Chapter 9

Three language-specific phonological interpretations of release bursts and short vowel-like formants

Silke Hamann[a] & Veronica Miatto[b]

[a]ACLC, University of Amsterdam [b]Stony Brook University

In this paper we challenge the assumption that there is one correct interpretation of auditory signals by showing that the mapping between perceptual cues and phonological categories is highly language specific. This is exemplified by three different languages and their different interpretations of final release bursts.

The first case is American English (AE), where a release burst in final position is always interpreted as a perceptual cue to a final plosive by AE listeners. The second case is the interpretation of a final release burst as a vowel, despite the lack of vowel-like formants following the burst, as in Korean. In the third case, exemplified by Italian, not only a release burst but also vowel-like formants are present word-finally, but the latter are not interpreted as phonological vowels.

These different phonological interpretations of the same phonetic material can only be formalized in a linguistic model that makes a clear distinction between phonetic and phonological representations, and allows for language-specific, non-universal mappings between the two. For this, we employ the OT version of the Bidirectional Phonetics and Phonology model, and show how different rankings of cue constraints in these three languages can account for the different phonological interpretations. This formalization accounts for both native and naive second language perception, and provides testable predictions for future experimental studies.

1 Introduction

Speech perception involves the interpretation of auditory information as phonological units, and is to a certain degree language specific, as early illustrations by e.g. Polivanov (1931/1974) have already shown. A familiar example of this language specificity is the phenomenon of so-called "illusory" vowels: listeners exposed to words with consonantal clusters that are illicit in their native language perceive a vowel breaking up these clusters, even though the acoustic signal does not contain any corresponding vowel-like formants, see e.g. Dupoux et al. (1999) for Japanese, Kabak & Idsardi (2007) and Durvasula & Kahng (2015) for Korean. Interpretations of this phenomenon mainly focus on the role of language-specific phonotactic restrictions and the absence of vowel formants, the latter leading to its name. Less attention is paid to the fact that these perceptual "illusions" seem to be an efficient way for listeners to interpret their native language, where the corresponding vowel categories are often realized as voiceless: In Japanese, for instance, the vowel /ɯ/ is usually devoiced if preceded and followed by voiceless consonants, therefore it helps Japanese listeners to perceive the absence of vowel formants in such a context as a vowel /ɯ/ to quickly retrieve the correct word form (see e.g. the results of the lexical decision task with Japanese listeners by Ogasawara & Warner 2009). In speech perception, language-specificity thus holds not only for phonotactic restrictions, but also for the interpretation of perceptual cues.

The assumption of a language-specific mapping between auditory information and phonological categories opposes the view of those generative phonologists that assume a universal phonetics-phonology interface, see e.g. Hale & Kissock (2007) and Hale & Reiss (2000) (cf. Hamann 2011 for a detailed discussion). In their view, sounds that share the same phonological feature (or feature value) across languages have identical auditory cues and/or articulatory gestures.

In the present article, we illustrate with the example of two possible cues for plosives, namely burst release and a following short period of vowel-like formants, how language-specific the interpretation of perceptual cues can be. There are, of course, many more language-specific plosive cues, such as voice-onset time (see, e.g., the seminal papers by Lisker & Abramson 1964, 1970), which we do not consider here for reasons of space and clarity. We restrict our discussion to velar plosives in three languages, American English, Korean and Italian, as summarized in Table 1. With respect to Italian, we focus on the Venetian variety and ignore variation in the interpretation of final schwa-like formants that have been reported for several other dialects (see e.g. Cavirani 2015 for the Lunigiana dialects).

Table 1: Cues for syllable-final velar plosives in three example languages

Cue interpretations	Example language
Closure release is an optional cue to a syllable-final plosive Short period of formants following the release is an obligatory cue to a following vowel	American English
Closure release is an obligatory cue to an onset plosive Short period of formants is a cue to a following vowel	Korean
Closure release is an obligatory cue to a syllable-final plosive Short[a] period of formants is an optional cue to a word-final plosive	Italian

[a]This period shows quite some variability in duration, as mentioned in Section 2.3. Nevertheless it is on average shorter than a full vowel.

The pattern of often-released coda plosives in American English can also be observed in the closely-related West-Germanic languages Dutch and German, and in Romance languages (where the coda release seems obligatory, see §2.3 below). The dominance of researchers with a West-Germanic or Romance native language in the linguistic field from the 16th until the mid 20th century, but also extensive, detailed phonetic studies of these restricted language groups might have led to the false assumptions that 1) the release of coda plosives is a common pattern in all languages, and 2) that such a release has to be interpreted as indicating a plosive, only. A lack of a plosive burst in syllable-final position, however, is obligatory in our second example language, Korean, and also in other languages spoken in Asia, such as Hong Kong Cantonese (Bauer & Benedict 1997), Thai (Abramson & Tingsabadh 1999), Vietnamese (Kirby 2011), and elsewhere, e.g., Karitiana (Storto & Demolin 2002), Ibibio (Urua 2004), and Efik (Cook 1969). Korean listeners interpret a coda burst as indicating a following (voiceless) vowel. Formant-like structures, on the other hand, are interpreted in many languages as indicating a vowel category, while in our third example language, Italian, they are a cue enhancing the release of a final plosive. A similar interpretation of short, vowel-like formant structures after coda consonants can be found in French (see, e.g., Flege & Hillenbrand 1987).

Besides a description of how the three example languages given in Table 1 interpret these two auditory cues, the present article also provides an explicit formalization, i.e. the mapping of these cues onto phonological categories (such as plosives and vowels). If this mapping were universal, language learners would not need to acquire it, and linguists would not need to model it. However, since we observe language-specific differences in the interpretation of these cues, we provide a model of these language-specific mappings to account for the knowledge that the learners of these three languages have acquired. To perform such a formalization, a model is needed that makes a strict distinction between phonetic realizations and phonological categories, and that allows for a language-specific mapping between the two. For this purpose, we employ Bidirectional Phonetics and Phonology (Boersma 2007, 2011, 2009), the only existing linguistic model to our knowledge that incorporates detailed phonetic realizations as well as phonological surface forms (categories but also other phonological structure like syllables, prosodic words, etc.). We will use this model to illustrate that by formalizing the details of L1 perception, i.e., which cues are mapped onto which phonological form, we can account for what listeners of these languages do when encountering an auditory form that does not exist in their L1 (i.e., we model the process of naive L2 perception).

This article is structured as follows. We briefly describe the relevant phonetic and phonological details of the three languages American English, Korean and Italian in §2. §3 provides the formalization, and §4 then discusses and concludes.

