

# Categorical phonotactic knowledge filters second language input, but probabilistic phonotactic knowledge can still be acquired

Language and Speech

1–27

© The Author(s) 2014

Reprints and permissions:

[sagepub.co.uk/journalsPermissions.nav](http://sagepub.co.uk/journalsPermissions.nav)

DOI: 10.1177/0023830914559572

[las.sagepub.com](http://las.sagepub.com)**Tomas O. Lentz and René W. J. Kager**

Utrecht Institute of Linguistics OTS, Utrecht University, Netherlands

## Abstract

Probabilistic phonotactic knowledge facilitates perception, but categorical phonotactic illegality can cause misperceptions, especially of non-native phoneme combinations. If misperceptions induced by first language (L1) knowledge filter second language input, access to second language (L2) probabilistic phonotactics is potentially blocked for L2 acquisition. The facilitatory effects of L2 probabilistic phonotactics and categorical filtering effects of L1 phonotactics were compared and contrasted in a series of cross-modal priming experiments. Dutch native listeners and L1 Spanish and Japanese learners of Dutch had to perform a lexical decision task on Dutch words that started with /sC/ clusters that were of different degrees of probabilistic wellformedness in Dutch but illegal in Spanish and Japanese. Versions of target words with Spanish illegality resolving epenthesis in the clusters primed the Spanish group, showing an L1 filter; a similar effect was not found for the Japanese group. In addition, words with wellformed /sC/ clusters were recognised faster, showing a positive effect on processing of probabilistic wellformedness. However, Spanish learners with higher proficiency were facilitated to a greater extent by wellformed but epenthesised clusters, showing that although probabilistic learning occurs in spite of the L1 filter, the acquired probabilistic knowledge is still affected by L1 categorical knowledge. Categorical phonotactic and probabilistic knowledge are of a different nature and interact in acquisition.

## Keywords

Speech perception, phonotactics, word recognition, cross-modal priming, second language acquisition

## Introduction

Second language (L2) learners have to face a ‘Catch-22’ situation that is potentially detrimental to successful acquisition of L2 phonotactic knowledge. They acquire a language while already in

---

### Corresponding author:

Tomas O. Lentz, Utrecht Institute of Linguistics OTS, Trans 10, 3512 JK Utrecht, Netherlands.

Email: [t.o.lentz@uu.nl](mailto:t.o.lentz@uu.nl)

possession of knowledge of another, namely their first language (L1). Details of the input that are not relevant to L1 phonological contrasts are filtered out (as will be explained below). This study is an empirical investigation into the L2 acquisition of phonotactic features that are filtered out by L1 knowledge. The first factor relevant to the question is the nature of the phonotactic filter on perception; the second factor is the nature of phonotactic acquisition. These will be discussed in this order below.

## 1.1 Phonotactic filters on perception

*1.1.1 How phonotactics affects perception.* The existence of phonotactic filters on perception was first proposed by Polivanov (1931), who reported anecdotal evidence of a Japanese native speaker's misreproduction of the Russian word /tak/, pronounced by Polivanov himself. The Japanese participant reproduced the word as [taku]; Polivanov claimed that the mismatch was not a misproduction of /tak/ as [taku], but rather a misperception of his own [tak] as /taku/ followed by then faithful production. The misperception is the result of the illegality of a /k/ in a coda position in Japanese; a Japanese listener therefore perceives [tak] as its closest legal alternative /taku/ instead of /tak/. In other words, the listener repairs the input to accommodate phonotactic constraints on what is acceptable as a legal word form.

Phonotactic filtering effects on perception were further explored and confirmed in many experiments under controlled conditions. Massaro and Cohen (1983) showed that perception is biased by phonotactic context: an ambiguous sound that is phonetically between English /l/ and /r/ is more likely to be perceived as the option that is phonotactically legal, as /l/ after /s/ and as /r/ after /t/. This finding can be interpreted as the effect of a filter against the illegal combinations \*/sr/ and \*/tl/. In another experiment by Massaro and Cohen (1983), English-speaking participants were asked to identify a two-phone sequence as 'dl', 'dr', 'bl' or 'br'. The first phone was ambiguous between /b/ and /d/ and the second between /l/ and /r/. Participants were less likely to perceive the sequence as the (illegal) cluster 'dl' than as one of the other (legal) clusters: the answer 'dl' was given less frequently than could be expected on the basis of the frequency of /d/ and /l/ answers when participants were given the legal combinations 'dr' and 'bl'. This bias against the 'dl' answer can be explained as application of knowledge that \*/dl/ is unlikely (or illegal).

Hallé, Segui, Frauenfelder, and Meunier (1998) used a gating experiment to show that French listeners' perceptions of consonant combinations are biased towards phonotactic legality for [tl] and [dl] stimuli. The first consonant is perceived faithfully when listeners hear the sound up to but excluding the /l/, but the perception of this same first consonant changes to a velar sound when given in fragments that include the /l/. The velar combinations /kl/ and /gl/ are legal in French, while \*/tl/ and \*/dl/ are not.

*1.1.2 Phonotactic perceptual effects can be language-specific.* The above experiments address the influence of phonotactics on speech perception, but they do not necessarily entail a problem for L2 acquisition. Though unlikely, the effects discussed above could have been caused by a general (universal) difficulty to perceive certain combinations. Such a universal difficulty might be the result of general or universal abstract phonotactic wellformedness differences that originate in an innate part of the language faculty; universal phonotactic wellformedness was advocated for by Berent, Steriade, Lennertz, and Vaknin (2007), who reported that both English and Russian native listeners have more misperceptions of phoneme combinations that are universally more marked, as compared to universally less marked combinations. All combinations were legal in Russian, but illegal in English. In contrast, Peperkamp (2007) argued that such universal tendencies are caused by phonetic difficulties in perceivability or producability of the phoneme transitions. For the

present study, phonotactic effects that are universal are not relevant, because transfer of universal knowledge from L1 to L2 cannot be detrimental to L2 acquisition, by virtue of the simple fact that universal knowledge should also hold for the L2.

However, phonotactic effects on perception that are specific to the native language of the perceiver do exist. Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) reported a double dissociation that provides evidence for the language-specificity of phonotactic influences on perception. They tested the perception of phoneme combinations that are absent from French but not from Japanese, as well as the other way around, with Japanese and French participants. The difference between a consonant cluster such as /bz/ in /ebzo/, which is illegal in Japanese but not in French, and a variant with a vowel between the consonants such as /ebuzo/, which is legal in both languages, was harder to distinguish for Japanese than for French participants. Vice versa, a vowel length contrast such as that between /ebuzo/ and /ebu:zo/, phonemic in Japanese but absent from French, was easier to distinguish for Japanese than French participants. This dissociation shows that phonotactically induced perceptual illusions can be language-specific, reflecting phonological illegality and not just phonetic properties of the misperceived phoneme combinations. Jacquemot, Pallier, LeBihan, Dehaene, and Dupoux (2003) found a neurological correlation of the same language-specific perceptual illusions: different brain regions are active in the processing of legal as opposed to illegal structures (see also Moro et al., 2001). In addition, Dupoux, Pallier, Kakehi, and Mehler (2001) showed that the process cannot be explained by a perceptual pull caused by lexical items (a Ganong effect, see Ganong, 1980); even if a legal phonotactic combination is normally present in many lexical items and an illegal combination in none, misperception of illegal phoneme combinations does not always end up at the nearest lexical item. For example, [mikdo] was perceived by Japanese listeners as the nearest legal non-word /mikudo/, not as the nearest word /mikado/. Hence, listeners applied abstract knowledge of the illegality of /kd/, not lexical knowledge.

Summarising, language-specific knowledge of phonotactics affects perception. This influence has the potential to cause misperceptions when applied to a language with different phonotactics. The question this article addresses is whether L1 filtering affects L2 acquisition. It is important to note that predicting L1 filters for every phoneme combination absent from an L1 will not be correct: some unattested combinations are perceived more successfully than others, depending on the phonological property they violated; Kabak and Idsardi (2007) showed this for different consonant clusters unattested in Korean – see also Berent et al. (2007), discussed above. In addition, as pointed out by Albright (2009), the unattestedness of sequences, such as the absence from English of the triphone /esp/ at the end of a word, is not a sufficient reason to consider such sequences as phonologically marked or illegal; /esp/ is perfectly acceptable to English listeners and one would not expect it to be filtered out by perceptual illusions resulting in triphones that are attested at word offsets such as /isp/ or /esk/ (as in ‘lisp’ and ‘desk’).

The absence of certain long phoneme sequences might not be salient, because the frequency of longer sequences is generally lower than the frequency of shorter ones. A zero frequency for a long sequence is therefore less surprising and possibly accidental (Pierrehumbert, 2001b), which means more input is needed to determine if absence is accidental for longer sequences than for shorter sequences. However, listeners sometimes show surprisingly specific phonotactic knowledge: Kager and Pater (2012) showed that Dutch listeners have implicit knowledge of the relative paucity of sequences of a long vowel and two consonants of which the second is not a coronal. This knowledge cannot be reduced to low wellformedness of the shorter component combinatons.

In short, phonotactic filters on perception exist, but not against all unattested phoneme combinations of a language. Hence, filtering effects cannot be explained only by lack of practice with perceiving unattested combinations; rather, filtering depends on the listener’s language-specific phonotactic knowledge. Unfortunately, theoretical accounts of abstract phonotactic representations

and phonotactic wellformedness are not conclusive. They diverge on the source of categorical and gradient differences between unattested structures, as well as on the calculations or formalisations involved (for the ongoing discussion, see Coleman and Pierrehumbert, 1997; Berent et al., 2007; Hayes and Wilson, 2008; Kabak and Idsardi, 2007; Coetzee, 2009; Albright, 2009; Berent, Lennertz and Balaban, 2011). In conclusion, phonotactic filters cannot (yet) be predicted. In the absence of a failsafe prediction, the existence of specific filters, while possibly suspected on theoretical grounds, needs to be corroborated by empirical data.

*1.1.3 Acquisition of L2 phonotactic wellformedness.* Theoretically, if L2 acquisition occurs on input that has passed through the L1 phonotactic filter, the filter can be too effective to be overcome. As an illustration, consider the L1 Japanese participants in the experiments by Dupoux et al. (1999). Compared to French listeners, these listeners (who did not speak French) have problems perceiving consonant clusters such as /bz/ in [ebzo]; Japanese phonotactic knowledge changes consonant clusters (by adding an illusory epenthetic vowel in the middle, leading to /ebuzo/), deleting them in the perceptual process. If an L1 Japanese listener tries to acquire French with input that passed through the L1 filter, they might not learn any L2 probabilistic phonotactics on consonant clusters.

