overSEAS 2020

This thesis was submitted by its author to the School of English and American Studies, Eötvös Loránd University, in partial fulfilment of the requirements for the degree of Bachelor of Arts. It was found to be among the best theses submitted in 2020, therefore it was decorated with the School's Outstanding Thesis Award. As such it is published in the form it was submitted in **overSEAS 2020** (http://seas3.elte.hu/overseas/2020.html)

ALAPSZAKOS SZAKDOLGOZAT

Kovács Petra

Anglisztika alapszak

Angol szakirány

2020

Eötvös Loránd Tudományegyetem

Bölcsészettudományi Kar

ALAPSZAKOS SZAKDOLGOZAT

Az anyanyelvi fonológia hatása az L2 beszédészlelésre

How native language phonology affects L2 speech perception

Témavezető:

Dr. Albert Ágnes Erzsébet

egyetemi adjunktus

Készítette:

Kovács Petra

anglisztika alapszak

angol szakirány

2020

A HKR 346. § ad 76. § (4) c) pontja értelmében:
„… A szakdolgozathoz csatolni kell egy nyilatkozatot arról, hogy a munka a hallgató saját szellemi terméke…"

NYILATKOZAT

Alulírott **Kovács Petra s.k.** ezennel kijelentem és aláírásommal megerősítem, hogy az ELTE BTK **anglisztika**.. alapképzés/alapszak **angol** .... szakirányán írt jelen szakdolgozatom saját szellemi termékem, melyet korábban más szakon még nem nyújtottam be szakdolgozatként/záródolgozatként és amelybe mások munkáját (könyv, tanulmány, kézirat, internetes forrás, személyes közlés stb.) idézőjel és pontos hivatkozások nélkül nem építettem be.

Budapest, 2020. 04. 08. ....

....
aláírás

A nyilatkozatot a szakdolgozathoz kell csatolni.

Abstract

Earlier research has shown that second language speech is perceived in terms of the phonological categories of the first language. The present paper reviews the literature to investigate how the phonology of the first language affects speech processing in a language learned later. Although the literature explains *what* influence the first language exerts, it fails to clarify the underlying mechanism of this influence: How it is exerted and why. The present thesis argues that this issue can be resolved by interpreting earlier findings in terms of a paradigm called the pattern recognition theory of mind. The paradigm has been used to explain native language speech processing, and this thesis demonstrates that its horizons can be expanded to account for the phenomena observed in bilingual speech perception, as well.

keywords: speech perception, second language acquisition, phonology, pattern recognition theory of mind

Table of contents

INTRODUCTION

Speech processing – the way humans do it – consists of a series of steps. Gósy (2005) differentiates four levels: hearing (decoding acoustic signals), speech perception (the recognition of speech sounds and their combinations), comprehension (which includes semantics and syntax), and finally, associations (recognizing connections between the processed utterance and earlier memories about the world.) Of these levels, this thesis is concerned with the second, speech perception, which takes place before – and as a prerequisite to – the comprehension of meaningful words.

When acquiring a first language (L1), infants start to recognize which speech sounds are contrastive in the language, and what contrasts they can ignore. For instance, Japanese infants are as attentive to the /l/–/r/ contrast as English ones at first, but eventually lose the ability to differentiate the two sounds (Segui, Frauenfelder & Hallé, 2001). In this way, infants acquire the phoneme inventory – the list of speech sounds which can contrast words – of their L1 and continue to utilize this system of mental representations whenever they process speech.

However, when learning a second language (L2) at a later age, the speech perceiver faces a problem: languages differ in their phonological systems, so the native phonology in the mind cannot fully facilitate the processing of L2 speech. The phonologies of two languages can differ in many aspects. First, the phoneme inventories might not completely overlap, as illustrated by the example of Japanese infants: In English, /l/ and /r/ are separate phonemes, while in certain other languages, they are not. Second, the phonotactic constraints – regularities as to how phonemes can combine in words – can differ. In some languages, word-initial /mb/ is allowed, but this is not the case in English. Last, the suprasegmentals of languages tend to be different, including the intonation patterns of utterances, or whether word stress is contrastive or predictable.

All these differences between one's L1 and L2 may have implications at the level of speech perception: They could hinder the recognition of phonemes, and eventually, the comprehension of speech. This is something that second language acquisition research has to face, and therefore, it is important to explore the ways in which the phonology acquired in infancy influences the perception of L2 speech. The present paper aims to accomplish this end with an emphasis on English as a second language.

The effect of the native language on L2 speech perception has been a topic of inquiry in the field of psycholinguistics. These findings are reviewed here from three aspects of phonology: the phoneme inventories themselves, the phonotactic constraints, and suprasegmental features. However, I argue that the L2 speech perception literature – although it presents the result of some process – does not provide a deep enough understanding of the mechanism that underlies this result: How and why does L1 phonology influence the perception of L2 speech?

