Macrothink Institute™

International Journal of Linguistics
ISSN 1948-5425
2022, Vol. 14, No. 4

The Acquisition of Italian Stops by Igbo Speakers

Patrizia Sorianello

University of Bari

Dept DIRIUM, Palazzo Ateneo, Piazza Umberto I, 70121, Bari, Italy

E-mail: patrizia.sorianello@uniba.it

Received: June 30, 2022 Accepted: August 1, 2022 Published: August 18, 2022

doi:10.5296/ijl.v14i4.20047 URL: https://doi.org/10.5296/ijl.v14i4.20047

Abstract

This research explored the acquisition of Italian singleton stops by Igbo speakers. The voiceless and voiced stop phonemes /p t k/ and /b d g/ are present in both Italian and in the Igbo languages. This creates a perfect condition to check the learning strategies implemented in Italian L2. The theoretical framework taken up in this study was the *Speech Learning Model*. According to this approach, the sounds of L2 can be perceived by learners as the same, similar or as new compared to those present in their L1, depending on the greater or lesser acoustic and perceptive distance established between them. A spectro-acoustic analysis achieved on the stops produced by Igbo learners of Italian showed the presence of sound categories with a different acoustic structure. In Italian L2 stops have undergone an important phonetic adaptation that makes them perceptually different from what is achieved by native Italians.

Keywords: VOT, Stop consonants, Italian L2, Acquisition, Igbo language

1. Introduction

Phonology is a very resistant component that involves a high risk of fossilisation even when learners achieve a high degree of competence in a language as L2. Adults encounter many difficulties in the production and perception of those sounds that are absent in their native language or have a similar articulation. Regarding this aspect, one of the most important variables is the age at which the first exposure to an L2 occurs; the formation of phonological categories in L2 has, in fact, an optimal development when the acquisitional process starts before the pubertal age, as it is favoured by greater neuronal plasticity (Lenneberg, 1967). The perception and production of sounds are the cardinal processes on which the main theoretical models are based. The success of the acquisition process only occurs when the learner can discriminate sounds on a perceptual basis and therefore he/she can represent the new categories by abstraction. The perceptual process precedes the production of sounds and

becomes responsible for the final outcome. Accordingly, when the perceptual classification of a sound fails, which is a condition that commonly occurs when two sounds are similar between each other, the mechanism of sound production will also fail. This happens frequently in two macro conditions: (1) the phonological systems of L1 and L2 present categories that are not equivalent in type or in number, and (2) the categories involved show phonological but not phonetic correspondence.

The *Speech Learning Model* (SLM) was adopted to interpret the results of our research since it fits better to the characteristics of the experimental situation examined. The model was theorised by James Emil Flege (1987, 1995, 1997, 2003); among the various conditions, it considers the speech production of adult learners, including immigrants, who have lived longer in a foreign country, determining a context of high permanence in the L2 territory and a low L1 use. *Length of Residence* (LOR), *Age of Arrival*, (AOA) and *Age of Learning* (AOL), in all their possible interactions, were also taken into consideration and have been shown to have a relevant role in defining the learner's profile. These aspects are, in fact, correlated: the phonological competence will be better when the AOA and the AOL indexes are low and the LOR index is high. The SLM assumes that the formation of new phonetic categories can also occur in the adult age (Flege, 2003; Flege *et al.*, 2003); however, if the first exposure to L2 (AOL) occurs in the post-pubertal age, the degree of conditioning of the phonological structures of the L1 will be greater, as the latter acts as a filter, altering the processing of sounds. In this complex dynamics, the development of the categories is also influenced by the level of similarity/dissimilarity with which sounds of L1 are placed with respect to those of L2. The range of possibilities is diverse and can be described by three different situations: i) two languages share the same phonological categories that also show the same phonetic characteristics (condition of equality), ii) two languages share the same phonological categories, but are differently produced on a phonetic level, iii) two languages present different phonological categories.

Sounds that are new or dissimilar to those present in the native language are easier to discriminate and to produce and they favour the development of new phonological categories, limiting the action of transfers and favouring the acquisition of a native pronunciation. Otherwise, the most complex situations for the learner are those in which the two languages (L1 and L2) share similar sounds that, just for their degree of similarity, are often the reason of a perceptual failure. In this case, transfer mechanisms are activated in order to bring the similar sounds of L2 in the closer phonological category of L1; the persistence of this condition over time increases the risk of fossilisation. According to Flege (2003), this determines an equivalence classification process, that is a condition in which the same phonological category is used to produce both a sound of L1 and of L2. In other words, similar sounds would be produced in L2 by exploiting the same articulatory modalities of native sounds, a mechanism through which several articulatory substitution processes arise that slow down the achievement of the phonological target. At the beginning, the phonetic categories of the two languages are positioned in a common phonological space and influence each other. Over time, the increase in exposure to L2 can cause both a *Category Assimilation* (CA) or a *Category Dissimilation* (CD). The first occurs when a sound of L2, not being sufficiently distant from that of L1, is brought back to the latter; in this intermediate category

(*merged category*) some acoustic features of the sound of L2 coexist with those of the sound of L1. This condition has a positive outcome over time, following a period of fossilisation, if the exposure to the language continues as new categorical formations will, in fact, be developed. On the other hand, CD occurs when a sound of L2 is perceptually discriminated, as characterised by a significant phonetic distance from the native sound and categorised as a new or dissimilar sound (Flege, 1995). However, the development of the new category is not immediate, given that it arises slowly. The greater the degree of phonetic dissimilarity between the sound of L2 and the closer one of L1, the greater the probability that a new category will be created (Flege, 2003). The international literature on the acquisition of L2 phonology is very wide. Nevertheless, to the best of our knowledge, studies focused on the acquisition of Italian by Igbo-speaking learners are still lacking. More specifically, in this research we focused on the phonetic production of voiceless and voiced stops, which are one of the most prototypical natural class of sounds among the languages of the world.

