(A)symmetries in the perception of non-native consonantal and vocalic length contrasts

Abstract

How well can non-native length contrasts be perceived, both in vowels and in consonants and is one more difficult than the other? Three listener groups (native Italian and German as well as advanced German learners of Italian) performed a speeded same-different task involving vocalic and consonantal length contrasts as well as segmental contrasts as controls. Phonologically, Italian, but not German, has a consonantal length contrast, while German, but not Italian, has a vocalic length contrast. Analysis of responses yielded a clear asymmetry: non-native vowel length contrasts were perceived as well as segmental contrasts and the native consonantal length contrasts. The perception of a non-native consonantal length contrast, however, was poor compared to the native group: Italians showed higher sensitivity for consonants than German learners of Italian, which in turn were better than German non-learners. Reaction time analyses indicated that even though Italians displayed a similar sensitivity to Germans with respect to vocalic length contrasts, their reaction times differed, suggesting different processes at play compared to the processing of a native length contrast.
Keywords: vocalic length, consonantal length discrimination, Italian, German, advanced learners, second language phonology, L2 suprasegmentals
1. Introduction

In the course of adult second language (L2) acquisition, the L2 sound system remains to be a big challenge for the vast majority of learners even after other grammatical (e.g., morpho-syntactic) facets of the target language might have been mastered. This usually results in a distinct and discernable non-native accent (cf. Ioup, 1984; Munro and Derwing, 1995), which has been shown to slow down speech comprehension in native listeners (Braun et al., 2011a, 2011b) and may even lead to lexical confusion. Dutch listeners of English, for instance, find it very hard to distinguish between English *cattle* and *kettle*, as the Dutch and German phoneme inventory only contain one half-open front vowel, [ɛ], instead of two (Escudero et al., 2008; Weber and Cutler, 2004).

In order to address this persistent feature of adult L2 learners, the focus of research in the area of second language acquisition over the past twenty years has shifted to investigations of mostly segmental interlanguage phonology, yielding a large body of psycholinguistic studies and the development of a number of phonology-specific theories (e.g., Perceptual Assimilation Model, cf. Best, 1995, Speech Learning Model, cf. Flege, 1995, Native Language Magnet, cf. Kuhl, 1993). Furthermore, over the last decade, there has been increasing experimental interest in the perception and production of suprasegmentals such as stress (e.g., Altmann, 2006; Dupoux et al., 2008; Kijak, 2009; Tremblay, 2008), tone (e.g., Chiao et al., 2011,
Wang et al., 2003; Gandour 1983), or segmental length (e.g., Hayes-Harb, 2005; Hisagi et al., 2010; McAllister et al., 2002).

Length is a suprasegmental feature that occurs with consonants as well as with vowels in the world’s languages (Ladefoged & Maddieson 1996). In that case, segmental length is unpredictable and signals lexical contrasts (e.g., the distinction between Staat, /ʃtat/ ‘state’ and Stadt /ʃtat/ ‘city’ in German). Contrastive length in consonants is found, among others, in Italian, Japanese, Arabic, Finnish, or Swiss German, contrastive length in vowels for example in Japanese, Swedish, German, or Finnish. Segments may also appear predictably lengthened or shortened based on their position in the word (e.g. lengthening of vowels in stressed syllables or shortening of unstressed nuclei, cf. Baumann et al., 2005; Braun et al., 2011b; polysyllabic shortening, cf. Lehiste, 1972; Port, 1981; White and Turk, 2010; phrase-initial strengthening, cf. Cho and McQueen, 2005; Fougeron and Keating, 1997), but

Non-target-like productions of L2 length for different class of segments are well documented in the literature, even by advanced L2 learners (Han, 1992; Kabak et al., 2011; McAllister et al., 2002; Mah and Archibald, 2003). As for consonantal length contrasts, Kabak et al. (2011) compared the duration of Italian singleton and geminate consonants produced by Italian natives, German learners of Italian and German non-learners. Results showed that the geminate-singleton ratio was more native-like in learners (2.0) compared to non-learners (1.8); yet it was still smaller
than that of Italian natives (2.4). Interestingly, group differences were found in the production of both singletons and geminates. McAllister et al. (2002) compared the production of Swedish long and short vowels by Swedish natives, as well as by highly advanced Estonian, English, and Spanish learners of Swedish. Analyses of V(:)C ratios showed two groups: Swedish natives and Estonian learners of Swedish (whose L1 has contrastive vowel length) produced larger differences between long and short vowels than English and Spanish learners of Swedish (whose L1 does not have contrastive vowel length). One explanation for such production problems may lie in non-target-like perception of length information in the L2 input.

In the literature, there are a number of studies on the perception of vowel length distinctions for L2 learners whose L1 lacks such a contrast, involving both behavioral as well as electrophysiological data (e.g., Hisagi et al., 2010; McAllister et al., 2002; Nenonen et al., 2003, 2005; Ylinen et al., 2005). On the other hand, substantially less work has been done on the L2 perception of geminate vs. singleton consonants (notable exceptions are Hardison and Saigo, 2010; Hayes-Harb and Masuda, 2008; Heeren and Schouten, 2008). Before we review some recent studies on the L2 perception of length by learners with different L1s (section 2.1), we will first provide information on the psychoacoustic perception of duration, the development of phonemic contrasts and the (prelexical) processing of durational information. Phonetically, length features as segment duration. Psychoacoustic studies have shown
that participants are able to reliably detect durational differences between pure tone
stimuli that are as short as 8.8 ms (Rammsayer and Lima, 1991). This indicates a high
perceptual sensitivity for durational contrasts in non-linguistic auditory input. Such
unbiased high sensitivity towards stimulus duration is not upheld for speech stimuli;
in that case perception of duration is influenced by the native phonological, an effect
that emerges early in development. In the course of first language acquisition, only
those phonetic differences are encoded that signal meaningful contrasts in one’s own
language. For instance, infants up to 6-8 months of age are still able to discern
acoustic detail in duration (e.g., differences in voice onset time), regardless of their
L1. A few months later, however, after 10-12 months of age only contrasts that are
active in the L1 system can be distinguished reliably (Werker and Tees, 1983, 2002).

