Is segmental foreign accent perceived categorically?

Rubén Pérez-Ramón[a,*], Martin Cooke[b,a], María Luisa García Lecumberri[a]

[a] *Language and Speech Laboratory, Universidad del País Vasco, 01006 Vitoria, Spain*
[b] *Ikerbasque (Basque Science Foundation)*

Abstract

Non-native speech is typically investigated at the level of utterances; as a consequence the segmental basis of foreign accent and its role remain unclear. A substantial part of the second language learning process involves acquisition of sounds that are disimilar to the sounds of a learner's native language, which nevertheless strongly influences learners' perceptions and productions. It is therefore of interest to study the role of foreign-accented productions at the level of individual segments. The principal issues addressed by the current study are whether accentedness at the segmental level is judged categorically by native listeners, and whether consonantal and vocalic segments are treated similarly. British English listeners judged as native or non-native a series of words in which a single vowel or consonant had been replaced by its Spanish-accented counterpart. The degree of segment accent was varied in equal amounts along a 21-step continuum using a blending technique based on native and non-native segments excised from words spoken by a balanced bilingual talker. Listeners assessed 24 distinct consonant and vowel continua. Across all vowel or consonant continua listeners' nativeness judgements varied with segment nativeness in a non-categorical fashion. However, most individual consonant continua, as well as those vowel continua that involved durational changes, were perceived categorically. These results suggest that while overall segment-level foreign accent might be considered to vary in degree, in reality at the level of individual segments –where second language learners' pronunciation awareness and control has to be focussed– small acoustic changes can convey a foreign accent.

Keywords: foreign accent, consonants, vowels, MOS, categorical perception

[*] Corresponding author
Email address: rperez.ram@gmail.com (Rubén Pérez-Ramón)

June 28, 2019

1. Introduction

Non-native pronunciation frequently exhibits properties which differ from those of native speakers. These differences, largely due to the influence of a non-native speaker's first language (L1), are perceived by listeners as a foreign accent (FA), principally caused by deviations from native norms at segmental and suprasegmental levels. However, the nature and role of individual speech characteristics such as accented realisations of specific phonetic segments are not well-understood. Instead, most FA research has employed holistic forms of assessment, in which listeners judge accent globally using speech materials for which higher-order linguistic levels provide potential confounds. Consequently, individual phonetic cues to foreign accent have been in the main unexplored, in spite of claims that global FA evaluations permit the identification of specific acoustic variables that lead to listeners' judgements (e.g., Southwood and Flege, 1999).

The focus of studies which have examined individual cues to FA has been on suprasegmental features such as speech rate (Munro and Derwing, 2001), duration (Tajima et al., 1997) and nuclear stress (Hahn, 2004) which are more easily isolated for analysis. An interesting exception is expt. 4 of Flege (1984) who used a cross-splicing technique to determine whether listeners could detect FA in individual phoneme-sized segments, although that study was limited to a single plosive consonant followed by one of two vowels. Some researchers (e.g., Anderson-Hsieh et al., 1992; Derwing and Munro, 1997; Munro et al., 2006) have correlated listener assessments of FA with error types such as segmental, prosodic, or grammatical errors. However, correlational methods cannot provide sufficiently detailed and direct information about the role of specific FA characteristics because they do not control individual cues. More recently, speech manipulations such as low pass filtering or backward speech have been used in order to focus on the phonetic characteristics in non-native speech. Munro et al. (2010) presented native and foreign-accented utterances in English to native English listeners in backwards speech and several masking conditions. Participants were asked to label each token either as foreign or native. Their results suggest that listeners are able to detect foreign-accented speech even when considerable prosodic variation is removed.

Holistic studies of foreign accent (e.g., Piske et al., 2001) have typically asked listeners to evaluate speech samples using rating scales for degree of foreign accent (DFA) . The underlying idea is that FA will appear with varying levels of strength between two extreme values (‘no FA’ to ‘very strong accent’). In a study comparing various DFA rating scales – viz. equal-intervals (Levi et al., 2007; Aoyama et al., 2008; Burda et al., 2003), continuous (Munro, 1993), and direct magnitude estimation (Brennan et al., 1975) – Southwood and Flege (1999) found that listeners were capable of partitioning FA into equal intervals, suggesting that accent varies by degree rather than in an all-or-nothing way. However, DFA judgements have typically been made using speech materials in which speaker-related variables such as age of arrival (Southwood and Flege, 1999) or length of residence (Oyama, 1976) form the axis along which DFA is judged. Consequently, the relationship between accent ratings and actual accent is, at best, indirect. No study has yet examined the effect on accent judgements of controlled acoustic manipulations along a non-native to native continuum. That is the purpose of the current investigation. Specifically, we examine the effect of Spanish-accented vowel and consonant segments inserted into English words.

The primary question concerns whether foreign accent is perceived categorically or not when applied at the level of individual segments. For the purposes of the current study we define ‘categorical’ to mean that *equal-sized changes in the acoustic stimulus result in unequal-sized changes in perceived nativeness*. Under this definition, a non-categorical relationship would manifest itself as a quasi-linear function linking stimuli on a continuum of equal acoustic steps to corresponding nativeness judgements. An operational definition of categorical/non-categorical patterning based on curve-fitting is given in section 4.3.

Classical categorical perception studies (e.g., Liberman et al., 1957) suggest that listeners tend to classify acoustic continua categorically, with sharp perceptual boundary changes between phonological categories, and little or no sensitivity to change within those categories. Categorical perception studies have involved many classes of speech sounds, including plosives (e.g., Liberman et al., 1961), nasals (e.g., Larkey et al., 1978) and fricatives (e.g., Formby et al., 1996). A few studies (Ikeno and Hansen, 2006; Flege, 1984; Park, 2013) have evaluated FA in a similar binary fashion i.e. by asking listeners whether a given sample is accented or non-accented. For instance, Park (2013) demonstrated that listeners’ detection of FA in monosyllables is dependent on segment production rather than on syllable structure. Ikeno and Hansen (2006) showed that accent detection (both regional and foreign) is dependent on listeners' familiarity with the accent and that significantly better accent detection results are obtained with longer speech samples (phrases vs. single words). The aforementioned study by Flege (1984), where experienced and naive listeners were asked to detect non-nativeness in stimuli containing segment-level material spliced from accented tokens, is the closest in spirit to what we propose here. However, rather than using only fully-accented segments, the current study takes the lead from work in categorical perception by generating a continuum from fully-accented to fully-native segments.

If listeners perceive accent at the segment level in a graded fashion, we will observe a non-categorical relationship between position along the continuum and judged nativeness. On the other hand, if listeners respond categorically to the question of whether a segment is accented or not, we anticipate a rapid change from non-native to native judgements for small movements along the continuum.

A second issue is whether foreign-accented vowels are perceived in the same manner as foreign-accented consonants. Differences between vowel and consonant perception have long been documented in a variety of experiments (e.g., Fry et al., 1962; Pisoni, 1973); see Cutler (2012) for a summary. Additionally, and with respect to English, there is much more variation across dialects and accents with respect to vowels than to consonants, to such an extent that even the number of vowels in the inventory may differ from one regional accent to another (Wells, 1982).

