SPOKEN WORD RECOGNITION BY ENGLISH-SPEAKING LEARNERS OF SPANISH

José María Lahoz-Bengoechea[1], *Alba Tuninetti*[2,3], *& Paola Escudero*[2,3]

[1] Universidad Complutense de Madrid
[2] MARCS Institute for Brain, Behaviour and Development (Western Sydney University)
[3] Australian Research Council Centre of Excellence for the Dynamics of Language
jmlahoz@ucm.es, a.tuninetti@westernsydney.edu.au,
Paola.Escudero@westernsydney.edu.au

ABSTRACT

Spoken word recognition is a hard task. As an aid, native listeners develop segmentation strategies efficiently attuned to phonological properties of their language, like the rhythmic unit (foot, syllable, or mora). If second-language (L2) learners persist in using their own unit, they may experience longer processing times and even miss word boundaries. Therefore, the question arises as to whether highly proficient L2-speakers can inhibit their segmentation habits.

Native Spanish subjects and English-speaking learners of Spanish took a word-spotting test. Participants heard nonsensical words and had to decide whether a real Spanish word or pseudoword was embedded. Some words and pseudowords were stress-initial; others were stress-medial. Different reaction times for both conditions would indicate foot-based segmentation. RTs showed non-significant differences across conditions for either L1 group. English speakers may interpret Spanish unreduced vowels as cues to foot beginning, with their foot-based segmentation having the same effect as syllable-based in this case.

Keywords: spoken word recognition; L2 acquisition; rhythmic units; Spanish.

1. INTRODUCTION

Learners of a second language (L2) typically complain that native speakers speak too fast. This perception has its source in the difficulties that learners have to segment the speech continuum as an aid to lexical access. Spoken word recognition is a complex task even for those who have that language as a mother tongue (L1), but these employ strategies that are efficiently attuned to the phonological properties of the language in question [4].

From infancy, we are biased to paying attention to phonological elements that occur periodically [5]. In some languages, like English, such rhythmicity is provided by the foot [14]; in others, like Spanish, the relevant unit is the syllable [14]; finally, some languages display a mora-based timing, like is the case in Japanese [1]. Native speakers of those languages learn to give more weight to lexical candidates that are aligned with the appropriate rhythmic unit. This procedure increases the efficiency in deciding between competing items and becomes a habit for speech processing that remains during adulthood.

Evidence for such behavior in the L1 has been provided by numerous studies ([8], [12], [13], among others). For example, Cutler & Norris [8] played nonsense sequences like *mintesh* and *mintayve* to their English-speaking subjects and asked them to identify whether a real English word was embedded (*mint* in these examples). Reaction times were faster in cases of the *mintesh* type because the second vowel is a schwa, whereas the full vowel in *mintayve* triggers foot segmentation, rendering [$_\Sigma$ min] [$_\Sigma$ tayve]. This splits the target word between two feet and changes the syllabic affiliation of the /t/. In contrast, [$_\Sigma$ mintesh] forms one only foot and /t/ remains ambisyllabic.

Some experiments have shown that the rhythm-based segmentation strategy does not adapt to the target language the subject is listening to when this belongs to a different rhythm type than their L1 [6]. Cutler et al.'s [7] study with English-French bilinguals shows that even such subjects display one dominant language. Indeed, it appears that a maximum of one rhythm-based strategy can be learned. However, bilinguals do not insist in using their dominant strategy when it turns out to be counterproductive in the other language. Rather, they learn to inhibit it and just rely on other (non-rhythmic) segmentation strategies, like transitional probabilities [15] or phonotactic constraints [17], among others.

The question is raised whether an experienced L2 speaker can also learn to inhibit the rhythmic segmentation when it does not pay off, and how much exposure to language is required to achieve that [3].

The present study examines the performance of English-speaking learners of Spanish with different levels of proficiency, who completed a word-spotting task in Spanish. While also being a syllable-timed language, Spanish allows a broader range of comparisons than French, due to the fact that the latter does not have contrasts based on stress position.

It is hypothesized that in sequences like guilperro (with an initially-stressed target word, perro), the foot boundary coinciding with the word boundary will favor word recognition. In contrast, word identification could be hindered in guilpapel (where papel is split in two feet).

It is also hypothesized that this difference according to stress position should disappear at least in higher-level students.

2. METHODS

2.1. Participants

20 subjects, aged 20–26 (mean 21.5), took part of this perception study (10 native Spanish speakers and 10 native English speakers). One of the English speakers was left-handed; all the other subjects were right-handed.

Among the English speakers, four had a B1 level of Spanish, two had a B2, three had a C1 and one had a C2, according to their Spanish teachers (and following the Common European Framework for Languages [2]).