1.1 The Voice Onset Time (VOT)

Stop sounds represent a universal manner of articulation; all 451 languages registered in the UPSID (*UCLA Phonological Segment Inventory Database*) present, in fact, this class of sounds (Maddieson, 1984; Henton *et al.*, 1992). Looking at the UPSID data, the most frequent stop is /k/ (89.6%) followed by /p/ (83.15%) and /t/, (40.13%) whereas /b/ (63.64%) is more frequent than /d/ (26.61%) and /g/ (56.10%) (Note 1 for more details).

The production of stops shows a precise dynamics: articulators establish a momentary blocking of the airflow in the vocal tract; this first stage, defined occlusion, is followed by a moment of articulatory release (explosion) induced by the rapid opening of the air passage. From an acoustical point of view, the moment of closure corresponds to the absence of a spectrographic noise in voiceless stops, while only a voice bar is present for voiced stops; the release starts with an abrupt vertical noise (burst) and gives rise to the Voice Onset Time (VOT).

The VOT was first identified in 1964 by Lisker and Abramson (1964). It coincides with the time elapsed between the articulatory release of the stop closure and the onset of the following sound. Acoustically-speaking, the VOT displays a short fricative aperiodic noise, which varies in length, intensity and spectral noise distribution.

The VOT represents one of the most important acoustic cues used by international literature to classify stop consonants. It allows to distinguish stops by place of articulation, voicing and degree of aspiration. Many studies have shown that voiceless stops have a longer duration and a more intense fricative noise than voiced stops. The place of articulation plays an important role on VOT's length: the VOT is, in fact, shorter in bilabial stops and progressively longer in alveolar and velar ones. This aspect, which has been subject of debate over the years, has been attributed to various physiological and supra-laryngeal mechanisms (see Cho & Ladefoged, 1999; Cho *et al.*, 2019). The VOT creates a continuum of phonetic and/or phonological contrasts. For this reason, the length of the VOT allows to distinguish not only voiceless stops from voiced ones, but also aspirated from unaspirated stops. According to Lisker and Abramson (1964, 1967), there are three VOT categories, depending on whether the vibration of the vocal folds begins before or after the release burst of the stop;

by convention, the latter represents the zero point (0 ms) to use as reference for measuring VOTs, so the VOT can have a negative or a positive value. More specifically, the following VOT categories can be distinguished: 1) *voicing-lead*: the vocal folds vibrate before the burst and therefore the VOT has a negative value; in many languages this configuration represents voiced stops; 2) *short-voicing lag*: the vocal folds vibrate after the burst, the VOT shows positive values, ranging from 0 to 25 ms, which is the typical configuration of voiceless unaspirated stops; 3) *long-voicing lag*: vocal folds vibrate following a long time interval, the VOT has positive values, ranging from 60 to 100 ms, which is the phonetic configuration of voiceless aspirated stops.

Many of the world's languages, such as Italian or French, only distinguish two phonemic categories of stops, voiced and voiceless, but in other languages, such as Korean or Bengali, aspirated stops establish a phonological contrast with the unaspirated ones. In this regard, Cho & Ladefoged (1999), who conducted a cross-linguistic study based on 18 languages, assumed that 'there is a continuum of possible VOTs from which languages may choose' (p. 226). In line with this assumption, they identified four different phonetic categories: 1) unaspirated stops: the VOT is between 0 and 30 ms (e.g. Italian, Spanish), ii) slightly aspirated stops: the VOT oscillates between 30 and 60 ms (e.g. English), iii) aspirated stops: the VOT has a duration between 60 and 90 ms (e.g. Apache), and iv) strongly aspirated stops: the VOT lasts over 100 ms (e.g. Navajo). Stops with a longer-lag VOT have a clear phase of aspiration characterised by a louder intensity and a greater subglottal pressure.

The VOT has been extensively studied, from a cross-linguistic perspective, in both the dimension of production and perception. From an acquisitional perspective, the VOT is a resistant speech cue, especially when the stops of a L2 language show different features compared to those of the L1 learners. Numerous studies have been carried out in this direction (among others, Flege & Eefting, 1987; Major, 1987; Flege, 1991; Artimonte Rocca & Marcelino, 1999; Chen *et al*., 2007; Nagle, 2017; Holliday, 2019). More in general, research has confirmed that the VOT sets phonetic contrasts between the native and the second language, causing the confluence of L1 and L2 stops within the same category. Usually, the VOT displays an intermediate duration between that of L1 and that of L2, which is an articulatory strategy that tends to reduce the difference of the consonants involved. Therefore, the process of perceptual identification of stops fails, which happens when L2 plosives are assimilated to these of the native language. With explicit reference to Italian as L2, the data collected thus far has provided further evidence in this direction. Speech samples of stops produced by eight Arabic speakers, eight Italians and eight Arabic-speaking learners were analysed by Mori and Barkat-Defrasas (2005) and Mori (2007). As demonstrated by the aforementioned authors, the acoustic measurements showed that in learners the VOT was higher than in native Italians, but lower compared to that of the Arabic language. In German learners of L2 Italian, voiced stops were shown to be achieved without burst and voice bar, while voiceless ones were characterised by a longer VOT than what was found for Italian plosives, demonstrating the presence of aspiration. In addition, the phonetic features of German stops tended to persist in the interlanguage even when the learning process progressively reached higher stages (Sorianello, 2008).

Our research shows two innovative points compared to most of the studies conducted on the production of stops in Italian as L2: first of all, it focused on adult speakers who learned Italian outside a school context, which is a specific learning condition supported by unavoidable motivations for social and work inclusion, but also strongly influenced by the local environment. Secondly, the production of Italian stops by Igbo learners has never been investigated thus far, to the best of our knowledge. Although the two languages show genealogical and typological distances, they share the same set of stop phonemes, giving rise to multiple possible learning scenarios, as explained later in the text (§ 2.1).