However, the scenario sketched above is at odds with experimental evidence on L2 phonotactics in perception. At least two studies show that successful acquisition of L2 wellformedness of L1 illegal clusters is possible. Trapman and Kager (2009) found that Spanish learners of Dutch can acquire the higher wellformedness of biconsonantal onsets that are legal in Dutch, as compared to onsets that are unattested in Dutch, even though all onsets were unattested in Spanish. Early-stage learners did not assign significantly higher wordlikeness ratings to non-words with Dutch-legal onsets, while advanced learners did, indicating that successful learning had taken place. The advanced learners were also faster at rejecting non-words with Dutch-unattested onsets in a lexical decision task, showing again that they noticed the mismatch with Dutch phonotactics and hence that they had acquired Dutch wellformedness. Boll-Avetisyan (2011) looked at the L2 acquisition of Dutch phonotactics by L1 Japanese and L1 Spanish learners. She used a short-term memory recall task with non-words that were phonotactically legal in Dutch but differed in the complexity of their syllable structure as well as in the probability (frequency) of the biphones of which they were composed. The results suggested that Dutch gradient phonotactic differences made the listeners faster at recalling words with high biphone frequencies, especially for words with complex syllable structures (CVCC and CCVCC) that are legal in Dutch but unattested and probably illegal in Japanese or Spanish. The results of both studies suggest that L1 unattestedness does not block access to L2 gradient wellformedness. However, it is possible that the L1-unattested structures were not subject to filtering out by L1 phonotactic knowledge.

Although it addresses the acquisition of L2 illegality instead of L2 wellformedness, a third relevant study on L2 phonotactic acquisition in perception is that of Weber and Cutler (2006). They showed that even advanced L1 German learners of L2 English are likely to assume a word boundary between not only two consonants that do not legally combine in English but do in German (e.g., /ʃl/), but also between consonants that do not legally combine in German but do in English (e.g., /sl/). Hence, these learners acquired knowledge of L2 phonotactic illegality, but did not 'unacquire' L1 illegality and continued using L1 illegality knowledge in their L2 perception. These results indicate that knowledge of L1 phonotactic illegality affects L2 perception for advanced learners. Of course, using knowledge of illegality in detecting word boundaries does not imply that the illegal phoneme combinations will always be filtered out, nor do these results rule out the possibility of unlearning L1 filters. They do suggest, however, that acquiring the L2 legality of combinations illegal in the L1 is not easy. German learners continued using L1 knowledge of German

illegality in their English perception even when they had already acquired knowledge of English (L2) illegality. In this case, unlearning of L1 knowledge (or learning of L2 legality for the L1 illegal consonant clusters) is harder than acquiring L2 illegality. Note that the acquired English illegality was not at odds with the German phonotactic knowledge, but complementary; if there is a German L1 filter, it would not have blocked acquisition of the English illegality because the combinations that were illegal in English are legal in German.

The three studies discussed above indicate that L1 filtering is not always a problem for L2 wellformedness acquisition. L2 acquisition of illegality and of gradient phonotactic wellformedness is possible. However, the studies do not answer the question about whether knowledge of L1 illegality, if it affects perception, also filters L2 input to such a degree that gradient phonotactic wellformedness cannot be acquired. A related question is whether gradient wellformedness is of the same nature as knowledge of categorical illegality. Gradient wellformedness typically has facilitatory effects such as faster processing, whereas illegality has negative effects such as filtering and misperception. The discussion above on the L1 filter might suggest that illegality has categorical effects on perception. Nevertheless, non-categorical phonotactics can also affect perception and effects of phonotactics do not have to be negative, but can also be positive (facilitatory).

Phonotactic knowledge can facilitate perception, because highly frequent phoneme combinations are more easily processed. Facilitating effects of high phoneme combination frequency have been found for a number of speech perception tasks. High frequency can be seen as a correlate of wellformedness (or vice versa; for a discussion, see Boersma, 1997; Coetzee, 2008, 2009). Vitevitch and Luce (1998) found that in tasks involving the lexical level, such as word repetition, an increase in phonotactic wellformedness of the stimuli occurs with a decrease in ease of processing. This decrease is caused by a confound variable, namely lexical neighbourhood density, which correlates positively with phonotactic probability but has a negative effect on word recognition due to competition of neighbours with the actually produced word. However, when nonwords or non-lexical tasks are used, this lexical effect is no longer decisive and phonotactic wellformedness emerges as facilitating processing. Luce and Large (2001) found the same facilitation of phonotactic wellformedness for words, when lexical neighbourhood density was controlled. Frisch, Large, and Pisoni (2000) showed that phonotactic wellformedness (in various measures) positively affects wordlikeness judgements. Bailey and Hahn (2001) confirmed this finding, but found that lexical neighbourhood density was more decisive than phonotactic wellformedness. In summary, although gradient phonotactic wellformedness is correlated with lexical neighbourhood density which can sometimes obscure its facilitatory effects, it facilitates speech processing. Highly wellformed units in the input are more likely to be perceived quickly and to be linked to lexical entries. In addition, wellformed phoneme combinations are also more likely to be registered in short-term memory (Gathercole, Frankish, Pickering, & Peaker, 1999; Boll-Avetisyan, 2011). Note that in the studies discussed above, phoneme combinations are usually hypothesised to be wellformed based on their high frequency. Even if this theoretical assumption is a simplification (Hayes and Wilson, 2008, Albright, 2009), gradient facilitatory effects of phonotactic wellformedness have been empirically confirmed. This positive effect of phonotactics on processing contrasts with the negative (filtering) effect of categorical phonotactic illegality.

Phonotactic effects on perception do not have to be categorical even when they are negative, that is, based on knowledge of illegality. The filtering of contrasts by phonotactic illegality is not necessarily categorical. Berent et al. (2007) suggested that there are gradient differences in illegality and that the filtering effect is relative to the degree of illegality. In fact, closer scrutiny of the experimental data on phonotactic filtering shows that participants are not categorically incapable of attending to phonetic differences that are not phonological. The Japanese listeners of Dupoux et al. (1999) could distinguish [ebzo] from [ebuzo] to some degree; even though they spotted /u/ in

the /u/-free production [ebzo] more often than French participants did, they did not perceive a /u/ all of the time: the /u/ was perceived in slightly over 70% of the cases. This is important for L2 acquisition, because if these listeners were to start learning a language with consonant clusters, they would on average perceive 30% of the consonant clusters faithfully as clusters.

Berent et al. (2011) showed that the nature of the experimental task matters for the ‘categoricalness’ of phonotactics-driven misperceptions, a finding that is potentially important to L2 acquisition. This study shows that in the case of phonotactic perceptual biases against illegal clusters, distortions in perception at one level do not mean that faithful information is not available at a lower level. The higher level is called ‘phonological’ by Berent et al. (2011) and the lower level ‘local phonetic’. In tasks assumed to be phonological, because they relate to the number of syllables, participants had more trouble distinguishing a less wellformed consonant cluster, namely [md], from an epenthesised version of the same cluster, ([məd]), as compared to distinguishing a more wellformed cluster from an epenthesised version, ([ml]–[məl]). This suggests that there is a perceptive bias caused by universal phonotactic knowledge of the wellformedness of the sonority profile of the clusters in question. In contrast, tasks assumed to be of a phonetic nature cause fewer misperceptions. When participants were asked to detect the presence of a phoneme, such as a schwa between two consonants that form an illegal cluster without this schwa, an effect of cluster wellformedness was not observed. Although absence of evidence does not prove evidence of absence, it is to be noted that mean error rates did not reach 50% in any condition. Thus, even if the study failed to reveal an actually existing difference in ease of phonetic processing between illegal phoneme combinations, a considerable amount of the information at the local phonetic level is faithful to the input. The L1 filter therefore does not necessarily stop input-based learning of L2 wellformedness.

To return to the example of a Japanese learner of French, whose L1 knowledge filters out consonant clusters, the question is whether such a learner can become better at processing those clusters that are more frequent in French. The theoretical possibility of exemplar-based learning of L2 gradient phonotactics must also be considered. According to exemplar theory, word forms are not necessarily completely normalised to a phonological (phonemic) representation (with L1 filtering in the case of L2 words) and episodic information can also be retained (Goldinger, 2003). Most exemplars in the learner’s mind are bound to be exemplars provided by native speakers, potentially allowing correct unfiltered L2 acquisition. However, it remains to be seen whether episodic information is indeed used for L2 acquisition. Although Goldinger (1998, 2003) argues that there are episodic effects on (repeated) word recognition, he does not deny effects of normalisation – that is, of representations containing abstractions. The same is argued by Pierrehumbert (2001a, p.139), who also states that ‘(...) the correct model must describe the interaction of word-specific phonetic detail with more general principles of phonological structure’. In brief, the fact that faithful input might be available at the phonetic or acoustic level provides a potential source for L2 phonotactic acquisition even if an L1 filter occurs at another level, but the question remains whether this potential source is indeed employed.

The existing evidence is not clear enough to decide whether knowledge of phonotactic illegality and knowledge of gradient wellformedness are of the same nature, or whether the filtering effects of illegality affect the input to all phonotactic learning. Both possibilities need to be considered, therefore. Second language acquisition in fact provides an excellent situation in which to determine whether knowledge of phonotactic illegality and of probabilistic wellformedness are of the same nature, because if they are it is impossible for a phonotactically illegal phoneme combination also to be wellformed. Within a single language, knowledge of illegality cannot be in conflict with knowledge of wellformedness, because one language cannot provide evidence for contradictory knowledge of the type ‘combination /xy/ is illegal, hence unattested’ and ‘combination /xy/

is wellformed, hence frequent/overrepresented'. However, this is not true for a combination of languages. If phonotactic knowledge of a first language is transferred to the second language, illegality knowledge applied to the L2 can be based on L1 experience, but gradient wellformedness on L2 input, in principle allowing the representation of contradictory knowledge.<sup>1</sup>

If illegality knowledge is of a different nature to that of gradient wellformedness, and categorical illegality does not block wellformedness acquisition, discrepant acquisition can occur. Equally, if illegality knowledge is actually of the same nature as knowledge of gradient wellformedness, acquisition of gradient wellformedness implies previous weakening and unlearning of phonotactic filters at odds with the wellformed structures. If an L1 filter can be unlearned, caution is needed if the L1 filtering effects are considered to be the result of strong, but gradient perception biases. If what is called 'illegality' is in fact just extremely low wellformedness, 'illegal' combinations are only highly likely to be misperceived, just as very wellformed phoneme combinations are extremely easy to be perceived. Misperception is then highly likely for 'illegal' phoneme combinations, but does not occur all the time, and acquisition of L2 phonotactic wellformedness is possible with brute force. After enough L2 input of phoneme combinations not allowed in the L1, some of this input must have survived the L1 bias. L2 learners then have a hold on the L1 phonotactic filter and can unlearn it, after which they can also acquire L2 gradient wellformedness differences between phoneme combinations that were first subject to the filter.