I propose that this issue can be resolved by interpreting the reviewed findings in terms of the pattern recognition theory of mind – a theory which originates from research in the field of artificial intelligence, but is used to explain the workings of human perception, including speech processing (Kurzweil, 2012). After presenting the theory, the findings of the L2 speech perception studies are analyzed in this paradigm in order to develop an appreciation of the difficulties which the language learner faces in processing L2 speech.

LITERATURE REVIEW

Differences in L1 and L2 phoneme inventories

As speech perception is the recognition of meaningless speech segments, the basic unit of speech perception is the phoneme. A sound is said to be a phoneme of a language if it can contrast words. Thus, /n/ and /ŋ/ are phonemes of English because they are responsible for the difference in the words *thin* and *thing.* In Hungarian, /ŋ/ does appear as a speech

sound in words like *munka, hanga,* but it is not a phoneme of this language because it only appears in predictable positions: before /k/ and /g/. Let us now see what happens when L1 and L2 phoneme categories do not exactly match.

Strange and Shafer (2008) mention that if a phoneme of the L2 does not exist (i.e., cannot distinguish words) in the L1, the recognition of the phoneme is likely to be incorrect. An example for this was presented by Broersma and Cutler (2011), who examined Dutch–English bilinguals' perception of the /æ/–/ɛ/ contrast, i.e. the vowels in TRAP and DRESS. (Note that the term *bilingual* is used in the sense of late bilingual unless otherwise stated, as this thesis is concerned with an L2 learned later in life.) These two phonemes of English are not contrastive in Dutch (much like in Hungarian); therefore, native speakers of Dutch, excluding infants, do not have two distinct categories for these sounds. Accordingly, this study found that the difference between the two vowels is not perceived, despite a high proficiency in English.

However, Strange and Shafer (2008) proceed to argue that incorrect perception does not always occur. For instance, Spanish–English bilinguals are reported to have no difficulty distinguishing between /iː/ and /ɪ/ (NEAR and KIT) or /ʌ/ and /ʊ/ (STRUT and FOOT), despite the fact that these vowels are not separate phonemes of their native language. Why this difference in perceptual difficulty? Strange and Shafer highlight that the methodology of the experiment itself influences participants' performance. On the other hand, it also seems to matter that certain differences in the acoustic signal are more salient than others. For instance, two vowels which differ in quality (as /æ/ and /ɛ/ do) are harder to discriminate than if they differ in length (/iː/ and /ɪ/), which indeed explains the above-mentioned discrepancy between Dutch and Spanish natives' perception of certain L2 vowels.

As for consonants, differences in voicing (/g/ vs. /k/) are more salient than differences in place of articulation (/p/ vs. /k/). However, voicing contrasts are realized in

various ways in different languages. In Hungarian, the difference between the syllables /pa/ and /ba/ is that in the first case, voicing (vocal fold vibration) starts when the articulation of the consonant is finished, whereas in the second case, voicing is already underway during the articulation of the consonant. This is not so in English: in /pa/, voicing starts later (this is called aspiration and indicated thus: /pʰa/), and /ba/ is articulated like Hungarian /pa/, with no voicing of the consonant itself. Other languages also have the fourth option, /bʰa/.

Gass (1984) examined how differences in the realization of voicing in one's L1 and L2 affect L2 speech perception. Here, participants heard instances of the syllable /pa/, modified in a way that voicing started at various points in time in different stimuli (they resembled either /ba/, /pa/, or /pʰa/). Participants were native speakers of different languages, and they were asked to judge whether they heard the English consonant /p/ or /b/. For these L2 speakers, many of the stimuli were ambiguous: Instances which resembled /pa/ were not categorically perceived as either /pa/ or /ba/. Instead, there was an area of ambiguity where categorization was uncertain. This was not the case for English native speakers, who had a clear idea of the /p/–/b/ difference.

The importance of acoustic salience in perceptual difficulty was also examined by Polka (1991), who emphasized that a sound's not being contrastive in the L1 as opposed to the L2 is not the only factor which affects L2 speech perception. Here, native English participants' ability to distinguish voicing contrasts in an unfamiliar language (Hindi, which uses all four contrasts mentioned above) was measured in order to explore what makes certain contrasts more difficult to perceive than others. Besides acoustic salience, the other candidates were phonemic status (Is the sound a phoneme in English?) and phonetic experience of the participants with the sound (Does the sound appear in English, even though it is not a phoneme – like /ŋ/ in Hungarian?). Acoustic salience (or the lack of it) proved to be the best predictor of perceptual difficulty in this study.

No matter how difficult it is to distinguish L2 phonemes, an error in speech perception is only relevant if the next level of processing, comprehension, is hindered (Broersma & Cutler, 2011). However, this is likely to happen if a phoneme inside an L2 word is misinterpreted. To understand why this is a problem, let us examine how a word is recognized by the listener.

This process is called lexical access, and according to current models, it works in the following way. As phonemes in the speech signal are recognized one by one – an extremely rapid process –, words which match the signal are activated, and words which do not are inhibited. In this way, many unlikely candidates are successfully ruled out, but a number of words still remain activated, competing for recognition. Eventually, the word which best matches the incoming signal is recognized, but the more competition there is from other lexical items, the slower the process (McQueen, 2005).