A short explanation of the notation we use is necessary. In order to distinguish three levels of representations, the present article employs square brackets for [auditory forms], slashes for /phonological surface forms/, and pipes for |underlying, lexical forms|.

2 The three example languages

2.1 American English

American English (henceforth: AE) is a language that has plosives both in onset and in coda position, as e.g. in *cake* |kʰeɪk|. In both onset and coda position, plosives can be followed and/or preceded by other consonants. Syllable-final coda plosives are often released with an audible burst, but can also be unreleased (e.g., Rositzke 1943, Crystal & House 1988, Davidson 2011). According to Kim (1998), this is a notable difference with Korean, where a plosive in coda position is never released, see elaboration in §2.2 below.

Whether coda consonants in AE have a burst cue depends on many factors: Plosives are more often released after tense vowels than after lax vowels (e.g., Parker & Walsh 1981), in phrase-final position plosives than in phrase-medial position (e.g., Crystal & House 1988), by female speakers than by male speakers (e.g., Byrd 1993), and velar plosives are more often released than coronal ones (e.g., Halle et al. 1957, Crystal & House 1988, Byrd 1993), to mention only the most often observed factors.

Though the presence of a burst cue is not necessary for AE listeners to perceive a coda plosive, its absence can be falsely interpreted as a syllable without coda plosive, as indicated by several perception studies (e.g., Householder 1956; Halle et al. 1957, who show that this is particularly the case for plosives following lax stops; and Lisker 1999, who shows that this holds for plosives, especially velars, following diphthongs).

AE has not been reported to employ short vowel-like formants after the release burst to enhance the burst. AE listeners perceive speakers of languages like Italian that use such vowel-like formants following a plosive release as producing a full vowel in this position (e.g., Hall 2006), and an Italian accent in English is therefore often caricatured with schwas or [e]-like vowels after each final consonant (see also Busà 2008).

2.2 Korean

Korean speakers do not release their syllable-final plosives (e.g., Martin 1951, Kim 1998, Kang 2003). Underlyingly aspirated plosives in this position also lose their aspiration, see, e.g., the neutralization of underlying |puʌkʰ| 'kitchen', which is realized as surface /.puʌk./, compared to /.pu.ʌ.kʰɛ./ 'in the kitchen' (Kang 2003: 224; notation adapted).

Due to the absence of bursts in syllable-final position, Korean listeners interpret a plosive release in the auditory signal as an indication that the plosive occurred in syllable-initial position and was followed by a vowel, even if there are no formants in the auditory signal to support the percept of a vowel. Since the high vowel /ɨ/ in open syllables is often devoiced before and after voiceless obstruents in Korean (Kim-Renaud 1987, cited by Kang 2003: 236), Korean listeners regularly interpret consonant clusters that are phonotactically illegal in Korean as being broken up by a voiceless vowel /ɨ/ (Kang 2003; see Durvasula et al. 2018 for a detailed account of the choice of vowel).

The effect of this perceptual interpretation can be extensively observed in vowel insertion both in the adaptation of loanwords and in experimental studies, as a large body of literature has shown (e.g., Durvasula & Kahng 2015, 2016,

Hutin 2014, Kang 1996, Kang 2003, Kabak & Idsardi 2007, Iverson & Lee 2006). While for velar plosives the perception with a following vowel is almost categorical, see e.g. English *spike*, which is borrowed into Korean as /.sɨ.pʰɑ.i.kʰi./, this is more variable for other places of articulation, see e.g. English *flute*, which is borrowed both as /.pʰɨl.lu.tʰi./ and /.pʰɨl.lut./[1] (examples from Kang 2003; notation adapted). The vowel insertion in such cases depends on many factors, amongst them the tenseness of the preceding vowel, the fact that Korean nouns show an alternation between /s/ and /t/ in final position, the actual realization of these words with a final release in the donor language (such as AE, cf. §2.1 above), the functional load of the word, etc. (see, e.g., Kang 2003, Chang 2018, Kim 2022, for extensive discussions). We will come back to this variation in §4.2.

2.3 Italian

In Italian, word-final consonants are always released, and are sometimes produced with vowel-like formants after the release, so a word like *jet* might be pronounced as [dʒɛtə] and a word like *tunnel* might be pronounced as [tunnelə] with lengthening of the last consonant. The presence of these word-final, schwa-like formant structures varies based both on inter-speaker (exposure to English) and intra-speaker variables (the type of word-final consonant, the number of repetition etc.), as explained below. In production, consonant-final words *can* therefore be followed by schwa-like formant structures, but for the majority of speakers this is not categorical (see Miatto et al. 2019, Miatto 2022). However, in perception, which is the focus of this paper, a consonant-final word like *jet* will be categorically perceived as consonant-final whether the consonant is followed by such vowel-like formants or not (see Miatto 2020 for a relevant study using nonce words).

Italian consonant-final words are relatively new and are mainly loanwords or acronyms, so these vowel-like formant structures have long been treated as vowels epenthesized to adhere to a phonological constraint that prohibits word-final codas (Bafile 2002, 2003, 2005 for Tuscan Italian, Passino 2008 for Abruzzese Italian, and Broniś 2016 for Roman Italian). The implications of this interpretation are that 1) the phonetic material is perceived and produced as a phonological vowel, and that 2) it constitutes the nucleus of a separate syllable. Although using different frameworks, these authors agree that *jet* in Italian is constituted by two syllables /dʒɛt.tə/, with the lengthened consonant serving as a geminated consonant.

---

[1] For this particular word, the borrowing with a coda consonant, thus without inserted vowel, occurs more frequently, as a Google search by a reviewer showed.

More recently, however, it has been proposed that these vowel-like elements are not syllabic, i.e., they do not constitute the nucleus of a syllable. Grice et al. (2018) propose that for Barese Italian it is a non-syllabic vowel that appears under certain phonological and prosodic pressures, and its characteristics are similar to Hall's (2006, 2011) intrusive vowels (see also Hall 2024 [this volume]). Similarly, Repetti (2012), followed by Miatto et al. (2019) and Miatto (2020, 2022) for Venetian Italian, propose that this is a vowel-like segment that is part of the release of the consonant. The latter is also the view we adopted here, and we will refer to the presence of these vowel-like formant structures after the release of a word-final consonant as "excrescent vowel". Based on studies on Venetian Italian, we are following the latter theoretical interpretation of excrescent vowels for the following reasons. Note that from here on, when we refer to "Italian" we specifically refer to the variety of Italian spoken in Veneto (North-East Italy).