English vowels are notoriously difficult to acquire for Spanish speakers: even advanced learners use a barely-modified Spanish vowel inventory when speaking English (Iverson and Evans, 2009). Indeed, none of the 5 Spanish vowels matches a unique English prototype (Hualde, 2005; Cebrián, 2019). Spanish vowels do not differ in duration as they do in English, and most can be assimilated to more than one English vowel category e.g., Spanish /o/ differs in F1 from both English /ɒ/ and /ɔː/; Spanish /i/ is shorter than English /iː/ and less peripheral (Bradlow, 1995); Spanish /e/ is higher than English /e/ (Hualde, 2005) and less centralised than English /ɪ/. Given these facts, Spanish-accented English vowels might introduce greater uncertainty than consonants, compounded by the fact that English regional variation might lead to more overlap between vowel categories. Accordingly, we hypothesise that a graded FA continuum is less likely to be perceived categorically for vowels than for consonants.

In the current study native English listeners judged English words, into which a single Spanish-accented consonant or vowel had been spliced, as native or non-native. Listeners classified stimuli from 24 continua (13 consonant and 11 vowel continua), where each continuum consisted of a sequence of 21 realisations of a single word in which the accentedness of one segment was modified from fully-native at one end of the continuum to fully-non-native at the other end. The amount of segment-level accent applied to each stimulus was controlled using a segment synthesis technique (section 2) which involved blending normalised native and non-native exemplars excised from target words. To eliminate speaker-related differences, and following García Lecumberri et al. (2014), speech material from a fully-balanced English-Spanish bilingual talker was used for both native and non-native accented tokens. Listeners also rated the quality of each stimulus using a mean opinion score (MOS) task.

2. Creating grades of foreign accent at the segmental level

The goal of the gradation technique is to generate different degrees of foreign accent, thereby eliciting a continuum of English words in which the pronunciation of a single segment is gradually modified from fully foreign-accented to fully native-accented, while the rest of the word remains unchanged. For example, the initial sound [h] of the English word 'house' would be replaced by a segment varying from [h] at the native end of the continuum to [x] at the non-native end.

The technique consists of the isolation of native and non-native segments followed by their recombination in which the segments are blended, using the process described below. Modified segments are then attached to the unaltered remainder of the word. Performing this manipulation at a range of different blending weights results in a continuum of words, each of which contains a different degree of acoustic foreign accent in only one segment.

2.1. Consonant continua

The gradation process is summarized in figure 1. The continuum is understood to have n equal steps in the interval $[0-1]$, where 0 represents the non-native end and 1 the native end. The derivations are given for a step at an arbitrary blending point λ in the range $[0-1]$.

1. **Excise native and non-native segments**. The gradation technique requires a native word and a non-native token, which may be a word or a nonword, in which the designated segment is in the same position

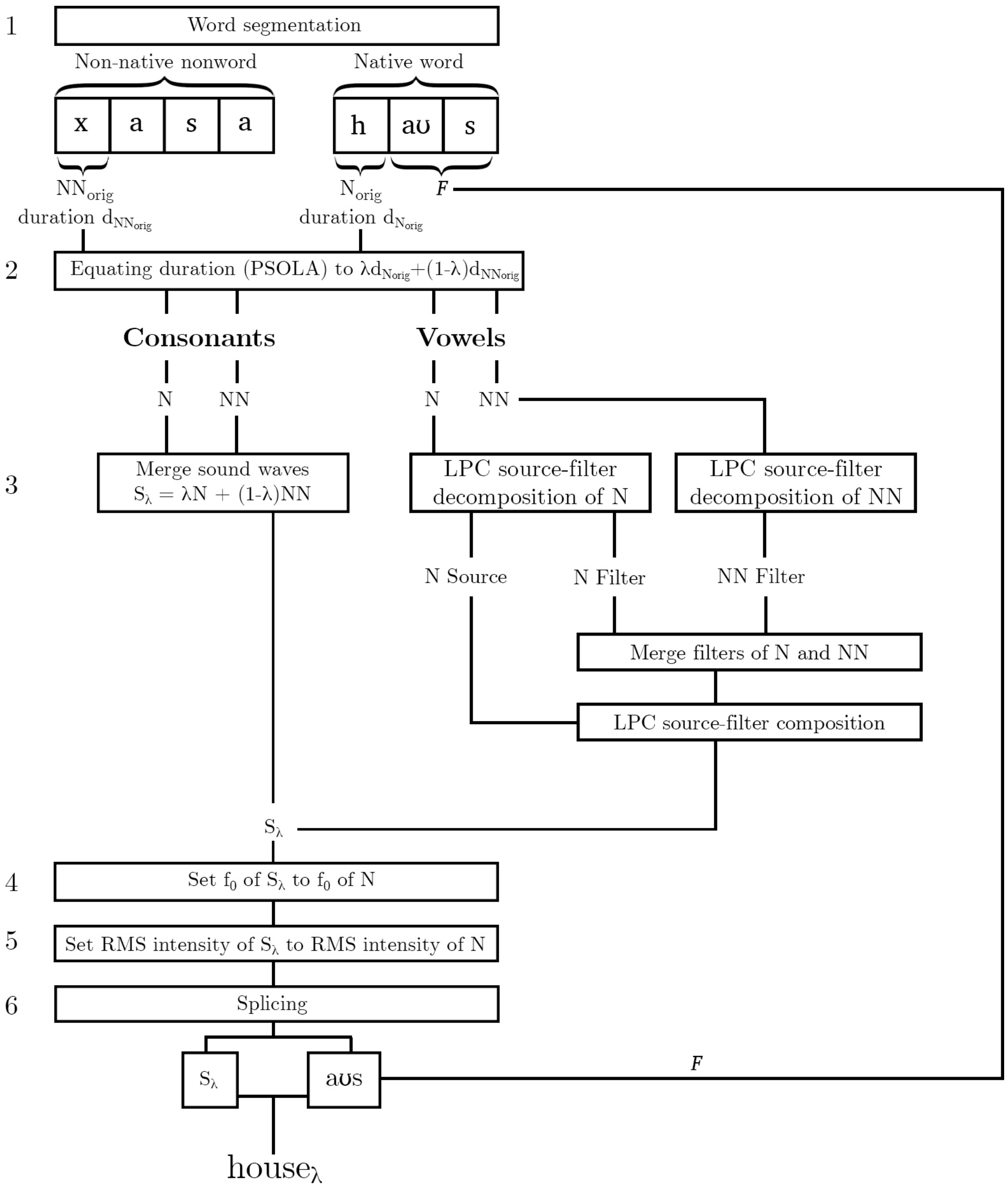

1
Word segmentation
Non-native nonword
Native word
x
a
s
a
h
aʊ
s
NN$_{orig}$
duration d$_{NN_{orig}}$
N$_{orig}$
duration d$_{N_{orig}}$
F
2
Equating duration (PSOLA) to λd$_{N_{orig}}$+(1-λ)d$_{NN_{orig}}$
Consonants
Vowels
N
NN
N
NN
3
Merge sound waves
S$_λ$ = λN + (1-λ)NN
LPC source-filter decomposition of N
LPC source-filter decomposition of NN
N Source
N Filter
NN Filter
Merge filters of N and NN
LPC source-filter composition
S$_λ$
4
Set f$_0$ of S$_λ$ to f$_0$ of N
5
Set RMS intensity of S$_λ$ to RMS intensity of N
6
Splicing
S$_λ$
aʊs
F
house$_λ$

Figure 1: The gradation process for step λ of the continuum illustrating how native and non-native segments are processed to generate an intermediate token. Note that the process applied to consonant and vowel continua differs at stage 3.

and same or similar phonetic context as the native segment. The native segment N_{orig} is segmented from the rest of the word, denoted F (frame); similarly, the non-native segment NN_{orig} is cut out from the host token.

