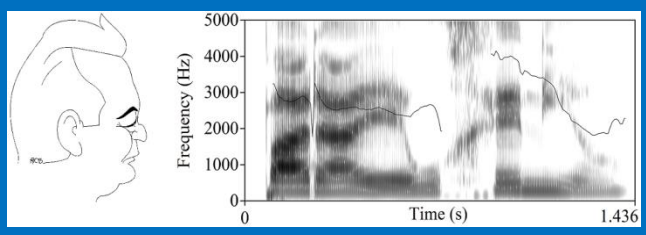
Investigating the role of listeners' reaction time in the assessment of L2 speech

Natalia Fullana



Fernández Planas, A. Ma. (ed.) (2016): *53 reflexiones sobre aspectos de la fonética y otros temas de lingüística*, Barcelona, págs. 363-371.
ISBN: 978-84-608-9830-6.

Investigating the role of listeners' reaction time in the assessment of L2 speech

Natalia Fullana
Universitat de Barcelona
nataliafullana@ub.edu

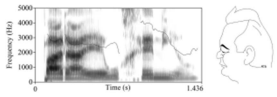
Para Eugenio, quien despertó en mí el interés por la fonética general y la fonética experimental al iniciar mi licenciatura y quien ha tenido las puertas del Laboratorio de Fonética siempre abiertas cada vez que lo necesité, lo que ha sido determinante para el desarrollo de mi trayectoria académica.

1. BACKGROUND

Second language (L2) speech acquisition research into the assessment of learners' speech production frequently involves asking native speakers of the target language (TL) to gauge L2 learners' production for degree of foreign accent (FA), intelligibility, comprehensibility, and/or fluency (for a review, see Munro and Derwing, 2011). Particularly, and when it comes to determining the degree of FA or accentedness, some of that research has been concerned with the role of listener factors in the evaluation of nonnative speech (e.g., Flege and Fletcher, 1992; Major, 2007; Munro, Derwing and Morton, 2006). The available findings so far indicate that range effects influence listener judgements – i.e., the larger the group of native speakers of the TL is, the more foreign-accented listeners (or judges) will rate learners' L2 speech production (Flege and Fletcher, 1992). However, listener factors such as familiarity with foreign-accented speech or with the learners' first language (L1), as well as expertise in assessing learners' production in the L2, have yielded differing results (cf. Bongaerts, van Summeren, Planken and Schils, 1997; Major, 2007; Munro et al, 2006; Thompson, 1991). Furthermore, studies on the evaluation of L2 speech have shown that individual variation among listeners may affect listener ratings. Those factors include, among others, the listeners' L1 background, gender, and their emphasis on certain scoring criteria over other sets of criteria (Kim, 2009; Eckes, 2008; and O'Loughlin, 2007, respectively; as cited in Isaacsa nd Trofimovich, 2011).

In order to account for this variability current research has considered other factors such as listeners' cognitive abilities. For instance, Isaacs and Trofimovich (2011) examined the extent to which individual differences in 60 raters' phonological memory, attention control, and musical ability had an effect on the accentedness, comprehensibility, and fluency scores assigned to a group of 40 French learners of English. The authors found that only musical ability significantly influenced listeners' evaluation of L2 speech, those majoring in a music degree being stricter in their FA ratings than those participants who were studying a degree other than music.

 ISBN 978-84-608-9830-6



An additional factor that could potentially affect listeners' ratings is reaction time (RT) or *the time a person takes to respond to a stimulus or a performing task* (Jiang, 2012:2). While RT data has often been considered from the learners' perspective (see studies cited in Jiang, 2012), the limited research on native and nonnative listeners' RT in the evaluation of nonnative speech suggests that it takes longer for raters to assess accented speech than native speech (Munro and Derwing, 1995).

Based on all of the above and in line with recent research on listener factors, this study aimed to explore the role of listeners' reaction time as a possible source of listener variation in the evaluation of L2 speech.

