DEVELOPMENT OF ACCENTUAL CATEGORIES IN JAPANESE AS A SECOND LANGUAGE

José Joaquín Atria, Valerie Hazan

Speech Hearing and Phonetic Sciences, UCL (University College London)
j.atria.11@ucl.ac.uk, v.hazan@ucl.ac.uk

ABSTRACT

Previous research shows that, Spanish speakers studying Japanese face significant difficulties when perceiving Japanese word accent contrasts, particularly between accented and unaccented words. We hypothesise that this is probably due to poorly developed accent categories in the target population. The results of a categorical perception test show great differences between native and non-native listeners in both category definition and boundary positions. Results also show differences between categories, in that some are more clearly defined than others. This partly explains response biases reported in the literature.

Keywords: Word Accent, Perception, Prosody.

1. INTRODUCTION

Both Japanese and Spanish have ways of marking word-level prominence, and in both cases this "accent"[1] has a contrastive role: it is a mark of verb tense in Spanish and distinguishes lexical contrasts; and as much as 14% of minimal sets in Japanese rely on accent placement for differentiation [15]. Japanese accent provides speakers with non-linguistic information about who the speaker is and where they stand due to a large degree of regional variation. Regardless, Japanese intonation in general and its accent in particular, are not often covered in depth in classes of Japanese as L2.

1.1. Acoustic differences

Both languages use different sets of acoustic cues to mark accent position: while in Spanish it is marked by a combination of F0, duration, and intensity [8, 9], in Japanese F0 is the only reliable cue [6]. Although in both cases pitch is the main perceptual cue for the accent, the relevant pitch movement is also different. In Spanish, the accent is marked by a pitch peak, roughly aligned with the end of the accented syllable but often displaced into the first half of the one following [10]. In Japanese, on the other hand, the accent is marked by a pitch fall immediately following the accented syllable[2] [16, 6].

These type of differences might suggest that L1 characteristics would not necessarily provide benefits for accent perception in Spanish and Japanese. However, this is not the case: a recent study by Kimura done on the opposite linguistic population (Japanese speakers studying Spanish, J1S2), showed native-like performances for non-native listeners with isolated words and words in declarative phrases [4]. And like the case stated above, classes of Spanish as L2 also do not normally focus on accent or intonation.

The comparative richness in Spanish accentual cues gives J1S2 an advantage, which accounts for part of the results in the cited study. But it also means that no individual cue needs to bear the entire burden of communicating the accent, which allows for more conflicting cues under certain conditions. This was also found in that study: in sentence contexts in which the accent in the target word was realised as a L* (instead of a more canonical L+H* or L*+H), the performance of J1S2 listeners dropped significantly, while that of native speakers showed no change.

By contrast, since Japanese has but one reliable acoustic cue for accent position, that cue has become very robust against changes due to sentence intonation [16]. This was tested by a recently reported study in which participants were presented with accentual

Table 1: Keywords used for stimuli generation

Trio	Accent position		
	First $\sigma^{\downarrow}\sigma$	Final $\sigma^{\uparrow}\sigma^{\downarrow}$	Unaccented $\sigma^{\uparrow}\sigma$
haɕi mo	chopsticks	bridge	edge
kaki mo	oyster	fence	persimmon

Table 2: Participant demographics

Group	n	Women	$\bar{x}_{age}$
Native	24	16	23.57
Non-native	34	26	22.27

trios in 3AFC categorical perception task, following the study by Kimura cited above. The results of that study showed no significant change in the performance of S1J2 when presented with different intonation patterns [1].

They did however find significantly poorer performance rates across the board, and in particular with certain accent types: S1J2 showed evidence of a bias towards perceiving words as accented (as opposed to *unaccented*), and as perceiving accented words as accented on the first syllable.

The primary role of F0 in both target languages is without question [8, 9, 2, 16, 6], as is the sensitivity of Spanish speakers to pitch differences. And while it is true that the Japanese accentual cues are few when compared to Spanish ones, it is also true that are more reliable. So why do S1J2 perform so badly across all contexts?

The response biases shown in [1] suggest it to be due to ill defined categories. This is all the more likely since the category that proved most difficult—unaccented words—does not exist in Spanish.

If this is the case, S1J2 should have significantly less-defined boundaries for the unaccented category than for the accented ones. And judging from the biases from the previous study, one might expect to find differences for different accent types, with the initial accent being the most well-defined.

Since the above cited studies have also found differences in the interpretation J1S2 made of pitch cues in Spanish [5], we also wanted to test if this was the case with S1J2 for Japanese.

2. METHOD

To test these hypotheses, we ran a 3AFC categorical perception test on a group of S1J2 and native Japanese speakers. The test used a synthetically generated continuum based on natural recordings.