Broersma and Cutler (2011) highlight that because L2 speech perception is often inaccurate, irrelevant words can be activated as competitors, so that the cognitive load of lexical access may be greater than in the case of L1 comprehension. To examine whether the phenomenon referred to as phantom word activation truly decelerates word recognition, Broersma and Cutler applied a priming task: Participants heard a spoken word, then saw a different word on a screen. They had to decide whether the visually presented stimulus is a real word or not. As mentioned earlier, these Dutch participants do not perceive the difference between the English phonemes /æ/ and /ɛ/. This means that when the word *lamp* was primed by the non-word *lemp* in the experiment, Dutch natives were faster to respond than the native English control group: *Lemp* activated *lamp* because the difference was not perceived.

The reason this is a problem is that strings of phonemes like *lemp* do appear in English when they are embedded in larger words or phrases (e.g. *evil empire* contains this

string). Broersma and Cutler (2011) found that such embedded instances of *lemp* activate *lamp* as well as when the string is heard in isolation. In this way, the irrelevant English word *lamp* competes for recognition, making the process of comprehension slower and less efficient than it is in the case of L1 speakers, who perceive the /æ/–/ɛ/ contrast accurately.

In conclusion, phoneme inventory differences between the L1 and L2 can affect the speed and accuracy of the language learner's listening comprehension. However, some L2 phonemes are easier to perceive than others because certain acoustic cues are sufficiently salient. Other L2 distinctions continue to be misperceived even with great L2 experience.

Differences in L1 and L2 phonotactics

The phonologies of two languages can differ in more aspects than just the phoneme inventories. How phonemes can combine inside a word is restricted by what phonologists call phonotactic constraints. For instance, word-initial /tl/ is illegal in English, as well as in Hungarian. However, languages may differ in their phonotactic constraints: While English does not allow the consonant cluster /mb/ word-finally, in Hungarian, word-final /mb/ is available (in words such as *domb, lomb*).

In bilingual speech production, there seems to be a clear influence of L1 phonotactics on L2 speech in many cases. Sperbeck and Strange (2010) mention Japanese learners of English, who insert an epenthetic vowel inside consonant clusters when producing English words which would be otherwise phonotactically illegal in Japanese (e.g. /spot/ becomes /supot/). Likewise, Hungarian learners of English frequently pronounce the final consonant cluster in English words such as *dumb.* It may well be that these transfers from the L1 are paralleled by native-like perception of the L2 input, but pronunciation errors in themselves are no proof of this. Does L1 phonotactics influence L2 speech perception?

To answer this question, it should first be established whether L1 phonotactics has any effect on speech perception in the L1 itself – it can also be that phonotactic constraints

are merely abstract linguistic facts, unutilized by the mind when processing speech. That this is not the case was demonstrated by Shatzman and Kager (2007). In this study, participants performed a lexical decision task where the non-words were either phonotactically legal or illegal in their L1. They rejected the illegal non-words with faster reaction times, which suggests that phonotactic information is indeed used when processing speech. Additionally, Segui, Frauenfelder and Hallé (2001) mention that infants who already acquired the phonological systems of their L1 are more attentive to phonotactically legal and probable non-words.

If this is the case, then the phonotactic constraints of the L1 could also play a role in the perception of speech in any language acquired later. This effect has been confirmed by a number of studies using various methodologies. Sperbeck and Strange's (2010) participants were Japanese learners of English. They heard pairs of non-words, and they had to judge whether each pair was the same or different. When one non-word had an initial consonant cluster (illegal in Japanese), and its pair had the same cluster separated by a schwa (the vowel /ə/), the participants could not differentiate the two as accurately as the control group of native English speakers. This suggests that the epenthesized vowel is not only present in L2 speakers' production of English speech, but in its perception, as well.

Similar results have been found in the case of Spanish learners of English. The main difference in Spanish phonotactics as compared to English is that word-initial s+consonant (#sC[1]) clusters are missing. Instead, Spanish turns these into #esC syllables – such as in the word *España –,* and Spanish accented English speech is often characterized by an epenthetic /e/ in this position. Again, this is merely a fact about speech production so far. However, Freeman, Blumenfeld and Marian (2016) demonstrated that an epenthetic /e/ also appears when Spanish learners of English perceive English speech. In this priming study,

---

[1] Where # signifies a word boundary, and C, any consonant.

participants showed an activation of /e/-initial words when primed by English words such as *strong*. This suggests that Spanish–English bilinguals not only produce but also perceive a vowel preceding a consonant cluster which would otherwise be illegal in their L1.

On the other hand, it seems that this effect diminishes as L2 proficiency grows. Carlson et al. (2016) examined Spanish–English early bilinguals, who differed in which one their dominant language was. Their task was to decide whether they heard a word-initial /e/ in a number of non-words. In #sC non-words with no initial vowel, Spanish-dominant participants perceived an /e/ in many cases, but English-dominant ones did not. Thus, experience with English affected the participants' speech processing such that consonant clusters illegal in Spanish but legal in English were not repaired during perception.