In languages that employ length contrasts (i.e., lexical distinctions between
minimal pairs containing short and long segments), prelexical and lexical
representations need to be tuned to durational contrasts so that they can activate the
appropriate lexical candidates (McQueen, 2005). On the contrary, when length
distinctions in the L1 are absent, perceptual sensitivity to duration is reduced
(Gottfried and Beddor, 1988). It should be noted, however, that perceptual sensitivity
to duration is needed in all languages as segment durations do not only signal lexical
contrasts but may also convey information about the prosodic structure and in the end,
only the combination of lexical and prosodic information allows us to decode the
utterance interpretation. For instance, phonemes are lengthened in phrase-initial and phrase-final position (Cho and McQueen, 2005; Fougeron and Keating, 1997; Wightman et al., 1992). A further reliable effect is polysyllabic shortening, i.e., the longer a word, the shorter its syllables (e.g., Braun and Geiselmann, 2011; Lehiste, 1992; White and Turk, 2010). Duration is further influenced by word stress, sentence accent, neighbouring segment type, etc (e.g., Baumann et al., 2005; Braun et al., 2011b; Cho and Keating, 2009). It is well documented that such fine phonetic durational detail is used in online speech comprehension (e.g., Salverda et al., 2003; Tagliapietra and McQueen, 2010), even in cases where length does not have phonemic status in a language.

2. Background

The current paper reports a test of the perception of segmental length contrasts which are phonemic in the non-native language (L2) but not in the native language (L1) and compares it to the perception of native length contrasts. In order to provide a holistic picture, the perception of both vowel and consonant length is investigated, which allows for a cross-linguistic comparison of the ability to perceive non-native length contrasts for different types of segments. Our focus lies on the perception of consonantal length. In addition, possible effects of experience with the L2 (as reported in Heeren and Schouten, 2008, 2010) can be traced by comparing naïve L1 speakers
with no experience with a given L2 (as a baseline) to advanced learners of that L2.

We will first give an overview of consonantal and vocalic length in L2 acquisition (section 2.1), followed by a review of studies suggesting improvement in L2 speech perception abilities over time (section 2.2). Finally, in section 2.3 we will outline the phonological properties of the two languages studied here, Italian and German.

### 2.1 Perception of L2 consonantal and vocalic length

The cross-linguistic perception of vocalic length supposedly poses little problems irrespective of the employment of durational differences for vowels in one's native language (e.g., Bohn, 1995; Cebrian, 2006; Flege et al., 1997; García Lecumberri and Cenoz, 1997). Bohn (1995) argues for a “general speech perception strategy” (p. 300), which claims that listeners utilize durational cues whenever spectral segmental information is not conclusive enough (cf. Desensitization Hypothesis, Bohn, 1995). In addition, it has been found that L2 learners whose L1 does not contrast vowel length may still use vowel duration to distinguish vowel pairs in the target language, even if native listeners of the target language make more use of spectral differences (e.g., Bohn, 1995; Kondaurova and Francis, 2008; Wang and Munro, 1999). That is, vowel duration appears to be more accessible than spectral information even for listeners whose native language does not have vowel length contrasts.
Nevertheless, there exist both behavioral and electrophysiological findings conflicting with this general positive view of durational cues for vowel perception and which indicate significantly better results for speakers whose L1 contains both long and short vowels. One such example is the cross-linguistics study conducted by McAllister et al. (2002) reported above. In addition, a mismatch negativity study (MMN) compared the perception of length contrasts for early Russian L2 learners of Finnish compared to Finnish native speakers. They were tested on vowels that could be mapped onto Russian phonemes and those that were too different to be mapped (Nenonen et al., 2003). The Russian learners, whose L1 does not contrast vowel length, revealed a lower mismatch negativity (decreased detection of the change) than Finnish natives but only in those sounds that could be mapped onto Russian vowels. For the vowels that could not be categorized, sensitivity to length contrasts was not different from Finnish natives. These findings suggest that the perception of vowel length is influenced by the L1 phonological system as well.

Unfortunately, there are only a few studies on the non-native perception of consonantal length, most of which focus on English learners of Japanese. Hayes (2002), for instance tested the perception of the Japanese singletons /k/, /t/ and /s/ and their corresponding geminates by English learners of Japanese as well as Japanese natives in a same-different task. Participants had to determine whether two stimuli that only differed in consonant length were the same or different. Not surprisingly, her
results showed that accuracy improved with proficiency. In 2005, Hayes-Harb reported an identification task (long vs. short consonant) in which the same consonants were manipulated into 13 durational steps. Native speakers showed a clear perceptual boundary while English participants with no exposure to Japanese (and thus no experience with consonantal length contrasts) yielded a near linear response curve. Results from English learners of Japanese were found to fall in-between.