2.0.1. Stimuli

We used the same set of words used in [1]: two sets of CVCV accentual minimal trios, shown in table 1. Since final-accented and unaccented words are indistinguishable in isolation [17, 16, 6], the conjunction particle *mo*—which does not alter the accent of the word it modifies [16, p.159]—was added for a total of 3 syllables per word.

The 12 keywords were framed in low-predictability carrier sentences, and 10 repetitions of each were recorded by two native Japanese speakers from Tōkyō (one female, one male; both of them in their late 20s). Recordings were rated by native speakers, and the most highly rated were chosen to generate the stimuli.

Figure 1: Schematic progression of the continua. F0 contours illustrate steps 1, 10, and 20 of the stimuli based on the female speaker's /haɕimo/.

Based on confusion patterns shown in the previous study [1], two contrasts were devised: one between first and final-accented words; and another between final-accented and unaccented words, as shown in fig. 1. Independent 10-step F0 continua were generated for each of these contrasts using Tandem-STRAIGHT [3], with the remaining features set to a fixed point.

The resulting 40 ($2 \times 2 \times 10$) stimuli were compressed as 192kbps MP3 files to be used in the testing platform, and each of them presented in a different random order per participant six times for a total of 240 stimuli.

In each trial, participants listened to one of the synthetically generated stimuli in a carrier sentence, and were presented with three buttons labelled with the individual items in the trio to which the word belonged. They were asked to click the one corresponding to the word they thought they heard. All instructions and button labels was presented in Japanese, using the standard Japanese orthography in *kanji* and *hiragana* script. Since Japanese orthography does not mark the position of the accent, the accented syllable was marked in red.

Participants were shown sets of six natural training items in a random order before beginning the test. Only participants who managed to get the entire set correct were allowed to continue. They had up to four training sets.

2.0.2. Participants

Participants were recruited remotely in Chile and Japan with the help of local assistants. The control group was composed of students at local universities in Tōkyō. The experimental group was recruited among university students in Santiago majoring in Japanese.

Before the test, participants were asked to fill a linguistic background survey providing broad information about their L2 proficiency and use. Self-reported assessment was fairly constant, and the large majority of participants claimed to be in a beginner-to-intermediate level ($\bar{x} = 2.02$ in a 5-point scale; $\sigma = 0.86$). Additional demographic information is found in table 2.

2.1. The Testing Platform

LimeSurvey [7] was used to present the items. Participants wore headphones and took the test in a quiet room under experimenter supervision. Each testing session lasted ~45 minutes including regular breaks.

3. RESULTS

3.1. Analysis

The plots shown in fig. 2 show the results for the perception test. Rows show the results for native (top) and non-native (bottom) listeners, and columns show the results for the individual contrasts: initial and final-accented on the left; final-accented and unaccented on the right. Responses for different categories are shown with differently-styled lines.

A steeper change in the number responses for a given category along the steps in the continuum would be evidence of a more clearly defined category boundary. To compare if both populations were judging pitch in similar ways, boundary values were also extracted. The boundary was defined as the point at which responses for a given category were expected to reach 50% despite the test had a 3AFC design. This because the continua were generated and analysed separately for each two-way contrast between members of the trio. That meant one would expect to see most responses belonging to two of the three categories, the third one being a sort of distractor.

Responses were analysed using a probit analysis [13, 11, 12] on the responses of each participant for each of the two contrasts: first and final, and final and unaccented. The fitted probit models were then used to predict the position of the category boundary and its slope. When responses remained below 50%, no category boundary was set.

These values were later analysed using 2-way ANOVAs with the response category as the dependent variable, and group (native or non-native) and step as the predictors. All tests were run using the `R` statistics package [14].

3.2. Japanese experiment

3.2.1. Category slopes

The results shown in fig. 2 illustrate dramatic differences between native and S1J2 listeners. Native listeners (top row) show very clear category boundaries and display ceiling and floor effects, which show the generated stimuli had no significant problems. On the other hand, responses of non-native listeners not only reach lower levels overall, even at their peak, but they also display shallower slopes, getting progressively more shallow the closer they get to the unaccented end.

This is confirmed by the data in table 3, which shows mean slope values for the response of each group (native or non-native) for each accentual category and contrast. Slopes for native listeners have clearly greater magnitudes than those of non-native speakers, and the table also shows the extent to which S1J2 listeners' responses approach a horizontal in the second contrast. This interaction between category and group was significant for both the first contrast ($F_{(2,336)} = 157.13$, $p < 0.001$) and the second ($F_{(2,336)} = 22.52$, $p < 0.001$).

The analysis also showed a significant main effect of category for both contrasts ($F_{(2,336)} = 697.29$, $p < 0.001$ and $F_{(2,336)} = 36.85$, $p < 0.001$, respectively). However, this result could be skewed by the presence of the third, non contrasted category. When this category (unaccented in the first contrast, first-accented in the second contrast) was removed, results of a new ANOVA showed no significance for the interaction in the first contrast, but did show it for the second contrast ($F_{(1,224)} = 276.56$, $p < 0.001$).