2. **Equate segment durations**. Since duration can be a cue for discriminating between segments, each step of the continuum is generated with a modified duration $d_\lambda = \lambda d_{N_{orig}} + (1-\lambda)d_{NN_{orig}}$. Both N_{orig} and NN_{orig} are scaled to have duration d_λ, resulting in two new segments N and NN.
3. **Blend segments**. Here the procedure differs for consonant and vowel constinua.

   Consonants. The duration-normalised segments N and NN are combined to produce a new segment computed as the weighted sum $S_\lambda = \lambda N(t) + (1-\lambda)NN(t)$.

   Vowels. Instead of blending N and NN segments directly at the waveform level, source-filter decomposition is performed independently on the N and NN segments, after which the filter components are blended according to the value of λ and recombined with the source component of the N segment. In the current study, source-filter decomposition was accomplished using the Burg algorithm (Burg, 1975; Press et al., 1992) as implemented in Praat (Boersma and Weenink, 2018). The overall effect is to create a continuum in the spectral (and hence formant) structure from the N to the NN segment.

4. **Match fundamental frequency to frame**. The f_0 of S_λ is replaced by that of N in order to avoid transition-zone artefacts when splicing S to the frame F.
5. **Normalise levels**. To accommodate any differences in recording level between the native and non-native material, the RMS energy of S_λ is scaled to match that of N.
6. **Splice new segment to frame**. The final stage of the gradation process is the concatenation of the generated segment S_λ to the frame F using an overlap-add technique.

Figure 2 provides spectrographic examples of two vowel and two consonant continua.

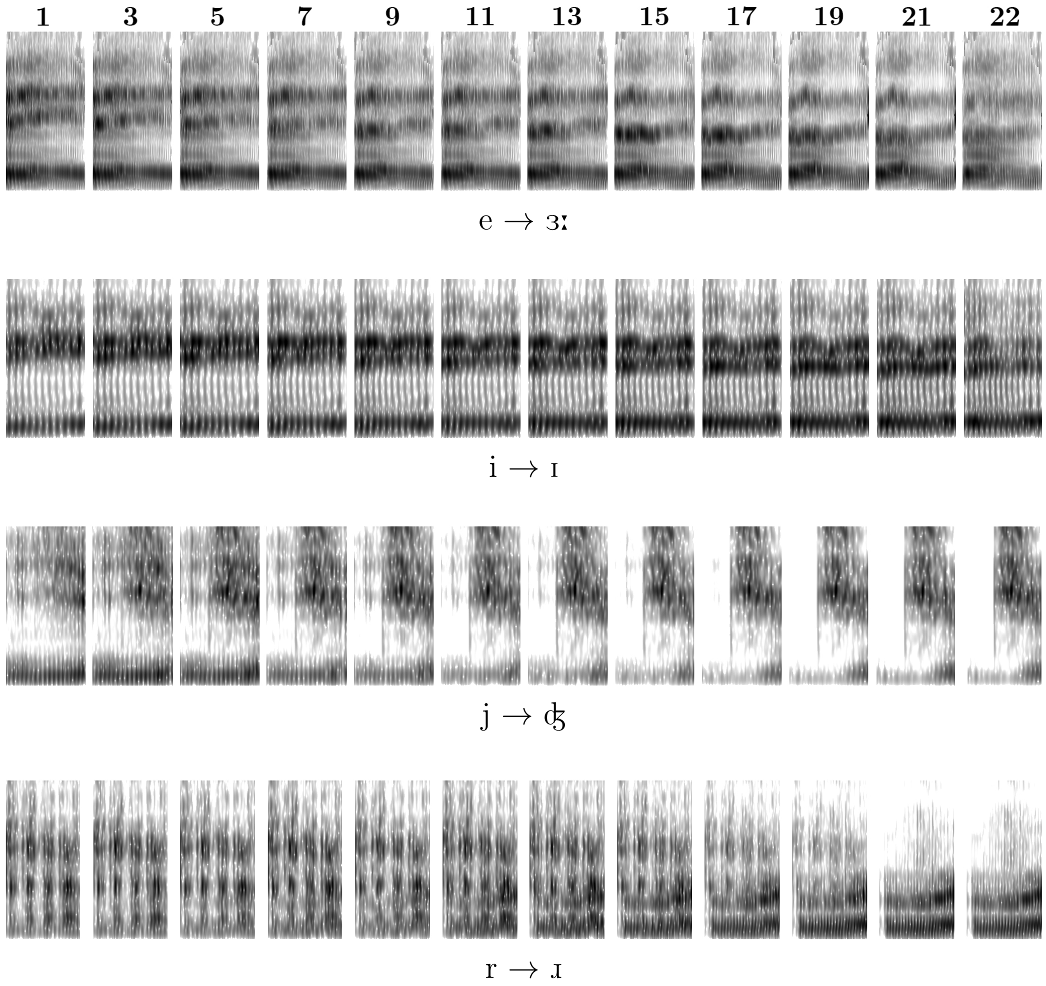


Figure 2: Spectrograms of generated tokens at the odd-numbered steps for two vowel and two consonant continua. Step 1 corresponds to the non-native end of the continuum, step 21 is the processed native token, while step 22 is the original (unprocessed) native token.

3. Experiment: nativeness categorisation of accent continua

3.1. Participants

A group of 22 native English listeners with ages between 18 and 35 (mean 21.3, s.d. 3.6 years) was recruited for this task. Recruitment took place at the Anglia Ruskin University in Cambridge, and all participants had a similar linguistic profile: native speakers of Southern British English, non-bilingual and no knowledge of, or regular contact with, Spanish. No hearing problems were detected among the participants following a pure-tone audiometric screening.

3.2. Materials

The gradation technique was applied to 24 English segments and their corresponding Spanish-accented realisations. For each of these pairs, 4 words were selected, leading to a total of 96 continua (table 1). All continua were generated in monosyllabic words in order to avoid a potential confound of syllable stress. The procedure described in section 2 was applied to each of the 96 continua, with values of λ chosen to produce 21 equal-sized steps. In addition, the unmodified word (i.e. the original native segment) was added to the continuum, in order to gauge any effect of the modification procedure, leading to 22 gradations per continuum.

For some of the selected English segments, the corresponding Spanish-accented segment is also a good exemplar of a different phonological category in English (e.g., in the [b]→[v] continuum, both ends are good exemplars of different phonemes in English). When this is the case, both ends of the continuum represent a real English word. In order to analyse whether the generation of minimal pairs had an effect on the perception of foreign accent, for each continuum we included at least one word that would generate a minimal pair (e.g., *ban*→*van*) and at least one that would not (e.g., *biew*→*view*) for those accented realisations which happened to coincide with one other English phonemic category. For instance, due to the acoustic characteristics of the plosives, as explained in Flege (1987) and Zampini and Green (2001) among others, the Spanish voiceless plosives [p], [t], [k] might be perceived as their voiced counterparts [b], [d], [g] by English listeners. Therefore, Spanish segments [t] and [k] were considered as possible realisations of English [d] and [g] respectively for the purpose of minimal pair selection. Spanish vowels can fall within the scope of several English vowel categories (Cebrián, 2019) and the Spanish pronunciation of English vowels is also strongly infuenced by orthography. Consequently, the presence of possible minimal pairs was not included as a variable for vowel continua or the consonant continua involving [x] or [r] as Spanish realisations. The possibility of minimal pairs exerting an influence was analysed for the remaining 10 consonant continua (section 3.5).