2.2. Stimuli

A male native Spanish speaker recorded the list of 288 stimuli in a soundproof booth. The stimuli consisted of nonsense sequences of three syllables. In half of the sequences, the last two syllables formed a real Spanish word. These were fairly frequent words in Spanish, usually learned in levels A1 or A2. Additionally, half of the sequences were stressed on the second syllable, and the other half on the third (i.e. word-initially or word-medially in the case of real words). 36 pseudoprefixes were used as the first syllable, in a balanced way across conditions (Real: yes/no, and Stress: initial/medial).

Here are some examples of stimuli, with real words underlined and stressed syllables in boldface: fum<u>**pe**lo</u>, clus<u>de**cir**</u>, dul**gar**so, blampi**bal**.

2.3. Procedure

Prior to the listening test, non-native subjects completed the Dialang [10] vocabulary-size test for Spanish, and scores were annotated.

All subjects were presented a word-spotting task built on PsychoPy, v. 1.90.2. They listened over headphones to the list of stimuli in a random order and had to press one key if they thought there was a real Spanish word embedded, or a different key if they thought it was a pseudoword. For each subject, answer keys (A and L) were randomly assigned to Real “yes” or “no” in order to balance the “yes” key with respect to their dominant hand. By chance, the only left-handed subject was assigned to answer “yes” with their left hand.

Subjects were instructed to give their answers as quickly as possible, and their answers were collected along with the reaction times, measured from the end of each stimulus.

2.4. Data analysis

2.4.1. Outlier exclusion

Cases with raw reaction times outside 2 SD of participants’ mean were discarded as outliers. Additionally, three words had over 40% incorrect answers among the English speakers and were excluded from the analysis.

2.4.2. Analysis of mean reaction times

Reaction times were analyzed by means of mixed-effects linear regression models. A model was run only on the subset of “hit” answers (namely real words that were correctly identified as such). Fixed factors were the main effects of L1 (Spanish or English), and Stress (initial or medial), and their interaction. The model was completed with random intercepts for Subject and Item, and the random effect of Subject on the slope of Stress, bearing in mind that Subject is nested in L1 and Item is nested in Stress [9].

Another model was run including answers of all the Types “hit”, “miss”, “false alarm”, and “correct rejection”. Fixed factors were L1, Stress, and Type, as well as their interactions. Random factors were as in the previous model, with the addition of the effect of Subject on the slope of Type.

L1, Stress, and Type were dummy-coded (reference levels being Spanish, initial, and hit, respectively), whereas Subject and Item were coded as deviation from the grandmean.

2.4.3. Analysis of d-prime

Following Signal Detection Theory [11], d-prime values were calculated for each subject to assess overall answer accuracy. Separate d-prime values were also obtained for initially- and medially-stressed words.

Table 1: Mean (and SD) RTs in seconds by Type of answer, L1, and Stress position.

		Initial	Medial	**Total**
HIT	Spanish	0.588 (0.390)	0.577 (0.382)	**0.582 (0.386)**
	English	0.557 (0.292)	0.542 (0.326)	**0.550 (0.309)**
	Total	**0.573 (0.346)**	**0.560 (0.356)**	**0.566 (0.351)**
MISS	Spanish	0.810 (0.372)	0.649 (0.453)	**0.742 (0.411)**
	English	0.849 (0.382)	0.819 (0.548)	**0.836 (0.460)**
	Total	**0.837 (0.377)**	**0.770 (0.524)**	**0.808 (0.446)**
FA	Spanish	0.874 (0.615)	0.781 (0.516)	**0.828 (0.568)**
	English	0.948 (0.513)	1.015 (0.616)	**0.983 (0.568)**
	Total	**0.920 (0.554)**	**0.930 (0.591)**	**0.925 (0.572)**
CREJ	Spanish	0.862 (0.498)	0.923 (0.559)	**0.892 (0.529)**
	English	0.787 (0.449)	0.795 (0.422)	**0.791 (0.436)**
	Total	**0.825 (0.476)**	**0.860 (0.501)**	**0.842 (0.488)**
TOTAL	**Spanish**	**0.728 (0.472)**	**0.748 (0.508)**	**0.738 (0.489)**
	English	**0.691 (0.407)**	**0.695 (0.429)**	**0.693 (0.418)**
	Total	**0.709 (0.441)**	**0.722 (0.471)**	**0.715 (0.455)**

d-primes for native and non-native Spanish speakers were compared by means of a t-test. A further t-test, paired by subjects, was used to compare d-primes in both Stress conditions. Similarly, paired comparisons were carried out for counts of hits, misses, false alarms, and correct rejections, each for both initially- and medially-stressed words (or pseudowords).