1.2 Stops in Italian and in the Igbo Languages

Nigeria presents a great linguistic variability; the official language is English, which is taught in schools and used throughout the nation with a vehicular function, often in its pidgin form, with different accents and pronunciations. Three major languages, namely Igbo, Housa and Yoruba, are also spoken in Nigeria as well as a great number of minor languages and dialects for a total of 400. Igbo is a tonal language spoken in a large area of South-Eastern Nigeria by about eighteen million people and characterised by a large linguistic fragmentation with a great deal of mutual intelligibility (Iloene, 2007; Nwaozuzu, 2008). For this reason, the literature generally refers to the standard Igbo. Considered a sort of fusion between the central Igbo and the Onitsha Igbo, it cannot be identified by a specific geographical setting, as it spreads over a vast territory called Igboland that includes the Eastern part of Nigeria and a large part of Southern Nigeria (Ikekeonwu, 1987). The phonemic inventory of standard Igbo consists of twenty-eight consonants and eight vowels (Ikekeonwu, 1991). Regarding stops, it presents /p b/, /t d/, /k g/ but also labial-velars /k͡p g͡b/ and labialised velar stops /kw gw/. Although standard Igbo does not establish a phonological contrast between aspirated and unaspirated stops, in some dialectal varieties, including Oshuoere, Umuchu, Isukwuato and Aku, the process of aspiration seems to have a contrastive value (Nkamigbo, 2014; refer to Note 2 for more details).

In standard Italian there are the following voiced and voiceless stops /p t k b d g/; they are always produced as unaspirated phonemes and can have a singleton or a geminate achievement, e.g. *note/notte* (notes/night), *fioco/fiocco* (dim/flake). In Italian voiced stops, the VOT has a very short duration, ranging on average between 4 to 12 ms (Landi, 1997), while in voiceless stops the VOT values range between 12 and 34 ms (Vagges *et al.*, 1978; Bortolini *et al.*, 1995; Landi, 1997; see Note 3). In addition, in read speech both voiceless and voiced stops show a high percentage of a release burst (from 78% to 99%), whereas in conversation the rates of burst and spike are lower and very often absent (Vagges *et al.*, 1978; Cerrato & Falcone, 1997; Landi, 1997; refer to Note 4).

2. The Research

This study focused on the production of Italian stops by Igbo speakers. A spectro-acoustic analysis was performed in order to identify any transfer phenomena between L1 and L2. To this end, voiced and voiceless stops were compared in native Italian and in Italian as a second language.

2.1 Research Hypothesis

Italian and the Igbo languages share the /p t k/ and /b d g/ stops, although, as mentioned earlier, in the latter the series of plosives is wider in number. Two different behaviours can arise, depending on whether or not the phonological categories share the same acoustic properties. In the first case (both phonetic and phonological identity), stops will be perceived by Igbo speakers as equal sounds; consequently, the production of stops between It-L2 and It-L1 will be equals on the phonetic and perceptual level. In the second case (phonological identity), stops of L1 and L2 will be characterised by partially different acoustic traits and will be classified as similar sounds, although they are coincident on the phonemic level; the presence of an insufficient phonetic distance will determine a perceptual failure leading to the activation of a transfer and to the production of stops with a phonetic structure that only partially coincides with that of the native Italian language.

2.2 Participants

Four male Nigerian speakers, aged between 26 and 32 (mean: 27.5, sd: 2.5) (hereinafter It-L2), took part in the research. Three of them attended lower secondary school in their country of origin, while only one attended upper secondary school. They currently live and work as agricultural workers in Acquaviva delle Fonti near Bari (about 30 km), the capital of the Apulian region and the most important city in South-Eastern Italy. The individuals came from the Southern area of Nigeria between Delta State and Port Harcourt and declared to have Igbo as first language and to be fluent in English. Since their arrival in Italy, they lived for six months at the Reception Centre for Asylum Seekers in Bari Palese and then moved to the urban area. They are currently attending an Italian language course aimed for migrants. With reference to the SLM indices, the participants were classified with a late *Age of Arrival* and a late *Age of Learning*; the age of the first exposure to Italian and the age of arrival did not coincide, the acquisition process, in fact, started after puberty. Moreover, the *Length of Residence* was relatively low (about 4 years), although it was reinforced by a high exposure to Italian (high-L2 use) and, consequently, by a low-L1 use. These subjects learned Italian mainly spontaneously, outside of a school context; they were exposed early to Italian for the obvious reasons of social and work integration, although they tended to live in small groups made up of compatriots. A control group (hereinafter It-L1), formed by four monolingual native Italian subjects, who were born and were based in Acquaviva delle Fonti (province of Bari) was also considered; they were young male workers, aged between 25 and 31 (mean: 26.5, sd: 2.5), with a high school diploma.

2.3 Method

The subjects (It-L2 and It-L1 groups) were invited to read a corpus made up of 54 target words inserted in as many carrier sentences. We selected high-frequency disyllabic and trisyllabic words, each containing intervocalic singleton stop consonants placed at the onset of the stressed syllable. All participants read the list of sentences in a random order in a silent room, unaware of the purpose of the research; they were audio-recorded with a high-quality digital instrument while reading the list of sentences (sampling frequency 44.100 Hz, 32 bit, .wav format). All subjects gave informed written consent to participate. Nigerian speakers

also produced sentences in their native language (I-L1); in this case, the selection of words happened with the help of the learners themselves. For three Igbo speakers, and in parallel for three control subjects, a sample of spontaneous speech was also collected for future comparisons. The data discussed here only concerned the read speech sample in both It-L2 and It-L1. The annotations and spectro-acoustic measurements were manually performed using the PRAAT software (Boersma & Weenink, 2015). All sentences were segmented and annotated by means of text grids. Overall, a corpus of 432 stops was analysed (54 stop stimuli x 4 participants x 2 groups). The acoustic structure of stops was analysed in order to assess the correspondence between the phonetic and the phonological production, paying particular attention to the presence/absence of the voice bar, burst and the VOT. Stops were classified evaluating both the acoustic waveform and the wideband spectrograms. At the same time, the following parameters were measured: 1) stop total duration, 2) stop closure duration, 3) VOT duration (ms), 4) ratio between the VOT and the stop total durations (VOT/C).