First of all, the occurrence rate of excrescent vowels in Italian is susceptible to experience with English, which does not display such vowel-like formant structures after coda releases (recall §2.1). Miatto et al. (2019) found that with increasing experience in spoken L2 English, speakers tended to produce fewer excrescent vowels after word-final consonants in Italian. This suggests that excrescent vowels are likely to be a phonetic phenomenon rather than a phonological repair, because the speakers' exposure to L2 English was probably too limited to have caused a change in their L1 phonology.

Second, following Hall (2006, 2011)'s diagnosis of epenthetic (i.e., phonological) vowels versus intrusive vowels (formant transitions between consonants), the characteristics of excrescent vowels in Italian are much closer to intrusive vowels than canonical epenthetic vowels. Excrescent vowels in Italian are not always present, they acoustically resemble schwa, and are highly variable in their duration. Moreover, they do not participate in certain phonological processes such as stress assignment (Repetti 2012) or syllabicity (Miatto 2020). In particular, Miatto's (2020) study on the perception of word-final excrescent vowels shows that Veneto speakers are not aware that they insert an excrescent vowel, and they do not perceive it. The participants of the study listened to nonce words such as /vik/, to which excrescent vowels of varying durations (incremental steps of 25 ms ranging from 0 to 100 ms) were inserted after the word-final consonant, and judged them as monosyllabic 93% of the time. Moreover, the duration of the word-final schwa did not influence significantly whether they would perceive the nonce word as a monosyllable or disyllable.

Third, Miatto's (2022) findings on factors that condition the presence of excrescent vowels after voiceless plosives can only be explained by referring to their perceptibility. In her study, excrescent vowels were more likely to appear

after labials and coronals than dorsals. It was argued that in Italian, since the overwhelming majority of consonants is followed by a vowel, formant transitions are extremely important cues for the perceptibility of plosives. Word-final labials and coronals, which have weaker bursts (Dorman et al. 1977), would then be less perceptible than dorsals, unless you had plosive releases that incorporated formant transitions. Another finding of Miatto's (2022) study is that with increased repetitions of the same plosive-final nonce word, the presence of excrescent vowels significantly decreased. Repetition is generally shown to have a negative effect on clarity and intelligibility of phonetic production (Fowler & Housum 1987), supporting the argument that the Italian excrescent vowel is a phonetic cue that aids the perceptibility of the final consonant, but appears less if the words are produced less carefully in later repetitions.

Finally, as stated in Miatto (2022), duration measurements were not consistent with a phonological analysis in which the excrescent vowel is syllabic and the nucleus of a separate syllable. In the study, she found that vowels in nonce words such as *fap* were short, and therefore obligatorily in a closed syllable due to their duration. Moreover, word-final consonants in nonce words were significantly shorter than control geminated consonants, which indicates that the consonants might not be geminated but only slightly lengthened, contrary to what was reported in previous literature.

3 Modelling the language-specific perception of burst release and vowel formants

In this section, we formalize the language-specific use of the plosive cues described in the previous section, making explicit the knowledge that the listeners of the three languages employ when listening to their native input. We also apply this knowledge in the form of a native perception grammar naively to input from the other two languages unknown to the listeners (naive L2 perception), and partly compare this to reported results from L2 perception and loanword adaptation (assuming adaptation took place via L2 perception, though alternative adaptations by bilinguals are also possible, see e.g. Paradis & LaCharité 1997).

As mentioned in the introduction, we employ Bidirectional Phonetics and Phonology (henceforth: BiPhon; Boersma 2007, 2011, Boersma & Hamann 2009) for our modelling, because it provides an explicit formalization of the phonetics-phonology interface by mapping auditory cues onto phonological surface categories, and vice versa. The relevant representations and constraint types to for-
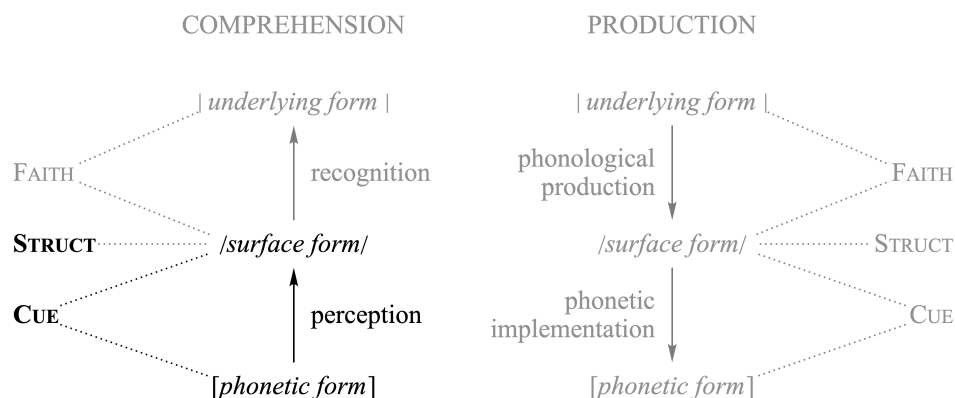
malize this mapping are given in Figure 1.[2]

COMPREHENSION   PRODUCTION

| *underlying form* | | *underlying form* |

FAITH   recognition   phonological   FAITH
                                 production

STRUCT   /*surface form*/   /*surface form*/   STRUCT

CUE   perception   phonetic   CUE
                                 implementation

[*phonetic form*]   [*phonetic form*]

Figure 1: The BiPhon model, with representations in italics and constraints in small capitals. The perception process is given in black. This mapping from auditory to surface representation and its reverse mapping in phonetic implementation form the phonetics-phonology interface.

The representations and mappings in BiPhon correspond to those in psycholinguistic models of comprehension (e.g., McQueen & Cutler 1997) and production (e.g., Levelt 1989). As the present article is only concerned with (native and nonnative) speech perception, the formalization is restricted to the perception process, any influences of lexical forms in speech comprehension are ignored, but are of course relevant in the perception of real words.[3]

The perception grammar consists of two types of constraints: Cue constraints, mapping the auditory onto the surface form, and Struct(ural) constraints, restricting the surface form. These constraints and their rankings are also used to model the production direction, see e.g. Boersma & Hamann (2009) for an illustration. The perception grammar is thus no additional device modelling only speech perception, but an integral part of a listener's/speaker's grammar. Our formalization is restricted to the voiceless velar plosive /k/ and a preceding front

---

[2]Note that Figure 1 makes no distinction between auditory and articulatory form, and summarizes both under [phonetic form]. As the present article is only about speech perception, this distinction is not relevant, but the interested reader is referred to Hamann (2011) and the discussion therein on the precise order of these forms in BiPhon and a comparison to alternative grammar models.