A schematic representation of the 11 vowel continua showing the English vowels and transitions to their Spanish-accented counterparts is shown in

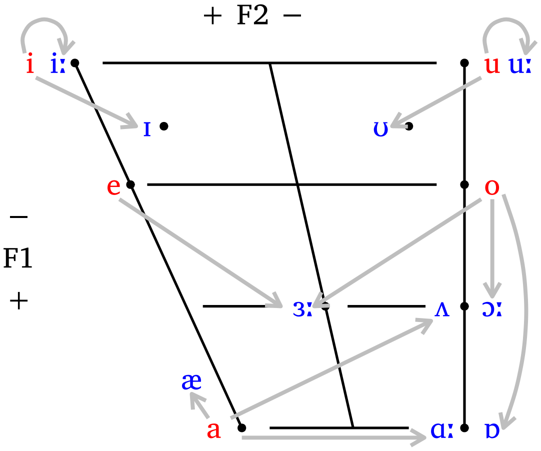
Continuum	Words			
x → h	help	hide	hole	house
r → ɹ	red	rent	rhyme	risk
k → k$^{\text{h}}$	cast	code	cold	kiss
t → t$^{\text{h}}$	town	tile	toad	tone
b → v	van	veil	valve	view
d → ð	than	that	this	thus
s → ʃ	shoe	short	shape	sharp
s → z	zap	zone	zoo	zoom
j → ʤ	jam	jaw	june	jest
ʤ → j	yak	yoke	years	youth
_f → _b	cab	nib	rib	crab
_θ → _d	god	load	dad	food
_x → _g	dog	frog	leg	smog
e → ɜː	bird	burn	firm	learn
o → ɜː	word	world	worm	worse
a → æ	back	cat	clap	pact
a → ɑː	fast	raft	shark	stark
a → ʌ	cut	drum	gun	nut
i → ɪ	clip	mist	pick	sin
i → iː	beam	seem	steam	team
o → ɒ	cost	dot	pot	spot
o → ɔː	clause	fall	orb	storm
u → ʊ	look	nook	put	should
u → uː	choose	mood	moon	spoon

Table 1: Continua endpoints for consonants in word onset and word coda (the latter prefixed with an underline symbol), and for vowels. The direction of the mapping is from the Spanish segment to the English segment. Square brackets have been dropped from segment symbols to avoid clutter.

figure 3. The transformations of [e]→[ɜː] and [o]→[ɜː] are special, as they represent the only case in which one English vowel can be mispronounced as two different vowels, namely [e] or [o], by Spanish speakers of English. The decision on which one to use is strongly linked to the spelling: Spanish speakers will choose [o] in cases in which the sound [ɜː] is orthographically represented by the letter ‘o’ (e.g., *word*, *worm*), and [e] in the rest of the cases (e.g., *bird*, *firm*).

3.3. Categorisation task

Using a two-alternative forced choice procedure, participants were asked to rate every step of each continuum as either foreign or native via an on-



English			Spanish		
Vowel	F1	F2	Vowel	F1	F2
æ	1032	1577	a	790	1597
ɑː	854	1320			
ʌ	882	1486			
ɜː	538	1653	e	577	2329
ɪ	473	2420	i	423	2742
iː	396	2865			
ɒ	508	1082	o	599	1054
ɔː	492	968			
ʊ	497	1078	u	256	726
uː	293	1054			

Figure 3: Left: Schematic representation of vowel continua. The grey arrows mark the transitions between Spanish vowels (in red) and English vowels (in blue) in an F1-F2 acoustic space. Right: Formant values in Hz for the vowels elicited by our speaker for this experiment, measured in the centre of the stable part of each vowel. The given values are an average derived from the four words in each continuum.

screen interface. Since in many consonantal cases both ends of the continuum represented valid English words (e.g., *van-ban* in the [b]→[v] continuum), the target word was presented orthographically to enable listeners to determine whether the auditory stimulus was an accented or non-accented realisation of the target word. Trials were presented in semi-random order with the only constraint being that no two steps of the same continuum appeared in consecutive trials. Participants were informed that each sound would only be played once and were encouraged to judge each stimulus individually and not to aim for a 50% rate for each of the two categories. The total number of trials was 2112 (24 continua × 4 words × 22 steps). Due to the large number of trials, they were presented in four equal-length blocks, each separated by a 1-minute break. On average, this task took nearly 46 minutes to complete.

3.4. Mean Opinion Score task

In a separate task which preceded the main categorisation task, participants were asked to rate the quality of a subset of the stimuli used in the categorisation task. Specifically, only steps 1, 6, 11, 16 and 21 of each continuum were rated, leading to a total of 480 trials (24 continua × 4 words × 5 steps). Listeners used a 5-point scale composed of the labels 1=BAD (very distorted), 2=POOR (fairly distorted), 3=FAIR (somewhat distorted), 4=GOOD (slightly distorted) and 5=EXCELLENT (no distortion). Participants were encouraged to use the full range. As in the categorisation task the target word was presented orthographically. The task took 13.2 minutes on average to complete.

In order for participants to get accustomed to both tasks, before the main part of each task a practice session was presented with five tokens from a [p]→[p^h] continuum. Participants were encouraged to ask any questions at the end of the practice session before commencing the main task.

3.5. Postprocessing

Following Rogers and Davis (2009), all responses whose reaction time was below 300 ms or above 5000 ms were removed ($< 1\%$).

To check whether participants responded differently to continua whose endpoints were minimal pairs in English, a generalized linear mixed effects model was constructed for each task with the package *lme4* (Elzhov et al., 2016), included in R (R Core Team, 2017). Fixed factors were *continuum* (10 levels), *step* (21 levels for the nativeness task, 5 levels for the MOS task) and the presence or not of a *minimal pair* (2 levels). This analysis indicated a significant minimal pair effect in only 5 cases (steps 9 and 13 of [ɟ]→[ʤ], steps 18 and 20 of the [_f]→[_b], and step 6 of the [t]→[t^h]. Since these represent a small fraction of potential cases, the *minimal pair* factor was not treated separately in subsequent analyses. For the MOS task, no significant differences were found for any step of any continuum between those continua that elicited a minimal pair and those that did not.

Differences among the four words of each continuum were also analysed. Significant differences were found in a very small number of cases, all but two involving the words ‘word’, ‘world’, ‘worse’ and ‘worm’ in the [o]→[ɜː] continuum (the other two cases being between *nib* and *rib* in steps 18 and 20 [$p < .01$] and *stark-raft* at step 19 [$p < .05$]). Consequently, *word* is not treated as a separate factor in subsequent analyses. No differences were found for words in the MOS task.

4. Results

4.1. Overall categorisation responses

Across all consonants or vowels, listeners produced a monotonically-increasing pattern of nativeness judgements as tokens became more native-like (fig. 4). For consonants, the relationship between token-wise nativeness and perceived nativeness is very close to linear, while for vowels the pattern shows asymptotic behaviour at the extremes. Mean perceived nativeness across consonants is slightly higher than across vowels [consonants: 68, vowels: 63; $p < .001$], but at the non-native end vowels are perceived as more native-like [consonants: 31, vowels: 39; $p < .001$], while at the native end vowels are perceived as less native-like [consonants: 96, vowels: 82; $p < .001$].