## 1.2 Summary and predictions

Summarising the scenarios for phonotactic learning and their predictions, two types of L2 phoneme combinations have to be considered. The types should differ on wellformedness in the L2, so that there is L2 wellformedness to learn (namely the difference), but both types must be filtered out by L1 knowledge. For these two types of L2 combinations, the following different acquisition scenarios can be identified:

1. Full Blocking: The L1 filter blocks all acquisition on the L2 combinations, by removing evidence for their legality and wellformedness/frequency from the L2 input before it reaches the stage of learning. This removal also effectively makes unlearning of the L1 knowledge underlying the filter impossible. This scenario does not imply that knowledge of illegality and probabilistic phonotactics are of the same nature, nor the opposite.
2. Leaky Filter: The L1 filter removes contrasts between the L2 structures and L1 repairs, but this is not so relevant to acquisition of L2 input as to prohibit the acquisition of L2 wellformedness. This scenario does imply that knowledge of illegality and probabilistic phonotactics are not of the same nature.
3. Bias Only: The L1 filter is in fact only a strong bias to perceive wellformed structures. Sufficient input containing the L2 combinations will lead to adaptations of phonotactic knowledge and hence make the filter disappear, allowing acquisition of L2 wellformedness. This scenario implies that knowledge of illegality and probabilistic phonotactics are of the same nature.

In the last two scenarios, the L1 filter is not so categorical that it blocks L2 wellformedness acquisition. If the filter only biases perception to the most likely forms, the least wellformed phoneme combinations are the least likely to be perceived. Gradient filtering behaviour also allows learning L2 phonotactics simply by attending to frequency of occurrence: a percentage of the L2 phoneme combinations will be perceived faithfully, allowing to start learning L2 phonotactics knowledge. The difference between the last two variants is thus that the L1 filter exists in scenario 2, but

**Table 1.** Predictions for different scenarios of phonotactic learning.

Scenario	Relation filter and wellformedness acquisition
1. Full Blocking	Filter → no acquisition
2. Leaky Filter	No relation between filter and acquisition
3. Bias Only	Acquisition → 'filter' disappears

separately from gradient wellformedness knowledge, while in Scenario 3, 'Bias Only', the filter does not exist separately from gradient wellformedness and the effects attributed to it are not categorical. The predictions for L2 phonotactic acquisition in each scenario are those as shown in Table 1 (arrows indicate 'if-then' implications). Note that if a filter or acquisition are observed, scenario 1, resp. 3, make a falsifiable prediction; if both are observed, only Scenario 2 fits.

In the operationalisation of prediction testing developed below, the L1 of L2 learners will be called Source Language and their L2 will be called Target Language, so that L1 and L2 listeners of the Target Language can be identified without confusion. As mentioned above, both the existence of the L1 filter and the existence of L2 differences in wellformedness between the aforementioned two types of phoneme combinations must be empirically established. For this purpose, a class of combinations filtered out in one language needs to be found; this language is then the Source Language. To find Target Language wellformedness at odds with the Source Language, a language with different types of wellformedness for phoneme combinations in the class just identified is also needed. The two types of wellformedness will be referred to as Neutral and Good, indicating that both are legal in the Target Language, but note that both are illegal in the Source Language.

Perceptual illusions caused by the Source Language filter reduce the contrast between these combinations and their nearest Source Language repairs. Gradient wellformedness causes faster lexical activation for words with Good phoneme combinations, in contrast to words with Neutral combinations. If a Source Language filter is observed, the three scenarios make different predictions about the acquisition of the Target Language wellformedness difference. In Scenario 1 'Full Blocking', the Neutral and Good combinations are not perceived and no acquisition takes place. In Scenario 3 'Bias Only', exposure to high numbers of Good clusters allows acquisition of their higher wellformedness (compared to Neutral clusters), but this implies that legality of the clusters in the Target Language is also acquired, because inconsistent knowledge of the Good clusters as being illegal and wellformed is not possible in this scenario. Hence, the effects of the (apparent) L1 filter are then predicted to disappear.

If the L1 filter is found to disappear, but L2 wellformedness is not acquired, Scenario 3 'Bias Only' is the most likely to be correct, but not completely confirmed: one cannot be certain whether L2 wellformedness is going to be acquired or, alternatively, that its acquisition is blocked by other unknown factors. Because Scenario 2 'Leaky Filter' does not predict a relation between the two types of acquisition, it would also fit a disappearing L1 filter without L2 wellformedness acquisition, but this scenario would really only be confirmed by a filter that continues to be active when L2 wellformedness is acquired.

Testing all of the above simultaneously requires using one set of experimental items and a common experimental methodology, to avoid the interpretation that any unexpected difference could be caused by a task or item effect (again, see Berent et al., 2011). The phonotactic wellformedness experiments and the perceptual illusions experiments mentioned above tapped into phonotactics at a different locus, namely in speech processing and in metalinguistic judgements respectively. Because speech processing ultimately maps input to actual words, looking at the effect of



**Table 2.** Ranking of sC-clusters according to *O/E*.

Cluster	Type	Observed	Expected	<i>O/E</i>
[sp]	good	5573	1783	3.12
[sx]	good	10762	3596	2.99
[st]	good	32173	15840	2.03
Total	good	48508	21219	2.28
[sm]	neutral	3548	3541	1.00
[sl]	neutral	3667	3304	1.11
[sn]	neutral	3742	3136	1.19
[sk]	neutral	2703	3241	0.83
Total	neutral	13660	13222	1.03

Note. *O/E* (Observed over Expected) ratios are calculated using frequencies of the phonetically annotated part of the Corpus of Spoken Dutch (CGN; Oostdijk, 2000). The values are based on the frequencies of the phonemes in consonant biphones.

phonotactic wellformedness in word recognition is preferable. L1 filter effects can be observed by the activation of a Target Language word by a Source Language version of the word, when this repaired version is presented to L2 learners of the Target Language. The priming effect of the repaired version can be compared to a baseline provided by L1 listeners of the Target Language. L2 priming has already been successfully employed by Broersma and Cutler (2008), who used priming to show that voicing in obstruents at the end of English words is not contrastive to L1 Dutch listeners.<sup>2</sup>

### 1.3 Case: Dutch /sC/ clusters

Dutch is the Target Language for the present study and the wellformedness contrast used is between two sets of legal Dutch consonant clusters starting with an /s/, referred to as /sC/ clusters hereinafter. The rationale behind this choice, apart from the practicality of availability of native Dutch participants and learners of Dutch, is that Dutch has a range of /sC/ clusters that can be hypothesised to have different degrees of wellformedness, reflected by their corpus frequency. In addition, /sC/ clusters are unattested in many languages, including Spanish and Japanese. Neither of these latter languages allows /sC/-clusters and they can (on the basis of loanwords and misproductions) be hypothesised to repair them as /esC/ and /suC/ respectively. These two languages are therefore used as the Source Languages in the present study. When Source Language listeners to Dutch suffer less from the epenthesis than native listeners, the hypothesised language-specific perceptual illusion is indeed present.

Dutch /sC/ clusters were assessed on their frequency in spoken, continuous Dutch to obtain an indication of gradient wellformedness differences.<sup>3</sup> Wellformedness is assessed with Observed over Expected (*O/E*) ratios (Pierrehumbert, 2003), calculated over consonant clusters in the *Corpus of Spoken Dutch* (CGN) (Oostdijk, 2000). *O/E* ratios are calculated by dividing the observed frequency *O* by the estimated frequency *E*. *E* is calculated by multiplying the probability of both components and the total number of tokens. This number (*E*) reflects how often two phonemes would be observed next to each other if there was no preference at all for a combination; if indeed there is no preference for a combination, *O* should equal *E*, yielding an *O/E* ratio of one. If the frequency of combinations can be predicted from the frequency of the constituent phonemes, *O/E* is equal to one, but this is not the case for many Dutch sC-clusters, as can be seen in Table 2 of all Dutch /sC/ clusters that occur word-initially in non-marginal quantities. In the case the *O/E* ratio of

some combination is higher than one, the combination is overrepresented, which is an indication of wellformedness. The group of clusters labelled Good contains clusters with a high *O/E*, while a group called Neutral contains clusters that occur roughly as much as expected if consonants in clusters were randomly combined; that is, that have an *O/E* of 1.

## 2 Method

### 2.1 Participants

Six groups of participants were recruited for all experiments reported in the present article. L2 participants were recruited using flyers, networks of previous participants and institutions that teach Dutch as a foreign language. These participants were either all involved in Dutch society (using the Dutch language regularly) or were still actively learning Dutch. Thirty-six participants, the Spanish group, had Spanish as their L1. The Spanish participants had been in the Netherlands for 7.6 years on average (*SD* 7.7, range 1–30) and rated their own listening proficiency in Dutch at 4.8 (*SD* 2.0) on a seven-point scale.

Thirty-three other participants, the Japanese group, had Japanese as their L1 and received a version of the experiment in which the epenthesis was of the [suC] type. The Japanese participants had on average been 10.3 years in the Netherlands (*SD* 8.5, range 1–24) and rated their listening proficiency in Dutch at 3.7 on a seven-point scale (*SD* 1.2).

As a native perception baseline, 71 native speakers of Dutch were recruited from the Utrecht Institute of Linguistics participant pool. Of these, 37 received the same version as the Spanish group and are referred to as the Native-Es group; 34 received the Japanese version and are referred to as the Native-Su group. Two control groups for L2 learners consisted of 31 participants with Dutch as L2 but an L1 that allowed /sC/ clusters.<sup>4</sup> This group is called Other-L1. Members of this group had on average been in the Netherlands for 2.2 years (*SD* 2.1, range 1–10) and rated their listening proficiency in Dutch at 4.5 (*SD* 1.4). Of the Other-L1 group, 17 participants received the es version and 14 the su version.

All participants reported not having problems with hearing stimuli over headphones nor with reading from a computer screen. Participants were paid a modest amount for their participation, which lasted up to one hour, because they also completed a questionnaire and a written language proficiency test (C-test).