[3]See, e.g., Boersma (2009) for the modelling of the Ganong effect in BiPhon with parallel evaluations of surface and underlying forms.

mid vowel /ɛ/, both occurring in all three languages. The formant cues corresponding to the vowel are summarized by the auditory form [ɛ], as the detailed cues are not of relevance in the present article, but even though the same symbol is used for the auditory as for the phonological form, the reader needs to keep in mind that [ɛ] represents values of the first three formants, duration, and other auditory information of a typical realization of the abstract phonological category /ɛ/. The auditory cues of the velar /k/ of relevance to our formalization are the following: a closure preceded by the vowel, represented as the sequence [ɛ ˺], where [˺] stands for the vowel transitions containing information on the velar place of the following plosive, and [_] for the silent closure during the voiceless plosive. Furthermore, there is the velar release burst, represented as [ᵏ]. In addition, we consider the excrescent vowel that can be found in Italian final plosives after the release of the burst, and will notate these vowel-like formants as [ᵊ], to represent their often very short duration. An inclusion of other cues and other places of articulation than velar would go beyond the scope of this article.

In the following sections, we formalize the language-specific interpretation of these auditory cues, starting with the minimal cues of an unreleased [ɛ ˺], and continuing by adding subsequently the burst and vowel-like formant cues to the auditory input of the perception tableaux. We will show how these inputs, even though not always native to the language in question, are dealt with by the native perception grammars of the three languages.

In the assumption that the raw auditory signal is perceived as an abstract category, and not stored as such, we depart from theories like Exemplar Theory (Pierrehumbert 2001), that presume listeners have a holistic memory trace of all the acoustic details of an auditory input that they encountered.

The constraint rankings that we employ in the following is strict, and therefore result in categorical behaviour, i.e., one candidate wins. It is well-known that humans do not exhibit such categorical behaviour, and that the percept depends on several linguistic and social factors. We will elaborate on how this variation can be integrated in our model in §4.2 below.

3.1 Perception of transitions into closure and the closure phase

The auditory form [ɛ ˺] is perceived as the phonological surface /ɛk/ due to the transitions into the closure and the silence during closure.[4] This is captured by

---

[4]A reviewer voiced scepticism about listeners' ability to perceive silence. Silence, i.e., the absence of periodicity or friction noise, and its relative duration, has been shown to be picked up by listeners and to be an important cue for distinguishing plosives from fricatives and affricates (e.g., Repp et al. 1978, Dorman et al. 1979).

Cᴜᴇ constraints like *[`_]/s/, which stands for "Do not map velar transitions and a silent closure onto an alveolar fricative in the surface form". Cᴜᴇ constraints are employed in BiPhon to map auditory information onto phonological categories, and are formalized negatively due to OT's exclusion mechanism.[5] Further, similar Cᴜᴇ constraints could be employed to exclude other consonantal candidates but we refrain from this in the interest of brevity and clarity. An antagonistic constraint, violated when these cues are mapped onto a phonological velar plosive, i.e., *[`_ ]/k/, though seeming counter-intuitive, is relevant and included in the following formalization. The constraint *[ `_]/ / avoids that these cues are simply ignored (mapped onto nothing in the surface form). Table 2 illustrates how these three constraints capture the correct mapping onto a velar voiceless plosive in Korean.

Table 2: Korean

[ε`_]	*[`_]/s/	*[`_]/ /	*[`_]/k/
☞ a. /εk/			*
b. /εs/	*!		
c. /ε/		*!	

In the perception tableaux in Table 2, as in the following perception tableaux, the input is an auditory form, and the output candidates are surface phonological forms, which are constructed by the listener to access underlying phonological representations in the mental lexicon (i.e., the intermediate stage in psycholinguistic models of speech perception and comprehension by, e.g., McQueen & Cutler 1997). The input is a native production of a Korean unreleased coda plosive.

Under ideal circumstances (without background noise), Korean listeners perceive [ε `_] as surface /εk/, i.e., as a velar plosive in coda position, because they are used to non-released plosives in their language and familiar with auditory inputs like these to be interpreted as surface representations of coda plosives. We therefore assume that the two first constraints in Table 2 are high ranked and the third low ranked in the perception grammar of Korean. AE listeners are also used

---

[5]With a set of Cᴜᴇ constraints that prohibits the mapping of all possible occurring auditory events onto all possible phonological categories, and input distributions of actually occurring auditory cue values for phonological categories in a language, one can simulate the acquisition of such a perception grammar and thus provide a stochastic model of the language-specific acquisition process of a child (see, e.g., Escudero & Boersma 2003). A more realistic model would assume positive connections between occurring values and phonological categories (as possible in Harmonic Grammar (Legendre et al. 1990), see, e.g., Zhou & Hamann 2024).

to perceiving unreleased plosives as coda consonants, and we therefore assume a similar ranking for now, though we learned in §2.1 that the absence of a burst can also be interpreted by AE listeners as a syllable without a coda plosive. We will return to this variability in AE perception in §4.2.

For Italian, unreleased plosives in coda position are not reported (recall §2.3). We therefore assume Italian listeners without any knowledge of languages like English or Korean are likely to perceive the non-native [ε ˺] as not containing a plosive. They require explicit cues for the existence of a plosive consonant, as we will see in the following sections. Rather, we speculate that Italians perceive this input as containing only a vowel, i.e., as the third candidate in Table 3. To the best of our knowledge, experimental evidence for these assumptions do not exist, and a future perception study will need to test them. However, our speculations are based on impressionistic evidence, namely the second author of this paper noticing that some native Italian participants were occasionally perceiving experimental nonsensical stimuli like [mip] as /mi/ when asked to repeat what they were hearing. It has to be noted that those particular stimuli had a very quiet bursts and no word-final schwa-like formants. This leads us to speculate that without a burst (or even with a very quiet one, as represented by these cases) Italian speakers do not perceive a final plosive consonant.

Based on our assumptions, we argue that in the Italian perception grammar, the third candidate wins, and therefore the constraints *[˺] / / and *[˺]/k/ have to have a reverse order than in Korean, see Table 3 for the Italian perception grammar:

Table 3: Italian

[ε˺_]	*[˺_]/s/	*[˺_]/k/	*[˺_]/ /
a. /εk/		*!	
b. /εs/	*!		
☞ c. /ε/			*

In Sections 3.2 and 3.3 we will not include these three constraints, nor the candidates two and three, but we will come back to them in §3.4.

3.2 Perception of a burst release

We continue by adding the cue of a burst release, which natively is used in AE for coda plosives. We thus consider how our three groups of native listeners would cope with a typical AE input (native perception for AE listeners, naive L2 perception for Italian and Korean listeners).