In another study on English (L1) and Japanese (L2) perception, Hardison and Saigo (2010) not only confirmed advanced L2 learners’ target-like perception abilities for short vs. long consonants, but also yielded different rates of success for the identification of different types of consonants: learners obtained significantly higher correctness scores for oral stops than for the fricative /s/. Hardison and Saigo (2010) further concluded that a larger sonority difference between a (singleton or geminate) consonant and the following vowel facilitates consonantal length perception. This suggests that – in addition to mere durational differences between singletons and geminates – there may be other factors that influence the perception of consonantal length (see also Payne, 2006, who describes data that show more articulatory effort for geminates than for singleton consonants in Italian).

In sum, there is conflicting experimental evidence as to whether or how the native phonological system affects the L2 perception of vocalic length. Regarding the perception of L2 consonantal length, there is no diversity of cross-linguistic studies or
of a larger variety of segments available. The few studies that exist overall suggest that it is difficult to perceive an L2 singleton-geminate contrast if it is absent in the L1.

### 2.2 Improvement with increased exposure

It must be noted that in most of the studies mentioned above, even when advanced learners did not reach target-like scores, learners at a more advanced stage still outperformed non-L2-learners and/or lower-level L2 learners from the same L1 background (e.g., Hayes, 2002; Hayes-Harb, 2005; Hardison & Saigo, 2010). From an acquisitional point of view this is very positive news as it indicates that the perception of an originally unfamiliar non-native length contrast is indeed learnable (cf. also the supporting results of training studies for Dutch learners of Finnish, e.g., Heeren and Schouten, 2008, for adults and Heeren and Schouten, 2010, for children). Furthermore, increased exposure to the L2, especially also to increased variation in the input, improves listeners' sensitivity (e.g., Sadakata and McQueen, 2011). Morrison (2002) also showed that learners may readjust their perceptual strategies with increased exposure to the target language: While Spanish beginning learners of English did not show a preferential use of mainly durational or of mainly spectral cues to distinguish between /i/ vs /u/, the same learners shifted towards relying on either spectral or durational cues for the English /i/ vs /u/ contrast after one month’s stay in
Canada. This suggests that even in cases where learners’ performance (as measured by discrimination or categorization results) approaches that of native speakers of the target language, the perceptual strategies of the two groups may differ.\(^1\) Such a readjustment of interlanguage phonological systems can not only be observed through improved perceptual performance on L2 contrasts, but also finds expression in longer reaction times compared to target language speakers (e.g., Minagawa-Kawai et al., 2005, for categorical perception of short vs. long vowels, as well as Dijkstra, 2005, for lexical decision times). This indicates that the processing cost for L2 learners is higher than for native speakers even when their behavioral data shows target-like performance.

2.3 The phonology and phonetics of Italian and German length contrasts

Standard German and Italian are both sensitive to quantity, although for different types of segments: German, but not Italian (Krämer 2009)\(^2\), employs a phonemic vocalic length contrast, e.g. [ban] 'ban' vs. [baːn] 'train', and Italian, but not German, distinguishes between geminate and singleton consonants, e.g. [fato] 'fate' vs. [fat:o] 'fact'. Thus, German and Italian are ideal languages for testing the perceptual

---

1. For insightful discussion and an elaborate OT-based model of similar differential auditory cue weighting for /i/ vs /ɪ/ in the interlanguage development of Spanish learners of two varieties of English, see Escudero and Boersma (2004).

2. It should be noted that ‘Italian’ for our purposes is intended to mean Standard Italian. We are aware that certain Italian dialects, e.g., Friulian (Baroni and Vanelli, 2000) or Milanese (Prieto, 2000), may contrast short and long vowels.
sensitivity to length contrasts in a different class of segments.

The German vowel inventory distinguishes not only between tense and lax vowels, but also between long and short ones (Wiese, 2000). Tense vowels, which are produced with more peripheral articulatory settings, on the one hand, appear long in stressed position and short in unstressed position. Lax vowels, the articulatorily more central counterparts, on the other hand, are generally short. There exist, however, two pairs of lax vowels for which only length is contrastive: /e/ vs. /e:/ and /a/ vs. /a:/ (Hall, 2000; Wiese, 2000). Since for these pairs length is the only distinguishing feature and independent of stress placement or vowel quality, segmental length is clearly important for German native speakers in their vocalic system. Acoustic measurements averaged over three speakers of Standard Northern German (Jørgensen, 1969) indicate that spectral information for long vs. short /a/ is identical, while the short front-mid vowel /e/ is somewhat more centralized compared to its long counterpart. A study on the perception of vowel length and quantity in German by Weiss (1976) showed that for [a] vs. [a:], durations less than 200 ms were classified as short and durations more than 200 ms as long; his data for [ɛ] and [ɛ:] suggest a boundary between 120 ms and 190 ms for this pair. In a similar

---

3 There are a number of mostly articulatory studies that argue that the teense-lax distinction in German reflects a difference in the "Silbenschnitt" and is hence structural rather than purely durational (Hoole and Mooshammer, 2002; Kroos et al., 1997; Vennemann, 1991). Irrespective of the underlying theory, the acoustic difference that the listener has to process for /a:/ vs. /a/ in German manifests itself only in durational terms (Jørgensen, 1969, Pätzold and Simpson, 1997).
investigation, Sendelmeier (1981) found that German native listeners purely rely on durational cues for /a/ vs. /a:/, but use a combination of duration and quality for other vowel distinctions.⁴

The Italian consonantal system contains numerous distinct phonemes that can occur both as singletons and geminates, namely /p/, /b/, /t/, /d/, /k/, /g/, /f/, /v/, /s/, /ʃ/, /ʒ/ , /m/, /n/, /l/ und /r/ (Muljačić, 1972). They appear contrastively almost exclusively in intervocalic position, but some geminates may also occur before liquids or glides (e.g. [pːr], [pːl], [pːj] or [bːw]). As to the remaining consonants in the inventory, /j, w, z, ʒ/ are always short and /ʃ, ñ, ñ, ts, ðʃ/ are ‘intrinsically long’ (Passino, 2008).

