Faculty of Arts Palacky University

Department of English and American Studies

Native and Non-Native Speech Perception in Noise

(Diploma Thesis)

Author: Bc. Tomáš Sedláček

Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.

Olomouc 2022

I declare that I elaborated the diploma thesis on my own, using solely the sources listed in the references.

In Olomouc on 4th of May 2022 Bc. Tomáš Sedláček

....

1. ANNOTATION

Native and Non-Native Speech Perception in Noise

(Diploma Thesis)

Author: Bc. Tomáš Sedláček

Field of Study: English and Spanish Philology

Faculty and Department: Philosophical Faculty, Department of English and American Studies

Supervisor: Mgr. Václav Jonáš Podlipský, Ph.D.

Number of characters: 219 338

Keywords: Speech perception, noise, L1 vs. L2 speakers, longitudinal immersion, phonetic training, Czech speakers of English

Description: The current thesis investigates native and non-native perception of speech in noise. It focuses on literature review which is divided into two sections, the first of which provides information about recognition of speech sounds, models of phonetic category formation, L2 learning factors and speech-in-noise specificity, the second one analyzes studies which focus on perception of non-native speech. The last content chapter, based on the acquired knowledge, gives advice towards future non-native perceptual research in noise with a special focus on Czech speakers of English in an immersion setting.

1.1 Anotace v češtině

Percepce rodilé a nerodilé řeči v šumu

(Diplomová práce)

Autor: Bc. Tomáš Sedláček

Studijní obor: Anglická a Španělská filologie

Fakulta a katedra: Filozofická fakulta, Katedra anglistiky a amerikanistiky

Vedoucí práce: Mgr. Václav Jonáš Podlipský, Ph.D.

Počet znaků: 219 338

Klíčová slova: Percepce řeči, šum, rodilí a nerodilí mluvčí, longitudální imerze, fonetický trénink, čeští mluvčí angličtiny

Charakteristika: Diplomová práce se zabývá percepcí řeči rodilých a nerodilých mluvčích v šumu. První obsahová kapitola věnující se přehledu literatury je rozdělena na dvě části. První část se zabývá obecnými fakty ohledně percepce řeči a šumu, představuje také modely vzniku fonetických kategorií, druhá část analyzuje fonetické studie, které se věnují především percepci nerodilé řeči v šumu. Poslední obsahová kapitola na základě získaných poznatků poskytuje doporučení, která by mohla usnadnit výzkum nerodilé percepce řeči v šumu. Zvláštní důraz se klade na případný výzkum, ve kterém by figurovali čeští mluvčí angličtiny.

Acknowledgement

I would like to thank my supervisor Mgr. Václav Jonáš Podlipský, Ph.D. for his guidance and useful advice throughout the creation of this thesis.

Table of Contents

2. INTRODUCTION

It is well documented that non-native perception can reach a similar success rate as native perception in quiet as suggested by e.g. Mayo et al. (1997), Febo (2003), Rosenhouse et al. (2006), Tabri et al. (2010) and Masuda (2016). A fact that is especially true for early bilinguals, non-native speakers, who learned their L2 early in their childhood (prior to 6 years of age) as suggested by e.g. Mayo et al. (1997) and Febo (2003). Moreover, even for late bilinguals, non-native perception can reach a high success rate or even a native-like level as suggested by e.g. Bongaerts (1999).

Even though it may not manifest itself when they perceive L2 speech in quiet, non-native speakers may rely on different perceptual cues or different cue combinations or the same cue combinations but weighting each cue differently, compared to the native speakers, as suggested e.g. by Kondaurova and Francis (2010). Nevertheless, it may be that the effects of such native vs. non-native differences appear when adverse conditions come into play. Those include different kinds of noise or reverberation of non-sound-attenuated rooms, in such cases we can see a significant detriment in the performance of the non-native listeners, compared to the native ones. This is also true for early bilinguals, for whom we would expect more robust, stable results, since they learn their L2 early in their childhood.

A question we would like to raise is whether non-native perception can be improved via intensive native input, e.g. participating in exchange programs in a foreign-speaking country. The current thesis reviews literature that concerns three areas, which could shed light on potential future research, aiming at presenting a template by means of which an experiment on non-native speech perception in noise in an immersion setting can be carried out. A special focus is devoted to Czech speakers of English whose particularities are mentioned in the future research section.

By looking at the provided evidence, the best way to perform a potential experiment would be to devise a longitudinal study testing the participants' level of non-native perception in adverse conditions before they leave the country and testing them again upon their return to the home country, which should provide enough evidence in favor of or against improving non-native speakers' perception through increased native input.