The addition of the burst to the auditory input, resulting in the input [ɛ ꞌ‿ᵏ] requires two separate Cᴜᴇ constraints. The first, *[ᵏ]/ /, is violated when the burst cue is ignored. This is not allowed in any of the three languages; hence the constraint must be highly ranked in all three. The second constraint avoids that the burst is interpreted as a coda consonant: *[ᵏ]/k./, where "." stands for a syllable boundary. In Korean, where the burst is an indication that the consonant occurred in onset position, this constraint is high ranked, in AE and Italian low.

We learned in §2.2 that in cases where there are no vowel cues, Koreans perceive an /ɨ/, as this vowel is often devoiced. The input [ɛ ꞌ‿ᵏ] would thus be mapped onto the phonological form /ɛ.kɨ/ by naive Korean listeners, a mapping that violates the Cᴜᴇ constraint *[ ]/ɨ/: "Don't map the absence of vowel cues onto a surface /ɨ/". This constraint is low ranked in Korean, see Table 4, due to the often-occurring devoicing, i.e., listeners are used to interpret a non-existence of formants in the auditory form as the vowel category /ɨ/ in Korean.

Table 4: Korean

[ɛꞌ‿ᵏ]	*[ᵏ]/ /	*[ᵏ]/k./	*[ ]/ɨ/
a. /ɛk/		*!	
☞ b. /ɛ.kɨ/			*
c. /ɛ/	*!		

For AE, we assume that the vowel that is closest to having no perceptual cues is the unstressed schwa, and hence the relevant candidate and constraint look slightly different, cf. Table 5. In contrast to Korean, AE does not allow listeners to perceive a vowel (even a schwa) in the absence of corresponding cues, as reflected in the high ranking of the constraint *[ ]/ə/.

Table 5: AE

[ɛꞌ‿ᵏ]	*[ᵏ]/ /	*[ ]/ə/	*[ᵏ]/k./
☞ a. /ɛk/			*
b. /ɛ.kə/		*!	
c. /ɛ/	*!		

Italian has similar candidates and a similar constraint ranking as AE in Table 5, since a velar release is in general very strong (Dorman et al. 1977) and therefore a sufficient cue for a plosive, also in Italian.

Silke Hamann & Veronica Miatto

3.3 Perception of short vowel-like formants following the burst

The last cue that we include in our formalization is the excrescent vowel that is typical in the production of coda consonants in Italian, resulting in the auditory input [ɛ ˈ ͜ ᵏə]. We will illustrate how Italian listeners perceive this input natively, and how the other two language groups deal with this non-native input.

Again, this cue requires constraints referring to it. The constraint *[ə]/ / is taking care that this cue is mapped onto a separate phonological surface vowel, and is high ranked in AE and Korean. In Italian, it is low ranked, and here the short vowel-like cue is interpreted together with the burst to indicate a velar occurring in final position; it enhances the burst cue. We can formalize this with the constraint *[ᵏə]/.k/, which is high-ranked in Italian, see Table 6.[6]

Table 6: Italian

[ɛˈ͜ᵏə]	*[ᵏ]/ /	*[ ]/ə/	*[ᵏə]/.k/	*[ᵏ]/k./	[ə]/ /
☞ a. /ɛk/				*	*
b. /ɛ.kə/			*!		
c. /ɛ/	*!				

Interestingly, this native Italian perception grammar predicts that naive Italian listeners without experience of a foreign language would perceive words in a foreign language that have final unstressed schwa categories as having no schwa (or any other vowel) in final position. In production, the Italian listeners would then realize such words with schwa-like formant structures due to the use of the same Cue constraint in the production direction (cf. Figure 1). AE listeners, on the other hand, perceive short vowel-like formants as a full vowel, see the respective constraint ranking in Table 7.

A similar ranking of *[ə]/ / above *[ᵏə]/.k/ in the Korean perception grammar accounts for the fact that Korean native listeners also perceive such a schwa as separate vowel.

---

[6]Non-velar plosives have in general weaker bursts (Dorman et al. 1977), and for them the cue of burst alone might not be sufficient to be perceived as a plosive in Italian. This could be modelled e.g. for an alveolar plosive with the following Cue constraints and their ranking:

$$*[ᵗə]/.t/ \gg *[ᵗ]/t./ \gg *[ᵗ]/ /$$

Table 7: AE

[ɛˀ_ᵏᵊ]	*[ᵏ]/ /	*[ ]/ə/	*[ᵊ]/ /	*[ᵏ]/k./	[ᵏᵊ]/.k/
a. /ɛk/			*!	*	
☞ b. /ɛ.kə/					*
c. /ɛ/	*!				

3.4 Summary of the constraint rankings

In Figures 2–4 the constraint rankings of the three languages are summarized in Hasse diagrams. In these diagrams, constraints in the upper stratum are higher ranked than those in the lower stratum to which they are directly connected by a line. For the three constraint sets that are not connected via lines, the ranking between these sets cannot be established (this holds for all three languages).
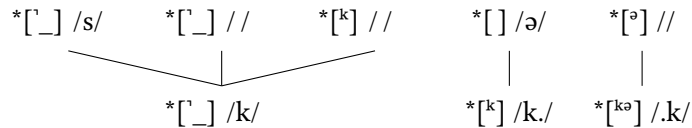

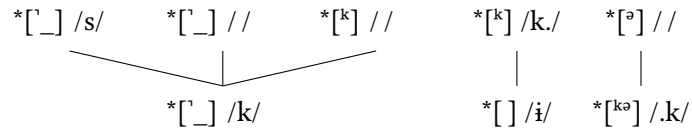Figure 2: Hasse diagram of the constraint rankings in AE.

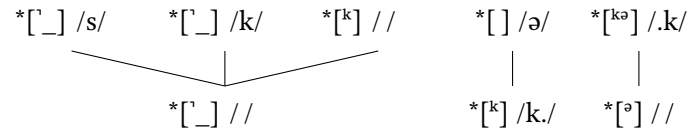Figure 3: Hasse diagram of the constraint rankings in Korean.

Figure 4: Hasse diagram of the constraint rankings in Italian.

As we can see, Korean (Figure 3) differs from AE (Figure 2) in the order of the two constraints responsible for the interpretation of the burst as coda plosive and of the silence as vowel (AE /ə/ and Korean /ɨ/), cf. the second, middle set

of constraints. Italian (Figure 4) differs from AE in the order of the constraints for the interpretation of the burst with short formants as onset plosive and of the short formants as nothing, cf. the third, right-most set of constraints, and from both AE and Korean in the order of the two constraints against interpreting formant transitions and silence cues as velar plosive or as nothing, cf. the middle of the first, left-most set of constraints.