Consonant closure duration was identified as the temporal interval between the last periodic wave of the previous vowel up to the release burst; the VOT was instead measured from the beginning of the release burst up to the first visible periodic vibration of the following vowel (see Figure 1). In the duration of the VOT we included both the spike and the following frication noise, given the difficulty to measure these two adjacent acoustic moments separately. Positive VOT values were found for voiceless stops and for most of the voiced stops.

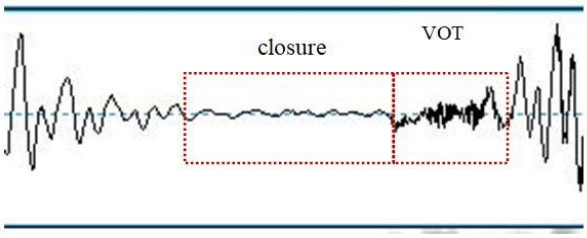

Figure 1. Acoustic representation of the closure and the VOT in voiceless stops

As expected, the speech of the L2 speaking subjects was slower than for L1 natives; for this reason, all duration measurements were normalised to speaker speech rate (duration/speaker speech rate; Tran *et al.*, 2017). This procedure cancelled the conditioning carried out by the speech rate, allowing to grasp the actual differences concerning the duration of the stops (Note 5). The mean and standard deviation were calculated for each acoustic cue. Data was exported to SPSS for statistical surveys; a one-way Anova with Tukey's post-hoc was performed. The significance level was set at 0.05.

2.4 Results

Overall, the stops produced in It-L2 did not always correspond to what was expected since they underwent some substitution processes. More in detail, shifts affected: 1) the place and manner of the articulation (2.2%) (e.g. /t/ > [ts]); 2) the voicing (3.5%) (e.g. /d/ > [t]), and 3) the length (2.3%) (e.g. /t/ > [tt]). Moreover, voiced stops were unreleased in 6.2% of cases

(see Note 6 for details). In It-L1 the cases of mismatch between the phonological target and its phonetic achievement concerned firstly voiced stops; more specifically, /b/ and /d/ were often pronounced without a clear release phase (22%), while /g/ (44%) and /d/ (18%) presented an approximant achievement. These cases were considered separately.

When conducting a more in-depth analysis, we observed that even when there were no articulatory shifts compared to the phonological target of the It-L1, L2 stops showed clear alterations that made them perceptually different from those of natives. This effect was the result of a longer duration, but also of a different acoustic structure, consisting of a stronger burst and a higher amplitude and length of the VOT. The stops produced in It-L2 were clearly longer than those of natives (It-L1) and all the differences were statistically significant [p=.000]; see the normalised durations depicted in Figure 2 and the spectrographic representation in Figures 3 and 4. In fact, in Igbo speakers the value of the VOT was clearly longer, about twice if compared with that of natives, while the duration of the closure in voiceless stops was more than double that in the native sounds.

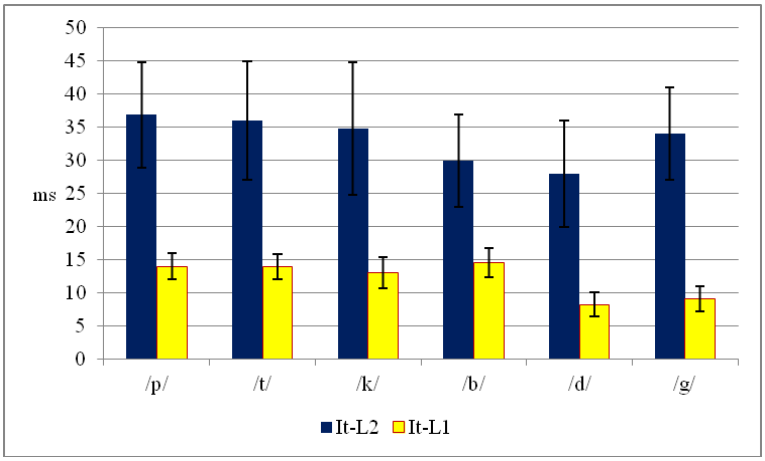

Figure 2. Mean normalised total duration (in ms) of stops in It-L1 and It-L2 (the error bars represent standard deviation values)

Moreover, the duration was on average triple, with the only exception of [b], which showed a lower rate of prolongation; in both groups, the shortest stops was [d]. In It-L2, durations also showed a greater dispersion, as documented by the higher values of the standard deviation. All voiceless stops were longer if compared to those produced by Italian natives [F(5,256)= 51.284, p=.000]. Nevertheless, the comparisons within groups never reached statistical significance. More in detail, in It-L2 the duration of /p/ vs. /t/ and /k/ did not show relevant differences ([p=.998] and [p=1.000], respectively); in the same way, no comparison reached the statistical significance threshold (/p/ vs. /t/, p=.946; /p/ vs. /k/, p=.976; /t/ vs. /k/, p=.1.000) in the native group.

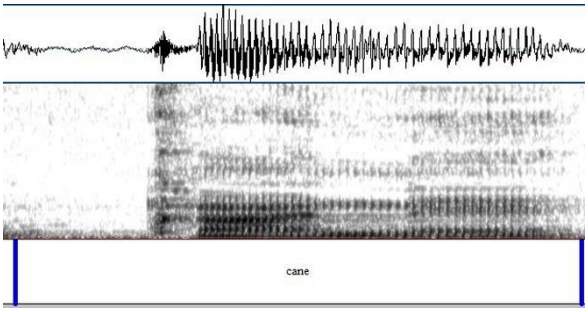

Figure 3. Acoustic waveform, wideband spectrogram and orthographic tier of the word *cane* (dog) produced by the learner It-L2_S; stop closure: 127 ms, VOT: 63 ms

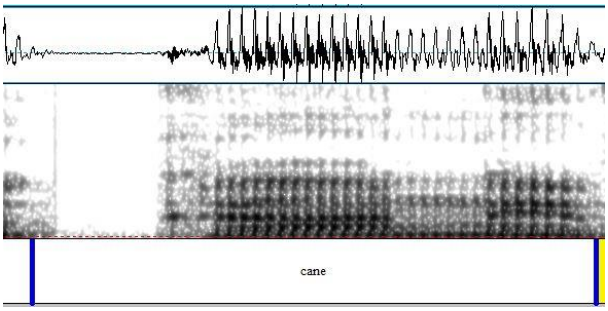

Figure 4. Acoustic waveform, wideband spectrogram and orthographic tier of the word *cane* (dog) produced by the Italian native It-L1_E; stop closure: 62 ms, VOT: 32 ms

As for voiced stops, /b/ and /d/ did not show significant differences [p=.959] in It-L2, while /g/ was longer than bilabial and dental voiced stops [p=0.38]. In the It-L1 group /b/ figured with the longest stop. This result can be explained by the reinforced pronunciation of /b/ in the intervocalic context or before a liquid consonant, which is a typical phonetic pronunciation of all varieties of Central-Southern Italian, for instance *abito* > ['abbito] dress, *cubo* > ['kubbo] cube. No significant differences were found for /d/ and /g/ [p=.953].