3. RESULTS

Less than 2% of the cases were excluded as outliers following the criteria explained above.

The results of the vocabulary test did not seem very reliable: for instance, some subjects in the C1 level scored lower than some subjects in the B1 level. Therefore, only the standardized levels (B1–C2) were taken into account. A more robust test should be used in future iterations of the experiment. Given the following results, non-native subjects do not appear to have any systematic bias towards knowing more of the stress-initial than the stress-medial words, or vice versa.

3.1. Reaction times

Mean reaction times (RTs) by Type of answer, L1, and Stress position are shown in Table 1. Unexpectedly, the English-speaking subjects (i.e. the non-natives) produce overall shorter RTs than the native Spanish speakers, and that cannot possibly be attributed to age differences [16], given the age homogeneity among groups, $t(15.6) = 1.71$, *ns*. By Type of answer, English speakers respond faster only when it is a correct answer (either hit or correct rejection), but more slowly when they produce incorrect answers (misses or false alarms).

In general, RTs are longer in the medial stress condition, but they are actually shorter for stimuli that are real words (i.e. hit or miss), for both language groups.

There appears to be a trend for shorter RTs in correct answers (especially hits), except that Spanish speakers' correct rejections are the longest type for this group.

Despite the trends that have been pointed out, the linear mixed model shows that neither the main effects of L1 and Stress, nor their interaction, are significant. L1: $F(1, 18.05) = 0.33$, *ns*. Stress: $F(1, 84.14) = 0.53$, *ns*. L1 x Stress: $F(1, 47.61) = 0.13$, *ns*. Substituting Level of Spanish for L1 does not produce any significant effect, either.

Including Type in the model does yield a significant main effect of Type: $F(3, 21.5) = 15.20$, $p < 0.001$, all types increasing RT with respect to reference level (hit). A post hoc comparison with Bonferroni correction shows that, on average, false alarms are coupled to significantly longer RTs than correct rejections; misses, instead, are not significantly different to false alarms or correct rejections with respect to RTs. Additionally, there is a significant triple interaction: L1 x Stress x Type, $F(3, 4734.8) = 3.95$, $p < 0.01$, manifested in an increased RT for English speakers' false alarms given as a response to medially-stressed words.

3.2. d-prime values

Table 2 shows d-prime values for each participant. In general, lower-level students (B1) are less accurate in their responses, while the highest d-primes are achieved by native speakers. However, there are some non-native individuals with better results than some native individuals.

Table 2: Stress-(i)nitial and stress-(m)edial hit and false alarm counts per subject ID (upper bound is 72 in each case). d-prime values are added for each Stress condition, as well as global d-prime, used to sort data from lower to higher overall accuracy of each individual's strategy.

ID	Level	Hit_i	FA_i	d'_i	Hit_m	FA_m	d'_m	d'
13	B1	68	29	1.79	62	29	1.74	1.77
11	B1	66	6	2.77	54	9	2.01	2.36
19	B1	63	13	2.10	64	15	3.07	2.38
12	B2	57	5	2.28	57	2	2.90	2.54
16	C1	68	10	2.67	61	12	2.71	2.69
05	native	68	10	2.64	68	10	2.97	2.78
18	B2	68	4	3.19	64	15	2.52	2.81
08	native	66	5	2.86	59	1	3.41	3.02
07	native	71	12	3.16	71	13	3.27	3.21
15	C1	72	7	4.39	64	13	2.85	3.31
03	native	69	6	3.11	70	5	3.71	3.35
04	native	69	3	3.46	65	3	3.38	3.42
01	native	70	3	3.65	71	8	3.34	3.43
14	B1	67	1	3.68	64	7	4.65	3.59
10	native	71	5	3.66	70	5	3.60	3.63
20	C2	70	4	4.68	67	6	3.51	3.90
17	C1	64	1	3.48	61	2	3.94	3.93
06	native	70	4	3.79	67	4	4.65	4.01
02	native	71	1	4.34	71	5	3.92	4.11
09	native	71	2	4.39	72	5	4.65	4.24

Regardless of the native language, there appears to be some subjects with higher scores in medial-stress condition, and others with higher scores stress-initially.

However, d-primes are not significantly different in the initial stress condition from the medial stress for either language group, as reflected by paired t-tests. Spanish: $t(9) = 1.42$, *ns*; English: $t(9) = 0.04$, *ns*.

A paired comparison of the count of each answer type shows that native Spanish speakers produce more correct answers stress-initially, whereas there is no Stress effect on the incorrect answers. Hit: $t(9) = 16.38$, $p < 0.001$. Miss: $t(9) = 1.96$, *ns*. False alarm: $t(9) = 0.12$, *ns*. Correct rejection: $t(9) = 3.80$, $p < 0.01$. The same finding is true for English speakers. Hit: $t(9) = 5.11$, $p < 0.001$. Miss: $t(9) = 0.85$, *ns*. False alarm: $t(9) = -0.74$, *ns*. Correct rejection: $t(9)= 3.88$, $p < 0.01$.