None of the three language-specific interpretations of the two cues involved any Struct constraints, i.e. is caused by restrictions on the syllable structure in the respective languages, though the reader needs to keep in mind that phonotactic restrictions such as on possible Coda consonants could play a role in perception and are therefore part of the perception grammar, where they evaluate the phonological surface forms (see Boersma & Hamann 2009 for an illustration from Korean).[7]

Though we have treated here the set of constraints as universal in order to facilitate a cross-linguistic comparison, we assume that the cues and the phonological categories used in a language are acquired on the basis of the input that the child receives (thus through statistical inference), and are not innate.

4 Discussion and conclusion

In §3 we illustrated how listeners of the three languages American English, Italian and Korean interpret the auditory cues of burst noise and a short period of vowel-like formants differently, and how this can be formalized by means of three perception grammars that differ in the (rankings of) Cue constraints. Only a grammar model that makes a systematic distinction between phonetic/auditory form and phonological representations, and which allows for a language-specific mapping between the two, such as BiPhon, can provide such a formalisation. Grammar models without such a distinction (such as, e.g., the two-level OT models by Flemming 2001 and Steriade 2001) need to introduce extra-grammatical devices (such as Steriade's p-map) to refer to possible auditory cues, while a universal mapping (as in the three-level models by e.g. Hale & Kissock 2007 and Hale & Reiss 2000) would not allow for any differences between languages.

---

[7]A reviewer wondered whether a Cue constraint like *[ᵏə]/.k/ does not incorporate structural information as it refers to the syllable boundary. Cue constraints like these express the fact that certain cues only occur in certain positions, but are inherently different from Struct constraints like NoCoda, which are cue-independent restrictions on the phonotactic structure of a language.

4.1 Why an explicit formalization?

A reviewer asked what a formalization like we performed in §3 can buy us. This is a valid question we would like to answer in this subsection. Any kind of linguistic formalization makes explicit the knowledge a speaker/listener has acquired about their language (in line with the general aim of a linguistic model). In contrast to a simple lists of this knowledge, an explicit formalization makes use of a restricted set of tools (such as, e.g., the CUE and STRUCT constraints in BiPhon, and the exclusion mechanism of OT), which force the scientist to also consider other logical possibilities (e.g., opposite CUE constraints or alternative candidates) that are often ignored when thinking in terms of simple lists, but which are necessary to fully capture all relevant information, e.g. language-specific knowledge.

Furthermore, an explicit formalization is able to make predictions. Our formalization of the Italian perception grammar and its application to non-Italian input, e.g., resulted in the prediction that naive Italian listeners without experience with a foreign language will perceive input that contains a final unstressed schwa in the surface form of the foreign language, and thus schwa-like formant structures in the auditory form, as having no schwa (or any other vowel) in final position. This prediction needs to be tested with perception experiments, and the experimental results can then inform us about the correctness of our assumptions. If a post-plosive final schwa in e.g. English will be perceived as separate phonological unit, then we were incorrect in assuming that schwa-like formant structures only function as enhancement of the burst cue, and as a result we would have to adjust the low ranking of the CUE constraint *[ᵊ]/ /.[8] For lack of experimental evidence, we assumed (based on acoustic descriptions of Italian plosives) that Italian listeners will perceive an auditory input without a final plosive burst, i.e., [ɛ ˺], as not containing a surface plosive. Future perception experiments will need to support or falsify this assumption, too, and might again lead to a possible adjustment of the native perception grammar: If the experimental data show that our assumption was incorrect and Italian listeners rely far less on the plosive burst than we suggested, then the CUE constraint *[˺]/k/ will need to be lower ranked, and Italian naive listeners are then predicted to have little problems in perceiving Korean unreleased final plosives as plosives.

---

[8]We actually expect the perception of a surface schwa by Italians to partly depend on the place of articulation of the plosive, with a very salient velar burst resulting in more schwa perceptions (as in this case the additional schwa-like formants are less important and probably less often occurring) than for less-salient alveolar or bilabial bursts.

As we can see from this, the formalization of a perception grammar is not only informed by experimental data, but it in turn informs experiments, by creating testable hypotheses for future perception experiments.

4.2 Speech perception is not that categorical

The perception grammars we set up in §3 resulted in categorical behaviour, with only one winning candidate per auditory input. In reality, listeners exhibit more variable behaviour, as the non-categorical results of perception experiments show. This variation can be due to several factors, as mentioned already in the discussion of the three languages in §2. Linguistic factors such as preceding vowel quality, stress placement, etc., could be directly implemented with separate Cue constraints that, e.g., differentiate between the mapping of a burst cue onto a surface plosive after tense and after lax vowels. Variation that is not due to such linguistics factors could be within and across speakers (see e.g. Zhou & Hamann 2020 for an illustration of both inter- and intra-speaker variation in L2 perception).

Intra-speaker variation can be dealt with by employing Stochastic OT (Boersma & Hayes 2001): Rather than a ranking order, constraints have values on a ranking scale, where two constraints that are closely ranked are likely to switch positions at evaluation time when noise is added onto the ranking values. This could, e.g., be used to implement the variable perception of coda plosives without burst release in AE. As shown in §3.1, the perception of an auditory input without burst results in a perceived coda consonant if the constraint *[˺]/ / is high and the constraint *[˺]/k/ low ranked, as in Korean, while the reverse ranking results in a percept without a consonant, as in Italian. A close ranking of these two constraints within one grammar, where the actual ranking values have to be determined with a computer simulation and actually-observed frequency distributions (see e.g. Boersma & Hayes 2001), could then account for variation between the two forms in AE.

Inter-speaker variation, on the other hand, could be handled by different constraints rankings, i.e. different grammars. A grammar with a high-ranked *[˺]/k/ could e.g. be employed for AE speakers who show a clear preference for released plosives in coda position, while a grammar where this constraint is low ranked could account for AE speakers who prefer unreleased coda plosives.

4.3 Default interpretation of cues

We hope that our study also drew attention to the problematic nature of the term "illusory vowel" that is used to refer to a perceptual interpretation that assumedly

departs from an expected, default interpretation, namely the absence of vowel-like formant structure in the auditory form as an absence of a surface vowel category. As linguists we need to be aware of the fact that there is no default in the interpretation of auditory material, but that this interpretation is strongly influenced by and optimized for the environment we are exposed to and grew up in (as is the case with any sensory input).

There seems to be, however, a cross-linguistic preference to interpret the presence of some cues as separate phonological units, in our example short vowel-like formant structures as full vowels. This idea has been captured for L2 perception as Recoverability Principle by Weinberger (1994; discussed by Jaggers & Baese-Berk 2020: EL512), according to which salient cues are "preserved" as their own phonological entity. In terms of Cᴜᴇ constraints, this would translate into the principle that each (salient) cue that is present in the auditory signal should be mapped onto a corresponding, separate phonological category. We saw that Italian is not in line with this principle when interpreting vowel-like formant structures. The principle would also be violated in cases where one phonological category has several auditory cues, as e.g. vowel transitions, silent closure and burst all cuing a single plosive. These examples illustrate that the Recoverability Principle can be a tendency in (L2) speech perception, at most.