The issue with longitudinal studies revolves around the fact that the researchers need a longer time span to perform such study, which might be the reason why surprisingly little research has focused on immersion experience in L2 environment and its effects on strengthening the robustness of L2 speech perception in noise, by, for instance, altering the participants' cue selection, cue-weighting or improving their cue extraction. The evidence for the future research section is thus derived from three types of studies with L2 participants: non-longitudinal non-native perceptual studies in noise, non-native longitudinal studies without noise and studies training non-native speech perception in a shorter time span, most of which do not employ noise and are carried out in laboratory conditions.

The current thesis offers a review of literature divided into two sections, the first of which, chapter 3, provides information about recognition of speech sounds, models of phonetic category formation, L2 learning factors and speech-in-noise specificity. The second one, chapter 4, analyzes experiments, which focus on non-native speech perception with/without noise. The last content chapter, based on the acquired knowledge, gives advice towards future perceptual research in noise with a special focus on Czech speakers of English.

3. LITERATURE REVIEW – ESSENTIALS

The current chapter is divided into subchapters: the first one focuses on how human speech is recognized. The following subchapter presents two models of speech learning, i.e. the Speech Learning Model by Flege (1995) and the Perceptual Assimilation Model by Best (1995) and their revised versions, hoping they would shed light on how L2 learners assimilate L2 sounds. We then present factors that might influence L2 learning and offer ideas about speech in adverse conditions.

3.1 Recognition of Speech Sounds

When talking about recognition of speech sounds, one must mention that we should look up to an idealized communication model, which, according to Miller (1951, 6-7), comprises of five key components: source, transmitter, channel, receiver and destination. As we are talking about human communication, the source signifies a person, someone who has information to pass to others, it can be another human, an animal or even a computer that can react to speech sounds. The sound emanates from the source with the help of the transmitter, be it the "human speech machinery," as Miller (ibid, 7) calls it, loudspeakers or headphones. Every sound has to be carried through a channel, a medium, most commonly the air, through which the sound travels to the receiver, in our case a person, an animal or a microphone. This receiver perceives the acoustic sound waves and either converts "them into nervous activity at their destination, the nervous system of the listener" or the microphone registers the sound waves and a computer program identifies them as certain speech sounds in the case of automatic speech recognition (ibid, 7).

English, same as all other languages, is a code, a set of symbols basically agreed upon by a community of talkers. What the transmitter does is called encoding of the source's message. This code can consist of many things, it can be a speech sound, a piece of a written language, a motion of someone's arms or any other convenient symbol (ibid, 10). On the other hand, when the message reaches the receiver, they are then faced with the reverse process, they have to extract from the code what the message was meant to be. We call such process decoding.

Such process, however, has to do with more than just speech perception, higher levels of language are also involved.

When perceiving speech sounds, it is important to mention that listeners have help in recognizing what kind of speech sounds they hear, unnecessary information is contained in the message that travels from the source to the receiver, more specifically, we are talking about redundancy of linguistic information which aids the listeners' successful recognition. The advantages of redundancy reside in being able to maintain communication even though parts of the message can be lost or deformed, as other adjacent portions provide enough information to recover the lost portions (ibid, 103). An important aspect of speech perception is that it is not only based on the 'bottom-up' sensory information about the incoming speech signal but also on the 'top-down' predictions and expectations based on prior linguistic experience (e.g. Cutler 2012).

Hermansky (2019, 112-114), in concordance with Miller, suggests that large amount of information is transmitted by the speech signal, much more than needed for successful communication, meaning that there is of course much information irrelevant to the message, more specifically, we are talking about, for example, speaker identity, gender, health, emotions, moods and other speaker-specific idiosyncrasies. Such information related to the speaker is secondary and does not usually assist the listener in decoding the message, nevertheless, it can still be of some use to them, if not for the successful decoding of the message, then for their psycho-social perspective, which is the reason those specificities are noticed. Hermansky lists two planes worth considering when talking about redundancy that could help the listeners perceive the message with better accuracy, in particular, those with regard to temporal and spectral domains. As for the former redundancy, he talks about vocal tract changes as results of transitions between each speech sound, we can thus say that there are certain delays, which create temporal redundancies. In the spectral domain, the redundancy resides in the fact that during speech production, the vocal tract constantly changes its shape, which undoubtedly influences the frequencies of the produced speech sounds. The idea of redundancy in the process of coding and decoding can be seen in Figure 1. Taking us closer to our topic is the fact that the majority of human communication takes place in an environment which contains quite a lot of background noise or reverberation and that is where those redundancies come into play, since they aid the listener in recognition of perceptual cues on which the listener bases the recognition of speech sounds, leading to successful understanding of the message.