These three studies with bilingual participants differ from Shatzman and Kager's (2007) study mentioned earlier in that they attest to the perception of illusory vowels to resolve phonotactic illegality, but they are not concerned with the speed at which phonotactically legal versus illegal words are processed by bilinguals. Segui, Frauenfelder and Hallé (2001) conclude that illegal consonant clusters are resolved in L2 speech perception by assimilating these clusters to legal ones in one of three ways: omitting the perception of a phoneme, perceiving an extra phoneme, or changing one phoneme in the cluster to another one. But is assimilation the only way in which L1 phonotactics can affect the perception of the L2?

Modern Hungarian, much like English, does not require vowel epenthesis to resolve consonant clusters. #sC clusters are rare in Hungarian, too (these are words like *sznob* and *sztár*), in favour of the fairly common #ʃC (e.g. *stég, sport*), which is in turn debatably existent in English (dictionaries do list *spiel* and *shtoom*). When perceiving #sC clusters in English speech, a Hungarian learner of English would – theoretically – not perceive illusory vowels, omit one of the initial phonemes, or change the /s/ to a /ʃ/, but that does not mean

8

that their perception of the L2 would remain unaffected by the L1. It might well be that Hungarian learners of English process #sC slower than #ʃC – similarly to Shatzman and Kager's (2007) participants, who showed different processing speeds for illegal and legal non-words –, and this effect may be weakened by growing L2 proficiency. These are hypotheses for an empirical study, but perceptual assimilation is not the only conceivable effect that L1 phonotactics may have on L2 speech perception.

To summarize, differences in L1 and L2 phonotactic constraints can affect the processing of L2 speech: Illegal consonant clusters may be perceived as legal ones, and their processing may require more cognitive effort. This has implications for ESL learners, since their comprehension of English speech can be hindered at the level of speech perception by seemingly abstract linguistic differences in their languages. These differences can also cause occasional speech segmentation errors: An illegal #CC cluster may be processed as a C#C sequence. This effect can only be reduced by sufficient experience in listening to L2 speech.

Differences in L1 and L2 suprasegmentals

So far, the discussion has centered around segmental information: phonemes and their combinations. However, speech processing also entails the perception of suprasegmental information (prosody), which is believed to occur parallelly to the perception of phonemes (Honbolygó et al., in press). Of the many features which constitute the prosody of an utterance, the present section is concerned with stress and intonation.

Stress is the relative prominence of one syllable over others (Honbolygó et al., in press). Its role differs from language to language: In English, for instance, stress can contrast words just like phonemes can (e.g. *súbject–subjéct*). In languages like Hungarian, French, and Polish, stress is not contrastive, but falls predictably on the first, last, and penultimate syllables, respectively (Peperkamp & Dupoux, 2002). If a bilingual's two languages differ in their use of stress, is L2 speech perception affected?

According to Honbolygó et al. (in press), the stress pattern of the native language (Hungarian in this study) is only activated during L1 speech perception, but not during the perception of an unfamiliar language. In the former case, participants rely on long term representations about native stress patterns, while in the latter, their brain activity is based on short term memories about stress patterns in the task at hand, ignoring the legal position of stress in Hungarian.

Does this mean that L2 stress contrasts are easier to perceive than L2 phoneme contrasts? Not necessarily. Honbolygó et al. (in press) mention that such long-term representations underlie the perception of phonemes, as well; stress perception should then pose similar problems. What these findings suggest is that the perception of stress patterns is a predictive process: The brain makes predictions about what is likely to occur next in the speech signal, and compares the actual input to these predictions – a process fully congruent with current theories about how perception in general works (see Blake & Sekuler, 2006).

What is implied in Honbolygó et al.'s (in press) findings is that the listener uses different strategies when predicting input in a familiar language as opposed to a less familiar one. In the latter case, predictions are made based on short term memory. This might be a useful strategy if the L2 has fixed stress (like French or Hungarian), but not in the case of English as a second language, as stress in this language is not predictable. The perception of L2 stress contrasts should then be problematic if stress is not contrastive in the L1.

Accordingly, Peperkamp and Dupoux (2002) found that native speakers of a number of languages lacking contrastive stress – namely French, Finnish, Hungarian, and Polish – show what the authors call stress "deafness": difficulty in distinguishing non-words which only differ in the position of stress. Although such minimal pairs exist in English, they are not abundant. Do these findings then bear any relevance to learners of English?

Most pairs of words distinguished only by stress are semantically related verb–noun pairs (such as *tormént–tórment*). In these cases, even if the place of stress is incorrectly perceived in the speech signal, the listener can infer the meaning from context with relative ease, so that comprehension is likely to remain intact. A somewhat harder task is to distinguish semantically unrelated minimal pairs (e.g. *ínsight–incíte, éssay–essáy*), especially if they are of the same part of speech (e.g. *canál–kénnel*).