In the two groups considered (It-L2, It-L1), the duration of voiceless stops was scarcely affected by the place of articulation. No important differences were, in fact, found in It-L2 concerning /b/ vs. /d/ [p=.092] and /d/ vs. /g/ [p=.067], but only for /b/ vs. /g/ [p=.000]; in It-L1 the comparison between /d/ and /g/ was not remarkable [p=.804].

The analysis of the total duration of stops did not allow to capture the relationship between the duration of closure and the VOT. In It-L2 the presence of longer VOTs contributed to increase the overall length of stops. In Igbo speakers, in fact, all stops presented longer VOTs than those found in natives (refer to Figures 5 and 6 for more details).

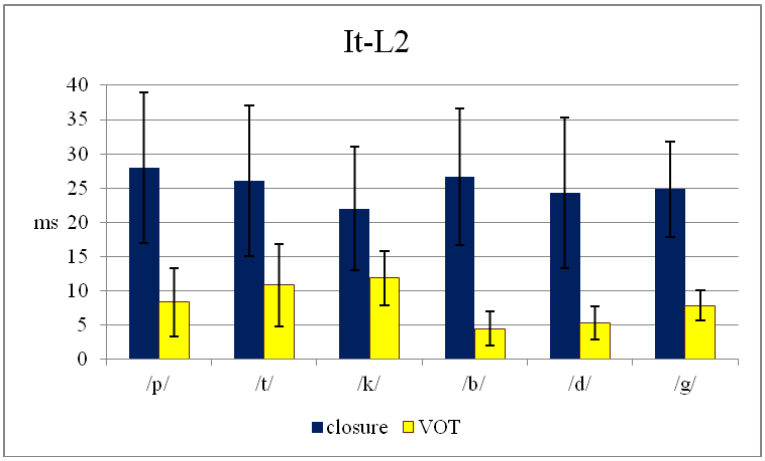

Figure 5. Mean normalised duration of closure and the VOT (in ms) for It-L2 individuals (the error bars represent standard deviation values)

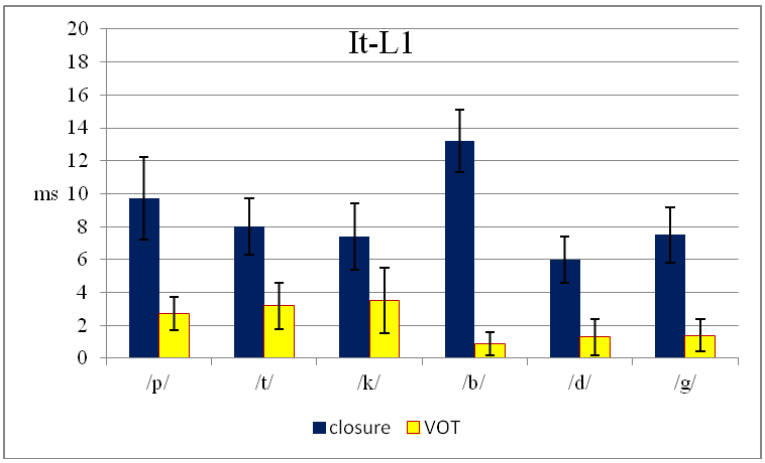

Figure 6. Mean normalised duration of closure and the VOT (in ms) for It-L1 natives (the error bars represent standard deviation values)

The differences were significant for both voiceless [F(2,252)=57.902, p=.000] and voiced stops [F(2,104)=51.653, p=.000]. As expected, in the two speech samples the VOT was shown to be conditioned by the place of articulation, being shorter in bilabial stops and progressively longer in dental and velar ones. Nevertheless, Tukey *post hoc* tests did not always reach statistical significance. As far as voiceless stops are concerned, in It-L2 the comparisons between /p/ and /t/ [p=.008] and /p/ and /k/, respectively, were significant, while the difference between /t/ and /k/ was above the limit of statistical significance [p=.067]. In the It-L1 group the comparison between the /t/ and /k/ VOT duration was not statistically significant [p=.804]. The comparison between groups showed that the voiceless stops produced in It-L2 were always significantly different from those of It-L1 [p=.000]. The same trend was also found in the sub-sample of voiced stops: in It-L2 the VOT was longer than that observed for the It-L1 [F(2,116)=18,842, p=.000]. The *post hoc* tests showed that all comparisons were significant, except for the one between /b/ and /d/ [p=.777] in the It-L2 group and between /b/ and /g/ [p=.998] in the It-L1 group (see also the spectrographic representations depicted in Figures 7 and 8).