2. METHOD

2.1. Participants

The participants in this study were part of the Barcelona Age Factor (BAF) Project (Muñoz, 2006). They were 232 Catalan/Spanish learners of English who differed in their age of onset of L2 learning in an instructed setting – ages 8, 11, 14, and 18+ – and in their exposure to English as a foreign language (FL) – 2.5 years, 4.5 years, and 7.5 years on average. A control group of 14 native English speakers (NSs) also took part in the study. Finally, seven native English listeners with a mean age of 26.1 years and with «a good ear» and skills for phonetic transcription were recruited to carry out the accentedness and vowel identification tasks (see below). (For further information on the participants, see Fullana, 2005)

2.2. Tasks and procedure

The learners of English and the control group were asked to repeat 34 words in English as presented via recorder. On the other hand, the seven English native listeners participated first in a FA rating task and then a vowel identification task containing the participants' oral productions of English words. The FA rating task consisted of rating English /i ɪ ɛ æ ɒ ʌ u/ on a 9-point scale of FA (*1=no FA; 9=strong FA*). The same vowel sounds were considered in the vowel identification task, whereby listeners had to identify the target sounds among 15 possible response options that included correct pronunciation and several potential mispronunciations for each one of the vowel sounds. In both tasks the inter-trial stimulus was 1.5 seconds and repeat listening to an item was possible. Also, following Southwood and Flege (1999), to examine intra-rater consistency a random 25% of the participants' total productions was added immediately after the 246 items (232 learners + 14 English NSs) of each task.

2.3. Analyses

Approximately 12050 accent scores and 12050 identification scores were obtained. As the main variable in the present study was reaction time, data screening was necessary to remove outliers with 2 standard deviations above the grand mean of the RT for

accentedness and that of vowel identification, separately. The main statistical analyses involved looking at the differences in listeners' reaction time for FA and identification scores through Friedman tests, the differences in listeners' reaction time for learners of English vs. English NSs by means of independent samples t-tests, and the relationship between RT and FA scores, on the one hand, and between RT and vowel identification scores, on the other, through Spearman rank-order correlations. In all cases, the alpha level was set at .05.

3. RESULTS

Results concerning the degree of FA and vowel identifications are reported in detail elsewhere (Fullana, 2005). Briefly, those findings indicated that learners were judged as having a medium degree of FA, while the control group was rated as being English NSs. Similarly, vowels produced by the control group were identified as correct vowel productions, whereas learners of English were often considered to have mispronounced English vowels, regardless of their age of onset of L2 learning and amount of experience in English. It was also observed that there was an acceptable degree of intra-rater agreement (ICC=.70-.80). Despite this, there was a lower and more varied degree of inter-rater agreement, which could not be attributed to the seven listeners' characteristics such as experience and/or familiarity with the learners' L1. Therefore, the results below will only focus on the potential effects of RT on listeners' lower degree of inter-rater agreement.

3.1. Listeners' RT for FA and identification scores

As noted above, listeners were originally given 1.5 seconds to assign an accent rating and to identify each of the seven target vowels (though it should be reminded that the task design allowed for replaying, so response/reaction time could be longer). Table 1 below shows the percentage range of accent ratings and vowel identifications per target vowel given by each listener within the original 1.5-second response time, where a high degree of variability across the listeners in accent ratings and identifications (%) assigned within 1.5 seconds can be observed. For example, in the accent rating tasks listeners were able to rate between 5.7% and 88.1% of the total items, while in the vowel identification task they identified between 16.0% and 87.0% of the total tokens within the original 1.5-second response time per item.