Acoustic studies investigating the phonetic properties of the singleton/geminate contrast for different consonant types in Italian (stops: Esposito and Di Benedetto, 1999; fricatives: Giovanardi and Di Benedetto, 1998; nasals: Mattei and Di Benedetto, 2000; affricates: Faluschi and Di Benedetto, 2001) revealed that duration is the main cue underlying the consonantal contrast. However, the duration of the vowel preceding the geminate or singleton consonant covaries to some degree: it is longer before a singleton and shorter before a geminate consonant (Esposito and Di Benedetto, 1999). Hence, vowel duration is allophonic and might be used as an acoustic cue to the perception of the singleton-geminate contrast. According to

⁴ See Lehnert-Le Houillier (2007) for a detailed overview of vocalic length studies for German.
Kingston et al. (2009), the allophonic variation in preceding vowel duration is comparatively small. Others have argued that the geminate-singleton discrimination in Italian depends less on the actual length of the consonant, but rather on the ratio between the consonant and the preceding vowel. Simply put, if the consonant is longer than the preceding vowel, it is perceived as a geminate, if it is shorter as singleton (Giovanardi and Di Benedetto, 1998), and any following segment has no influence. Thus, the ratio between the duration of a vowel and its postvocalic consonant is also highly relevant.

In sum, since Italian and German exploit length contrastively in one class of segments but not in the other, this pair of languages is ideally suited to investigate the native and non-native length perception in vowels and consonants. The overview of previous studies above has shown mixed results for the perception of vocalic length, while the empirical data on the perception of consonantal length is fairly restricted to one language pair (English learners of Japanese). Our experiment will therefore provide cross-linguistic data, supplemented by reaction times to compare processing difficulty.

3. Experiment

In the current study, we used a speeded AX-discrimination-task since it does not
demand lexical knowledge of the language in question and can hence be performed by learners and non-learners alike. We chose an inter-stimulus-interval of 1600ms, which forces listeners to access prelexical representations to store the first token for comparison.

The research questions under investigation in the current study are the following:

(1) Are there differences in the discrimination of non-native consonantal and vocalic length contrasts? Can listeners with a length contrast in one class of segments (e.g., vowels) employ durational cues for the discrimination of a length contrast in another class of segments (e.g., consonants)?

(2) Does the ability to perceive consonantal length improve with increased exposure to a language with such a contrast? Does the perception of consonantal length contrasts become native-like?

(3) Do the mechanisms for native and non-native vocalic and consonantal length perception differ?

Given that both kinds of length contrast are signaled by segment durations and given that the durational differences between short and long segments are well beyond the just noticable difference for duration, one might predict that there are no differences in the non-native perception of consonantal and vocalic length contrasts. On the other
hand, following Bohn's (1995) Desensitization Hypothesis, non-native vowel length contrasts should be easier to perceive than non-native consonantal length contrasts.

3.1 Methods

Participants

Three groups of participants participated voluntarily in the experiment, 10 speakers of Standard German with no training in Italian (henceforth non-learners), 10 native Germans who were proficient learners of Italian (henceforth German-Italian learners), and 10 native speakers of Italian with no knowledge of German. None of the participants had any self-reported speech or hearing deficits.

The non-learners consisted of 3 male and 7 female speakers, aged between 21 and 28 (mean age: 23.3). None of them had learned Italian or another language with a phonemic consonantal length contrast.

The learners included 9 females and 1 male speaker, aged between 20 and 35 (mean age 23.1). They had all learned Italian for at least 11 months at a University (up to a maximum of 11 years). Except for two, they had all lived in Italy for at least 4 months (to a maximum of 3 years).

The Italian group was comprised of 4 male and 7 female speakers, aged between 19 and 29 (mean age 24.1). None of them had learned German or stayed in Germany for a longer period, but all except for one had learned English as an L2.
Materials

The stimulus material consisted of 41 phonotactically legal mono- or disyllabic non-words taken from the Italian GEMMA project (Di Benedetto, 2000). They were phonotactically legal in German as well. Four experimental conditions involving segmental and length contrasts, both in vowels and consonants, were created (for an overview see Table 1). For the consonantal length condition, we used VCV and VC:V minimal pairs with a contrast between geminates and singletons. For the consonantal segment condition, we used VCV minimal pairs with singletons consonants that only differed in the place of articulation (same manner of articulation and voicing). We used a subset of consonants that occur in German and which can occur as singletons and geminates in Italian. The consonants were chosen so that words in the consonantal length condition consisted of two fricatives, two nasals or two stops, both unvoiced and voiced (/f, s, m, n, t, b/). They were combined with the vowels /a, i, u/, a subset of the Italian and German vowel inventories. In the vowel length condition, we used CVC and CV:C minimal pairs with a contrast between long and short vowels. The vowels were restricted to /a/ /a:/, /e/ and /e:/ since only for these two vowel phonemes there is a length contrast in German (see section 2.3). In the vowel segment condition, we used CVC minimal pairs with the short vowels /a/ and /e/. The
surrounding consonants had the same place and manner of articulation⁵.