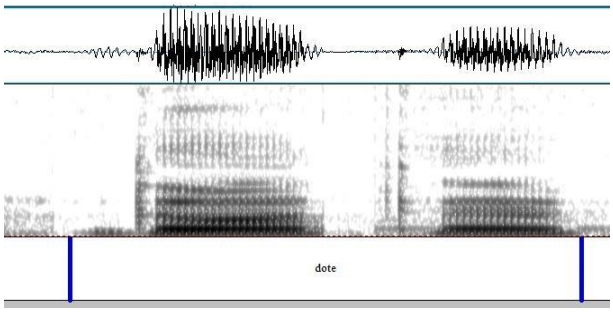

Figure 7. Acoustic waveform, wideband spectrogram and orthographic tier of the word *dote* (dowry) produced by the learner It-L2_O; stop closure of /d/: 52 ms, VOT: 20 ms

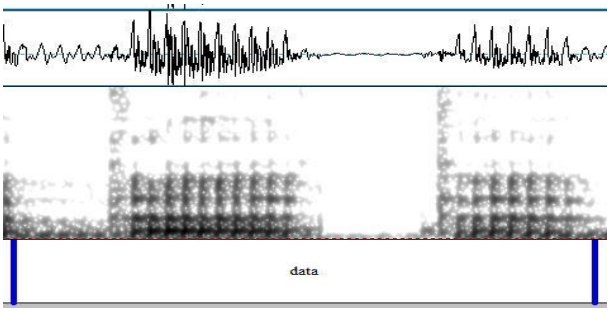

Figure 8. Acoustic waveform, wideband spectrogram and orthographic tier of the word *data* (date) produced by the learner It-L1_E; stop closure of /d/: 39 ms, VOT: 10 ms

In order to achieve a more detailed picture, the mean normalised ratio between the VOT duration and the absolute stop duration was also computed. The results obtained for the two groups show that the mean ratio was also different ([F(2,256)=1.912, p=.000]; see Figure 9).

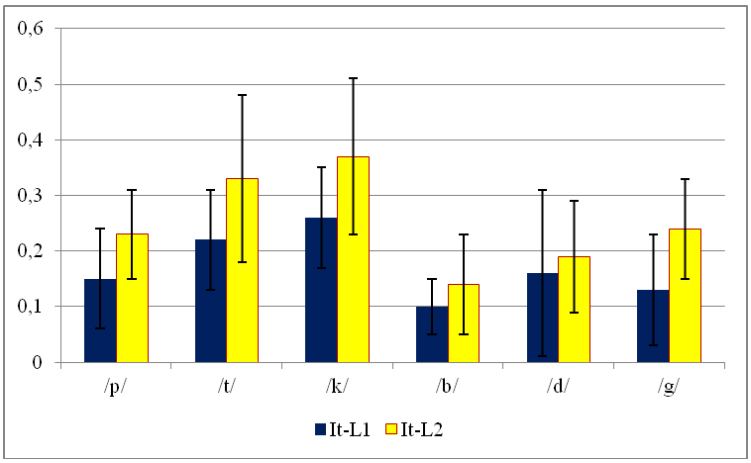

Figure 9. Mean ratio between normalised VOT and the stop total length in It-L2 and It-L1 (the error bars represent standard deviation values)

Overall, the mean ratio was progressively higher the greater the VOT length. It figures as a reliable parameter since it provides the percentage of time occupied by the VOT with respect to the consonant total duration. For the voiceless stops, all the comparisons within and among groups were significant, with the only exception of the normalised ratio between /t/ vs. /k/ both in It-L1 and in It-L2 ([p=.315] and [p=.293], respectively), whereas for the voiced

consonants the comparisons across groups were always significant. In contrast, the comparison within groups was weak, the ratios /b/ vs. /g/ in It-L1 and /d/ vs. /g/ in It-L2 were, in fact, non-significant ([p=.999] and [p=1.00], respectively). In more general terms, in native Italian the VOT occupied around a quarter of the stop duration in /t/ and /k/; in the remaining stops, the temporal degree occupied by the VOT was lower, while in /b/ and /g/ it was equals only to 10-15% of the total length. In It-L2 a different trend was observed for both voiceless and voiced stops. Regarding voiceless stops, the VOT/C ratio was always higher and accounted for over 30% of the consonant total length in /t/ and /k/. As far as voiced stops, the VOT/C ratio was almost 20% in /d/ and /g/, whereas in both groups /b/ showed the lowest ratio. All differences between It-L1 and It-L2 were statistically significant, with the only exceptions of /b/ [p=0.07] and /d/ [p=0.69].

In addition to the duration ratios, it should be noted that in It-L2 the VOT, especially in voiceless stops, was characterised by a clear and very intense fricative noise (see Figure 10), in contrast to native stops where the fricative noise presented a diffuse nature and a lower intensity.

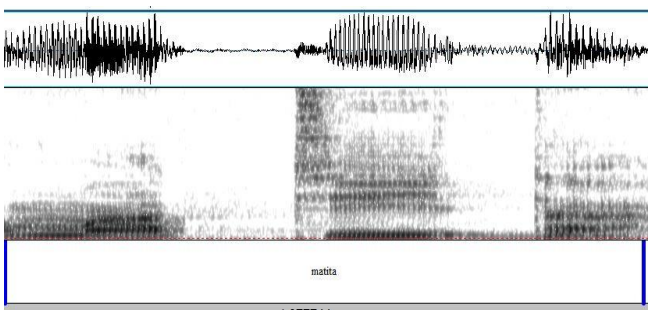

Figure 10. Acoustic waveform, wideband spectrogram and orthographic tier of the word *matita* (pencil) produced by the learner It-L2-A; stop closure of the first /t/: 160 ms, VOT: 50 ms

This last aspect was confirmed by a preliminary analysis, still in progress, which was carried out by extracting the average values of the amplitude (burst amplitude) and of the spectral form (spectral moments) obtained in correspondence with the median section of the release of the stops.

3. Discussion

The experimental findings highlighted that the Italian stops produced by Igbo-speaking learners differed from those of Italian natives. Although the phonological categories of voiceless and voiced stops are present in both languages, Igbo learners were shown to process Italian stops in a different way. In It-L2 stop consonants, especially voiceless ones, had a longer closure duration and a longer and more intense VOT. The learners produced most Italian voiceless stops with a long-lag VOT and voiced stops with short-lag VOT values. This suggests that, in It-L2, voiceless stops were mostly achieved as aspirated sounds. Consider that in standard Igbo, aspirated stops have not been reported. However, as mentioned before, standard Igbo is a language label that contains many varieties, some of which have shown the presence of aspirated stops. Another point that should be taken into consideration is that,

during their schooling, the participants of our research learned English, which may be considered a sort of 'lingua franca' for them, i.e. a language used as a medium of communication by individuals that have different native languages and, therefore, that is employed most of the time with a vehicular function. For this reason, Italian represented a third language for Igbo speakers, also with respect to the chronological order of learning. Thus, the production of Italian stops was the combined result of the overlap between the phonetic features of their variety of Igbo (L1) and the variety of English that they were able to manage (L2).