### 2.2 Materials

The experimental items were 60 words starting with /sC/. Two types of fillers were employed. Word fillers, 60 words not starting with /sC/, prevented targets with /sC/ being the only words. Non-word fillers, 60 non-words starting with /sC/ and 60 non-words not starting with /sC/ were used to maintain the same balance in the trials where the correct answer was 'no'. Thus, properties of the prime did not give information about the correct answer to the trial over the total set of 240 trials. The experimental items were nested in the Phonotactic condition, with 30 words starting with Good clusters and 30 with Neutral clusters. Most words were monosyllabic; none had more than two syllables. All words were singular nouns or infinitives, selected from the most frequent words starting with sC clusters. Word frequencies were estimated by the equally weighed average of the frequency in the CGN Corpus of Spoken Dutch (Oostdijk, 2000) and the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). The weighted average frequency ensured both databases had equal influence on the frequency measure, in spite of their different size. Because the most frequent 30 words with Neutral clusters were less frequent than those with Good clusters,

there was a word at the Good level for every word at the Neutral level that was picked to match its frequency as close as possible. It was not possible to strictly control for lexical neighbourhood density, but the average number of neighbours for the Good items was 13.24, while it was 11.30 for the Neutral items. This difference was not significant in a Welch two-sample *t*-test,  $t(52.79) = 1.62$ ,  $p = 0.1091$ . The 50/50 balance between Good and Neutral clusters in the experimental items was also used for filler words and filler non-words starting with an /sC/ cluster. All non-words were phonotactically legal in Dutch.

The experimental items were used as visual targets and preceded by auditory stimuli of three Priming types: Unrelated, Epenthesised or Faithful. In the first case, the prime was a different word, not related to the target (and hence not expected to actually prime the target). In the last case, the stimulus was a correct pronunciation of the target word, expected to prime it. The crucial Epenthesised condition allowed the perceptual illusions of L2 listeners caused by L1 filters to be tested; the prime was manipulated with the insertion of an epenthetic vowel, hypothesised to correspond to the L1 perceptual illusions of Spanish and Japanese. For Japanese, a cluster starting with /s/ is assumed to be transformed by adding a /u/ after the /s/. This manipulation will be referred to as the *su* manipulation. In Spanish, a short vowel [e] is commonly present before the cluster (and inserted in loanwords in absence of another vowel). Because Dutch does not have a short /e/, but a long /e:/ and a short /ɛ/, the latter was used; it is not contrastive with Spanish /e/ because Spanish has only the five vowels /a/, /i/, /e/, /o/ and /u/. This manipulation will be referred to as the *es* manipulation.

The Unrelated and Faithful primes were recorded by a native speaker of Dutch who was unaware of the purpose of the study. Epenthesised primes were constructed using the Faithful recordings of the targets. The [s] of these recordings was deleted and the remainder was pasted after a fragment of [ɛs] or [su] carved out of other words pronounced by the speaker, in such a way that the transition was as natural as possible. The manipulation was performed using PRAAT (Boersma, 2001), combining two fragments at zero-crossings of the amplitude wave, conserving the amplitude gradient. For the Faithful primes, the same splicing was performed, now pasting a production of [s] spliced from another word, so that these primes, as far as possible, matched the epenthesised primes' possible unnaturalness at the splice point. Three native speakers of Dutch, all linguists, confirmed the naturalness of all items when taken as (fragments of) Dutch. The duration of the [ɛ] was around 100 ms; the [u] was around 80 ms.

Visual word recognition entails having orthographic influences, normally mirroring the phonotactics (Bailey and Hahn, 2001). Dutch orthography does indeed closely follow the phonemes of the clusters, except for /sx/, which is written 'sch'.

The 60 experimental items were divided in three groups of 20 and each participant received a version in which one of the groups was presented with Unrelated primes, one with Epenthesised primes and one with Faithful primes; the group assignment was rotated in a Latin Square design. The three groups of 60 filler items described above were each also divided in three groups of 20 that were assigned to Unrelated, Epenthesised or Faithful primes, which were fixed and not rotated.

## 2.3 Procedure

The test was administered in a sound-isolated cabin at the university laboratory, or in a quiet room at the participant's home, using a computer running Ubuntu LINUX 6 with Xenomai real-time support. Participants were (individually) seated by the experimenter and received instructions according to a fixed protocol. They were told that they would hear a Dutch word or non-word and see a Dutch word or non-word on the screen, possibly different. They were instructed to decide whether the word on the screen was a real Dutch word as quickly as possible. Responses were to

be given by pressing the 'yes' or the 'no' button on a button box, which were marked with the corresponding Dutch words. The yes button was placed at the side of the dominant hand as reported by the participant. Participants were told they would occasionally be asked to write down the last auditory stimulus (this happened four times, after every 60 trials), in order to avoid them stopping to pay attention to the auditory primes. The written stimuli provided for checking as to whether or not they were listening to the auditory stimuli, but the results showed virtually no errors, only spelling oddities, mainly in non-words. The process of writing down also allowed the participants to take a brief, self-timed break. No information was given about the nature of words used in the experiment. Three practice trials were presented and repeated if necessary until participants showed understanding of the procedure.

Auditory stimuli were presented over Beyerdynamic DT250/80 headphones; visual targets on a 43cm computer screen. After the auditory stimulus finished playing, there was a 500 ms interval, after which a fixation cross (+) was shown for 500 ms in a white background to announce the appearance of the target. The target was shown for 750 ms; participants had 2500 ms from the start of the target presentation in which to respond. The words were shown in lower case black letters on a white background. The targets were always visible during the answer window. After every trial, feedback was presented on screen ('correct', 'wrong', or 'too late' in Dutch).

The order of the 240 trials (60 experimental trials, 60 with filler words and 120 with filler non-words as targets) was rerandomised anew for each participant. Epenthised primes were either of the form [esC] (es version) or [suC] (su version). In order not to confuse the type of epenthesis with the native language of the participants, both the native language and the version of the experiment were collapsed into one condition Group. Two Groups of Dutch native listeners received either the Spanish or Japanese epenthised primes; these groups are called Native-Es and Native-Su. Participants with Dutch as a second language were given the epenthesis associated with their own native language. These Groups are therefore simply called Spanish and Japanese. The control participants with other first languages than Japanese and Spanish were randomly assigned to the OtherL1-Es or OtherL1-Su groups, which received the indicated type of Epenthised primes.

After the actual experiment, participants filled in a questionnaire about their language backgrounds and took a C-test to assess their general proficiency level in Dutch (for a discussion on the validity of C-tests, see Eckes and Grotjahn, 2006). The C-test, from Keijzer (2007, p. 359), consisted of five Dutch texts, ranging from informal to formal. Participants had to complete twenty words in each text, from which words were removed. Their proficiency was expressed in the number of correct answers they gave. Only words that were orthographically correct and morphologically and semantically possible in the context were scored as correct.

## 2.4 Analysis

Accuracy scores of more than 134 out of 240 good answers are less than 5% likely for participants answering randomly. Participants scoring less than 134 correct answers could have responded randomly or else knew too few Dutch words and were excluded. For the Japanese group, 30 participants were left (-3), for the Spanish group, 35 (-1). From the Other-L2 control group, 27 participants were left (-4). No native Dutch speakers were removed. Only experimental items were used in the following analyses; due to an implementation error, one item from the Good level of Phonotactics had to be discarded. Of the remaining data, only those trials in which the word was recognised – that is, the participant pressed the 'yes' button – were used for latency analyses.

The dependent variable, the reaction time in ms, was logarithmically transformed to improve normality of the distribution. Responses were defined as outliers and removed when they were more than three times the interquartile range away from the quartiles for each participant. There

proved to be only outliers that were very fast (below 50 ms), not very slow ones (trials without button presses after 2500 ms had been recorded as timed out). Outlier removal removed 0.7% of the observations in the Japanese Group, 1.4% for the Spanish Group, 0.7% for the Native-Es group, 0.6% for the Native-Su group and 0.4% for the Other-L1 group.

The sphericity assumption necessary to use repeated measures ANOVA is not to be expected to be true in groups differing in L1 and L2. To avoid the incorrect sphericity assumption and allow correcting for noise more strictly, the data were analysed in a mixed-effects linear model with crossed random effects for participants and items (following Quené and Bergh, 2008; see also Baayen, Davidson, & Bates, 2008).

Contrasts in the Group condition were planned: Native-Es vs. Native-Su, Native-Es vs. Spanish and Native-Su vs. Japanese. Japanese and Spanish groups were not compared, because comparing a different L1 (Japanese and Spanish) and a different type of epenthesis is not informative for the purpose of this study and only three contrasts are possible between four levels. For the effect of Priming, the Unrelated primes were taken as a baseline; both the estimate for Faithful and Epenthesis Priming (if negative) indicate a priming effect.

The C-test score was used as a proficiency measure that may be assumed to predict learning of illegality and/or wellformedness. Additional predictors added to the model were the length of stay in the Netherlands, previous reaction time, correctness of previous answer, trial number and word frequencies from the CGN Corpus of Spoken Dutch and CELEX; these nuisance predictors are assumed to have the potential to contribute to noise elimination, leading to better estimates for the predictors under scrutiny.

Collinearity between C-test Score and Group, caused by higher proficiency of Dutch participants, was removed by residualisation of C-test Score to Group, meaning that Group was used to predict the C-test Score and that the divergence from the predicted value (group mean) was used. Thus, the independent influence of proficiency after the influence of Group was modelled. The average score for Dutch participants in the Native-Es group was 91.9 (*SD* 8.1, range 74–99) and in the Native-Su group 90.2 (*SD* 9.8, range 81–97). The Spanish group scored on average 31.0 (*SD* 18.2, range 1–79) and the Japanese group 26.0 (*SD* 14.6, range 3–66). For the same reason the number of years spent in the Netherlands was also residualised to Group. CGN and CELEX lexical frequencies were residualised to Phonotactic condition, allowing these predictors to correct for frequency effects without disturbing the estimate for the Phonotactic condition.

Model fitting started with the most elaborate model, containing interactions of all experimental predictors (Phonotactics, Priming, Group with planned contrasts, C-test Score), as well as nuisance predictors and random effects for item and participant. Predictors with small influences were removed, after which the old and the simplified model were compared. Nuisance predictors were not removed. If the old model was not significantly better, the simplification was deemed to be warranted, leading to a model that, if possible, corrects for confounding nuisance predictors (see Baayen et al., 2008). With the model after simplification, *p*-values were conservatively estimated for predictor estimates by assuming the degrees of freedom to be the lowest of either the number of items or of participants, minus 1. For 95% Confidence Intervals, the *mcmc* function from the *languageR* (1.4) package was used (see Baayen et al., 2008). This package also returns (less conservative but possibly more accurate) *p*-values, but these were not used.<sup>5</sup> The significance of priming effects could be argued to be one-tailed, but the reported values are always two-tailed.

