Objective versus perceived proficiency in pronunciation assessment of phonetic detail in advanced L2 English learners

Joaquín Romero[a], Leticia Quesada Vázquez[b]

[a]Rovira i Virgili University, Spain; [b]Universidad Nebrija, Spain
joaquin.romero@urv.cat, lquesada@nebrija.es

ABSTRACT

Assessment of pronunciation in a foreign language can be performed in two main ways: objectively, i.e., using acoustic analysis, or indirectly, by relying on native speaker perception of notions like comprehensibility and accentedness. This study investigates the mismatch between acoustic measurements and perception ratings of subphonemic aspects of English pronunciation, such as stop aspiration and devoicing by advanced learners of the language who are training to become English teachers. Results show that, while the analysis of VOT data evidenced significant differences in comparison to native speaker productions, perception ratings seemed to ignore these differences and consistently judged the productions with very high scores. This indicates that, while phonetic detail may not be crucial for effective communication in the L2, it can be an indicator of a high level of oral proficiency that should be included as part of an effective pronunciation instruction program.

Keywords: Pronunciation assessment, phonetic detail, comprehensibility, accentedness.

1. INTRODUCTION

1.1. Goals of L2 pronunciation instruction

Of the different areas of foreign language teaching/learning that are commonly included in standard L2 programs, pronunciation remains a particularly controversial and difficult skill for both instructors and learners [1]. Lack of training and insecurity on the part of the teachers often results in pronunciation being excluded from language classes completely. On the other hand, learners often feel that the effort needed to acquire anything close to native-like pronunciation outweighs its potential benefits. In this sense, recent research [2] has advocated for abandoning the goal of native-like pronunciation and recommends that instruction should concentrate on achieving a degree of comfortable intelligibility instead. Like with any other aspect of foreign language acquisition, however, the specific goals of the learning process very much depend on the typology of the learner. When the pronunciation instruction is part of a teacher training program, that is, students who are preparing to become models of the language themselves, it might be worth reconsidering the convenience of a native-like goal in pronunciation.

1.2. Assessment of L2 pronunciation

Regardless of the objectives of a teacher/learner, assessment of L2 pronunciation remains a complicated task, not only because of the difficulty in determining when a specific pronunciation feature has been acquired satisfactorily, but also because of the problems finding methodologies that can provide a complete, systematic, and reliable picture of a speaker's pronunciation proficiency. In this sense, two main approaches have traditionally been used. On the one hand, a more objective approach focuses on phonetic (acoustic and/or articulatory) analysis [3, 4] of an L2 speaker's oral productions and the comparison of these productions with those of native speakers. On the other hand, a more communicative-based approach relies on perception by native speakers and judgments based on constructs such as comprehensibility, accentedness, etc. [5].

In recent years, many studies have followed a multimodal approach that combines the objective precision obtained from phonetic analysis with the perceptual functionality of native speaker rater judgments [6, 7], the goal being to obtain a more complete and realistic picture of an L2 speaker's pronunciation proficiency. A potential problem arises, however, when the results of the two methodologies do not align, that is, diverging or contradictory assessments are obtained from the two methodological approaches. At the root of this mismatch is often a lack of agreement as to what the goals of a specific pronunciation instruction are and, more particularly, between the goals and models of the learners and those of the evaluators.

The current study uses a mixed approach to evaluate one such situation with a group of advanced L2 English teacher trainees with a high degree of motivation to improve their pronunciation. The purpose of the study is to test whether their objective pronunciation proficiency, as evaluated via acoustic analysis, matches the perceived output, as judged by trained native English-speaking teachers.

2. METHOD

2.1. Participants

A total of nine subjects participated in the study, seven females and two males who were all students in a 4-year undergraduate program in English at Universitat Rovira i Virgili, in Tarragona, Spain and were an average of 21 years of age. They were all native speakers of Spanish and Catalan with slight varying degrees of dominance for one of the two languages. At the time of the study, the students had just completed their 3rd year of the program, during which they had all taken an obligatory two-semester course on the Sound System of English.
The nine subjects that participated in the study were classified according to their overall performance in the Sound System of English course, based on a quartile distribution. Thus, five of the nine subjects belonged to the fourth quartile, while the remaining four were distributed between the second and the third quartiles. Only the five subjects in the fourth quartile were actually part of the analysis, while the remaining four were used as distractors.