Acknowledgements

References

Abramson, Arthur S. & Kalaya Tingsabadh. 1999. Thai final stops: Cross-language perception. *Phonetica* 56(3-4). 111–122.

Bafile, Laura. 2002. Il trattamento delle consonanti finali nel fiorentino: Aspetti fonetici. In *Atti delle XXIII Giornate G. F. S. November 28-30*, 205–212. Pisa: ETS.

Bafile, Laura. 2003. Le consonanti finali nel Fiorentino e nel Napoletano. *Rivista Italiana di Dialettologia. Lingue, Dialetti, Società* 26. 149–178.

Bafile, Laura. 2005. Struttura sillabica e consonanti finali in varietà italiana. *Quaderni del Dipartimento di Linguistica - Università di Firenze* 15. 1–25.

Bauer, Robert S. & Paul K. Benedict. 1997. *Modern Cantonese phonology*. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110823707.

Boersma, Paul. 2007. Some listener-oriented accounts of h-aspiré in French. *Lingua* 117(12). 1989–2054. DOI: 10.1016/j.lingua.2006.11.004.

Boersma, Paul. 2009. Cue constraints and their interactions in phonological perception and production. In Paul Boersma & Silke Hamann (eds.), *Phonology in Perception*, 55–110. Berlin: De Gruyter.

Boersma, Paul. 2011. A programme for bidirectional phonology and phonetics and their acquisition and evolution. In Anton Benz & Jason Mattausch (eds.), *Bidirectional Optimality Theory*, 33–72. Amsterdam: Benjamins. DOI: 10.1075/la.180.

Boersma, Paul & Silke Hamann. 2009. Loanword adaptation as first-language phonological perception. In Calabrese Calabrese & Willem Leo Wetzels (eds.), *Loan Phonology*, 11–58. Amsterdam: Benjamins. DOI: 10.1075/cilt.307.02boe.

Boersma, Paul & Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1). 45–86.

Broniś, Olga. 2016. Italian vowel paragoge in loanword adaptation. Phonological analysis of the Roman variety of Standard Italian. *Italian Journal of Linguistics* 28. 25–68.

Busà, Maria Grazia. 2008. Teaching prosody to Italian learners of English: Working towards a new approach. In C. Taylor (ed.), *Ecolingua: The Role of E-corpora in Translation, Language Learning, and Testing*, 113–126. Trieste: Edizioni Università di Trieste.

Byrd, Dani. 1993. 54,000 American stops. *UCLA Working Papers in Phonetics* 83. 97–116.

Cavirani, Edoardo. 2015. *Modeling phonologization: Vowel reduction and epenthesis in Lunigiana dialects*. Utrecht: LOT.

Chang, Charles B. 2018. Perceptual attention as the locus of transfer to nonnative speech perception. *Journal of Phonetics* 68. 85–102.

Cook, Thomas L. 1969. Efik. In Elizabeth Dunstan (ed.), *Twelve Nigerian Languages.* 35–46. London: Longmans Green.

Crystal, Thomas H. & Arthur S. House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83(4). 1553–1573. DOI: 10.1121/1.395911.

Davidson, Lisa. 2011. Characteristics of stop releases in American English spontaneous speech. *Speech Communication* 53(8). 1042–1058. DOI: 10.1016/j.specom.2011.05.010.

Dorman, Michael F., Lawrence J. Raphael & Alvin M. Liberman. 1979. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustic Society of America* 65(6). 1518–1532. DOI: 10.1121/1.382916.

Dorman, Michael F., Michael Studdert-Kennedy & Lawrence J. Raphael. 1977. Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics* 22(2). 109–122.

Dupoux, Emanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier & Jacques Mehler. 1999. Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25. 1568–1578. DOI: 10.1037/0096-1523.25.6.1568.

Durvasula, Karthik, Ho-Hsin Huang, Sayako Uehara, Qian Luo & Yen-Hwei Lin. 2018. Phonology modulates the illusory vowels in perceptual illusions: Evidence from Mandarin and English. *Laboratory Phonology* 9. 7. DOI: 10.5334/labphon.57.

Durvasula, Karthik & Jimin Kahng. 2015. Illusory vowels in perceptual epenthesis: The role of phonological alternations. *Phonology* 32. 385–416. DOI: 10.1017/S0952675715000263.

Durvasula, Karthik & Jimin Kahng. 2016. The role of phrasal phonology in speech perception: What perceptual epenthesis shows us. *Journal of Phonetics* 54. 25–34. DOI: 10.1016/j.wocn.2015.08.002.

Escudero, Paola & Paul Boersma. 2003. Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. In Sundha Arunachalam, Elsi Kaiser & Alexander Williams (eds.), *Proceedings of the 25th Annual Penn Linguistics Colloquium (Penn Working Papers in Linguistics 8.1)*, 71–85.

Flege, James Emil & James Hillenbrand. 1987. A differential effect of release bursts on the stop voicing judgments of native French and English listeners. *Journal of Phonetics* 15(2). 203–208. DOI: 10.1016/S0095-4470(19)30548-0.

Flemming, Edward. 2001. Scalar and categorical phenomena in a unified model of phonetics and phonology. *Phonology* 18(1). 7–44.

Fowler, Carol & Jonathan Housum. 1987. Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26. 489–504. DOI: 10.1016/0749-596X(87)90136-7.

Grice, Martine, Michelina Savino & Timo B. Roettger. 2018. Word final schwa is driven by intonation: The case of Bari Italian. *The Journal of the Acoustical Society of America* 143(4). 2474–2486. DOI: 10.1121/1.5030923.

Hale, Mark & Madelyn Kissock. 2007. The phonetics-phonology interface and the acquisition of perseverant underspecification. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 81–101. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199247455.013.0004.

Hale, Mark & Charles Reiss. 2000. 'substance abuse' and 'dysfunctionalism': Current trends in phonology. *Linguistic Inquiry* 31(1). 157–169. DOI: 10.1162/002438900554334.

Hall, Nancy. 2006. Cross-linguistic patterns of vowel intrusion. *Phonology* 23. 387–429. DOI: 10.1017/S0952675706000996.

Hall, Nancy. 2011. Vowel epenthesis. In Marc van Oostendorp, Collin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, 1576–1596. Oxford: Wiley-Blackwell. DOI: 10.1002/9781444335262.wbctp0067.