The interactions involving Group were removed, to see if the epenthesis types and participants in the different groups could be collapsed into an L1-L2 difference, but this removal was not licensed,  $\chi^2(38) = 219.12, p < 0.001$  \*\*\*. The data were therefore analysed separately for the es and su versions of the experiment; the factor Group thus only contrasts native and non-native listeners and not the version of the experiment.<sup>6</sup> The Other-L1 group of the control experiments was divided

**Table 3.** Latency model es version: predictor estimates.

Predictor	Estimate	SE	Significance		
			t(58)	p	95% CI
Intercept (Unrelated/Native-Es)	6.461	0.031			[6.40, 6.53]
Phonotactics	-0.054	0.026	-2.09	0.046*	[-0.10, -0.00]
Priming: Epentheseised	-0.196	0.016	-12.38	< 0.001***	[-0.23, -0.16]
Priming: Faithful	-0.227	0.016	-14.40	< 0.001***	[-0.26, -0.20]
Spanish	0.282	0.063	4.46	< 0.001***	[0.17, 0.39]
C-test	-0.001	0.005	-0.19	0.390	[-0.01, 0.01]
Epentheseised $\cap$ Spanish	-0.055	0.023	-2.38	0.025*	[-0.10, -0.01]
Epentheseised $\cap$ Phonotactics $\times$ C-test	0.012	0.005	2.28	0.032*	[0.00, 0.02]
Epentheseised $\cap$ Phonotactics $\cap$ Spanish $\times$ C-test	-0.016	0.006	-2.78	0.010**	[-0.03, -0.00]

Note. Interactions between experimental predictors that are not shown were not near significance; that is, they had a  $p$ -value above 0.10. All nuisance variables are omitted from the table. The intercept represents Group = Native-Es and Priming = Unrelated. Intersection signs  $\cap$  mean the intersection of two levels of different conditions, i.e. all the trials that conform to both parts of the intersection. The log of the likelihood was -519.5.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$ .

up based on the version of the experiment and separately compared to the corresponding Native group.

Accuracy models were made with the same predictors as the selected latency models, with logit of correctness as the dependent variable (see Jaeger, 2008). These models were used to check for speed/accuracy trade-offs; no hypothesis testing was based on the accuracy.

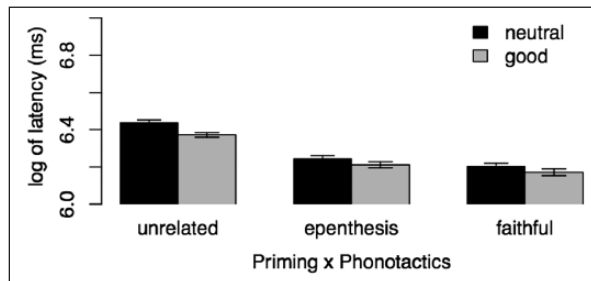
## 2.5 Results

**2.5.1 Es version.** For the es-groups, the four-way interaction Phonotactics  $\times$  Group  $\times$  Priming  $\times$  C-test-score cannot be removed  $\chi^2(2) = 11.09$ ,  $p = 0.004^{**}$ . Thus, neither this interaction nor lower interactions or main predictors were removed. In Table 3, the estimates of all main effects and significant interactions are given (for latency). Figure 1 and Figure 2 show latency data for the two groups.

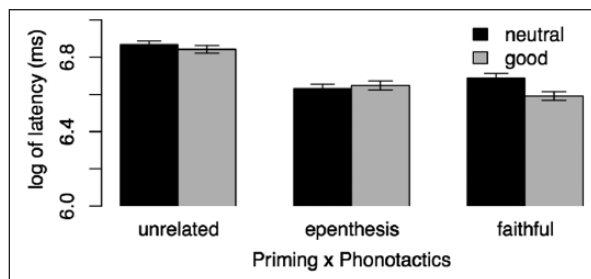
Reactions of native listeners were significantly faster for Good items than for Neutral ones,  $t(58) = -2.09$ ,  $p = 0.046^*$ ; L1 Spanish listeners did not significantly differ from the Native-Es group with respect to Phonotactics, i.e. the difference between Good and Neutral is not significantly different between the groups. Both Epentheseised and Faithful primes cause a significant faster reaction, compared to the baseline of Unrelated primes,  $t(58) = 12.38$  /  $t(58) = 14.40$ ,  $p < 0.001^{***}$ , for the Native-Es group. L1 Spanish participants were generally slower,  $t(58) = 4.46$ ,  $p < 0.001^{***}$ . The priming effect of Epentheseised primes was larger for the Spanish group than for the Native-Es group,  $t(58) = -2.38$ ,  $p = 0.025^*$ .

For Epentheseised primes of Good items, higher proficiency correlated with slower recognition (less priming) for the Native-Es group,  $t(58) = 2.28$ ,  $p = 0.032^*$ , but for the Spanish group, the correlation was in the other direction and significantly different from the Native-Es group,  $t(58) = -2.78$ ,  $p = 0.010^{**}$ . Figure 3 shows the two correlations.

For all significant latency effects, the accuracy effect was in the other direction (faster processing was associated with higher accuracy, or vice versa); hence, a speed/accuracy trade-off can be ruled out: see Table 4 for the model estimates for accuracy.



**Figure 1.** Latencies for Native-Es group by Priming  $\times$  Phonotactics. Error bars indicate  $\pm$  1 standard error.



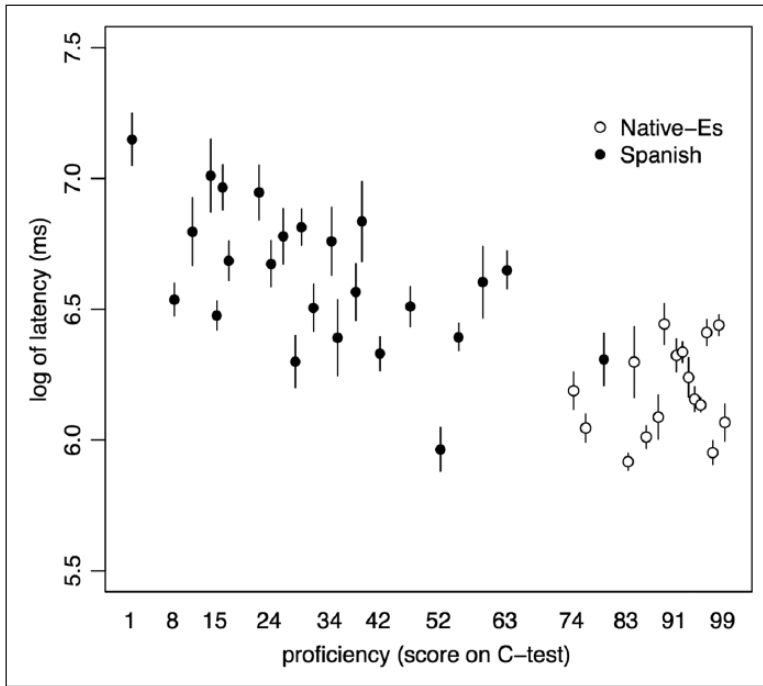
**Figure 2.** Latencies for Spanish group by Priming  $\times$  Phonotactics. Error bars indicate  $\pm$  1 standard error.

**2.5.2 *Su* version.** Simplification of the latency model of the *su* version was licensed for removal of the four-way interaction and the three-way interaction Phonotactics  $\times$  C-Test Score  $\times$  Group. Table 5 shows all significant predictors and Figure 4 and Figure 5 show latencies for the two groups. Phonotactics did not reach significance for the Native-Su group. For this group, both Epenthesised and Faithful primes lead to faster responses to the target,  $t(58) = -5.55 / t(58) = -12.44, p < 0.001^{***}$ , and higher proficiency correlated with the priming effect of Faithful primes,  $t(58) = -2.86, p = 0.008^{**}$ .

The L1 Japanese group was overall slower,  $t(58) = 7.48, p < 0.001^{***}$ , but this effect was attenuated by a priming effect for both Epenthesised and Faithful primes that was larger than for the Native-Su group,  $t(58) = -2.79, p = 0.010^{**} / t(58) = -3.29, p = 0.003^{**}$ . The effect that participants with greater proficiency were primed more by Faithful primes was significantly smaller for the Japanese group than for the Native-Su group,  $t(58) = 2.65, p = 0.013^{**}$ . Given the estimated confidence interval, the effect was probably absent from the Japanese group— that is, proficiency probably had no influence on the priming strength of Faithful primes in the Japanese group. Apart from the above, the Japanese group's latencies did not significantly differ from the Native-Su group.

For all significant main effects on latency, faster processing was accompanied by higher accuracy, or vice versa. However, for all interactions that had a significant effect on latency, a speed/accuracy trade-off is possibly present (see Table 6). Of these possible trade-offs, only the increase in speed for Faithful primes for Japanese participants as compared to Native listeners was accompanied by an almost significant drop in accuracy,  $z = -1.95, p = 0.051$ . The other interactions that were significant for latency were not significant for accuracy.

**2.5.3 Control groups.** Separate models were fitted to compare the control group (learners of Dutch with another L1) to the Dutch native speakers also included in the main analysis. The



**Figure 3.** Latencies for Epenthesised primes with Good clusters by Proficiency. Each dot represents a C-test Score and the associated average reaction time to targets with good clusters primed by primes with epenthesis, for the Spanish and Native-Es groups. Error bars represent  $\pm 1$  standard error.

**Table 4.** Accuracy model es version: predictor estimates.

Predictor	Estimate	SE	Significance	
			z	p
Intercept (Unrelated/Native-Es)	3.128	0.262		
Phonotactics	0.637	0.434	1.47	0.142
Priming: Epenthesised	1.256	0.409	3.07	0.002 *
Priming: Faithful	2.418	0.654	3.69	< 0.001 **
Group: Spanish	-0.849	0.385	-2.20	0.028 *
Epenthesised $\cap$ Spanish	-0.065	0.463	-0.14	0.888
Epenthesised $\cap$ Phonot. $\times$ C-test	0.007	0.164	0.04	0.967
Epenthesised $\cap$ Phonot. $\cap$ Spanish $\times$ C-test	-0.022	0.166	-0.14	0.893

Note. The estimates are linear components of the log-odds. Experimental predictors that are not shown were not near significance; that is, they had a  $p$ -value above 0.10. Nuisance variables are all left out of the table. The log of the likelihood was -310.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$ .