<table>
<thead>
<tr>
<th>C length</th>
<th>C segment</th>
<th>V length</th>
<th>V segment</th>
</tr>
</thead>
<tbody>
<tr>
<td>[afa]</td>
<td>[afa]</td>
<td>[faf]</td>
<td>[f af]</td>
</tr>
<tr>
<td>[asa]</td>
<td>[afa]</td>
<td>[f af]</td>
<td>[f af]</td>
</tr>
<tr>
<td>[usa]</td>
<td>[asa]</td>
<td>[f af]</td>
<td>[f af]</td>
</tr>
<tr>
<td>[usa]</td>
<td>[asa]</td>
<td>[f af]</td>
<td>[f af]</td>
</tr>
<tr>
<td>[mam]</td>
<td>[mem]</td>
<td>[mem]</td>
<td>[mem]</td>
</tr>
<tr>
<td>[mem]</td>
<td>[mem]</td>
<td>[mem]</td>
<td>[mem]</td>
</tr>
<tr>
<td>[bep]</td>
<td>[bap]</td>
<td>[bap]</td>
<td>[bap]</td>
</tr>
<tr>
<td>[bap]</td>
<td>[bap]</td>
<td>[bap]</td>
<td>[bap]</td>
</tr>
<tr>
<td>[vef]</td>
<td>[v ef]</td>
<td>[vef]</td>
<td>[vef]</td>
</tr>
<tr>
<td>[vef]</td>
<td>[vef]</td>
<td>[vef]</td>
<td>[vef]</td>
</tr>
<tr>
<td>[nan]</td>
<td>[n an]</td>
<td>[n an]</td>
<td>[n an]</td>
</tr>
<tr>
<td>[nan]</td>
<td>[n an]</td>
<td>[n an]</td>
<td>[n an]</td>
</tr>
<tr>
<td>[gak]</td>
<td>[g ek]</td>
<td>[g ek]</td>
<td>[g ek]</td>
</tr>
<tr>
<td>[gak]</td>
<td>[g ek]</td>
<td>[g ek]</td>
<td>[g ek]</td>
</tr>
</tbody>
</table>

Table 1: Minimal pair contrasts used in the four conditions

⁵ Voicing of onset and coda consonants differed in some pairs as German exhibits syllable-final devoicing of obstruents.
Furthermore, we created three practice pairs, two of which differed in consonantal and vowel length (\([ulu – ul:u], [dap – da:p]\)) and one with no contrast (using two different tokens of the same type, \([iri1 – iri2]\)).

Experimental and practice stimuli were read in isolation and recorded twice by a female speaker who was bilingual in German and Italian. Recordings took place in a sound-attenuated cabin at the University of Konstanz. Data were directly digitized using a DAT recording with a sampling rate of 44.1kHz and a resolution of 16Bit.

To test the durational differences of the stimuli in the length conditions, long and short consonants and vowels were annotated using Praat (Boersma and Weenink, 2009), following standard segmentation criteria (Turk et al., 2006). More specifically, for plosives, we annotated closure duration (starting from a clear drop in the amplitude of the waveform and a drop in higher frequency energy, especially F2, in the spectrogram up to the release burst) and for fricatives the duration of friction (as determined by aperiodicity in the waveform). For nasals, measurement started when the amplitude in the waveform dropped and the waveform showed less high frequency components (drop in high frequency energy in spectrogram). Results of an acoustic analysis showed that on average, long vowels were 2.3 times longer than short vowels, while geminates were on average 1.9 times longer than singleton consonants (see Table 2). The vocalic duration measurements ensured that the acoustic criteria for

---

6 The number stands for a particular token of that type, i.e., iri1 and iri2 represent different recordings.
the length distinction in vowel and consonants were met.
Table 2: Average durations and standard deviations (in ms) for short and long segments.

Furthermore, for vowels, the first and second formants at the mid point of the vowel
were automatically extracted. Results of a paired t-test showed a significantly higher F2-value for short /a/ compared to long /a/ (1435.6 Hz vs. 1322.1 Hz, t(5) = 4.4, p < 0.01) but no difference in F1 (on average 956 Hz, p > 0.4). For /e/, F1 and F2 did not differ for long and short vowels (p > 0.05). Average F1 was 624 Hz, average F2 2209 Hz.

To exclude the possibility that hearers could make the critical consonantal length judgment based on the duration of the vowel preceding the consonant, this vowel was held constant. For each vowel /a, i, u/ we chose a token that was always shorter in duration than any geminate consonant and longer than any singleton consonant. Using Praat, the selected instances of /a/, /i/ and /u/ were spliced out and every vowel preceding a consonant was then replaced by the vowel splice. This way, the first vowel of every non-word was always constant across consonantal length conditions.

**Procedure**

Since every word was recorded twice, there were 8 versions of every word pair: For pairs without a contrast (i.e., with different tokens of the same type), there were four possible combinations (e.g., afa1-afa2, afa2-afa1, af:a1-af:a2, af:a2-af:a1), for pairs with a contrast there were eight possible combinations (e.g. afa1-af:a1, af:a1-afa1, afa1-af:a2, af:a2-afa1, afa2-af:a1, af:a1-afa2, afa2-af:a2, af:a2-afa2).
We created two experimental lists, each containing all four versions of the pairs without a contrast and four of the eight versions of the pairs with a contrast (order of combinations in the pairs with a contrast was counterbalanced across lists, e.g., af:a1-af:a1 in list 1 and af:a1-afa1 in list 2). Each list hence contained 192 trials, 24 word pairs x 8 versions (4 with and 4 without a contrast). We created four blocks, containing 48 stimuli each. The experiment started with a training block, consisting of the three practice pairs. Participants were randomly assigned to one of the lists (five participants for each list).