	/i/	/ɪ/	/ɛ/	/æ/	/ɒ/	/ʌ/	/u/
FA	37.6 (1) 83.3 (6)	31.8 (4) 83.7 (2,6)	**5.7** (5) **88.1** (6)	24.0 (1) 76.4 (2)	27.4 (5) 87.0 (2)	20.1 (5) 83.2 (6)	18.7 (5) 73.6 (2)
VW ID	18.4 (1) 71.0 (2)	28.6 (4) 70.2 (3)	41.0 (1) 86.5 (2)	21.5 (1) **87.0** (3)	26.0 (4) **87.0** (2)	**16.0** (5) 70.1 (3)	34.1 (1) 69.9 (3)

Table 1. *Percentage range of accent ratings (FA) and identifications (VW ID) per target vowel given within 1.5 seconds. Note: (number) indicates listener 1, 2, 3, 4, 5, 6, or 7.*

 ISBN 978-84-608-9830-6

In spite of the variability observed in the table, some trends can be outlined. In the first place, three listeners rated or identified target sounds in a smaller percentage within 1.5 seconds: listeners 1 (FA + VW), 4 (VW), and 5 (FA). By contrast, listeners 2 (FA + VW), 3 (VW), and 6 (FA) rated or identified target sounds in a higher percentage within 1.5 seconds. Friedman test results indicated that each listener took their own time (RT) to rate and/or identify vowels (X^2 51.465–257.294, *df* 6, $p < .05$) (see Table 2). Consequently, listeners' RT values were examined separately as far as differences between English NSs and learners of English are concerned, in addition to the correlations between accent scores and RT and between vowel identifications and RT.

	/i/	/ɪ/	/ɛ/	/æ/	/ɒ/	/ʌ/	/u/
FA	1.28 – 3.06	1.16 – 3.11	1.09 – 4.57	1.40 – 3.30	0.96 – 3.45	1.26 – 3.74	1.44 – 3.62
VW ID	1.46 – 3.36	1.53 – 3.08	0.85 – 1.99	0.95 – 3.19	1.60 – 2.80	1.41 – 3.59	1.58 – 2.72

Table 2. *Mean range of RT (in sec) for accent (FA) scores and vowel identification (VW ID) scores per vowel across 7 listeners.*

3.2. Differences in listeners' RT between Catalan/Spanish learners of English and English NSs

When comparing RT for native vs. nonnative speech, it was observed that on average English NSs' productions were rated and identified at shorter RT rates than those of learners of English (see Figure 3a and 3b), yielding 19 and 29 significant differences (out of 49 comparisons) for accent ratings and vowel identification scores, respectively. However, it should be noted that in the accentedness task instances of shorter RT for learners of English vs. English NSs were found on the part of listener 2 for /i/ and /ɒ/, listener 3 for /æ/, /ʌ/ , and /u/, and listeners 5 and 7 for /u/.

(a) (b)

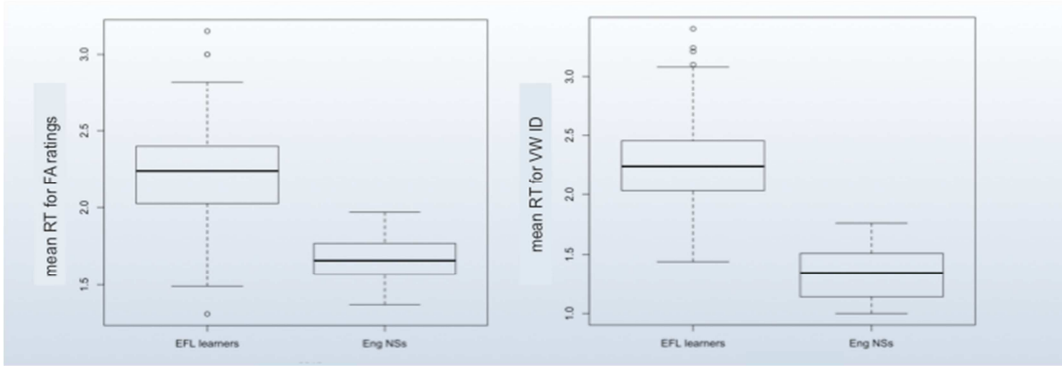


Figure 3. *Mean RT for accent scores (a) and vowel identification (b) in nonnative vs.native speech (left and right boxplots in 3a and 3b, respectively).*

3.3. Relationship between RT and listeners' judgements

Concerning the relationship between listeners' RT and accent scores and vowel identifications, a higher number of significant correlations between RT and vowel identifications was evident, in contrast to RT and accent scores. Tables 3 and 4 below further illustrate this finding, as well as the following outcomes. First, listeners 2 and 7's RT was nearly always correlated with higher accent scores at significant rates, i.e., longer RT involved assigning a higher degree of FA. As for vowel identifications, listener 2's identification scores were also significantly correlated with RT values. Specifically, it took that specific listener significantly longer to identify mispronunciations than expected native-like/correct vowel productions.