The results are in line with the predictions made of the SLM. The situation described above generates a complex dynamics of acquisition in which it is plausible to assume the presence of a classification of equivalence. A single phonological category in native and non-native languages, in our case stops, creates different phonetic categories whose sounds, which are equivalent on the abstract level, show instead distinct spectro-acoustic features in phonetic substance. It is a condition in which there is no phonetic correspondence between the sounds that belong to the same phonological category. This generates a condition of perceptual assimilation of sounds, given that L1 and L2 stops are similar in production, but not equals between each other. According to the data collected, Igbo learners produced aspirated stops in It-L2; the VOT values of voiceless and voiced stops were significantly longer than those of natives. With reference to the classification of Cho & Ladefoged (1999), voiceless but not voiced stops were considered slightly aspirated, as their VOT ranked between 30 and 60 ms. By true, the production was quite varied, even the same learner showed short and significantly long VOTs, at the same time; and there were several cases of /t/ and /k/ VOTs over 60 ms. On the other hand, the stops produced by Italian natives were never aspirated. By comparing It-L1 and It-L2, it emerged that stops were clearly different in terms of phonetic structure and perceptual salience. This was particularly clear for voiceless stops, while for voiced stops the picture was less clear. The latter, in fact, could not be considered aspirated consonants, since their VOT values were always below 30 ms. However, also in this case, the VOT figured as a differentiating element, given that its absolute duration was on average twice than that of Italian voiced stops. In addition, the VOT was accompanied by a higher percentage of bursts, an aspect that indicates the presence of a greater degree of muscular and articulatory effort. Therefore, Igbo speakers appeared to transfer the phonetic traits present in their L1 stops into Italian generating a condition of similarity between sounds, which is a very common condition in the interlanguage of learners and a frequent cause of slowdown of the learning progress. The Italian stops were, in fact, not perceived as different sounds from those of their language, consequently they converged in the category of L1. This situation was also confirmed by the fact that the substitution processes detected in It-L2 were quite limited, which means that Igbo speakers identified the phonological category of stop sounds and reproduced the salient features, such as voicing, manner and place of articulation. The differences concerned firstly the relationship between the duration of stops and of release, and secondly the spectrographic structure of the VOT that, in addition to being longer, also showed a relevant frication noise. This suggests the presence of stops, which were produced with a greater articulatory force and, overall, with a greater degree of tenseness. However, one aspect cannot be ignored: the sentence-reading task may have led Nigerians to adopt a

controlled pronunciation of carrier sentences with consequent emphasis of some phonetic features. Therefore, in the future, the comparison with the semi-spontaneous speech corpus could turn useful to ascertain the strategies taking place during the learning process.

4. Conclusions

The comparative analysis showed that voiced and voiceless stops produced by Igbo speakers are different from those of Italian natives. More specifically, in It-L2 voiced stops have higher VOT values, which are almost always accompanied by a visible spike, while voiceless stops have significantly longer VOTs and a more intense aspiration noise. In the two languages the stop phonemes do not correspond to each other, determining the activation of transfer processes and the forming of similar, but not identical, phonetic categories in It-L2. The presence of voiceless aspirated stops represents a specific pronunciation trait in Igbo learners, a transfer of their native language, as it does not belong to the Italian Apulian territory in which they live. This research, which is the first focused on the acoustic realisation of stops by Igbo learners of L2 Italian, confirms the distinctive value of the VOT. Although promising, the research presented here is incomplete and more detailed studies should be conducted. For instance, the sample of participants should be expanded as well as the dataset of the phonemes analysed. In addition, the analysis of the spectral components of the VOT or the perceptual weight of the phonetic features of the It-L2 stops should be promoted and strongly advised in future research.

Acknowledgments

The author would like to thank the It-L1 and It-L2 participants who kindly took part in this experimental research and Arcangelo Martielli who contacted and recorded the Igbo speakers.

References

Anyasi, B. N., Babarinde, O., & Iloene, G. (2020). Acoustic analysis of obstruent in some Igbo dialects. *Theory and Practice in Language Studies*, *10*(12), 1510-1527. https://doi.org/10.17507/tpls.1012.02

Artimonte Rocca, P., & Marcelino, M., (1999). Some characteristics of VOT in plosive produced by speakers of English and Portuguese. In J. J. Ohala., Hasegawa, Y., Ohala, M., Granville, D. & Bailey, A. C. (Eds.), *Proceedings of the 14th International Congress of Phonetic Sciences* (1999, vol. 3, pp. 1425-1428). University of California, San Francisco.

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer* [Computer program]. Version 5.4.09. Retrieved 1 June 2015, from http://www.praat.org/

Bortolini, U., Zmarich, C., Fior, R., & Bonifacio, S. (1995). Word-initial voicing in the productions of stops in normal and preterm Italian infants. *International Journal of Pediatric Otorhinolaryngology*, *31*, 191-206. https://doi.org/10.1016/0165-5876(94)01091-B

Cerrato, L., & Falcone, M., (1997). Il burst nelle occlusive in sequenze VCV e VC:V dell'italiano: Un'analisi acustica. In P. M. Bertinetto, & Cioni L. (Eds.), *Unità fonetiche e fonologiche: produzione e percezione* (pp. 29-40). Atti delle VIII Giornate di Studio del

Gruppo di Fonetica Sperimentale. Scuola Normale Superiore di Pisa, Pisa: Stamperia SNS.

Chen, L. M, Chao, K. Y., & Peng, J. F. (2007). VOT productions of word-initial stops in Mandarin and English: a cross-language study. In *ROCLING 2007 Poster Papers*, 303-317.