German participants (non-learners and learners) were tested on the same-different task at the University of Konstanz, either alone or in groups of two to three. The experiment was programmed using in-house software. Stimuli were presented via headphones (Sony MDR-CD570) by means of an external player (M-audio microtrack II). Three beeps (1 kHz, 300 ms) signaled the beginning of a block and participants could take breaks between blocks. Each trial started with a 300 ms, 1 kHz sinusoid beep and 500 ms of silence, followed by the first stimulus. The two members of a pair were separated by a silence of 1600 ms. This long ISI was chosen in order to prevent solely acoustic discrimination. Participants were then given a maximum of 1600 ms before timeout and no visual aids were provided. All answers and reaction times were recorded using a button box. Participants used their dominant hand for a 'same' response and their non-dominant hand for a 'different' response.
The Italian participants were tested in Rome. As the experiments used for German participants could not be shipped in time, the Italian AX-task was programmed in presentation. Due to a miscommunication between testers, the experimental procedure differed slightly from the one for the German participants. Participants first saw the trial number on screen for 1000 ms. Then the screen turned blank and a 300 ms, 1 kHz sinusoid beep was played followed by 500 ms of silence. The ISI between the first and second stimulus was again 1600 ms. At the offset of the second stimulus, the text clicca ora 'click now' appeared on screen. Reaction times were measured relative to the appearance of the text, up to a timeout of 1600 ms.

3.2 Results

In total, 5760 data points were recorded (30 participants x 192 trials). From these, 239 data points had to be excluded due to timeout (4.1% of the data). Timeout data were analyzed using a binomial logistic regression model with timeout (yes/no) as dependent variable and participant group (German non-learner, learner or Italian), type of contrast (length or segment), trial type (without or with contrast) and segment class (consonant or vowel) as fixed factors and participants and items as crossed random factors (Baayen, 2008; Baayen et al., 2008). Results showed only a main effect of participant group (p < 0.05): there were more timeouts in the German-Italian L2 group compared to the Italian and German groups (69 timeouts for non-learners,
Furthermore, we excluded 94 responses in which participants responded too early (1.6% of the data). Trials with such early responses were distributed evenly across conditions (all p-values > 0.3). This left 5427 data points for analysis. We will first present the analysis of d’ scores, a measure of sensitivity to the kinds of contrasts tested and then turn to reaction times.

**d’ analysis**

We calculated d’ scores (Macmillan and Creelman, 2005) for each participant for each of the four conditions (consonantal and vocalic length contrasts as well as consonantal and vocalic segment contrasts). d’ scores reflect a participants' sensitivity for differences in the stimuli based on hits and false alarms. A general linear model with d’ scores as dependent measures and *participant group, segment class* (C or V) and *type of contrast* (length or segment) as fixed factors showed a significant 3-way interaction (p < 0.05).

To investigate the nature of this interaction, data were split by segment class. In the vocalic conditions, there were no effects of *participant group, type of contrast* and no interactions (all p-values > 0.05). The average d’ score was 3.14.

In the consonantal conditions, there were main effects of *type of contrast* (p < 0.0001), *participant group* (p < 0.005), and an interaction between the two factors (p
< 0.05), see Figure 1. More specifically, for consonantal length contrasts, d' scores for German non-learners (d' = 1.31) were significantly lower than those for learners (d' = 1.95, p < 0.05), which in turn were lower than those for Italians (d' = 3.04, p < 0.005). For consonantal segment contrasts, learners (d' = 3.26) did not differ from German non-learners (d' = 3.30, p > 0.5); the difference between learners and Italians (d' = 3.44) was significant (p < 0.05). However, similar to vocalic segment contrasts, all three participant groups' d'-scores were close to ceiling (between 3.26 and 3.44).

Figure 1: Average d’-scores for the discrimination of consonantal length and segmental contrasts by the three participant groups, based on the statistical model. Whiskers represent standard error.
To test whether the lower d' scores for German non-learners and learners in the consonantal length contrast was caused by difficulties in the 'same' or 'different' trials, we also analyzed the accuracy in this condition. To this end, we calculated a binomial logistic regression model with accuracy as dependent variable and participant group and trial type ('same' or 'different') as fixed factors. The results showed main effects of participant group (p < 0.05), of trial type (p < 0.05) and a significant interaction between the two factors (p < 0.05). To investigate the nature of the interaction between participant group and trial type, we looked more closely at the reactions to 'same' and 'different' trials, respectively. For 'different' trials, results showed significant differences across all three groups: Accuracy was higher for German learners of Italian than for German non-learners (z = 2.0, p < 0.05), and higher for Italians than for German learners (z = 3.9, p < 0.001). In trials without a contrast, there were no significant differences across participant groups (all p-values > 0.05). Thus, the difference in d’ scores for consonantal length contrasts between participant groups can be accounted for by their differential performance on ‘different’ trials.
Reaction time analysis

The reaction time (RT) analysis was performed to investigate task difficulty for the different length conditions (native and non-native), which provides information on the underlying processes involved. Therefore, we only analyzed RTs in trials with a vocalic and consonantal length contrast\(^7\). To account for participant-specific RT-differences, we normalized the raw RT data in the following way: RTs for correct trials without a length contrast, task difficulty cannot be investigated as we cannot decide whether participants reacted on the lacking length contrast or the lacking segmental contrast.