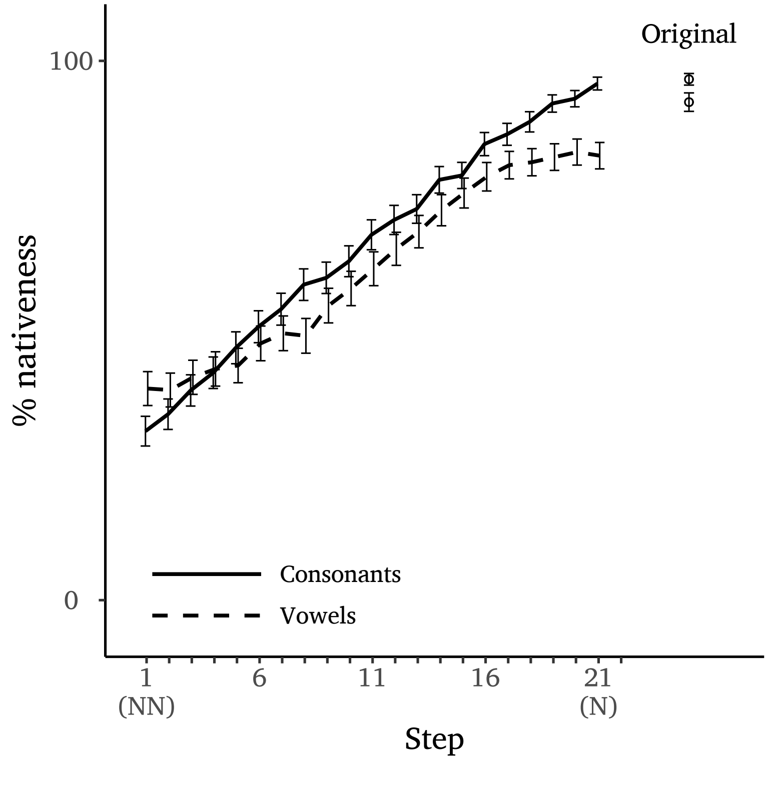


Figure 4: Mean nativeness judgements across all consonants or vowels for each step of the NN-N continuum, together with the original unmodified native stimuli. Error bars here and elsewhere indicate ±1 standard errors.

For consonants, the gradation procedure applied at the native end of the continuum (step 21) did not lead to any change in mean nativeness judgements compared to unmodified stimuli [$p = 0.30$]. However, there was a reduction for the vowels [$p < .001$] which might account for the overall lower nativeness judgements at the native end of the continuum relative to that of consonants.

4.2. Categorisation responses for each continuum

Figure 5 plots listeners' categorisation results for each individual continuum, revealing that the aggregated responses shown in fig. 4 are built up of a diverse range of patterns. Among the consonants, the only continuum exhibiting a non-categorical (quasi-linear) pattern is [_θ]→[_d], although [d]→[ð] shows little change along the length of the continuum. Many continua undergo a rapid change from non-native to native in the early steps of the continuum ([k]→[k^{h}], [s]→[z], [s]→[ʃ], [t]→[t^{h}], [b]→[v]), while others are more balanced ([x]→[h], [ʤ]→[j], [j]→[ʤ]) or tend to nativeness in the later part of the continuum ([r]→[ɹ], [_x]→[_g], [_f]→[_b]). In contrast, vowel continua tend to be either flat or sigmoidal in shape, similar to the aggregated response.

Continua in fig. 5 are ordered by decreasing nativeness rating at step 1 (i.e. the non-native end), revealing that the degree to which a Spanish heavily-accented segment is considered native-like in English varies substantially across continua, and suggesting that accent saliency (as measured by one minus the proportion of nativeness judgements at step 1) is not uniform across different segments. Accent saliency appears to be highest for [j]→[ʤ], [ʤ]→[j] and the six vowel continua involving durational changes between native and non-native tokens.

4.3. A classification scheme for categorical/non-categorical patterns

To facilitate comparison of different patterns of nativeness judgements, we introduce a set of classes (fig. 6) and a decision procedure for choosing the best-fitting class for each continuum. Categorical patterns are characterised by a rapid change in perceived nativeness for a small acoustic change and are distinguished by whether the change occurs near to the non-native segment region of the continuum (N BIASED) or towards the native segment (NN BIASED), or somewhere in the middle (BALANCED). The non-categorical patterns are either PROPORTIONAL or correspond to situations where acoustic manipulation has little or no effect (FLAT N and FLAT NN).

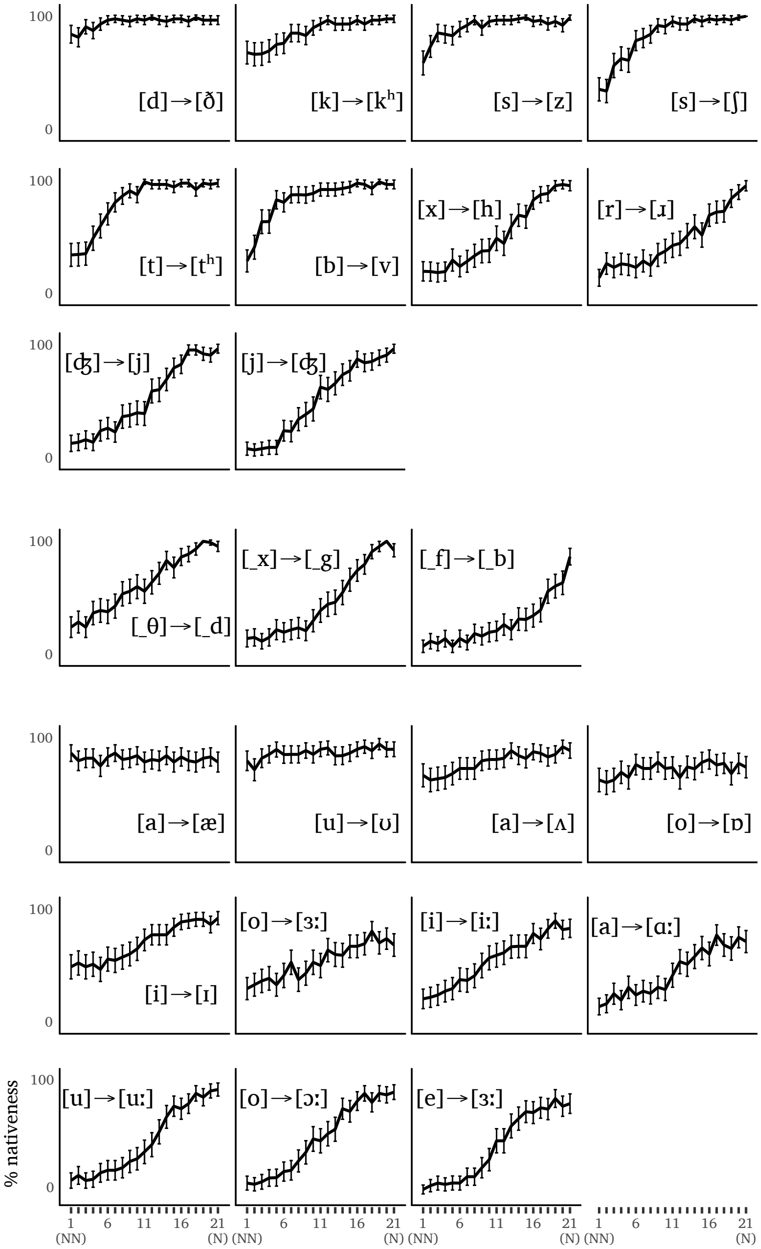


Figure 5: Individual nativeness judgements for word-initial consonants (rows 1-3), word-final consonants (row 4) and vowels (rows 5-7). Responses are ordered by decreasing nativeness in step 1.