3.2.2. Category boundaries

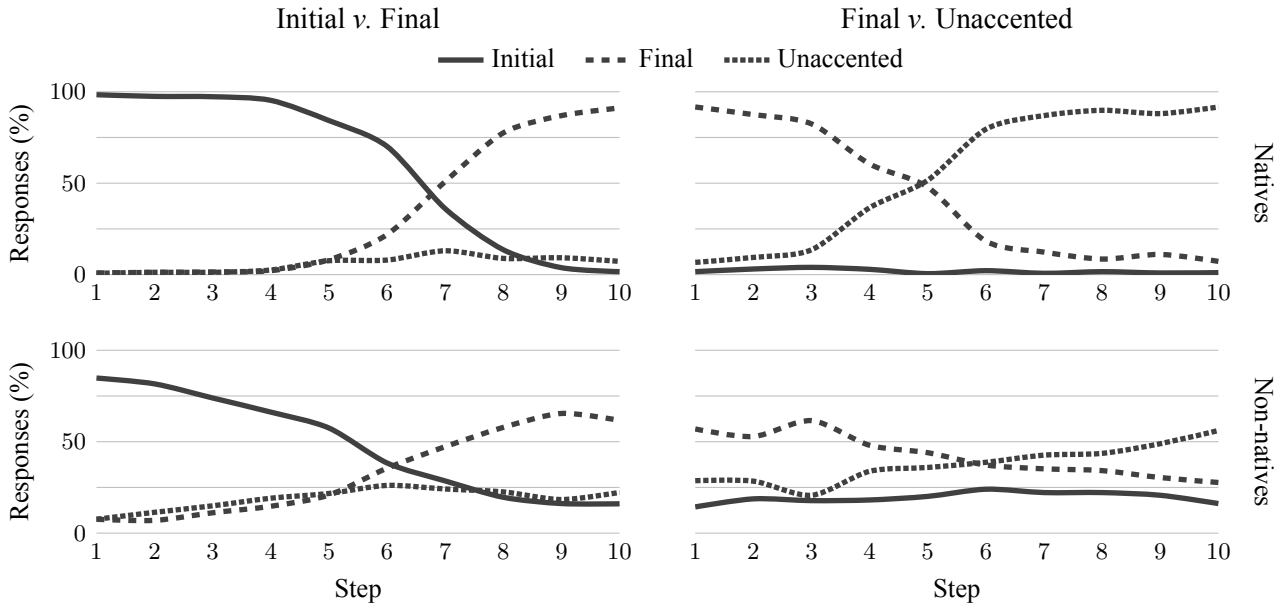
Since J1S2 responses in the second contrast do not clearly cross the 50% mark in the second contrast, boundary positions could not be significantly estimated for this contrast. This is of course telling in itself.

Responses for the first contrast, however, show much clearer boundary positions for the S1J2, which are also shifted from those of the native listeners (meaning that non-native listeners started changing their responses sooner in the continuum than their native counterparts). Part of this difference might be due to a greater number of unaccented responses by S1J2, but even if those responses are discarded the point where the other two curves cross is a whole step further to the left of the native responses. This group difference is confirmed in the ANOVA ($F_{(1,205)} = 11.90$, $p < 0.001$), which also showed significant results for the interaction between group and category

Table 3: Mean slope values. Larger absolute magnitudes show sharper category boundaries.

Category	Initial *v.* Final		Final *v.* Unaccented	
	N	NN	N	NN
Initial	−0.849	−0.320	−0.146	0.012
Final	0.749	0.294	−0.510	−0.108
Unaccented	0.112	0.068	0.540	0.067

Figure 2: Responses for the Japanese experiment



($F_{(1,205)} = 19.31$, $p < 0.001$).

A main effect of category was also found to be significant ($F_{(2,205)} = 50.84$, $p < 0.001$).

4. DISCUSSION

Results from the experiments confirm what previous studies had suggested: the poor performance of S1J2 can at least partly be explained by their comparatively poorer development of accurate L2 accentual categories. Furthermore, the categories that they *have* developed are not all created equal, and they seem to follow the pattern shown in the results reported in [1], with significantly greater slope values for the accented categories, and an almost completely flat response for the unaccented category.

Still, however low the slope values (particularly for unaccented words) they do show a continuously rising trend towards the expected end of the continuum. This is evident if the plots in fig. 2 are taken horizontally as a single continuum, since the line corresponding to unaccented responses rises slowly but steadily across. Values in table 3 also show this: low though the values may be, they do have the right sign.