2.2. Stimuli

The stimuli were made up of target words embedded in a total of six sentences (Table 1) that participants had recorded individually and submitted as part of their class assignments during the course. The sentences focused on one of the six stop consonants of English, respectively. In addition to different VOT settings, the sentences also included other instances of allophonic variation of the consonants, such as unreleased allophones in word final position, flapped /t, d/, glottalized /t/, etc. A total of 54 sentences (6 sentences x 9 subjects) were used.

	labial	alveolar	velar
Voiceless aspirated	Penny	ten	candy
	pay	twenty	cause
	Portugal	fifteen	complications
Voiced unaspirated	Bob	dance	Gary
	bus	Donna	got
	bound	drugs	game
Voiceless Unaspirated	spent	start	school

Table 1: Target words included in the design.

2.3. Acoustic Analysis

Prior to the acoustic analysis, all 54 sentences were normalized for amplitude in Praat using a standard script. In addition, the 30 sentences corresponding to the fourth quartile students were also checked for potential deviations in overall speaking rate and duration. No such deviations were identified. Subsequent acoustic measurements were obtained exclusively from the 5 fourth quartile students.
Using waveforms and spectrographs in Praat, measurements were obtained of 3 key words from each of the six sentences, as indicated in Table 1. For the purposes of this paper, only those measurements involving VOT values in word-initial stops will be presented, amounting to a total of 21 words per subject. VOT was measured as the acoustic duration between the release of the stop burst and the onset of vocal fold vibration for the following vowel. For the voiceless stops /p, t, k/, all subjects produced exclusively instances of voice lag and therefore only positive VOT values were obtained. For the voiced stops /b, d, g/, on the other hand, while a majority of cases of voice lead were produced, resulting in negative VOT values, some occasional instances of short lags were also identified. A total of 105 VOT readings were obtained (21 words x 5 subjects). Figure 1 illustrates instances of positive VOT for /t/ (top) and negative VOT for /g/ (bottom).

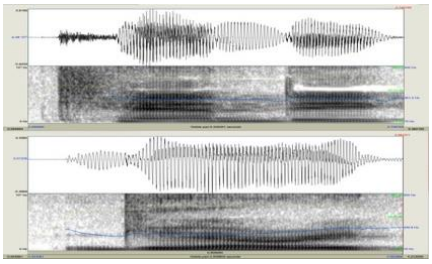

Figure 1: Examples of VOT productions.

2.4. Ratings

For the more qualitative assessment of the subject's pronunciation, 5 native English speakers participated as raters in a listening experiment where they were asked to judge the productions of all 9 subjects. Of the 5 participants, 4 were speakers of North American English and 1 was a speaker of Southern British English. All raters were trained graduates from a master's degree program in Teaching and Learning English as a Foreign Language at Universitat Rovira i Virgili. As part of the program, they had all taken a course on Teaching Pronunciation in the EFL class and had extensive experience working with native Spanish/Catalan.
Raters were asked to listen to the sentences produced by the 9 subjects and rate them based on two 5-point Likert scales. The stimuli were presented to them in a random order using Google Forms. On each page, a link was available to play an audio file. After listening to the audio file, they were asked to evaluate it according to two parameters, i.e., foreign accent and

comprehensibility. A brief description of what each of these notions mean was provided for guidance: **foreign accent**: from not accented at all (1) to heavily accented (5); **comprehensibility**: from not comprehensible at all, i.e., you find it impossible to tell what is being said (1) to highly comprehensible, i.e., you have to make no effort at all to understand what is being said (5). In total, raters listened to 54 audio files (30 from the five target speakers and 24 from the remaining 4 speakers as distractors) and the entire experiment took approximately 25 minutes.