---

\(^7\) For trials without a length contrast, task difficulty cannot be investigated as we cannot decide whether participants reacted on the lacking length contrast or the lacking segmental contrast.
responses were averaged for each participant and condition. Then, the (thus averaged) RTs in the segmental control conditions (vowel and consonants) were subtracted from the (averaged) RT in the respective length conditions (vowels and consonants) for each participant. Figure 3 illustrates the average RTs for the three participant groups in the two length conditions. In what follows, we will describe the results of a univariate ANOVA with normalized RT as dependent variable and participant group and segment class (consonant or vowel) as fixed factors.

Normalized RTs to trials with a length contrast were significantly shorter for vowel length contrasts than for consonantal length contrasts (F(1,60) = 34.7, p < 0.001). There was no effect of participant group (p > 0.1) but an interaction between participant group and segment class (F(2,56) = 7.2, p < 0.005). More detailed analyses showed that for Italians, there was no difference in RTs for consonants and vowels (F(1,18) < 1, p > 0.8), while for both German groups, reactions were significantly shorter for vowel length contrasts than for consonantal length contrasts (F(1,18) = 22.7, p < 0.001 for non-learners, F(1,18) = 17.5, p = 0.001 for German-Italian L2). Furthermore, normalized RTs in the consonantal length condition were significantly longer for the two German groups compared to the Italian group (both p-values < 0.05), while the two German groups did not differ from each other (p > 0.9).

---

8 Reaction times for segment conditions did not differ as a function of participant group or segment class (all p-values > 0.2).
9 Since the data were aggregated, we calculated a general linear model instead of a linear mixed effect regression model as done before.
The difference in normalized RTs in the vowel length condition was also significant (F(2,27) = 2.8, p < 0.05). German non-learners and learners did not differ from each other (p > 0.8), but were significantly faster than Italians (p < 0.05).

![Normalized reaction times (RT for length condition - RT for segment condition) for trials with a length contrast as a function of segment type and participant group. Whiskers represent standard error.](image)

**Figure 3.** Normalized reaction times (RT for length condition - RT for segment condition) for trials with a length contrast as a function of segment type and participant group. Whiskers represent standard error.

### 4. General Discussion

Analyses of participants' responses in this same-different task clearly showed an
asymmetry in the ability to perceive length contrasts that are not phonemic in one's L1. Vowel length contrasts, on the one hand, appeared to pose little problems, irrespective of whether vowel length was phonemic in ones' native language or not: they were perceived as successfully as segmental contrasts. On the other hand, the perception of consonantal length contrasts proved to be difficult for non-native listeners, even for the advanced German learners of Italian. Despite a considerable improvement compared to the group without exposure to a language with consonantal length contrasts, their perception was (still) clearly different from that of native Italian listeners.

The learners in this study have had ample experience with Italian (most of them had lived in Italy for at least several months), which is most likely the source for their improved sensitivity to the consonantal length contrast. In terms of theories of speech perception, there are two possible explanations that can account for the results of the perception of non-native consonantal length contrasts: On the one hand, the learners’ phonological system might have restructured in a way to allow for long alternants in addition to short exemplars for each such individual segment that they observe in the target language (i.e., upon encountering enough instances of a geminate /tː/ or /kː/, they establish this new category, and this happens for each single segment independently). Such a process would not be very efficient, though. Alternatively, a more general concept of geminate might have been established (we are not suggesting
any details on the mechanism of this process), which applies to the whole class of consonants and thus can easily be transferred to any new consonant. Based on the results of the current study, using materials that contained consonantal length contrasts in a broad range of stimuli (plosives, fricatives, nasals, and liquids, voiced as well as voiceless) rather than just individual consonant pairs, the latter hypothesis seems more sensible. Hence learners seem to have started to adapt their prelexical processing (McQueen, 2005) so as to process and integrate this (for them) new type of contrast into the L2 system. This enhanced prelexical processing in turn explains why their performance lies in between the performance of Italian natives and participants who have no experience with a language with consonantal length contrasts. This replicates the findings of Heeren and Schouten (2008, 2010) regarding the learnability of consonantal length differences.

In terms of phonological L2 learning models, the one that can account for this improved perception of consonantal length by the L2 learners best can be considered to be Flege’s (1995) Speech Learning Model (SLM). Most crucially, the SLM allows for readjustment of native categories and the formation of new categories – in our case the development of a new suprasegmental contrast (length) for a whole class of segments (consonants). German speakers with no exposure to a consonantal length contrast obviously do not associate the durational differences in consonants with

---

10 The materials were not designed, however, to adjudicate between these two hypotheses.
different phonological categories, as indicated by their low discrimination scores. Thus, they map the consonants – irrespective of their duration – onto one and the same category, the only one they have available in their L1 inventory. This could be interpreted as a native language magnet-like effect in terms of Kuhl (1995) or as perceptual assimilation (Best, 1995). This interpretation is strengthened by the fact that naïve German listeners mostly classified trials with consonants of different length as 'same', although the two stimuli to compare represented different acoustic tokens. German speakers with no consonantal length experience detected even less differences between such pairs than L2 learners did, indicating a stronger magnet effect (or more perceptual assimilation) for the non-learners.