This is not evidence of a category. But it does seem to be evidence of a trend, of some degree of sensitivity to the relevant cue (pitch), and perhaps of a general sense that, whatever feature is being perceived, it is paired with another as-of-yet non-category.

Nothing in the current results can account for the difference in the placement of the boundary positions (when they exist). This interaction between group and category in the position of category boundaries was one of the significant results we found, but the question of what explains that interaction remains open. We stated in the introduction that a lower number of acoustic cues provided at the same time a poorer stimulus and a more constant one, since that single cue carried the burden of communicating the contrast by itself. This might not be the case in every case, but it certainly is the case in Japanese.

However, when we say an acoustic cue is not available, or not relevant, it does not mean that the acoustic correlates of that cue disappear from the signal; it only means that differences in that cue are not informative for the identification of the contrast in the target language. It is perfectly possible that other acoustic cues, which are naturally disregarded by the native speakers, are playing a more subtle role and appearing as conflicting cues for the L2 listeners.

This is probably not the case with duration, since there are little if any duration differences in Japanese accentual minimal pairs, and certainly we found none in the stimulus we were using. But intensity might still be responsible. Indeed, the target words used in this study and shown in table 1 are susceptible to high-vowel devoicing (a common phenomenon between voiceless consonants), in addition to the natural intensity differences between open and closed vowels, like the ones in both /kaki/ and /haɕi/. The heightened peak responses for initial-accented words could be, then, an effect of these intensity differences inherent in our stimuli, and not in fact due to greater categorical development. Further study is needed to resolve this question.

5. REFERENCES

[1] Atria, J. J., Hazan, V. 2013. Problemas en la percepción del acento japonés para hablantes de castellano [Issues in the perception of Japanese accent for Spanish speakers]. 20º Congreso de la Sociedad Chilena de Lingüística.

[2] Hualde, J. I. 2005. *The Sounds of Spanish with Audio CD*. Cambridge University Press.

[3] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., Banno, H. 2008. Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation. *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008* 3933–3936.

[4] Kimura, T., Sensui, H., Takasawa, M., Toyomaru, A., Atria, J. J. 2012. Influencia de la entonación oracional sobre la percepción del acento español por estudiantes japoneses. (21), 11–42.

[5] Kimura, T., Sensui, H., Takasawa, M., Toyomaru, A., Atria, J. J. 2013. Jōshō intonēshonka deno kyōsei no onseiteki jitsugen [Phonetic phenomena affecting the stress in a rising intonation]. 59º Congreso de la Asociación Japonesa de Hispanistas.

[6] Labrune, L. 2012. Accent. In: *The Phonology of Japanese*. Oxford University Press 178–266.

[7] LimeSurvey Project Team/Carsten Smith, 2012. *LimeSurvey: An Open Source survey tool*.

[8] Llisterri, J., Machuca, M. J., de la Mota, C., Riera, M., Ríos, A. 2003. Algunas cuestiones en torno al desplazamiento acentual en español. In: Herrera Z., E., Martín Butragueño, P., (eds), *La tonía: dimensiones fonéticas y fonológicas*. El Colegio de México 163–185.

[9] Llisterri, J., Machuca, M. J., de la Mota, C., Riera, M., Ríos, A. 2005. La percepción del acento léxico en español. In: *Filología y lingüística. Estudios ofrecidos a Antonio Quilis* volume 1. 271–297.

[10] Llisterri, J., Marín, R., de la Mota, C., Ríos, A. 1995. Factors affecting f0 peak displacement in spanish. *Eurospeech'95. Proceedings of the 4th European Conference on Speech Communication and Technology* volume 3. ISCA 2061–2064.

[11] Mayo, C., Turk, A. 2004. Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions. *The Journal of the Acoustical Society of America* 115, 3184.

[12] Mayo, C., Turk, A. 2005. The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *The Journal of the Acoustical Society of America* 118(3), 1730–1741.

[13] Nittrouer, S., Studdert-Kennedy, M. 1987. The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech, Language and Hearing Research* 30(3), 319.

[14] R. Core Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

[15] Sibata, T., Shibata, R. 1990. Akusento wa dōongo wo dono teidō benbetsu shiuru ka [Is word-accent significant in differentiating homonyms in Japanese, English and Chinese?]. 17(7), 317–327.

[16] Vance, T. 2008. *The sounds of Japanese*. Cambridge University Press.

[17] Vance, T. J. 1995. Final accent vs. no accent: utterance-final neutralization in tokyo japanese. *Journal of Phonetics* 23, 487–499.

---

[1] Traditionally the word used when talking about Spanish has been "stress", but since the reasons for this are not immediately relevant the same term will be used indistinctly. Furthermore, the term "accent" will always be used in the sense of "word-level accent".

[2] The distinction between moras and syllables as the accented units is, likewise, not relevant to this discussion.