124 | 98 | 60 | 57 | 53 |
| F1 | 265.3 | 265.8 | 307.7 | 220.5 | 231 | 323.6 |
| (Hz) | (55) | (49.6) | (92.4) | (26.3) | (34.1) | (276.6) |
| Hold | 105.7 | 104.8 | 101.6 | 70.3 | 64 | 64.4 |
| (ms) | (32.9) | (31.1) | (31.1) | (21.2) | (22.7) | (22.9) |
| Release | 12.2 | 13.7 | 23.9 | 11.8 | 14.5 | 25.5 |
| (ms) | (5.7) | (5.4) | (11.2) | (5.1) | (5.6) | (9.2) |

**Table 3.** Voiced stop realization: F1, duration of the hold phase, and duration of the release phase by speakers of both groups.

A significant difference was observed between the two groups in the duration of the segment corresponding to the phase in which the articulators are in contact: learners always produce a longer duration than native speakers (see Table 3).

The differences were significant for all three places of articulation: labial stops ($F[1, 159] = 43.878$, $p = 3.495\text{e-}11$ *), dental stops ($F[1, 180] = 60.09$, $p = 9.062\text{e-}15$ *), and velar stops ($F[1, 150] = 49.002$, $p = 2.557\text{e-}12$ *).

In terms of the duration of the release phase, the results were completely different from those mentioned above, as illustrated in Table 3. The duration was similar for both groups of speakers. The differences were not significant in any of the cases: for labial stops, $F[1, 145] = 0.042681$, $p =$

0.83; for dental stops, $F[1, 172] = 1.5141$, $p = 0.21$; and for velar stops, $F[1, 137] = 1.3718$, $p = 0.24$.

### 4.4. Voiceless stops

A total of 779 cases of voiceless stops were analysed, with 179 cases produced by the ES group and 600 cases produced by the CH group. All native speakers correctly produced these consonants.

### 4.4.1. Error categorization

The learner speakers presented very few errors for producing these consonants, only 12 cases were categorized as mistakes (2%), with four referring to the labial (33.4%), two to the dental (16.6%), and six to the velar (50%). In fact, only speakers M02

and M06 made mistakes, and they were all due to voicing.

### 4.4.2. The effect of attention

To carry out the statistical analysis based on the learners' focus of attention (i.e. on liquids versus stops), the two abovementioned speakers were selected. The results indicated no effect of attention in any of the three places of articulation: labial stops, $\chi^2(1) = 0$, $p = 1$; dental stops, $\chi^2(1) = 0.50505$, $p = 0.47$; and velar stops, $\chi^2(1) = 3.6597e-32$, $p = 1$.

### 4.4.3. Acoustical analysis

The only acoustic parameter analysed in these phonemes was the duration of the VOT since this consonant was in word-initial position. The realizations by the natives were compared to the correct realizations by the learners. The Kruskal-Wallis test indicated that there was a difference between the groups depending on the place of articulation; thus, while no significant differences were observed for labial stops ($F[1, 255] = 2.762$, $p = 0.09$), the results for dental stops ($F[1, 257] = 17.877$, $p = 2.356e-05$ *) and velar stops ($F[1, 253] = 33.702$, $p = 6.424e-09$ *) differed significantly between the native and learner groups (see Table 4).

| | /p/ | | /t/ | | /k/ | |
|---|---|---|---|---|---|---|
| | $N$ | VOT (ms) | $N$ | VOT (ms) | $N$ | VOT (ms) |
| CH group | 196 | 18.9 (17.5) | 198 | 19.4 (8.6) | 194 | 34.6 (14.1) |
| ES group | 60 | 22.1 (12.1) | 60 | 23.8 (10.4) | 60 | 53.6 (20.4) |

**Table 4.** Duration of the VOT in stop realizations of voiceless stops by speakers of both groups.

## 5. Discussion

The findings of this experiment provide further proof of the importance of focusing on phones in the target language and providing learners with proper education on the mechanisms of articulation in the case of new phonemes not found in their native language's phonic system. This instruction helps learners in being aware of the articulatory movements they must produce to articulate that phoneme (the concept of attention in Schmidt, 2001, 2010) and be able to articulate it naturally. Once learners have internalized this mechanism, whether or not they focus on a phoneme when producing a certain word should have no bearing on their ability to correctly produce these phonemes (Schmidt's concept of understanding). However, the fact that better results are obtained when there is an attention effect on the production of a phoneme may indicate that learners have not assimilated these articulatory motions required for accurate realization. In fact, it is important to differentiate between phonemes that learners are already familiar with because they exist in their native language and those that are new to

them (Loewen, 2015). In this regard, different results were obtained depending on whether the focus was on a new phoneme or a familiar phoneme. When learners were told to pay attention to the liquids, mistakes errors increased by 40% in the case of stop consonants, new phonemes. By contrast, in voiceless stop phonemes, which were shared by both languages, the error rate was just 2%, and these errors were observed in only two speakers. The same holds for liquid phonemes; the lateral was the phoneme that learners already knew in their mother tongue, while the tap was the new phoneme. The results for lateral consonants indicated that the learners produced 6% fewer errors when they paid attention to this phoneme; however, as mentioned in the results section, this difference was not significant. When it came to the new phoneme, however, there was a 17% increase in errors in the production of rhotics according to whether or not the learners focused their attention on the phoneme being produced.