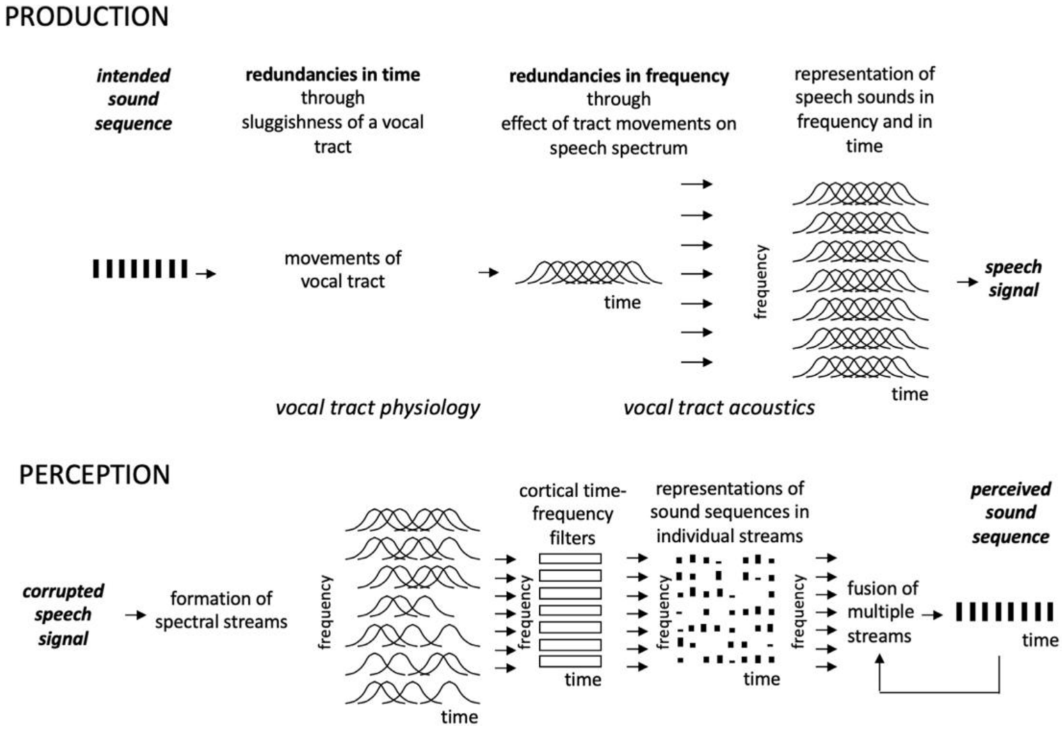

PRODUCTION
intended sound sequence
redundancies in time through sluggishness of a vocal tract
redundancies in frequency through effect of tract movements on speech spectrum
representation of speech sounds in frequency and in time
movements of vocal tract
time
frequency
speech signal
time
vocal tract physiology
vocal tract acoustics
PERCEPTION
corrupted speech signal
formation of spectral streams
frequency
time
cortical time-frequency filters
frequency
time
representations of sound sequences in individual streams
frequency
time
frequency
fusion of multiple streams
perceived sound sequence
time

Figure 1 represents the idea of redundancies generated both in time and in frequency (adapted from Hermansky 2019, 113).

3.2 Models of Phonetic Category Formation

We chose to review the models predicting speech learning, because they could clarify, how an immersion experience might influence foreign language (FL) perception in noise, the models will give us more information about FL perception and make us understand, how L2 listeners acquire novel L2 sounds. In addition, they will also shed some light on the difference between early and late learners. Moreover, such models could help predict which vowel and consonants will cause trouble to the non-native listener. In this section, we shall look at two different models predicting phonetic category formation: Flege's Speech Learning Model (SLM) and Perceptual Assimilation Model for L2 (PAM-L2) by Best and Tyler. Revised versions of the models, SLM-r and PAM-L2 for foreign language acquisition will also be considered.