Listener	/i/	/ɪ/	/ɛ/	/æ/	/ɒ/	/ʌ/	/u/
1		+	+		+		+
2	+	+	+	+	+	+	+
3			+			-	
4	+		+	+		-	+
5	+	+	+				
6	+	+	+				
7	+	+	+	-	+	+	+

Table 3. *Correlations between RT and accent scores. Note: + significant positive correlation (rho range: .166 to .632); - significant negative correlation (rho range: -.177 to -.206).*

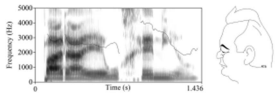
Listener	/i/	/ɪ/	/ɛ/	/æ/	/ɒ/	/ʌ/	/u/
1	+	+	+	+		+	
2	+	+	+	+	+	+	+
3	+		+	+	+	+	
4		+	+		+	+	+
5	+			+	+		
6	+	+	+	+		+	+
7	+	+	+	+	-	+	

Table 4. *Correlations between RT and vowel identifications. Note: + significant positive correlation (r_{pbi} range: .134 to .552); - significant negative correlation (r_{pbi}: -.223).*

4. DISCUSSION AND CONCLUSIONS

The aim of the present study was to examine the role of listeners' RT as a possible source of listener variation in the assessment of L2 speech, as listeners had previously exhibited

 ISBN 978-84-608-9830-6



a lower degree of inter-rater agreement that could not be attributed to often-reported factors such as familiarity with the learners' L1 or range effects. The results pointed to inter-listener variability in RT in the ratings of participants' productions for FA and identification of target vowel sounds. Although generally longer RT values were moderately correlated with increasingly accented scores, the same and/or different listeners were found to also rate more foreign-accented productions at shorter RT rates as a function of the target vowel being assessed. Additionally, the overall finding of native English listeners' faster RT for NS-produced vowels than for nonnative speech is in line with previous studies (Jongman and Wade, 2007; Munro and Derwing, 1995). Finally, the fact that there were more significant correlations between RT and vowel identifications than between RT and accent scores might suggest task effects (Derwing, Rossiter, Munro and Thomson, 2004) since it might be easier for listeners to choose among a given set of response options than assign an accent rating. As for the latter, this outcome further supports the notion that accentedness is a more subjective dimension, as is comprehensibility (Isaacs and Trofimovich, 2012).

The observed differences in RT therefore contribute to research on possible sources of listener variation (Flege and Fletcher, 1992; Isaacs and Trofimovich, 2011; Major, 2007). Furthermore, the findings of reaction time values in the evaluation of L2 speech might be interpreted in light of Munro's (2008) «reconceptualized model» of accentedness, intelligibility and comprehensibility. According to this model, a learner's final accent score (or comprehensibility or fluency score) results from a combination of stimulus properties (e.g., learners' deviations in segmentals, prosody, grammar, etc.), listeners' factors (e.g. familiar topic, familiarity with speakers, type of accent, etc.), contextual factors, and an error component. The list of features of each component in the model is open to further additions (see formula in Munro, 2008, p. 205). Based on the results of this investigation, it is hereby suggested that reaction time be included as one of the features of the listener factor component in Munro's model and its contribution to learners' final accent/comprehensibility/intelligibility score be further examined.

To conclude, a number of pedagogical implications can be drawn from the reported findings. Taking into account Derwing and Munro's (2005) call for more collaborative work between researchers and practitioners/teachers, the present study could be placed within current lines of investigation that apply methods of more experimental/laboratory research to classroom settings (Gass, Mackey and Ross-Feldman, 2011; Hummel and French, 2010). Moreover, this study highlights the potential contribution of listeners' RT, and by extension teachers' RT, as another possible source of bias in assessing L2 learners' pronunciation, along with musical ability as suggested by Isaacs and Trofimovich (2011).