Regarding the acoustic parameters and taking voiceless stops as a reference, we observed that the

native speakers did not display the VOT values reported by the different authors consulted (Castañeda, 1986; Asensi et al., 1997; Villamizar, 2002; Martínez Celdrán & Fernández Planas, 2007). The duration of bilabial stops ranged from 14 to 17.5 ms (22.2 ms in our data); for dental stops from 19 to 20.7 ms (23.8 ms in our data); and for velar stops from 32.2 to 35.4 (53.6 ms in our data). This finding may support researchers' claims that speakers learning other languages modify the acoustic characteristics of their native system due to crosslinguistic influence from the languages they are learning (Flege & Efting, 1987; Flege, 1995, among others). However, there is still a relationship between the three places of articulation, in both voiced and voiceless stops: the duration of the release phase is always shorter in labial stops and longer in velar stops.

With regard to liquid phonemes, although the rhotic was deemed a new phoneme and the lateral the familiar phoneme, the F2 values in learners were consistently much lower than in native speakers. It should also be mentioned that, as with the characteristics of stops in the native group, the F2 values for lateral phonemes (1818 Hz) deviated greatly from the data reported in the studies consulted: 1554 Hz for Quilis et al. (1979), 1514 Hz for Almeida & Dorta (1993) and 1474 Hz for Martínez Celdrán & Fernández Planas (2007). These data once again support the theory that native speakers may have modified the acoustic features of the phonemes shared with other languages, resulting in a merged category. With respect to both stops and liquids, the results obtained for native speakers were closer to the values reported in the articles referenced (see Table 2 for F2 of the lateral phoneme and Table 4 for the duration of the release phase in voiceless stops).

It should also be noted that there was a significant variation in the duration when new phonemes were produced; foreign language learners always produced consonants with a longer duration than native speakers. In this regard, Liu (2019) reported the same phenomenon when comparing native speakers and foreign language learners.

## 6. Conclusions

The main objective of this study was to determine whether learner speakers' focus of attention when reading a pseudoword modifies their realizations of Spanish stop and liquid phonemes. According to our findings, the phoneme on which learners focused their attention had a significant impact on their realizations. This task may facilitate the teaching of new phonemes to foreign language learners. However, the results also pointed to enormous intra-speaker variability and suggested that, despite having the same language proficiency, they may not have had the same ability to produce a sound naturally when not focusing their attention on it; some speakers are more likely to make mistakes when they do not focus on the phonemes being pronounced.

However, not all phonological categories were affected in the same way regarding the focus of attention. Here, only two categories were examined (liquids and stops), but stops had a larger percentage of errors than liquids. This may be due to the fact that the allophonic distribution of Spanish stops is more complex than liquids. Another possible explanation could be the allophonic variations of rhotics, which are also speaker-dependent, and all of them are correct in Spanish. In this regard, differences were also observed in the phonetic category since native Spanish speakers preferred approximant realizations, whereas learners favoured a stronger realization, which this study referred to as a flap.

The second objective of this study was to offer data on the acoustic parameters of the phonetic categories analysed. It is important to keep in mind that our comparison between the two groups applied only to realizations categorized as correct. Comparison of these parameters gave rise to varied results. First, even though the learner speaker's

correct realizations were categorized as canonical according to the spectrogram's characteristics, they had different values than those of native Spanish speakers. However, as mentioned above, the results of native Spanish speakers differed from those reported in the literature.

Except for voiceless stops and the lateral, when the two groups of speakers presented relatively similar values, the duration was always longer in the learners' group for all the acoustic parameters analysed. It is important to note that these were precisely the phonemes shared by the languages spoken by the learners, i.e. Chinese and Spanish, and hence phonemes with which the learners were already familiar. With respect to new phonemes, the learners felt the need to hyper-articulate them for a correct production, resulting in a longer duration. Thus, we might assume that learners use different strategies to assimilate the phonological system of a new language; when it comes to similar phonemes, they are less likely to exaggerate the values in the foreign language since they have already acquired them in their mother tongue.

In summary, this study confirmed that focusing attention on a certain phoneme improves production, regardless of whether or not speakers have lexical access, but that improvement may depend on the complexity of the phonological category or even each sound. Further research should explore other phonological categories to compare the production of new phonemes and similar phonemes between native speakers and learners. It would also be useful to continue analysing sounds produced by multilingual speakers, as the acoustic data collected from these analyses might differ from that gained in previous studies with monolingual speakers.

## Acknowledgements

## References

Almeida, M., & Dorta, J. (1993). Datos acústicos de las líquidas españolas. In C. Díaz Ayalón (Ed.) *Homenaje a José Pérez Vidal* (97-110). Litografía.