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207-229. https://doi.org/10.1006/jpho.1999.0094

Cho, T., Whalen, D. H., & Docherty, G. (2019). Voice onset time and beyond: Exploring laryngeal contrast in 19 languages. *Journal of Phonetics*, *72*, 52-65. https://doi.org/10.1016/j.wocn.2018.11.002

Flege, J. E. (1997). The role of phonetic category formation in second-language speech learning. In New Sounds 97, *Proceedings of the Third International Symposium on the Acquisition of Second Language Speech*, University of Klagenfurt, 79-88.

Flege, J. E., Schirru, C., & Mackay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467-491. https://doi.org/10.1016/S0167-6393(02)00128-0

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47-65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*, 395-411.

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

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. Schiller, & Meyer, A. (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185-202. https://doi.org/10.1016/0167-6393(87)90025-2

Henton, C., Ladefoged, P., & Maddieson, I. (1992). Stops in the world's languages. *Phonetica*, *49*, 65-101. https://doi.org/10.1159/000261905

Holliday, J. J. (2019). The perception and production of word-initial Korean stops by native speakers of Japanese. *Language and Speech*, *62*(3), 494-508. https://doi.org/10.1177%2F0023830918785649

Ikekeonwu, C. I. (1987). Igbo dialect cluster: A classification. Paper presented at the *Departmental Seminar Series of the Department of Linguistics/Nigerian Languages*, University of Nigeria: Nsukka.

Ikekeonwu, C. I. (1991). Igbo. *Journal of International Phonetic Association, 21*(2), 99-101. https://doi.org/10.1017/S0025100300004473

Iloene, G. O. (2007). *Igbo Phonology*. In O. Yusuf (Ed.), *Basic linguistics for Nigerian languages teachers*. Port Harcourt: M. & J. Grand Orbit Communications Ltd and Emhai Press.

Landi, R. (1997). Le consonanti occlusive in stile diversi di parlato. In F. Cutugno (Ed.), *Fonetica e Fonologia degli stili dell'italiano parlato* (pp. 143-155). Atti delle VII Giornate di Studio del Gruppo di Fonetica Sperimentale, C.I.R.A.S.S., Universita degli Studi di Napoli 'Federico II', Roma: Esagrafica.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830

Lisker, L., & Abramson, A. S. (1967). Some effects of context on Voice Onset Time in English stops. *Language and Speech, 10*(1), 1-28. https://doi.org/10.1177%2F002383096701000101

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, I. (2013). Voicing and gaps in plosive systems. In M. S. Dryer, & Haspelmath, M. (Eds.), *The World Atlas of Language Structures Online, Max Planck Institute for Evolutionary Anthropology*, Leipzig. Retrieved June 11, 2021, from http://wals.info/chapter/5

Major, R. C. (1987). English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics, 15*, 197-202. https://doi.org/10.1016/S0095-4470(19)30560-1

Mori, L., & Barkat-Defradas, M. (2005). Acoustic properties of foreign accent: VOT variation in Moroccan-accented Italian. *Interspeech 2005*, Lisbon, 2909-2912.

Mori, L. (2007). *Fonetica dell'italiano L2. Un'indagine sperimentale sulla variazione nell'interlingua dei marocchini*. Roma: Carocci.

Nagle, C. (2017). A longitudinal study of Voice Onset Time development in L2 Spanish stops. *Applied Linguistics, 40*(1), 86-107. https://doi.org/10.1093/applin/amx011

Nkamigbo, L. C. (2014). Nasalization, aspiration and labialization in Igbo dialect phonology. *SKASE Journal of Theoretical Linguistics*, The Slovak Association for the Study of English, 59-68.

Nodari, R., Celata, C., & Nagy, N. (2019). Socio-indexical phonetic features in the heritage language context: VOT in the Calabrian community in Toronto. *Journal of Phonetics, 73*, 91-112. https://doi.org/10.1016/j.wocn.2018.12.005

Nwaozuzu, G. I. (2008). *Dialects of the Igbo language*. University of Nigeria Press: Press Ltd, Nsukka.

Sorianello, P. (1996). Indici fonetici delle occlusive sorde nel cosentino. *Rivista Italiana di*

Dialettologia, *20*, 123-159.

Sorianello, P. (2008). Persistenza di categorie fonologiche nell'italiano L2. In L. Costamagna, & Marotta, G. (Eds.), *Processi fonetici e categorie fonologiche nell'acquisizione dell'italiano* (pp. 121-146). Pisa: Pacini.

Tran, T. T. H., Vallée, N., & Granjon, L. (2017). Effects of word position on the acoustic realization of Vietnamese final consonants. *Journal of Phonetics*, *76*(1), 1-30. https://doi.org/10.1159/000485103

Vagges, K., Ferrero, F., Magno Caldognetto, E., & Lavagnoli, C. (1978). Some acoustic characteristics of Italian consonants. *Journal of Italian Linguistics*, 3-4, 69-81.

Notes

Note 1. http://web.phonetik.uni-frankfurt.de/upsid.html. On the /p t k b d g/ distribution, see also Maddieson (2013).

Note 2. For a detailed description of obstruents in the Arọ and Mgbo dialects of Igbo, see Anyasi *et al.* (2020).

Note 3. This is true for read and controlled speech. In spontaneous corpora, most Italian voiced stops do not show a clear moment of release; as a consequence, in such cases, they are associated with negative VOT values.

Note 4. In many regional varieties of Italian, stops undergo various phenomena of weakening, such as spirantisation, lenition, sonorisation, or strengthening; for instance, in Central and Southern varieties /b/ is always prolonged in the intervocalic context and before liquid consonants. In several dialects spoken in Central Calabria, post-sonorant singleton and geminate voiceless stops are aspirated especially when placed in a post-stress context (Sorianello, 1996; Nodari *et al.*, 2019).

Note 5. The presence of a low speech in L2 learners is a very typical aspect, which is observable in the interlanguage, especially when the proficiency of the target language is not high. Moreover, in our corpus, Igbo speakers also showed a poor ability to read Italian out loud, given that their process of Italian acquisition took place outside of the school context.

Note 6. Several reduced stops were noticed in spontaneous speech.

Copyrights