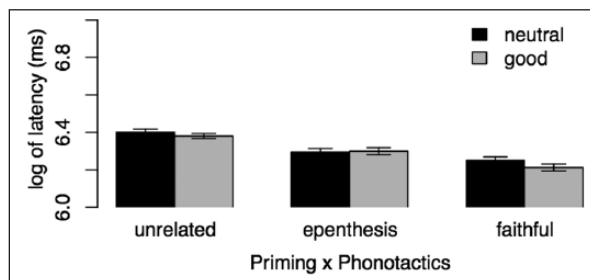
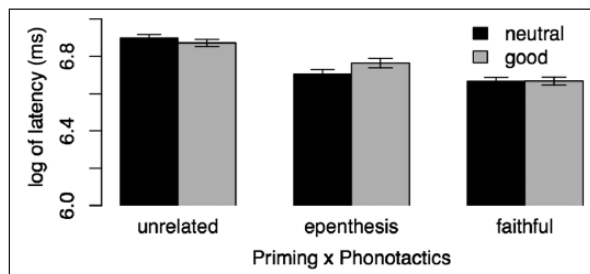
C-Test Score was not used as a predictor as it was not obtained for all participants with another L1 than Dutch, Spanish or Japanese. In the same fashion as with the Spanish and Japanese groups, the control participants that received the es or su versions of the experiment were compared to the Native-Es and Native-Su participants, respectively. The relevant estimated values



**Table 5.** Latency model su version: predictor estimates.

Predictor	Estimate	SE	Significance		
			t(58)	p	95% CI
Intercept	6.387	0.038			[ 6.32, 6.45]
Phonotactics	-0.022	0.026	-0.82	0.283	[-0.07, 0.03]
Priming: Epenthesis	-0.079	1.433	-5.55	< 0.001***	[-0.11, -0.05]
Priming: Faithful	-0.180	0.014	-12.44	< 0.001***	[-0.21, -0.15]
Japanese	0.463	0.062	7.48	< 0.001***	[ 0.36, 0.57]
C-test	-0.013	0.007	-1.85	0.073.	[-0.02, -0.01]
Faithful × C-Test	-0.009	0.003	-2.86	0.008**	[-0.02, -0.00]
Epenthesis $\cap$ Japanese	-0.060	0.021	-2.79	0.010**	[-0.10, -0.02]
Faithful $\cap$ Japanese	-0.071	0.022	-3.29	0.003**	[-0.11, -0.02]
Faithful $\cap$ Japanese × C-test	0.009	0.003	2.65	0.013*	[ 0.00, 0.02]

Note. Interactions between experimental predictors that are not shown were not near significance; that is, they had a  $p$ -value above 0.10. All nuisance variables are left out of the table. The intercept represents Group = Native-Su and Priming = Unrelated. Intersection signs  $\cap$  mean the intersection of two levels of different conditions, i.e. all the trials that conform to both parts of the intersection. The log of the likelihood was -213.5.  
 $P < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$ .

**Figure 4.** Latencies for Native-Su group by Priming × Phonotactics. Error bars indicate +/- 1 standard error.**Figure 5.** Latencies for Japanese group by Priming × Phonotactics. Error bars indicate +/- 1 standard error.

are shown in Tables 7 and 8. The non-native listeners were overall slower, es:  $t(42) = 5.56$ ,  $p < 0.001$ \*\*\*; su:  $t(46) = 5.92$ ,  $p < 0.001$ \*\*\*. No significant difference between native and control group was found for the effect of Epenthesis primes of either version, although the su version

**Table 6.** Accuracy model su version: predictor estimates.

Predictor	Estimate	SE	Significance	
			z	p
Intercept	3.449	0.276		
Phonotactics	0.964	0.484	1.99	0.047*
Priming: Epenthised	0.770	0.411	1.87	0.061.
Priming: Faithful	1.361	0.421	3.23	0.001**
Japanese	-0.726	0.370	-1.96	0.049*
C-test	0.012	0.048	0.25	0.806
Faithful × C-Test	-0.062	0.096	-0.64	0.525
Epenthised ∩ Japanese	-0.248	0.472	-0.53	0.599
Faithful ∩ Japanese	-0.932	0.478	-1.95	0.051.
Faithful ∩ Japanese × C-test	0.052	0.098	0.53	0.593

Note. The estimates are linear components of the log-odds. Interactions between experimental predictors that are not shown were not near significance; that is, they had a *p*-value above 0.10. Nuisance variables are all left out of the table. The intercept represents Group = Native-Su and Priming = Unrelated.

\*\*\* *p* < 0.001, \*\* *p* < 0.01, \* *p* < 0.05, . *p* < 0.1.

**Table 7.** Latency model for control experiment es: predictor estimates.

Predictor	Estimate	SE	Significance		95% CI
			t(42)	p	
Intercept	6.437	0.030			[ 6.38, 6.49]
Group: OtherLI-Es	0.323	0.058	5.56	< 0.001***	[ 0.22, 0.42]
Phonot. ∩ OtherLI-Es	0.095	0.044	2.18	0.039*	[ 0.01, 0.19]
Epenthised ∩ OtherLI-Es	-0.038	0.029	-1.29	0.171	[-0.10, 0.02]
Phonot. ∩ Epenthised ∩ OtherLI-Es	-0.101	0.060	-1.69	0.096.	[-0.21, 0.01]
Phonot. ∩ Faithful ∩ OtherLI-Es	-0.110	0.060	-1.81	0.079.	[-0.11, -0.23]
Phonot. ∩ Epenthised. ∩ OtherLI-Es × C-test	-0.001	0.005	-1.83	0.076.	[-0.02, 0.00]

Note. The predictors that were significant in the main models or have a *p*-value is below 0.10 in this model are shown, except predictors that only refer to the baseline group Native-Es, already reported in Table 3. The model was equal to the one used for the main experiment. The log of the likelihood was -213.9.

\*\*\* *p* < 0.001, \*\* *p* < 0.01, \* *p* < 0.05, . *p* < 0.1.

shows a trend,  $t(46) = 1.67, p = 0.010$ . The control group in the es experiment was slightly more sensitive to Phonotactics than the Native-Es group,  $t(42) = 2.18, p = 0.039^*$ . Accuracy was modelled in the same way as above for the main experiments; no significant speed/accuracy trade-offs were found.

## 3 Discussion

### 3.1 Experiment validity

Faithful and Epenthised primes activate the target, because reaction times were significantly shorter than for Unrelated primes. For native listeners, words with Good clusters were recognised

**Table 8.** Latency model for control experiment su: predictor estimates.

Predictor	Estimate	SE	Significance		
			t(46)	p	95% CI
Intercept	6.415	0.030			[ 6.37, 6.47]
Group: OtherL1-Su	0.323	0.055	5.92	< 0.001***	[ 0.24, 0.42]
Epenthesised $\cap$ OtherL1-Su	-0.048	0.029	-1.67	0.100.	[-0.10, 0.02]
Faithful $\cap$ OtherL1-Su	-0.009	0.029	-0.30	0.379	[-0.07, 0.04]
Faithful $\cap$ OtherL1-Su $\times$ C-test	0.009	0.003	2.63	0.015*	[ 0.00, 0.02]

Note. The predictors that were significant in the main experiments or have a  $p$ -value is below 0.10 in this experiment are shown here, except predictors that only refer to the baseline group Native-Su, already reported in Table 5. The model was equal to the one used for the main experiment. The log of the likelihood was -200.2.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , .  $p < 0.1$ .

more quickly. This effect of wellformedness was only overall significant for the es group. The reason for the lack of significance in the Native-Su group might be the visible tendency of the epenthesis to make the Good targets suffer more than the Neutral ones, although the corresponding interaction of Phonotactics and Priming did not reach significance. The possibility that Good items suffer more from es epenthesis than Neutral items is consistent with the idea expressed by Vitevitch and Luce (1999) that these clusters are represented sublexically. It is likely that the epenthetic vowel breaking up the clusters in the es Epenthesised primes causes more damage to Good clusters than to Neutral ones, because the Good clusters then no longer match their sublexical representations, while such representations do not exist for Neutral clusters.

Although it is possible that the group of words starting with Good clusters are intrinsically easier to recognise than the words starting with Neutral clusters, by coincidence – that is, for non-phonotactic reasons – such effects are likely to be corrected for by the item random effect and the fixed effect for word frequency. Note that although the Native-Su group does respond significantly more accurately to Good targets, no conclusion can be based on this finding, because accuracy data were not collected for hypothesis testing and so drawing conclusions on accuracy would therefore amount to capitalising on chance.

### 3.2 Phonotactic filters and acquisition

**3.2.1 Spanish L1 and es: filter.** The Spanish group differed from the native listeners, because Epenthesised primes activated target words to a greater degree. The priming effect for Epenthesised primes is very similar in size to the effect for Faithful prime, suggesting that the contrast between the clusters and repaired versions is absent. In addition, the control experiment suggests the filters are specific to Spanish as L1. Importantly, the es version of the experiment shows a partial Target Language gradient effect and a Source Language filtering effect in the same context, /sC/ clusters, in the form of a correlation between proficiency and facilitation of probabilistic phonotactics for epenthesised clusters. First, the increase with proficiency of facilitation for Dutch phonotactics for Epenthesised primes makes it unlikely that Spanish L1 learners of Dutch apply a priori knowledge of (universal) wellformedness of the Good clusters. However, to determine whether transfer from Spanish is at all possible, the  $O/E$  ratios of the /sC/ clusters were calculated for continuous Spanish, using the CORLEC corpus (Moreno Sandoval, 2003). Table 9 shows Spanish  $O/E$  ratios for the different clusters. The differences are far less pronounced in Spanish than in Dutch.

The perseverance of perceptual illusions for Epenthesised primes and the emergence of a facilitating role of L2 wellformedness suggests that a representation of wellformed clusters is emerging

**Table 9.** Ranking of sC-clusters on O/E in Spanish.

Cluster	Type	O	E	Spanish O/E	Dutch O/E
[sp]	good	2645	1602	1.65	3.12
[sx]	good	158	173	0.91	2.99
[st]	good	8529	4732	1.80	2.03
Total	good	11332	6507	1.74	2.28
[sm]	neutral	1616	1095	1.48	1.00
[sl]	neutral	1197	1703	0.70	1.11
[sn]	neutral	731	504	1.45	1.19
[sk]	neutral	4434	2473	1.79	0.83
Total	neutral	7978	5775	1.38	1.03

Note. O/E (Observed over Expected) values are based on frequencies found in the CORLEC Corpus of conversational Spanish; the conversational part has been used, comprising 196,917 words. Utterances were separated and otherwise words were joined. The close relation between orthography and phonology in Spanish allowed obtaining a (poor man's) phonemic transcription, though any assimilations not annotated by the transcribers have been lost. The expected values are again based on the frequencies of the phonemes in consonant biphones; using the frequency over the whole corpus gives even more pronounced differences. For ease of comparison, Dutch O/E values from Table 2 are repeated here.

with growing proficiency. However, this representation does not stop misperceptions: the well-formed clusters facilitate word recognition but are still subject to illusionary epenthesis.