It is also questionable how pervasive stress "deafness" is. Honbolygó et al.'s (in press) findings contradict those of Peperkamp and Dupoux (2002): In this study, Hungarian natives could in fact distinguish between non-words differing only in stress. The difference is in the methodology: While Honbolygó et al. used a relatively simple task – participants had to decide whether pairs of non-words are the same or different –, Peperkamp and Dupoux's participants performed a complex short term memory task, where they had to reproduce strings of syllables they heard by transcribing them with the help of numbers on a keyboard. The perception of stress contrasts, then, is also a function of the cognitive demand imposed by the task.

What these studies both acknowledge is that the presence of stress itself – i.e. the relative loudness of some syllable compared to others, regardless of its contrastiveness – is perceived by listeners of any native language. Whether a distinction can be made based on stress only also depends on factors other than a phonological difference between the L1 and L2.

Let us now turn to a further suprasegmental feature of speech: intonation. This term refers to the changes in pitch, that is, the height at which the syllables of an utterance are spoken (Nádasdy, 2006), which changes result from variation in the fundamental frequency of speech. Although intonation is also affected by factors which lie outside of language (such

as the age, sex, or emotions of the speaker), this section is concerned with the way intonation can change the meaning of a sentence.

Like stress, the fact of pitch changes is perceived by every human with a healthy hearing. However, languages attribute different meanings to certain intonation patterns, and not every pattern has an equivalent in all languages. As an example, Table 1 summarizes the uses of intonation patterns in English (Nádasdy, 2006) and Hungarian (Varga, 1983).

Table 1.

Meaning of intonation patterns in English and Hungarian

Intonation		English		Hungarian
fall	\	exclamations tag questions	WH-questions statements commands	Y/N questions with the particle *-e*
rise-fall	∧	–		Y/N questions echo questions
descend	⌒	–		exclamations
fall-rise	\/	but-implication		
low rise	⁄	indifference		questions (monosyllabic or topic-only)
high rise	⁄	Y/N questions echo questions		–

It is apparent that some intonation–meaning connections are shared between these two languages. For instance, but-implications are expressed in a similar way, as illustrated by these examples:

(1) The restaurant is \/<u>closed</u>… (but we can go to the café)
(2) Az étterem \/<u>zárva van</u>… (de a kávézóba beülhetünk)

On the other hand, not all intonation patterns are used in both languages: Rising-falling and descending intonation is missing in English, while the high rise is not present in Hungarian. Further, a number of meanings are expressed differently, such as echo questions ("please-repeat" questions):

(3) (I was born on the North Pole.) /<u>WHERE were you born?</u>
(4) (Az Északi-sarkon születtem.) Hol ∧<u>születtél</u>?

Another important difference is in the use of low rising intonation. While in English it is not typically used for questions, in Hungarian, it is a questioning intonation pattern in the case of monosyllabic utterances (5) and topic-only questions (6):

(5) Zöld?
   'Is it green?'

(6) A nővéred? (… Hogy van?)
   'And your sister? (…How is she?)'

So far, these are merely observations about the linguistic structure of the two languages, and they do not reveal anything about the perception of intonation patterns. Although such an empirical comparison between English and Hungarian is lacking, intonation perception has been examined in both monolingual and bilingual contexts.

Firstly, Wales and Taylor (1987) report that L1 questions and statements can indeed be identified by native English listeners based on the pitch changes at certain points in the utterance, so if the intonation is modified to rise sufficiently at the end, the utterance is perceived as a (yes-no) question, whereas if intonation falls sufficiently, the stimulus is judged to be a statement.

As regards L2 intonation, Lengeris (2012) suggests that language learners' accent and intelligibility is influenced more by L1-like production of prosody than of phonemes. However, training the perception of pitch changes in L2 utterances results in improved, more comprehensible speech production. Such training can be effectively implemented by computer programs which allow the learner to see visual representations of the pitch changes while listening to utterances spoken by native speakers, and then compare them to visualizations of the learner's own production of the utterances. In this way, learners become aware of the way L2 intonation is realized, and their perception as well as production of suprasegmentals develops. What this implies, however, is that late L2 learners' perception of L2 intonation is by default suboptimal, affected greatly by the L1, and it does not improve automatically with L2 experience.

In this way, the differences in English and Hungarian intonation presented above may cause a problem in the perception of L2 English speech for Hungarian learners (or vice versa) as the modality of certain utterances may be inaccurately perceived. Admittedly, syntax, meaning and context also aid comprehension, so a breakdown of communication is not inevitable; however, depending on the learner's proficiency, such cues may not be perfectly available in certain communicative situations. Learners of English may then profit from training their perception of L2 intonation patterns.

Auditory perception and the pattern recognition theory of mind

Any discussion about speech perception can benefit from a basic understanding of the big picture: the way auditory information is processed. In this section, a brief review of only the key literature of this broad topic will follow. This will set the scene for the next chapter, *Analysis*, where the relevance of this literature will readily become apparent.