As indicated above already, the hypothesis of an obvious effect of the native phonological system on non-native durational contrasts, especially on inexperienced participants, cannot be upheld for vowels. In the vowel length condition, all participant groups achieved a high level of accuracy in the current study, regardless of the native language. As illustrated by the d’-scores as well as by the accuracy scores, the non-native length contrast in vowels did not pose any problems for the Italian listeners whose native language system does not have a vowel length contrast. We see two possible explanations for this finding: First, our data can be taken to support a general speech perception strategy such as the one proposed by Bohn (1995), who claims that cross-linguistic vowel length perception (but not consonantal length
perception) is good independent of prior experience with it in the L1. Furthermore, vowels carry a high phonetic informational load. For instance, rhythm and intonation are mainly implemented by suprasegmental features realized on vowels. Furthermore, vowels also contain cues to neighboring segments (e.g., formant transitions) and may convey speaker identity. Therefore, vowels might just be processed more efficiently than consonants in general (cf. Cutler and Mehler’s, 1993, ‘periodicity bias’ for infants). This might well mask a potential L1 disadvantage for the perception of non-native vowel length contrasts for Italian listeners.

Second, the sensitivity of the Italian group to the non-native vowel length contrasts could also be attributed to the use of allophonic vowel length: In Italian, stressed open syllables especially in penultimate position are considerably lengthened (Braun and Geiselmann, 2011; Krämer, 2009). Since such positionally conditioned phonetic variation occurs very often, Italians have frequent phonetic experience with long and short vowels in the L1 (cf. Kondaurova and Francis, 2008). If experience with allophonic vowel duration were the main predictor for the ability to perceive length contrasts, however, then German listeners would be expected to perform equally well as Italian listeners since they also have experience with allophonic consonant length. For instance, when the same phoneme occurs at the end of one word or morpheme and again at the onset of the next word or morpheme (as in *mit Tim* ‘with Tim’ or *mitteilen* ‘to convey’), a ‘quasi geminate’ is produced. According to
acoustic analyses (Mikuteit, 2007), such 'quasi geminates' are on average 1.8 times longer than an otherwise identical singleton consonant, a ratio that mirrors the duration ratio in the current consonantal length condition. Consequently, German listeners should be expected to have little problems with the perception of the non-native consonantal length contrast. This is however not what we found. Conceivably, allophonic consonant duration in German occurs much less frequently than allophonic vowel duration in Italian. As Kondaurova and Francis (2008) have shown, the frequency of occurrence of allophonic durational differences predicts how strongly listeners rely on duration in a non-native categorization task. In future studies it will be interesting to investigate more closely under which conditions allophonic durational differences are processed efficiently in the native language (for example, if there is a frequency threshold) and whether this reliably predicts success in length distinctions in a second language.

Our main argument for an asymmetric processing rests on the observation that consonantal length contrasts are more difficult to perceive in an L2 compared to vocalic length contrasts. A closer look at the processing costs (reaction times) provides supplemental evidence for an asymmetric performance. To be precise, the reaction times in the vowel length condition were longer for Italians than for both German groups, suggesting that the non-native length contrast was more difficult to process. A comparable increased processing cost for non-native contrasts was also
found for the perception of consonantal length. Here, the two German groups were clearly slower than the Italian group. Such a difference in RTs between native and non-native length contrasts is particularly revealing in the case of vocalic length, in which all participant groups’ behavioral performance was symmetrical, in this case nearly at ceiling. Thus, symmetrical (i.e., comparable) response data may be reached through asymmetric (i.e., different) processing paths. We therefore conclude that the underlying mechanisms leading to successful discrimination still differ between speakers with and without a given phonemic length contrasts in their native language. In other words, L2 interlanguage may approach target-like behavior on the surface but learners’ underlying processes still remain different from those of native speakers of the target language.

5. Conclusion

Listeners' durational sensitivity to non-native vowel length contrasts seems to be high even when vowel length is not contrastive in one's native language (the task may only be harder to do). On the contrary, consonantal length contrasts remain to be a challenge even after prolonged exposure to a language that employs such contrasts. It remains unclear how important (the frequency of occurrence of) allophonic experience with durational contrasts for a given class of segments (i.e., either consonants or vowels) in the L1 may be. Conceivably, the observed cross-linguistic
asymmetry in the ability to perceive vowel and consonantal length may have its roots in the informational value carried by vowels and consonants, respectively.

On a more general note, it is striking that Italians, despite the demonstrated perceptual sensitivity to durational information in vowels, appear to experience difficulties in adequately producing duration in a non-native language, such as English. A prototypical pronunciation error of Italian L2 learners of English is the non-distinct production of pairs such as *sheep* and *ship* (which are both pronounced as *[ʃip]*)). Future research therefore will have to show whether those learners suffer less from difficulties in the perception of that contrast but potentially more from difficulties on the level of lexical encoding (cf. Hayes-Harb and Masuda, 2008), in generating the correct prosodic plan for their L2 utterances (cf. Levelt, 1993) or in an adequate synchronization of the different articulators in L2 speech on the level of production.

References


Braun B and Geiselman S (2011) Italian in the no man's land between stress-timing and syllable-timing? Speakers are more stress-timed than listeners.


Fougeron C and Keating PA (1997) Articulatory strengthening at edges of prosodic


phonological status of the tense-lax distinction in German. *Forschungserichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (IPKM)* 35, 17-25.


http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-opus-78588


Munro MJ and Derwing TM (1995) Foreign accent, comprehensibility, and


Vennemann T (1991) Syllable structure and syllable cut prosodies in Modern