If Spanish L1 learners of Dutch access the cluster representations ignoring the epenthetic vowel, an increase with proficiency of the facilitation of Faithful primes should also have been observed. Such a correlation has not been found, although this absence might be due to a similar trend for Dutch native listeners obscuring a difference between groups. However, the more cautious explanation is that the Spanish L1 listeners require the epenthetic vowel to benefit from wellformedness.

**3.2.2 Japanese L1 and su: no filter.** L1 Japanese listeners are primed more by Epenthesised primes than native listeners but also by Faithful primes. In other words, the Japanese group benefits more from priming, but cannot be concluded to be less sensitive to epenthesis than native listeners. The absence of an effect of a Japanese L1 filter could mean that a pre-existing filter was unlearned, but because no evidence for this explanation is available it remains tentative. Because no learning of probabilistic phonotactics was observed either, the Japanese data cannot be used to rule out any of the three scenarios.

The results do not follow predictions that one could make based on the findings of Dupoux et al. (1999). In the latter study, Japanese participants had trouble discriminating between consonant clusters and epenthesised versions of those clusters, unlike the Japanese group in the present study. However, Dupoux et al. investigated non-lexical processing and their participants were unlikely to have been aware of the relevance of consonant clusters or epenthesis for the task prior to testing. In the present experiments, the task was to recognise Dutch words, something the participants face in their everyday lives. Most, if not all, participants had received language teaching in Dutch or English and are therefore likely to be aware of their unfaithful perception and/or production, possibly also for epenthesis in consonant clusters.

It is interesting that the Japanese group is significantly faster for Faithful primes compared to the Native-Su group, but also less accurate. With increasing proficiency, the speed for the Faithful primes drops significantly compared to the Native-Su group, but accuracy rises (almost significantly). If there is indeed a possible speed/accuracy trade-off, one tentative explanation is that

matching between spoken and written words is different for participants with a different writing system such as Japanese.

The control experiment of the *su* version is not relevant due to the absence of a Japanese-specific effect.

**3.2.3 Comparison of *su* and *es* results.** The *es* and *su* versions of the experiment cannot easily be statistically compared, but it is interesting to note that the effect for the *es* version and the Spanish group was not also found for the Japanese group and the *su* version. There are a number of explanations for this difference.

Compared to Spanish, the expected Japanese perceptual illusion is more across-the-board, because it is not restricted to *sC* clusters. Japanese has almost no consonant clusters at all, while Spanish has quite a few, just not of the */sC/* type. In addition, Dutch has no word pairs that lose their contrast with epenthesis to *[esC]* only (as the English pair */e'stet/–/'stet/*, 'estate' – 'state'). It is therefore quite possible that the Spanish group was never required to focus explicitly on the contrast to learn two words differing only in epenthesis, allowing the filter to persist, whereas the Japanese listeners are likely to have had problems with contrast loss due to epenthesis, such as in the pair */ku'lant/–/'klant/*, 'lenient' – 'customer'. There is even a handful of contrasts based on */sC/–/suC/*, e.g. */su'pe:rə/–/'spe:rə/*, 'to have supper' – 'spears', although these never involve frequent words. Such pairs are relevant in the light of the findings of Davidson, Shaw, and Adams (2007), who found that the contrast between a non-native *CC* cluster and an epenthesised version of that cluster of the form *CC* is easier to learn if a minimal pair is given. In short, a Japanese L1 filter, when present, would be more severe and its nefarious effects more noticeable than those of the Spanish filter. The lack of L1 filter effects for Japanese as compared to the presence of the Spanish filter might therefore be caused by unlearning at a very early stage of acquisition.

Another possible reason for the lack of a Japanese filter is that the epenthetic vowel */u/* was contrastive with the illusory vowel in the perceived 'repaired' */sC/* cluster. Japanese listeners distinguish between different epenthetic vowels (Dupoux et al., 2001) and Japanese speakers sometimes produce very reduced or devoiced vowels, crucially between voiceless consonants as in the present study (Han, 1962). In contrast, Ogasawara (2013) reports that voiced vowels, in contexts where devoiced vowels were expected to occur, do not reduce lexical access speed.

The *su* epenthesis might be acoustically more salient than the *es* one, because the *es* epenthesis still contains the cluster as surface sequence. Fleischhacker (2001) concluded, on the basis of cross-linguistic and experimental data, that anaptyxis (epenthesis within a cluster) is most damaging to sibilant-stop clusters (such as */st/*, */sp/*, */sk/*), sibilant-nasal clusters (*/sm/*, */sn/*) and sibilant-liquid clusters (*/sl/*), compared to prothetic epenthesis. The perceived difference to the faithful version is large compared to the same epenthesis in other consonant clusters, so Fleischhacker posits that the anaptyxis into sibilant-stop clusters as found in Japanese loanwords is a choice (over prothetic epenthesis) forced by the highly ranked constraint that no consonants can occur in non-prevocalic position (except in case of consonant doubling or nasals). Tentatively, one might extend this theory to predict that the larger contrast between anaptyctic epenthesis and unepenthesised clusters makes it easier to unlearn the L1 filter, when the Japanese constraint is disregarded; however, disregarding this constraint is probably tantamount to unlearning the L1 filter.

Whatever the reason why a perceptual illusion was not found for the participants in the Japanese group, the null result means that this group cannot be used to test predictions of the unification of categorical illegality and gradient wellformedness. The Japanese group did not differ significantly from the native listeners with regard to wellformedness, but as the Native-Su group failed to show

a significant effect of wellformedness, this does not imply that the Japanese listeners are sensitive to Dutch wellformedness differences.

### 3.3 General discussion and conclusion

The three scenarios of Table 1 can be assessed for the es experiment. An L1 filter has been found for learners of Dutch, as well as quicker recognition for the Good cluster words (not different from native listeners). With increasing proficiency the learners significantly differ from native listeners: they become faster for primes with epenthesised wellformed clusters. The most cautious interpretation is that L2 wellformedness has been acquired through the L1 filter. In other words, a warped version of Dutch gradient wellformedness is acquired.

Scenario 1 'Full Blocking' has to be rejected because it predicts acquisition of probabilistic wellformedness through the filter to be impossible. Wellformedness of sC clusters, while subject to a filter, is not *completely* blocked for acquisition. Scenario 3 'L1 Bias Only' is ruled out because the filter did not become less strong with higher proficiency, even though the acquisition of the probabilistic difference shows that enough input to acquire the legality of /sC/ clusters was available. This elimination leaves scenario 2 'Leaky Filter', because partial acquisition is possible, while an L1 filter actively reduces contrast. The scenario is also compatible with the su results, albeit tenuously given that no filter was convincingly detected.

The metaphorical seeping of L2 gradient wellformedness through the filter reveals itself because the wellformedness acquired by the Spanish group includes the illusionary epenthetic vowel, meaning that there are no faithful representations of actual /sC/ clusters, only of illusionary /esC/ clusters. The question is how this is possible; Dutch input does not contain many [ɛsC] or [esC] clusters (and hardly any Dutch words have an /esC/ or /esC/ onset) and in any case it contains more /sC/ clusters (including every occurrence of /esC/ or /esC/, unless syllabification intervenes). If the limited amount of /esC/ clusters in Dutch input makes its wellformedness learnable, the wellformedness of /sC/ is at least as learnable. Thus the observed /esC/ wellformedness cannot have been learned from faithful perception of Dutch. L1 Spanish knowledge has to be invoked to explain the shape of the 'acquired' knowledge.

The influence of L1 illegality of /sC/ clusters can be formalised as a constraint on phonological structure. The general larger priming effect of epenthesised primes for the Spanish group shows that this constraint operates on the perceptual input as used for lexical access. It might also affect the input for probabilistic phonotactic acquisition if words that enter the lexicon are subject to the L1 filter. The L2 representation of Dutch words would then differ from native representations of these words, in line with conclusions drawn by Pallier, Colomé, and Sebastián-Gallés (2001, p. 448), who posit that 'lexical representations consist of abstract language-specific phonological representations' and that 'it seems that this abstract phonological code, once acquired, is hard to modify'. Their conclusions were based on the finding that Spanish dominant Spanish-Catalan bilinguals show priming for Catalan words that differ from the target words in only a vowel contrast that Spanish does not have. Spanish dominant bilinguals were primed by /neta/, 'granddaughter', for /neta/ 'clean', while Catalan dominant bilinguals were not. Citing Pallier et al. (2001, p. 448) again, 'a lack of sensitivity to difficult L2 phonemic contrasts (...) extends to the way L2 words are represented in the mental lexicon'. It is possible that the same lack of sensitivity to contrast for the Spanish group in this experiment, not between two phonemes but between the phonotactic structures /esC/ and /sC/.

A Spanish-filtered L2 Dutch lexicon would contain many more epenthesised Good clusters than epenthesised neutral clusters, explaining the acquisition of the wellformedness difference among similar lines as for unepenthesised clusters in a native Dutch lexicon. The absence of acquisition of

facilitation for unepenthised primes with Good clusters has then to be explained, because after the L1 filter, they should also contain the /esC/ clusters that the more advanced L2 learners consider wellformed. It is possible that such acquisition took place but was not detected because of a lack of statistical power.

If probabilistic phonotactic knowledge is not acquired from the lexicon, but rather from continuous speech (as suggested by Adriaans and Kager, 2010), this speech must be assumed to be subject to L1 filtering before being passed on to the acquisition system. A third possibility is that the categorical constraint applies to acquired representations. In any case, the persistence of the categorical L1 filter is unexplained if categorical phonotactics would be a special case of probabilistic phonotactics: more evidence for the legality of /sC/ clusters is available than evidence for the wellformedness of Good /sC/ clusters. Unlearning L1 constraints has to be assumed to be more difficult than acquiring probabilistic phonotactic knowledge (which is more specific and based on arguably less evidence). Onishi, Chambers and Fisher (2002) show that new constraints can be acquired even after brief exposure to syllables and just one /sC/ cluster would strictly speaking already disprove a categorical constraint against /sC/.

Apparently, the learner does not apply evidence against the L1 filter (in the form of sC clusters in the input) as readily as using high frequencies of the (illusionary epenthised) wellformed clusters in the filtered input to allow these to facilitate word recognition, indicating that the representations underlying the filter are of a different nature to those governing the gradient effects of probabilistic phonotactics. The present results are therefore incompatible with a reduction of categorical phonotactic (il)legality to probabilistic phonotactics: illegality (constraints) is not simply (extremely) low wellformedness or probability. The filtering action of constraints precedes probabilistic acquisition: if the latter had taken place on more phonetically faithful representations of the input or on exemplars, epenthised /sC/ clusters would not have been acquired as being wellformed.