2.5. Data Analysis

For the analysis of the acoustic data, the VOT duration measurements obtained from the recordings of the 5 subjects were compared with standard VOT values for English stops available in the literature [8], which were included in the design as control data. In order to test for potential differences between the observed values and the standard ones, the data were submitted, separately by voicing, to a linear mixed-effects model analysis with *Group* (Observed values, Control data), and *Point of articulation* (Labial, Alveolar, Velar) and the *Group* x *Point of articulation* interaction as fixed effects, and *Subject* as a random effect. Table 2 below provides the VOT values from [8] used here. All the analyses were carried out in JASP.

	labial	alveolar	velar
voiceless	58	70	80
voiced	1	5	21

Table 2: Target VOT values for English stops.

As far as the data from the perception ratings are concerned, the results for foreign accent and comprehensibility from each rater were extracted from the Google Forms score sheet, classified by voicing of the target consonant, and averaged by subject. In addition, interclass correlation coefficient (ICC) were calculated in order to obtain an estimate or rater reliability. The results for this section will be presented qualitatively and discussed in the light of the quantitative VOT results.

3. RESULTS

3.1. VOT

The results of comparing the VOT values of the 5 subjects that participated in the study with the standard values established in the literature are shown in Table 3. As can be seen, there is no significant difference for *Group* when the consonant target is a voiceless stop, indicating that the subjects are generally hitting the expected VOT target for these consonants. There is, however, a significant effect of *Point of Articulation*, which is due to the fact that, as a general rule, subjects are able to reproduce VOT values for alveolar and velar stops more accurately (M= 87.653 ms and 89.627 ms, respectively, compared to the 80.00 ms reported in Lisker and Abramson) than for labials, which often show insufficient voice lag (M= 65.493 ms).

Effect	df	F	p
voiceless stops			
group	1, 8.00	3.088	0.117
POA	2, 76.00	16.086	< .001
group ✻ POA	2, 76.00	0.806	0.451
voiced stops			
group	1, 8.10	41.772	< .001
POA	2, 25.04	1.206	0.316
group ✻ POA	2, 25.04	0.703	0.504

Table 3: Results of tests of fixed effects (*Group*, *Point of articulation*) for VOT comparison.

The more interesting comparison, however, involves the positive VOT values, for which a clear significant effect of *Group* is obtained, whereas the effect of *Point of articulation* is not significant. This confirms the observations that these subjects systematically use long voice leads in their productions of English voiced stops independently of the point of articulation.

3.2. Ratings

3.2.1. Foreign accent

Figure 2 below displays the results of the ratings of foreign accent by the 5 native speakers of English.

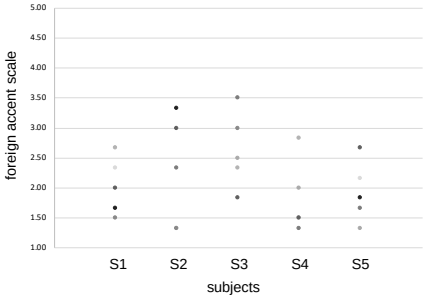

Figure 2: Results for foreign accent ratings. For each subject, the vertical dots show the score of each of the five raters.

As can be observed, with the exception of a few isolated cases, a majority of ratings seem to indicate

the presence of little or moderate foreign accent. The interclass correlation coefficient values for this part of the experiment were generally quite low (ICC <.5) indicating poor rater reliability in general. For the most part, however, the results for foreign accent point at a general agreement that the subjects are producing good or very good renditions of the English sentences that they were reading.

3.2.2. Comprehensibility

Regarding the second parameter, comprehensibility, the results of the ratings are displayed in Figure 3.

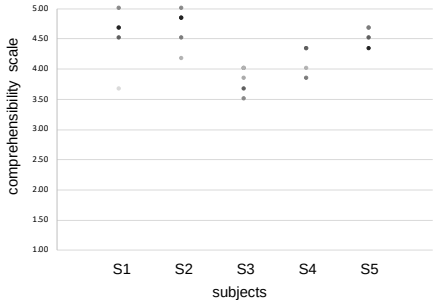

Figure 3: Results for comprehensibility ratings. For each subject, the vertical dots show the score of each of the five raters, which overlap in some instances.