The processing of auditory information as described by Blake and Sekuler (2006) can be straightforwardly related to the suprasegmental features of speech discussed above, but not as easily to phonemes. This is because what we call stress and intonation in speech are generally equivalent to loudness and pitch, respectively, which in turn can be connected to physical measures: sound intensity for loudness (stress) and frequency for pitch (intonation). As the modulations in intensity and frequency can be measured by machines in decibels and hertz, so can the nerves of the auditory system register and interpret these measures. Some sounds only consist of one frequency – these are pure tones such as the sound of a tuning fork –, but even if a sound comprises multiple frequencies – as is the case for speech, but also noise –, the abundant amount of nerves can successfully monitor all of them by dividing the work, with different nerves responding to the changes of different frequency bands.

However, the recognition of segmental information – phonemes – is a much more complex task because there is no one-to-one correspondence between a constellation of frequencies and a phoneme. One speaker does not produce the same sound twice in exactly the same way; two speakers' production of the same sound are also different; and one sound is pronounced differently depending on what sounds surround it (/s/ is different when it is followed by /u/ as opposed to /i/ because the rounding or spreading of the lips already starts while articulating the consonant.) This means that the frequencies which make up the sound differ significantly, not only in absolute but also in relative terms. Blake and Sekuler (2006) conclude that the processes of auditory perception are not sufficient to explain speech perception in the sense of phoneme recognition. For that, other processes need to operate, including feature detection: extracting from the acoustic signal the features of phonemes, such as voicing, and place and manner of articulation. It remains unclear from this explanation how such an extraction is achieved.

The claim that auditory perception and feature detection rely on distinct processes is debatable. The processing of sounds – speech or otherwise – can be explained by Kurzweil's (2012) pattern recognition theory of mind in a unified way. According to this theory, the brain processes information with the use of pattern recognizers[2]. Any input that the brain receives is comprised of hierarchically organized patterns. In our example, where the input is (L2) speech, the acoustic signal is a pattern, the features to be extracted from it (voicing etc.) are also patterns made up of the acoustic signal, and the phonemes themselves are patterns made up of the features. Recursion, after all, is a fundamental characteristic of language.

As for the recognition of phonemes despite the great variability of their realizations, the theory posits that any pattern is stored with an amount of redundancy: A category such

---

[2] A pattern recognizer is an actual physical entity: a group of nerves.

as a phoneme is represented by multiple examples, and there are also several pattern recognizers for one pattern. For this redundancy to be attained, the pattern has to be encountered and recognized many times. This enables a variety of sounds to be accepted as an instance of a certain phoneme.

Further, there is communication between the pattern recognizers at any two levels of the hierarchy, and this communication is bidirectional[3]. If the pattern "voicing" is recognized at the level of feature detection, the next level, phoneme recognition will prepare for recognizing voiced phonemes and rejecting voiceless ones. Meanwhile, the following level, word recognition can inform the phoneme recognizer that a certain phoneme is likely to occur soon, and then that phoneme becomes more likely to be recognized. This is what makes perception a predictive process. It has to be such because information is flowing in so rapidly that its processing is not viable without predictions.

Finally, the pattern recognition theory of mind explains the findings mentioned earlier about the differences in the acoustic salience of phonetic features (Strange & Shafer, 2008; Polka, 1991). Namely, it has been shown that the length of vowels is more salient than their quality, while the voicing of consonants is more salient than their place of articulation; in other words, temporal features are easier to recognize. Kurzweil (2012) argues that the size of a pattern – in the case of speech, its temporal duration – is the most reliable parameter in its recognition. However much one would like to believe that the difference in vowel quality between the phonemes /iː/ and /e/ is as important as their difference in length, this is not the case from the viewpoint of speech perception.

In summary, speech perception is a process of recognizing higher and higher levels of patterns: basic physical measurements for suprasegmentals, then features and phonemes. This process is made possible by the high rate of redundancy at which the information is

---

[3] Communication between two pattern recognizers is realized through the connections between the nerves which constitute them.

stored. Importantly, speech perception is predictive. Next, let us see how this seemingly broad description is immensely relevant for the way L1 phonology affects L2 speech perception.

<center>**ANALYSIS**</center>

Having reviewed the findings of the L2 speech perception literature, it is apparent that the various levels of L1 phonology play a role in the perception of L2 speech. So far, this thesis has established the *outcome* of a process – namely, that speech perception is hindered to varying degrees by certain factors. It is now time to analyze the mechanisms which underlie this process: How – and why – is it that such an outcome is attained?

The acquisition of L1 phonology is essentially a period of collecting redundant copies of patterns. These patterns include phonemes, possible combinations of phonemes, and stress and intonation systems. Infants must learn to associate these with patterns further down the hierarchy: acoustic signals. Once this is achieved, not only these patterns can be recognized during speech perception, but predictions can also be made as to what other patterns are likely to occur next, increasing the efficiency of perception. Notably, predictions can only be made based on the patterns already stored with great redundancy, which is achieved by experiencing the pattern several times.