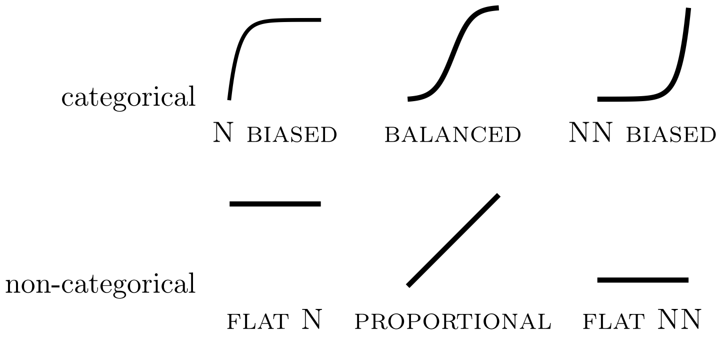


Figure 6: Schematic patterns relating nativeness judgements to acoustic continua.

To determine which pattern best describes mean listener responses for each continuum, a two-step curve-fitting procedure was employed using the 4-parameter sigmoid (equation 1):

$$f(x) = d + \frac{a - d}{1 + \left[{}^{x}\!/_{c}\right]^{b}} \tag{1}$$

where a and d denote the lower and upper asymptotes, c is the point of inflection, and b is related to the gradient of the curve at c.

The first step uses an *unconstrained* fit and is used solely to identify whether listener responses are best characterised as FLAT N or FLAT NN: if both a and d are in the upper tercile, the pattern is classified as FLAT N; similarly, if a and d are in the lower tercile, the pattern is deemed FLAT NN. If neither condition holds, in the second stage a *constrained* fit is carried out to select between the remaining patterns. If the point of inflection is within the NN tercile (steps 1-7) the pattern is classified as N BIASED; similarly, if it is within the N tercile (steps 15-21) it is defined as NN BIASED. The two remaining cases are distinguished in terms of the slope parameter, b. Moderate slopes lead to a classification as PROPORTIONAL while steeper slopes correspond to a BALANCED classification. Table 2 summarises these constraints. Sigmoid parameters were computed using the `nlsLM` command included in the package `minpack.lm` (Elzhov et al., 2016) for R. The best fit was selected using Akaike's Information Criterion (AIC, Sakamoto et al., 1986).

Figure 7 shows the best-fitting patterns for each of the 24 continua. Of

Fit	a	b	c	d	Outcome
unconstrained	upper tercile	-	-	upper tercile	FLAT N
	lower tercile	-	-	lower tercile	FLAT NN
constrained	-	-	NN tercile	-	N BIASED
	-	> 3	middle tercile	-	BALANCED
	-	1-3	middle tercile	-	PROPORTIONAL
	-	-	N tercile	-	NN BIASED

Table 2: Decision procedure for nativeness judgements. The upper and lower terciles correspond to divisions of the nativeness axis, while the N and NN terciles correspond to steps 15-21 and 1-7 of the continuum respectively. In the constrained fit, parameters a and d were limited to the intervals $[-\infty, 2/3]$ and $[1/3, \infty]$ respectively.

the 13 consonant continua, only one is best-classified as PROPORTIONAL, in contrast to the clear linear proportional trend seen in fig. 4 when all consonants are considered together. The most-frequent classification is N BIASED, indicating that for many continua, a small change in the direction of the native category is sufficient for listeners to classify the sound as native-like. In contrast, four vowel continua are classified as PROPORTIONAL, and four other vowel continua are BALANCED, confirming the visual impressions of relatively gradual changes in the proportion of tokens judged as native in fig. 5 as stimuli become acoustically closer to the native token.

4.4. Mean opinion scores

Listeners rated the mean quality of tokens on the vowel continua at 4.16, somewhat higher than the mean rating of 3.86 for consonantal continua [$p < .001$]. Both values are close to a qualitative GOOD or ‘slightly distorted’ rating on the MOS quality scale. Quality scores increased monotonically from 3.55 to 4.26 [$p < .001$] for consonants and 3.93 to 4.31 [$p < .001$] from the non-native end to the native end of the aggregated continuum (fig. 8). Similar increases were observed for almost all individual continua (fig. 9). MOS scores were positively correlated with higher nativeness categorisation responses [$r = 0.95$, $p < .05$].

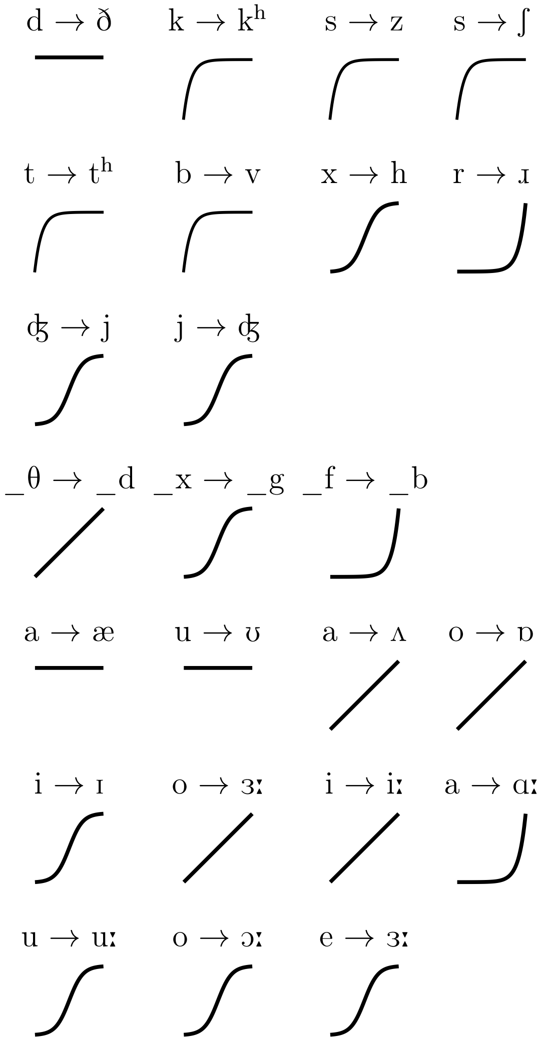


Figure 7: Best-fitting schematic patterns for the continua of fig. 5.