The picture in this case indicates an even more positive assessment than in the case of foreign accent. Overall, the ratings, except for S3, are very good to excellent, with a few instances of perfect scores. Thus, these results indicate that raters were generally perfectly capable of understanding the productions of the subjects and that, in many instances, they had to make no effort at all to do so. As far as the interclass correlation coefficient values are concerned, rater reliability was generally higher (ICC .5 – .7) here than for foreign accent.

4. DISCUSSION AND CONCLUSIONS

The results of the acoustic analysis revealed existing inaccuracies in the production of English stops by the Spanish learners. Specifically, insufficient VOT length was identified for voiceless labial stop /p/ and, especially, for voiced stops /b, d, g/, indicating a pervasive difficulty in achieving native targets of positive VOT in these consonants. Thus, it seems that the subjects have developed a deeper awareness of the VOT settings in voiceless stops, especially /t/ and /k/, than in voiced stops. This may be due to the fact that, at least intuitively, a feature like aspiration is perceptually more salient than pre-voicing and, thus, L2 English learners can easily distinguish between an English aspirated stop category and a Spanish unaspirated one, whereas distinguishing between an English voiced stop with no pre-voicing and a Spanish one with pre-voicing may require an additional degree of awareness.

Still, as illustrated by the results from the qualitative analysis of native speaker rater judgments, these inaccuracies observed in the acoustic analysis do not seem to weigh heavily on the overall evaluation of the subjects' productions. While values for foreign accent were general quite good, the assessment of comprehensibility indicates that the subtle differences in advanced phonetic features such as aspiration and devoicing are not relevant in their evaluation of the subjects' productions. These advanced phonetic features seem to fall under the radar of the raters almost completely and they appear to play no role in how intelligible the speakers are to their ears.

While these detailed phonetic features may not be particularly important for effective communication in the L2, especially given that variation can play a role in how a native speaker perceives foreign accented speech, they can be indicators of a high level of oral proficiency. How to evaluate them, however, does not seem to be an easy task. As shown in this study, even trained language teachers who are familiar with the characteristics of these particular L2 speakers seem to ignore phonetic detail in their evaluation of the subjects' productions. Thus, it may be necessary to reassess the use of general perceptual constructs such as comprehensibility and foreign accent and find mechanisms to find-tune them by making them more precise, selecting raters to fit specific requirements or making the rating process much more focalized.

Regarding the objectives of L2 pronunciation teaching/learning, the findings from this study are interpreted as providing support for the convenience of adapting the goals to the specifics of the situation, in particular, to the typology and interests of the learners. While comfortable intelligibility may be a desirable goal in a majority of learning environments and for a large number of learners, specific populations, such as prospective teachers, should be able to exploit their potential to achieve nativelikeness. Ultimately this may require a degree of concentration on pronunciation that would probably be impractical or excessive for most other populations or learning environments, However, there seems to be no real reason why, just like most learners aim to achieve native-like levels of grammar and vocabulary use, for example, comparable levels of pronunciation should not be attainable by using a combination of awareness, knowledge of the sound system of the language, command of speech articulation, and intensive practice and corrective feedback.

5. REFERENCES

[1] Gilakjani, A., Ahmadi, S., Ahmadi, M. 2011. Why is pronunciation so difficult to learn. *English Language Teaching 4*(3), 74-83.

[2] Levis, J. M. 2005. Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*(3), 369–377.

[3] Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C., Molholt, G. 2005. The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, *26*(2), 227–247.

[4] Gick, B., Bernhardt, B., Bacsfalvi, P., Wilson, I. 2008. Ultrasound imaging applications in second language acquisition. *Phonology and Second Language Acquisition 36*, 315-328.

[5] Isbell, D. 2018. Assessing pronunciation for research purposes with listener-based numerical scales. In O. Kang, A. Ginther (eds), *Assessment of Second Language Pronunciation*. New York: Routledge, 89-112.

[6] Derwing, T. M., & Munro, M. J. 2005. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379-397.

[7] Saito, K., Plonsky, L. 2019. Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning 69*(3), 652-708.

[8] Lisker, L., Abramson, A. 1964. A cross-language study of voicing in initial stops: acoustical measurements. *Word 20*(3), 384-422.

Acknowledgements This work was supported by grant PID2020-117804GB-I00 from the Spanish Ministry of Science and Innovation.