6. REFERENCES

BONGAERTS, T; C. VAN SUMMEREN; B. PLANKEN and E. SCHILS (1997): «Age and ultimate attainment in the pronunciation of a foreign language», *Studies in Second Language Acquisition*, 19, pp. 447-465.

DERWING, T. M. and M. J. MUNRO (2005): «Second language accent and pronunciation teaching: A research-based approach», *TESOL Quarterly*, 39, pp. 379-397.

DERWING, T. M; M. J. ROSSITER; M. J. MUNRO and R. I. THOMSON (2004): «Second language fluency: Judgments on different tasks», *Language Learning*, 54, pp. 655-679.

ECKES, T. (2008): «Rater types in writing performance assessments: A classification approach to rater variability», *Language Testing*, 25, pp.155–185.

FLEGE, J. E. and K. L. FLETCHER (1992): «Talker and listener effects on degree of perceived foreign accent», *Journal of the Acoustical Society of Americ*a, 91, pp. 370-389.

FULLANA, N. (2005): *Age-related effects on the acquisiton of a FL phonology*, PhD dissertation, Universitat de Barcelona.

GASS, S; A. MACKEY and L. ROSS-FELDMAN (2011): «Task-based interactions in classroom and laboratory settings», *Language Learning*, 61, pp. 189-220.

HUMMEL, K. M. and L. M. FRENCH (2010): «Phonological memory and implications for the second language classroom», *The Canadian Modern Language Review*, 66, pp. 371-391

ISAACS, T. and P. TROFIMOVICH (2011): «Phonological memory, attention control, and musical ability: Effects of individual differences on rater judgments of second language speech», *Applied Psycholinguistics*, 32, pp. 113-140.

ISAACS, T. and P. TROFIMOVICH (2012): «Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings», *Studies in Second Language Acquisition*, 34, pp. 475-505.

JONGMAN, A. and T. WADE (2007): «Acoustic variability and perceptual learning: The case of non-native accented speech», in O.-S. Bohn and M. J. Munro (eds.): *Language experience in second language speech learning*, Amsterdam, John Benjamins, pp. 135-150.

KIM, Y.-H. (2009): «An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach», *Language Testing*, 26, pp. 187–217.

MAJOR, R. C. (2007): «Identifying a foreign accent in an unfamiliar language», *Studies in Second Language Acquisition*, 29, pp. 539-556.

MUNRO, M. J. (2008): «Foreign accent and speech intelligibility», in J. G. Hansen Edwards and M. L. Zampini (eds.): *Phonology and second language acquisition*, Amsterdam, John Benjamins pp. 193-218.

MUNRO, M. J. and T. M. DERWING (1995): «Processing time, accent, and comprehensibility in the perception of foreign-accented speech», *Language and Speech*, 38, pp. 289-306.

MUNRO, M. J. and T. M. DERWING (2011): «The foundations of accent and intelligibility in pronunciation research», *Language Teaching*, 44, pp. 316-327.

MUNRO, M. J; T. M. DERWING and S. L. MORTON, (2006): «The mutual intelligibility of L2 speech», *Studies in Second Language Acquisition*, 28, pp. 111-131.

MUÑOZ, C. (2006): *Age and the rate of FL learning*, Clevedon, UK, Multilingual Matters.

O'LOUGHLIN, K. (2007): «An investigation into the role of gender in the IELTS oral interview» in L. Taylor and P. Falvey (eds.): *IELTS collected papers: Research in speaking and writing assessment*, Cambridge, Cambridge University Press, pp. 63-97.

SOUTHWOOD, M. H. and J. E. FLEGE (1999): «Scaling foreign accent: direct magnitude estimation versus interval scaling», *Clinical Linguistics & Phonetics*, 13, pp. 335-349.

THOMPSON, I. (1991): «Foreign accents revisited: The English pronunciation of Russian immigrants», *Language Learning*, 42, pp. 177-204.