Asensi, L., Portolés, S., & del Río, A. (1997). Barra de explosión, VOT y frecuencia de las oclusivas sordas del castellano. *Estudios de Fonética Experimental*, *9*, 221-242.

Barquero, A. C. (2020). *Lengua, cultura, interculturalidad. El tratamiento de la competencia pragmática como parte de la competencia comunicativa en los libros de texto de ELE en el ámbito escolar alemán*. Universitätsverlag Potsdam.

Blecua, B. (2001). *Las vibrantes del español: Manifestaciones acústicas y procesos fonéticos* [PhD dissertation]. Universidad Autónoma de Barcelona.

Boersma, P., & Weenink, D. (2017). Praat: Doing phonetics by computer [Computer program]. Version 6.0.36. 5. Retrieved 27 April 2021 from http://www.praat.org.

Castañeda, M. L. (1986). El V.O.T de las oclusivas sordas y sonoras españolas. *Estudios de Fonética Experimental*, 2, 91-110.

Council of Europe Council for Cultural Co-operation Education Committee Modern Languages. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.

Cortés Moreno, M. (2014). Dificultades lingüísticas del español para los estudiantes sinohablanes y búsqueda de soluciones motivadoras. *SinoELE*, (10), 173–208.

Duanmu, S. (2007). *The Phonology of Standard Chinese* (2nd ed). Oxford University Press.

Flege, J., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, 15, 67-83.

Flege, J. (1995). Second language speech learning: theory, findings and problems. In W. Strange,

(Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research* (pp. 233-272). York Press.

Gil, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Arco/Libros.

Ishihara, N., & Cohen, A.D. (2010). *Teaching and Learning Pragmatics*. Longman.

Kang, O., Thomson, R., & Murphy, J. (Eds). (2017). *The Routledge Handbook of Contemporary English Pronunciation*. Routledge.

Keating, P. (1984). Phonetic and phonological representation of stop consonant voicing. *Language*, 60(*2*), 286-319.

Lemus Sandoval, J. E. (2002). VOT como rasgo distintivo de las consonantes oclusivas españolas e inglesas. *Científica*, 4, 49-58.

Lan, Y., & Wu, M. (2013). Application of Form-Focused Instruction in English Pronunciation: Examples from Mandarin Learners. *Creative Education*, 4(9), 29-34.

Liberman, A., Delattre, P. D., & Cooper, F. S. (1958). Some cues for the distinction between voiced and unvoiced stops in initial position. *Language and Speech*, 1(3), 153-167.

Liu, Z. (2019). *Análisis de las obstruyentes en chino y en español como L3. Estudio acústico y perceptivo para la categorización de errores* [PhD dissertation]. Universidad Autónoma de Barcelona.

Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge University Press.

Loewen, S. (2015). *Introduction to Instructed Second Language Acquisition*. Routledge.

Machuca, M. J. (1997). *Las obstruyentes no continuas del español: Relación entre las categorías fonéticas y fonológicas en habla espontánea* [PhD dissertation]. Universidad Autónoma de Barcelona.

Martínez Celdrán, E. (1993). La percepción categorial de /b-p/ en español basada en las diferencias de duración. *Estudios de Fonética Experimental*, 5, 223-239.

Martínez Celdrán, E., & Fernández Planas, A. M. (2007). *Manual de fonética española. Articulaciones y sonidos del español*. Ariel.

Quilis, A. (1981). *Fonética acústica de la lengua española*. Gredos.

Quilis, A. (1993). *Tratado de fonología y fonética españolas*. Gredos.

Quilis, A., Esgueva, M., Gutiérrez Araus, M. L., & Cantarero, M. (1979). Características acústicas de las consonantes laterales españolas. *Lingüística Española Actual*, 1(2), 233-344.

R CoreTeam. (2019). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*. https://www.R-project.org/.

Ruiz de Zarobe, L., & Ruiz de Zarobe, Y. (2013). *Enseñar hoy una lengua extranjera*. Octaedro.

Saito, K. (2014). Experienced teachers' perspectives on priorities for improved intelligible pronunciation: The case of Japanese learners of English. *International Journal of Applied Linguistics*, 24, 250–277.

Santamaría, E. (2015). Percepción y evaluación de la pronunciación del español como L2. *Revista Española de Lingüística, 45*(1), 175–207.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 3-32). Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Istanto, M. Nagami, J. W. Sew, T. Suthiwan, & I. Walker, *Proceedings of CLaSIC 2010*, Singapore, December 2-4 (pp. 721-737). National University of Singapore, Centre for Language Studies.

Shintani, N. (2016). *Input-based Tasks in Foreign Language Instruction for Young Learners*. John Benjamins.

Villamizar, T. (2002). Caracterización acústica de las consonantes oclusivas en el español de Venezuela. El parámetro duración (silencio-barra de explosión-VOT). *Lengua y Habla*, 7(1), 140-157.