5. Discussion

The main aim of the current study was to investigate whether listeners judge segments categorically as either foreign- or native-accented or whether they perceive degrees of foreign accent in a graded manner, proportional to acoustically-implanted changes in accent. For this purpose, foreign accent continua spanning from a Spanish to a native English realisation were generated for 24 English segments. Unlike previous studies, degree of foreign

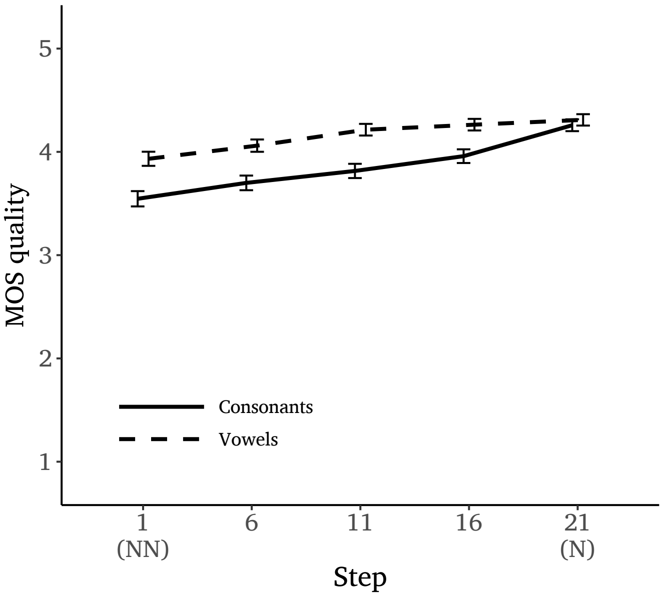


Figure 8: Across-continua MOS values measured for consonants and vowels.

accent was not obtained by using stimuli from speakers with different proficiency levels. Rather, grades of accentedness were created through acoustic manipulations of the speech of a bilingual speaker, who produced native English and Spanish segments. This technique ensured that the steps of the continua were consistently equally-sized, while controlling for inter-speaker differences.

Aggregated responses for both consonants and vowels showed a quasi-linear increase in perceived accentedness with increasing degree of acoustic distance from the native category, in agreement with holistic FA studies. However, responses to individual segments were largely not linear: 16 out of 24 patterns for individual segments were classified as categorical. Since continua were generated using equal-sized acoustic steps, we conclude that (i) for many segments, small acoustic accent-related changes cannot be detected reliably and judged in terms of accent strength; and conversely that (ii) small acoustic changes can result in a rapid change in nativeness judgement. These outcomes suggest that degrees of foreign accent in longer utterances may arise due to the superposition of multiple segment accentness patterns which themselves are individually quite distinct and categorical.

A secondary aim of this study was to determine whether graded FA for

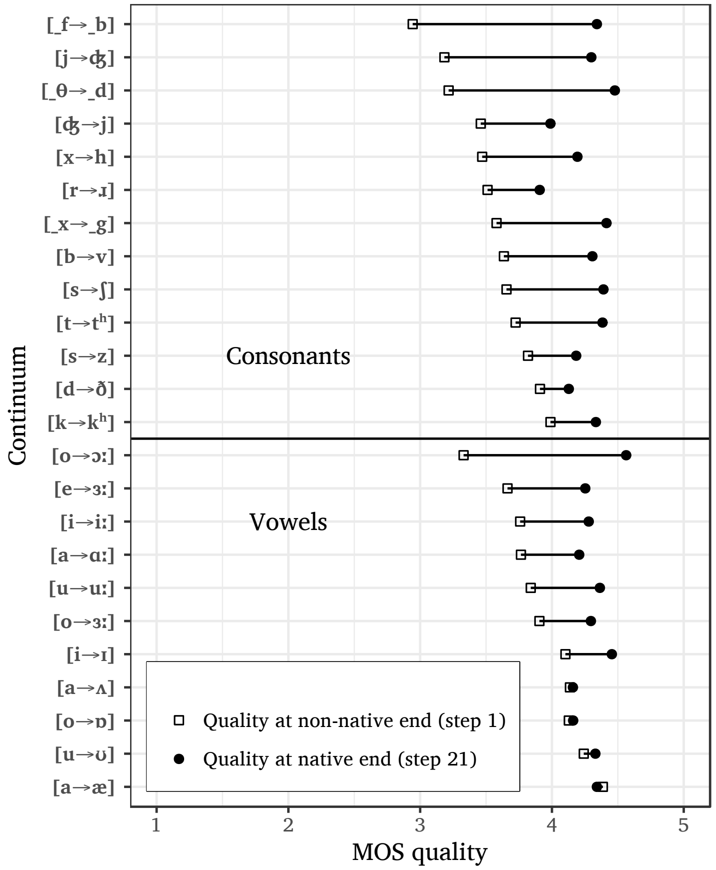


Figure 9: MOS values for the endpoints of each continuum. Consonant and vowel continua are separately ordered from low-to-high MOS in step 1.

vowels is perceived less categorically than for consonants, in line with native language categorical perception studies (see, e.g., Cutler, 2012). Our results support this prediction: while consonants were perceived mostly categorically, vowels showed a more mixed picture, with 6 non-categorical patterns amongst the 11 continua. Additionally, the non-native end is perceived as less-accented for many vowel continua compared to consonantal continua. This outcome supports the idea that listeners have less clear boundaries for their vowel categories – the "category insecurity or flexibility" notion (Cutler, 2012). In the case of English, this flexibility is augmented by the large degree of regional variation amongst vowels compared to consonants, which might explain why there is more tolerance to non-prototypical tokens. Additionally, Spanish vowels may fit as exemplars (even if non-ideal) within several English categories, increasing the uncertainty for English listeners when faced with a Spanish-like vowel realisation of an English target.

Continuum	Non-native	Native	Difference
e → ɜː	77	343	266
o → ɜː	106	247	140
a → æ	92	146	54
a → ɑː	98	277	180
a → ʌ	94	137	43
i → ɪ	73	107	34
i → iː	86	228	142
o → ɒ	87	142	55
o → ɔː	97	320	224
u → ʊ	81	136	55
u → uː	88	291	202

Table 3: Durations of native and non-native vowel segments for each vowel continuum, alongside their difference (all in ms).

Interestingly, vowel continua which include durational changes show a more categorical pattern, similar to that of most consonants. Previous studies (e.g., Minagawa-Kawai et al., 2002; Mugitani et al., 2009) suggest that for Japanese, which has a vowel system in which duration is distinctive, listeners perceive vowel duration continua in a categorical manner. In our data, changes in vowel duration (table 3) between continua endpoints for long English vowels are at least twice as large as those for shorter English vowels, and comparable to the difference between long and short vowels in systems where duration is contrastive. Further tests in which vowel duration is isolated from spectral changes are necessary to ascertain whether listeners were showing categorical perception for these continua because of the durational contrast or because duration differences were compounded with spectral changes.

Finally, with any stimulus generation technique the possible presence of processing artefacts needs to be considered. Mean opinion scores show a modest increase in signal quality across the continua as the token becomes closer to the native exemplar (fig. 8). The fact that listeners undertook the MOS task prior to the main categorisation experiment may have led participants to interpret increasing distance from the native exemplar as a modest reduction in quality. We do not favour an alternative explanation in terms of artefacts in the stimulus generation process since the strongest artefacts would be predicted in the middle of the continuum where the stimuli are furthest from the native or non-native exemplars. However, there is evidence of processing-related artefacts at the native end of the vowel – but not consonant – continua (fig. 4). Since the vowel generation procedure differed from that for consonants in step 3 (fig. 1), it appears that minor artefacts may have been introduced during source-filter decomposition or resynthesis.