Hall, Nancy. 2024. Intrusive and epenthetic vowels revisited. In Ji Yea Kim, Veronica Miatto, Andrija Petrović & Lori Repetti (eds.), *Epenthesis and beyond: Recent approaches to insertion in phonology and its interfaces*, 167–197. Berlin: Language Science Press. DOI: 10.5281/zenodo.14264542.

Halle, Morris, George W. Hughes & J.-P. A. Radley. 1957. Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America* 29. 101–116. DOI: 10.1121/1.1908634.

Hamann, Silke. 2011. The phonetics-phonology interface. In Nancy Kula, Engbert Doede Botma & Kuniya Nasukawa (eds.), *Continuum Companion to Phonology*, 202–224. London: Continuum.

Householder, F. W. Jr. 1956. Unreleased PTK in American English. In Horace Lunt, Hugh McLean & Cornelis H. van Schooneveld (eds.), *For Roman Jakobson*, 235–244. The Hague: Mouton.

Hutin, Mathilde. 2014. Can release burst be responsible for epenthesis? A case study of Korean. In *Proceedings of the 5th International Conference on Phonology and Morphology*. Chonnam National University, Gwangju.

Iverson, Gregory K. & Ahrong Lee. 2006. Perception of contrast in Korean loanword adaptation. *Korean Linguistics* 13. 49–87. DOI: 10.1075/kl.13.03gki.

Jaggers, Zachary & Melissa M. Baese-Berk. 2020. Investigating a bias for cue preservation in loanword adaptation. *JASA* 147(6). EL511–EL516. DOI: 10.1121/10.0001375.

Kabak, Bariş & William J. Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech* 50(1). 23–52. DOI: 10.1177/00238309070500010201.

Kang, Hyunsook. 1996. English loanwords in Korean. *Studies in Phonetics, Phonology and Morphology* 2. 21–47.

Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. *Phonology* 20. 219–273.

Kim, Hyunsoon. 1998. A phonetic characterization of release and nonrelease: The case of Korean and English. *Language Research* 34. 347–368.

Kim, Jungyeon. 2022. Multinomial regression modeling of vowel insertion patterns: Adaptation of coda stops from English to Korean. *Linguistics* 60(6). 1989–2017. DOI: 10.1515/ling-2021-0031.

Kim-Renaud, Young-Key. 1987. Fast speech, casual speech and restructuring. *Harvard Studies in Korean Linguistics* 2. 341–359.

Kirby, James. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41(3). 381–392. DOI: 10.1017/S0025100311000181.

Legendre, Géraldine, Yoshiro Miyata & Paul Smolensky. 1990. Harmonic grammar – A formal multi-level connectionist theory of linguistic well-formedness: An application. In *12th annual conf. CSS pod*, 884–891. Psychology Press.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge: MIT Press.

Lisker, Leigh. 1999. Perceiving final voiceless stops without release: Effects of preceding monophthongs versus nonmonophthongs. *Phonetica* 56(1-2). 44–55.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20. 384–422. DOI: 10.1017/S0025100311000181.

Lisker, Leigh & Arthur S. Abramson. 1970. The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the 6th International Congress of Phonetic Sciences*. Prague.

Martin, Samuel. 1951. Korean phonemics. *Language* 27. 519–533.

McQueen, James & Anne Cutler. 1997. Cognitive processes in speech perception. In William J. Hardcastle & John Laver (eds.), *The Handbook of Phonetic Sciences*, 566–586. Oxford: Blackwell.

Miatto, Veronica. 2020. Perception of word-final inserted vowels and syllabicity in Italian. In *Proceedings of the 11th International Conference of Experimental Linguistics (ExLing 2020)*, 137–140.

Miatto, Veronica. 2022. Between intrusivity and epenthesis: A case of Italian word-final inserted vowels. Submitted.

Miatto, Veronica, Silke Hamann & Paul Boersma. 2019. Self-reported L2 input predicts phonetic variation in the adaptation of English final consonants into Italian. In Sasha Calhoun, Paola Escudero, Marija Tabain & Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS 2019)*, 949–953. Melbourne: International Phonetic Association.

Ogasawara, Naomi & Natasha Warner. 2009. Processing missing vowels: Allophonic processing in Japanese. *Language and Cognitive Processes* 24. 376–411. DOI: 10.1080/01690960802084028.

Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33. 379–430.

Parker, Frank & Thomas Walsh. 1981. Voicing cues as a function of the tense/lax distinction of vowels. *Journal of Phonetics* 9. 353–358. DOI: 10.1016/S0095-4470(19)30969-6.

Passino, Diana. 2008. *Aspects of consonantal lengthening in Italian*. Padua: UniPress.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Joan Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 137–157. Amsterdam: Benjamins.

Polivanov, Evgenij Dmitrievič. 1931/1974. La perception des sons d'une langue étrangè. In Evgenij Dmitrievič Polivanov (ed.), *Selected works: Articles on general linguistics*, 223–237. The Hague: Mouton.

Repetti, Lori. 2012. Consonant-final loanwords and epenthetic vowels in Italian. *Catalan Journal of Linguistics* 11. 167–188.

Repp, Bruno H., Alvin M. Liberman, Thomas Eccardt & David Pesetsky. 1978. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance* 4(4). 621.

Rositzke, Harry. 1943. The articulation of final stops in General American speech. *American Speech* 18. 39–42. DOI: 10.2307/487266.

Steriade, Donca. 2001. Directional asymmetries in place assimilation: A perceptual account. In Elizabeth Hume & Keith Johnson (eds.), *The role of speech perception in phonology*, 219–250. San Diego: Academic Press.

Storto, Luciana & Didier Demolin. 2002. The phonetics and phonology of unreleased stops in Karitiana. *Annual Meeting of the Berkeley Linguistics Society* 28(1). 487–497.

Urua, Eno-Abasi E. 2004. Ibibio. *Journal of the International Phonetic Association* 34(1). 105–109.

Weinberger, Steven. 1994. Functional and phonetic constraints on second language phonology. In Mehmet Yavas (ed.), *First and second language phonology*, 283–302. San Diego: Singular.

Zhou, Chao & Silke Hamann. 2020. Cross-linguistic interaction between phonological categorization and orthography predicts prosodic effects in the acquisition of Portuguese liquids by L1-Mandarin learners. *Proceedings of Interspeech* 2020. 4486–4490.

Zhou, Chao & Silke Hamann. 2024. Modelling the acquisition of the Portuguese tap by L1-Mandarin learners: A BiPhon-HG account for individual differences, syllable-position effects and orthographic influences in L2 speech. *Glossa* 9(1). 1–39. DOI: 10.16995/glossa.9692.