The above formalisation implies that L1 illegality knowledge that is transferred to an L2 protects itself by filtering out L2 counterevidence against itself (such as lexical contrastiveness). This prediction would be proved incorrect if an L1 filter was found that disappears with increasing proficiency. The Japanese group of the present study does not show this to be the case. The persistent use of knowledge of German phonotactic illegality in English by highly proficient German learners of English reported by Weber and Cutler (2006) does not prove that L1 filtering cannot disappear, but fits the prediction, while the results of Trapman and Kager (2009) and Boll-Avetisyan (2011) are compatible with this view as well as the opposite.

If phonotactic knowledge is to be formalised as consisting of constraints on phonological representations on the one hand and probabilistic knowledge on the other, the latter also needs formalising. Vitevitch and Luce (1999) propose that there are sublexical representations for clusters with high probability of occurrence that behave as phonemes, mapping auditory input to higher level chunks such as words. Under this formalisation, the sublexical representations of Good clusters that the more proficient participants in the Spanish group have acquired must be assumed to contain the epenthetic vowel. As discussed above, the effect of the constraints could be on the learning input (lexicon or continuous speech), or on the representations. The latter is less likely to be correct if wellformed clusters are represented similarly to phonemes, because such sublexical representations, containing epenthetic vowels for L2 learners, are expected to be subject to perceptual learning when they are similar to phonemes (see for an overview Kraljic, Brennan, & Samuel, 2008; Samuel and Kraljic, 2009). Every time the cluster is encountered without the epenthesis and in a lexically unambiguous context, the listener uses the information from the lexicon to update his or her representation. Spanish learners exposed to native Dutch should often hear and recognise words containing an /sC/ cluster produced without epenthesis, in contrast to their representation of that word, thus

inciting perceptual learning that erodes the epenthetic vowel in the representation of the cluster. However, erosion was not observed. Because perceptual learning has the power to adapt existing representations, it must be assumed the L1 filter blocks perceptual learning as well as affecting representations. A more cautious alternative is that the L1 filter simply applies to incoming speech before it is passed on to lexical activation or acquisition processes. If indeed the Spanish learners do not acquire the probabilistic difference between Good and Neutral clusters without epenthesis, the sublexical representations must be assumed to apply to the input after the filter, allowing the epenthesised Good clusters to facilitate target word activation.

In summary, phonotactic illegality knowledge constraints straighten the input into native categories and structures, after which the more wellformed of these structures are easier to recognise and/or process than the others. L1 filters transfer to the L2 and are then tenacious and can affect the acquisition of L2 phonotactics. L2 learners are not excluded from some benefits of L2 wellformedness, because they can build up representations which, while not necessarily native-like, capture some of the wellformedness differences in the Target Language and allow for more efficient processing of frequent phoneme combinations.

### Acknowledgements

The authors thank Keren Shatzman, Hugo Quené, Silke Hamann, Jacques Mehler, Arthur Samuel, Frank Wijnen, Paul Luce, Wim Zonneveld, Frans Adriaans, Natalie Boll-Avetisyan and an anonymous reviewer for helpful comments on earlier versions and Hayo Terband, Frits van Brenk, Joost Bastings, Tim Schoof, Bart Penning de Vries, Annita IJff, Mirjam Trapman and Theo Veenker for assistance with the experiments.

### Funding

This work was funded by the Netherlands Organisation for Scientific Research (NWO) [grant numbers 277-70-001, 360-89-030].

### Notes

1. The opposite possibility is less likely because unlearning illegality seems to be hard (Weber and Cutler, 2006). The results of the three studies on L2 phonotactics that were discussed above (Weber and Cutler, 2006; Trapman and Kager, 2009; Boll-Avetisyan, 2011) indicate that (categorical) L1 illegality is transferred to an L2 and that L2 probabilistic phonotactics can be acquired.
2. For these listeners, English words with voiced and voiceless unvoiced codas are primed by both a faithful version of those words and by a version in which the coda voicing is changed. Dutch has no voicing contrast in (word final) codas, because Dutch coda obstruents are always voiceless.
3. Continuous speech is expected to be the best bet for the source of successful L2 phonotactic acquisition, as L2 lexicons might be incomplete and suffer from distortions due to misperceptions. Lexical statistics show the same, but even more pronounced, differences in wellformedness.
4. Thanks are due to Jacques Mehler for suggesting this control group.
5. A digital bundle containing data and models can be obtained from the authors and is accessible with R version 2.15.3 (2013-03-01), with the packages `lme4_0.999999-0` and `languageR_1.4` (and possibly with similar versions).
6. Thanks are due to Arthur Samuel for suggesting that an analysis split by groups might be more insightful.

### References

- Adriaans, F., & Kager, R. (2010). Adding generalization to statistical learning: The induction of phonotactics from continuous speech. *Journal of Memory and Language*, 62(3), 311–331.
- Albright, A. (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology*, 26(1), 9–41.



- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database (release 2)*. CD-ROM. Philadelphia, PA: Linguistic Data Consortium.
- Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods. *Journal of Memory and Language*, 44(4), 568–591.
- Berent, I., Lennertz, T., & Balaban, E. (2011). Language universals and misidentification: A two-way street. *Language and Speech*, 55(3), 311–330.
- Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, 104(3), 591–630.
- Boersma, P. (1997). How we learn variation, optionality, and probability. In *Proceedings* (Vol. 21, pp. 43–58). Amsterdam: Institute of Phonetic Sciences, University of Amsterdam.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Boll-Avetisyan, N. (2011). Second language probabilistic phonotactics and structural knowledge in short-term memory recall. In N. Danis, K. Mesh, & H. Sung (Eds.), *Proceedings of the 35th annual Boston University Conference on Language Development* (Vol. 1, pp. 60–72). Somerville, MA: Cascadilla.
- Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, 36(1), 22–34.
- Coetsee, A. W. (2008). Grammaticality and ungrammaticality in phonology. *Language*, 84(2), 218–257.
- Coetsee, A. W. (2009). *Grammar is both categorical and gradient*. In S. Parker (Ed.), *Phonological argumentation: Essays on evidence and motivation* (pp. 9–42). London: Equinox.
- Coleman, J., & Pierrehumbert, J. B. (1997). Stochastic phonological grammars and acceptability. In J. Coleman (Ed.), *Third meeting ACL special interest group in computational phonology: Proceedings of the workshop* (pp. 49–56). Somerset, NJ: Association for Computational Linguistics.
- Davidson, L., Shaw, J., & Adams, T. (2007). The effect of word learning on the perception of non-native consonant sequences. *Journal of the Acoustical Society of America*, 122(6), 3697–3709.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1568–1578.
- Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, 16(5/6), 491–505.
- Eckes, T., & Grotjahn, R. (2006). A closer look at the construct validity of C-tests. *Language Testing*, 23(3), 290–325.
- Fleischhacker, H. (2001). Cluster-dependent epenthesis asymmetries. In A. Albright & T. Cho (Eds.), *Papers in phonology 5* (pp. 71–116). Los Angeles: UCLA.
- Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42(4), 481–496.
- Ganong, W.F. III (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, 6(1), 110–125.
- Gathercole, S. E., Frankish, C., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 84–95.
- Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105(2), 251–279.
- Goldinger, S. D. (2003). Puzzle-solving science: the quixotic quest for units in speech perception. *Journal of Phonetics*, 31(3), 305–320.
- Hallé, P. A., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: a case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, 24(2), 592–608.
- Han, M. S. (1962). Unvoicing of vowels in Japanese. *Onsei no Kenkyuu*, 10, 81–100.
- Hayes, B., & Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3), 379–440.

- Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., & Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: A functional magnetic resonance imaging study. *Journal of Neuroscience*, 23(29), 9541–9546.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
- Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, 50(1), 23–52.
- Kager, R., & Pater, J. (2012). Phonotactics as phonology: knowledge of a complex restriction in Dutch. *Phonology*, 29(1), 81–111.
- Keijzer, M. C. J. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. LOT Publications, Utrecht.
- Kraljic, T., Brennan, S. E., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, 107(1), 54–81.
- Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16(5), 565–581.
- Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception and Psychophysics*, 34(4), 338–348.
- Moreno Sandoval, A. (2003). La evolución de los corpus de habla espontánea: la experiencia del LLI-UAM. In A. J. Rubio Ayuso (Ed.), *Actas de las II jornadas en tecnologías del habla 2002*. Granada: Universidad de Granada.
- Moro, A., Tettamanti, D., Perani, D., Donati, C., Cappa, S., & Fazio, F. (2001). Syntax and the brain: Disentangling grammar by selective anomalies. *NeuroImage*, 13(1), 110–118.
- Ogasawara, N. (2013). Lexical representation of Japanese vowel devoicing. *Language and Speech*, 56(1), 5–22.
- Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, 83(1), B13–B23.
- Oostdijk, N. (2000). The Spoken Dutch Corpus. Overview and first evaluation. In M. Gravididou, G. Carayannis, S. Markantonatou, S. Piperidis, & G. Stainhauer (Eds.), *Proceedings of LREC-2000, Second international conference on language resources and evaluation* (Vol. 2, pp. 887–894). Paris: ELRA.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychological Science*, 12(6), 445–449.
- Peperkamp, S. (2007). Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition*, 104(3), 631–637.
- Pierrehumbert, J. B. (2001a). Exemplar dynamics: word frequency, lenition, and contrast. In J. Bybee & P. J. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (Vol. 45, pp. 137–157). Amsterdam: Benjamins.
- Pierrehumbert, J. B. (2001b). Why phonological constraints are so coarse-grained. *Language and Cognitive Processes*, 16(5/6), 691–698.
- Pierrehumbert, J. B. (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay, & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 177–228). Cambridge, MA: MIT Press.
- Polivanov, E. (1931). *La perception des sons d'une langue étrangère [Perception of sounds of a foreign language]*. In *Travaux du cercle linguistique de Prague: Vol. 4. Réunion phonologique internationale tenue à Prague (18–21/xii 1930)* (pp. 79–96). Prague: Jednota Československých Matematiků a Fysiků.
- Quené, H., & Bergh, H. van den. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425.
- Samuel, A. G., & Kraljic, T. (2009). Perceptual learning for speech. *Attention, Perception, & Psychophysics*, 71(6), 1207–1218.
- Trapman, M., & Kager, R. (2009). The acquisition of subset and superset phonotactic knowledge in a second language. *Language Acquisition*, 16(3), 178–221.
- Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science*, 9(4), 325–329.

- Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408.
- Weber, A., & Cutler, A. (2006). First-language phonotactics in second-language listening. *Journal of the Acoustical Society of America*, *119*(1), 597–607.