What happens if the input is L2 speech? Initially at least, patterns will be recognized and predictions made based on the stored information – i.e., based on the L1 phonology. This is not entirely disadvantageous; in fact, it is highly beneficial because this is the only way to recognize non-native speech as speech, and start learning the language. L2 perception will be less optimal in this way, but it will still be sufficiently optimal – much more so than

having to acquire another phonology from scratch. From the point of view of human evolution, this mechanism is not a bad arrangement at all.

From the point of view of the language learner on the other hand, the minor inaccuracies in pattern recognition at the level of phonology can exert their influence on the recognition of patterns on the next higher level, or on the predictions as to what is going to happen soon on the previous level. With enough L2 experience, the accuracy of perception may improve, but not inevitably. Let us now interpret the findings about the perception of L2 phonemes, phonotactics, and suprasegmentals in light of Kurzweil's (2012) pattern recognition theory of mind.

Broersma and Cutler's (2011) Dutch participants perceived the English phoneme /æ/ as /e/. This is because these speakers did not acquire the pattern /æ/, their native language lacking such a phoneme. However, because they store the pattern /e/ with great redundancy, the acoustic signal of the sound /æ/ was recognized as an instance of this phoneme. The /æ/ – /e/ distinction being one in quality, its acoustic salience is low, which is why this difference continues to be difficult to perceive even for proficient L2 speakers of English: A new /æ/ pattern can hardly be established.

Not any two speech sounds are, of course, similar enough to be recognized as the same pattern, but the distinction must be salient to avoid such assimilation. This is accomplished if the difference is temporal (such as the /iː/ – /ɪ/ distinction in Strange & Shafer, 2008), but non-temporal cues may also be salient enough (the /ʌ/ – /ʊ/ distinction in the same article).

Concerning phonotactics, the predictions made by the pattern recognizers may play an important role. Take as an example a language like Japanese, where two consonants cannot normally follow each other. In this case, once the pattern for a consonant is recognized, all of the other consonant pattern recognizers receive an inhibitory signal: a

warning that a consonant is very unlikely to appear next. Recognizers for vowel patterns on the other hand will receive the opposite message: They should lower their activation thresholds because a vowel will probably occur. This is why the Japanese learners of English in Sperbeck and Strange's (2010) study did in fact perceive the vowel /ə/ in CC clusters: The activation threshold for this vowel was so low that essentially any acoustic signal evoked its recognition.

What does this theory predict for cases when the phonotactics of the L1 and L2 is more similar than that of Japanese and English – so for instance, in the case of Hungarian and English? There are still some differences, for example word-initial /mj/ is unattested in Hungarian, but found in English (in words like *mule, music*). In this case, the Hungarian learner of English – in the initial stages of learning – does not store a pattern for word-initial /mj/, so the pattern recognizers will find it unlikely that such a string is present in the input. If word-initial /m/ is detected, the activation threshold for the /j/ recognizer will be raised. The pattern /j/ might still be recognized in this case, but with more effort and time. However, growing experience with the L2 means that such strings will be stored with greater and greater redundancy. Therefore, the processing of such clusters should not cause any problems for proficient speakers.

Finally, the perception of L2 suprasegmentals can be explained with this theory. If stress in the L1 is predictable, then the pattern recognizers learn to rely on this characteristic with a great amount of certainty. In Hungarian, for example, if a pattern for a word is recognized, then the threshold for the stress recognizer will lower considerably: Surely, the next syllable is stressed, as it is the start of a new word. Vice versa, if a word was not successfully recognized, but a stressed syllable was detected, the most active word recognizer will be notified, and the word will then be accessed.

However, if the L2 has lexical stress like English, the predictions about L2 speech input will continually fail. The perceptual system will then adapt to these circumstances by making the predictions (setting the activation thresholds) based only on the current speech input. This is why the brain activity of Honbolygó et al.'s (in press) participants was influenced more by short term patterns.

In real life situations, however, the Hungarian L1 system will not be able to detect reliable stress patterns in L2 English speech, so it will have a hard time making predictions. Without predictions, the pattern recognition mechanism is less efficient. It is also not surprising that in experiments, the cognitive load of the task modifies the results: The more patterns there are to recognize at a given time, the more likely it is that some of them will remain undetected.

Regarding intonation, it can be said that modulations in the fundamental frequency of speech are recognized as patterns, after which they activate higher level patterns such as "Wh-question", "indifference", and the like. The L1 and L2 can differ in how the perceptual system interprets the pitch changes. For instance, the pattern "frequency rises slightly" activates the higher level pattern "low rising intonation" in both Hungarian and English, but the next higher level pattern differs between the two languages: In Hungarian, the pattern "yes-no question" is recognized, even if the English input was intended as "indifference."