6. Conclusions

When presented with English words in which a single vowel or consonant segment had been replaced with a Spanish-accented counterpart, listeners' overall nativeness classification showed a near-linear (non-categorical) dependence on acoustic proximity to the native segment. However, the relationship between degree of acoustic manipulation and perceived nativeness for individual segments was best described as categorical for most consonants and for vowels involving durational differences between native and non-native exemplars, with relatively small acoustic changes leading to a switch in nativeness judgement. Vowel continua lacking a durational cue to nativeness were generally perceived as native-like throughout the continuum. These outcomes indicate that while accentedness judgements appear graded when all segments are taken together, accent salience is highly-dependent on the individual segment.

One of the main foci of effort for foreign language speakers is the production of sounds that differ from those of their L1. While an interlocutor may well form an overall impression of foreign accent at the level of phrases and sentences, a foreign language learner is in the main forced to deploy their productive skills at the level of segments. The current study highlights the importance of understanding both accent salience and productive allowances at the level of individual segments as a step towards the goal of reducing foreign accent.

Acknowledgements: This work was funded by the Spanish MINECO project DIACEX (FFI2012-31597). The authors thanks Dr. Bettina Beinhoff for help in hosting the listening experiment.

Bibliography

Anderson-Hsieh, J., Johnson, R., Koehler, K., 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segments, prosody, and syllable structure. Language Learning 42, 529–555.

Aoyama, K., Guion, S., Flege, J., Yamada, T., Akahane-Yamada, R., 2008. The first years in an L2-speaking environment: A comparison of Japanese children and adults learning American English. Int. Rev. Applied Linguistics 46, 61–90.

Boersma, P., Weenink, D., 2018. Praat: doing phonetics by computer [computer program]. version 6.0.43, retrieved 8 September 2018 from http://www.praat.org/.

Bradlow, A. R., 1995. A comparative acoustic study of English and Spanish vowels. The Journal of the Acoustical Society of America 97 (3), 1916–1924.

Brennan, E., Ryan, E., Dawson, W., 1975. Scaling of apparent accentedness by magnitude estimation and sensory modality matching. Journal of Psycholinguistic Research 4, 27–36.

Burda, A., Scherz, J., Edwards, C. H. H., 2003. Age and understanding speakers with Spanish or Taiwanese accents. Perceptual and Motor Skills 97, 11–20.

Burg, J. P., 1975. Maximum entropy spectral analysis. Ph.D. thesis, Stanford University.

Cebrián, J., 2019. Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. The Journal of the Acoustical Society of America 145, EL52–EL58.

Cutler, A., 2012. Native Listening. MIT Press.

Derwing, T. M., Munro, M. J., 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s. Studies in Second Language Acquisition 19, 1–16.

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., Bolker, B., 2016. minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. R package version 1.2-1. URL https://CRAN.R-project.org/package=minpack.lm

Flege, J. E., 1984. The detection of French accent by American listeners. The Journal of the Acoustical Society of America 76, 692–707.

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.

Formby, C., Childers, D. G., Lalwani, A. L., 1996. Labelling and discrimination of a synthetic fricative continuum in noise: A study of absolute duration and relative onset time cues. Journal of Speech, Language, and Hearing Research 39 (1), 4–18.

Fry, D. B., Abramson, A. S., Eimas, P. D., Liberman, A. M., 1962. The identification and discrimination of synthetic vowels. Language and Speech 5 (4), 171–189.

García Lecumberri, M. L., Barra-Chicote, R., Pérez-Ramón, R., Yamagishi, J., Cooke, M., 2014. Generating segmental foreign accent. In: 15th Annual Conference of the International Speech Communication Association (Interspeech 2014). pp. 1302–1306.

Hahn, L. D., Jul. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. TESOL Quarterly 38 (2), 201.

Hualde, J., 2005. The Sounds of Spanish. Cambridge University Press.

Ikeno, A., Hansen, J. H. L., 2006. Perceptual recognition cues in native English accent variation: Listener accent, perceived accent and comprehension. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing. pp. 401–404.

Iverson, P., Evans, B. G., 2009. Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. The Journal of the Acoustical Society of America 162, 866–877.

Larkey, L. S., Wald, J., Strange, W., 1978. Perception of synthetic nasal consonants in initial and final syllable position. Perception & Psychophysics 23 (4), 299–312.

Levi, S. V., Winters, S., Pisoni, D., 2007. Speaker-independent factors affecting the perception of foreign accent in a second language. The Journal of the Acoustical Society of America 121, 2327–2338.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., Bastian, J., 1961. An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. Language and Speech 54, 175–195.

Liberman, A. M., Harris, K. S., Hoffman, H. S., Griffith, B. C., 1957. The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54 (5), 358–368.

Minagawa-Kawai, Y., Mori, K., Furuya, I., Hayashi, R., Sato, Y., 2002. Assessing cerebral representations of short and long vowel categories by nirs. Neuroreport 13 (5), 581–584.

Mugitani, R., Pons, F., Fais, L., Dietrich, C., Werker, J. F., Amano, S., 2009. Perception of vowel length by Japanese-and English-learning infants. Developmental Psychology 45 (1), 236.

Munro, M. J., 1993. Productions of English vowels by native speakers of Arabic: acoustic measurements and accentedness ratings. Language and Speech 36, 39–66.

Munro, M. J., Derwing, T. M., 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech. Studies in Second Language Acquisition 23 (4), 451–468.

Munro, M. J., Derwing, T. M., Burgess, C. S., 2010. Detection of nonnative speaker status from content-masked speech. Speech Communication 52 (7-8), 626–637.

Munro, M. J., Derwing, T. M., Morton, S. L., 2006. The mutual intelligibility of L2 speech. Studies in Second Language Acquisition 28, 111–131.

Oyama, S., 1976. A sensitive period for the acquisition of a nonnative phonological system. Journal of Psycholinguistic Research 5 (3), 261–283.

Park, H., 2013. Detecting foreign accent in monosyllables: The role of L1 phonotactics. Journal of Phonetics 41, 78–87.

Piske, T., MacKay, I., Flege, J., 2001. Factors affecting degree of foreign accent in an L2: A review. Journal of Phonetics 29, 191–215.

Pisoni, D. B., 1973. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics 13 (2), 253–260.

Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., 1992. Numerical Recipes in Fortran 77: The Art of Scientific Computing. Cambridge University Press.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL `https://www.R-project.org/`

Rogers, J., Davis, M., Jan. 2009. Categorical perception of speech without stimulus repetition. Proceedings of Interspeech, Brighton, 376–379.

Sakamoto, Y., Ishiguro, M., Kitagawa, G., 1986. Akaike Information Criterion Statistics. Springer Netherlands.

Southwood, M. H., Flege, J. E., 1999. Scaling foreign accent: direct magnitude estimation versus interval scaling. Clinical Linguistics & Phonetics 13 (5), 335–349.

Tajima, K., Port, R., Dalby, J., 1997. Effects of temporal correction on intelligibility of foreign-accented English. Journal of Phonetics 25 (1), 1–24.

Wells, J. C., 1982. Accents of English. Cambridge: Cambridge University Press.

Zampini, M., Green, K., 2001. The voicing contrast in English and Spanish: The relationship between perception and production. In: Nicol, J. (Ed.), One Mind, Two Languages: Bilingual Language Processing. Malden, MA: Blackwell, Ch. 2, pp. 23–48.