4. DISCUSSION AND CONCLUSIONS

English-speaking subjects show no different RTs from Spanish speakers in the word-spotting task, and no different d-primes according to Stress position either, which could be taken to mean that they have already learned to inhibit foot-based segmentation, at least starting from level B1. Further studies with a larger sample are needed to confirm this and to see what happens with learners of the lowest levels, A1 and A2, which were not represented in this study.

If they are really not using the foot for segmentation, they might be resorting to other strategies, like phonotactic constraints. Consonantal transitions between the pseudoprefixes and the words were divided into four Groups, according to whether they are legal within words in Spanish, English, both, or none (illegal sequences should necessarily signal a word boundary and facilitate segmentation). There is found to be no significant effect on RTs of either Group, $F(3, 26.9) = 1.52$, or the interaction Group x L1, $F(3, 26.9) = 2.05$.

A subtle difference between this design and that of previous studies could also explain these results. Here the nonsense sequences (*fumpelo*, *clusdecir*, etc.) had an extra initial syllable instead of final like in *mintesh* and *mintayve*. In the stress-medial condition, even if it is true that the target word is split between two feet, there is no change in the syllabic affiliation of any segment, as happened with *mintayve*. A future version of this experiment should compare between logatomes including the extra syllables at the beginning and at the end.

An alternative explanation has to do with Spanish having no vowel reduction in unstressed syllables. Should English-native listeners not pay attention to the stress position itself, but take those full vowels to mean that every syllable is foot-initial, in that case the foot-based strategy would have the same effect as syllable-based segmentation. Future studies with a syllable-timed language that nonetheless has vowel reduction, like Catalan, may cast light on this.

Spoken word recognition is a phenomenon of paramount importance in language processing, and poses a lot of troubles for L2 learners. However, more evidence is needed to know all the factors affecting the cognitive processes that lie behind it, especially in the case of the processing of non-native languages.

5. ACKNOWLEDGMENTS

This work has been partially funded by the Erasmus+ Programme of the European Union and by the Australian Research Council Centre of Excellence for the Dynamics of Language.

The authors are grateful to Anne Cutler for her helpful comments on this paper. Our thanks also go to the anonymous reviewers, whose insightful suggestions have helped to improve the manuscript. Finally, we want to thank Enrique Santamaría for his help in recruiting participants.

6. REFERENCES

[1] Bloch, B. 1950. Studies in Colloquial Japanese IV: Phonemics. *Language* 26, 86–125.

[2] Council of Europe, 2001. *Common European framework for languages: Learning, teaching, assessment*. Strasbourg: Language Policy Division.

[3] Cutler, A. 2000. Listening to a second language through the ears of a first. *Interpreting* 5, 1–23.

[4] Cutler, A. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT.

[5] Cutler, A., Mehler, J. 1993. The periodicity bias. *Journal of Phonetics* 21, 103–108.

[6] Cutler, A., Mehler, J., Norris, D., Segui, J. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25, 385–400.

[7] Cutler, A., Mehler, J., Norris, D., Segui, J. 1992. The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology* 24, 381–410.

[8] Cutler, A., Norris, D. 1988. The Role of Strong Syllables in Segmentation for Lexical Access. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113–121.

[9] Gelman, A., Hill, J. 2007. *Data analysis using regression and multilevel / hierarchical models*. Cambridge: Cambridge University Press.

[10] Lancaster University. n. d. *Dialang*. https://dialangweb.lancaster.ac.uk/

[11] MacMillan, N. A., Creelman, C. D. 1991. *Detection theory: A user's guide*. Cambridge: Cambridge University Press.

[12] Mehler, J., Dommergues, J. Y., Frauenfelder, U., Segui, J. 1981. The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior* 20, 298–305.

[13] Otake, T., Hatano, G., Cutler, A., Mehler, J. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32, 258–278.

[14] Pike, K. L. 1945. *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.

[15] Saffran, J. R., Newport, E. L., Aslin, R. N. 1996. Word segmentation: The role of distributional cues. *Journal of Memory and Language* 35, 606–621.

[16] Thompson, J. J., Blair, M. R., Henrey, A. J. 2014. Over the hill at 24: Persistent age-related cognitive-motor decline in reaction times in an ecologically valid video game task begins in early adulthood. *PLoS ONE*, *9*(4), e94215.

[17] Weber, A., Cutler, A. 2006. First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, *119*(1), 597–607.