A further example is if the L2 English input uses high rising intonation, which is characteristic of echo questions in this language. The activation of the pattern recognizer for "echo question" will be inhibited for L1 Hungarian speakers, because this is not a typical intonation for that purpose in Hungarian. The meaning of the utterance may still be recognized based on other patterns, but this will be a slower process. Training the perception of L2 intonation with the method mentioned in Lengeris (2012) works in as much as it increases the redundancy of L2 intonation patterns in the mind of the language learner.

CONCLUSION

The present paper aimed to explore the way in which the phonology of the L1 influences the perception of L2 speech. The existing literature provided answers to some of the questions connected to the topic, but so far a comprehensive examination was lacking.

In summary, the L2 speech perception literature demonstrates the role which the L1 plays in L2 speech perception. Namely, L2 phonemes are likely to be perceived based on L1 categories, depending on how salient the difference between the two categories is. If it is not salient enough, perception continues to be inaccurate even when L2 proficiency grows. This is not the case for the perception of phoneme combinations: L1 phonotactic constraints modulate L2 perception initially, but L2 proficiency may override this effect. As for the perception of suprasegmentals, the influence of the L1 is more robust in the case of intonation than in the case of stress, but even intonation perception can be improved with conscious effort.

The how and why of the L1 influence remained unexplained in this literature. Even in broader, neuroscientific accounts of stimulus perception, the recognition of phonemes is explained away by theorizing distinct processes for the perception of speech and non-speech auditory stimuli. Suspiciously, the suprasegmental aspects of speech could, in this account, be regarded as non-speech stimuli, which in reality they are not.

The solution proposed in this paper is to interpret the observed L1 influence with reference to the pattern recognition theory of mind. This theory has already been used to account for speech processing. The present thesis applied the theory to reveal the mechanisms behind the language learner's speech perception difficulties. It can be concluded that the effect of L1 phonology is exerted in a process of recognizing higher and higher levels of patterns in the L2 speech signal. The reason the L1 influences this process

is that the L2 patterns are stored with low redundancy or not at all, so the input is analyzed

by the system in terms of L1 patterns.

<center>**REFERENCES**</center>

Blake, R., & Sekuler, R. (2006). *Perception.* New York, NY: McGraw Hill.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *The Quarterly Journal in Experimental Psychology, 64*(1), 74–95. https://doi.org/10.1080/17470218.2010.499174

Carlson, M. T., Goldrick, M., Blasingame, M., & Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Language and Cognition, 19*(5), 939–954.

Freeman, M. R., Blumenfeld, H. K., & Marian, V. (2016). Phonotactic constraints are activated across languages in bilinguals. *Frontiers in Psychology, 7,* 702. https://doi.org/10.3389/fpsyg.2016.00702

Gass, S. (1984). Development of speech perception and speech production abilities in adult second language learners. *Applied Psycholinguistics, 5,* 51–74.

Gósy, M. (2005). *Pszicholingvisztika.* Budapest, Hungary: Osiris.

Honbolygó, F., Kóbor, A., German, B., & Csépe, V. (In press). Word stress representations are language-specific: Evidence from event-related brain potentials. *Psychophysiology.* 2020;00:e13541. https ://doi.org/10.1111/psyp.13541

Kurzweil, R. (2012). *How to create a mind.* Richmond, UK: Duckworth.

Lengeris, A. (2012). Prosody and second language teaching: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.), *Pragmatics and prosody in English language teaching* (pp. 25–40). Dordrecht, The Netherlands: Springer.

McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. L. Goldstone (Eds.), *Handbook of cognition* (pp. 255–275). London, UK: SAGE Publications.

Nádasdy, Á. (2006). *Background to English pronunciation.* Budapest, Hungary: Nemzeti

    Tankönyvkiadó.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. In C.

    Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 203–240). Berlin,

    Germany: Mouton de Gruyter.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and

    acoustic contributions. *Journal of the Acoustical Society of America, 89*(6), 2961–

    2977. https://doi.org/10.1121/1.400734

Segui, J., Frauenfelder, U., & Hallé, P. (2001). Phonotactic constraints shape speech

    perception: Implications for sublexical and lexical processing. In E. Dupoux (Ed.),

    Language, brain and cognitive development (pp. 195–208). Cambridge, MA: The MIT

    Press.

Shatzman, K. B., & Kager, R. (2007, August). *A role for phonotactic constraints in speech*

    perception, 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.

Sperbeck, M., & Strange, W. (2010). The perception of complex onsets in English:

    Universal markedness? *University of Pennsylvania Working Papers in Linguistics,*

    16(1), 195–204.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The

    re-education of selective perception. In J. G. H. Edwards & M. L. Zampini (Eds.),

    Phonology and second language acquisition (pp. 153–191). Amsterdam, The

    Netherlands: John Benjamins Publishing Company.

Varga, L. (1983). Hungarian sentence prosody: An outline. *Folia Linguistica, 17*(1–4),

    117–152. https://doi.org/10.1515/flin.1983.17.1-4.117

Wales, R., & Taylor, S. (1987). Intonation cues to questions and statements: How are they

    perceived? *Language and Speech 30*(3), 199–211.