3.2.1 Flege (1995) and His Speech Learning Model (SLM)

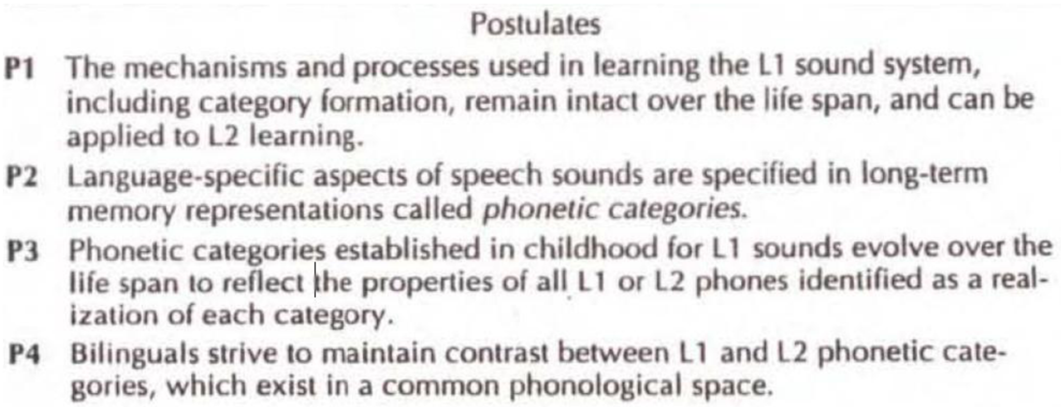
This model was proposed by James Emil Flege in a chapter "Second language speech learning theory, findings, and problems" published in 1995 as part of a book *Speech perception and linguistic experience: issues in cross-language research* edited by Winifred Strange. In this chapter, Flege considers how foreign accent affects the L2 speaker, investigates its cause and repercussions on speech production. The author presents his Speech Learning Model and also summarizes empirical research.

Intrinsically connected with L2 learning is the presence of foreign accent (FA), which usually comes with some undesirable consequences, and as such, if the L2 learner is aware of his FA, he has a desire to dispose of it. Flege (1995) mentions the difficulty to perceive L2 contrasts, intensified by unfavorable listening conditions, which are rather common, considering where and under which conditions most human speech takes place. There could also be the listener's prejudice arising from group stereotypes connected with the speaker's accent. Ura et al. (2015) provided evidence for such stereotypes, more specifically, they found foreign accents to influence discriminatory judgments, as well as linguistic prejudice to negatively affect attitudes towards immigrants and foreigners.

Flege also looks into reasons why FA arises (ibid, 234). He mentions neural plasticity reduction as a result of ageing leading to weakened senso-motoricity when producing L2 sounds. This concept is connected to a hypothesis that has received a lot of attention over the years being known as the critical period (CP), the supporters of which speak of the impossibility of perfect attainment of new forms after certain age. Nevertheless, Flege is against the idea of CP, stating in postulate one that the mechanisms and processes used in learning L1 are still applicable to learning L2, as we can see in Table 1. In addition, FA could also be caused by inaccurate L2 perception, namely when the L2 listeners are unable to assess accurately the distinctive properties differentiating L2 sounds from those differentiating L1 sounds. In addition, inadequately stored and structured information of such properties could also lead to less accurate L2 perception. Other causes of FA include inadequate phonetic input, lack of motivation; psychological block to sound more native-like, as doing so might jeopardize the speaker's social group togetherness. Furthermore, poor habits might have been formed during the beginnings of speaker's L2 attainment and as such might influence his performance.

Flege's Speech Learning Model (ibid, 237-243) aims at a complete attainment of L2 pronunciation, which means that it does not focus on beginner learners of second language, but rather on bilinguals with long experience with a given language. The model also tries to account for the production of L2 vowels and consonants, considering the problems related to age, which seem to limit the possibility of sounding accent-free. As we acquire our L1, we learn to distinguish all phonemic contrasts, we become attuned to sounds which are distinctive between categories, called phonemes, and those that are distinctive within each category, represented by allophones or free variation. Nevertheless, it is not as easy with the L2, a learner "may fail to discern the phonetic differences between pairs of sounds in the L2, or between L2 and L1 sounds" due to the fact that either a pair of contrasting L2 sounds is classified to a single L1 category equivalently, or L1 phonology serves as a filter to L2 features but only to those that are phonetically important (ibid 238). The model talks about there being a substantial effect of accurate perception targets guiding the senso-motoric L2 learning. If there is a lack of such targets, no flawless L2 production is expected. It is not to say that "all L2 production errors are perceptually motivated [... nevertheless] many L2 production errors have a perceptual basis" (ibid, 238). Flege's model thus suggests a strong perception-production link with perception driving (un)successful production.

Flege bases his model on several postulates which he takes as a starting point to formulate his seven hypotheses for L2 sound acquisition, the postulates are presented in Table 1 below. Flege's first hypothesis focuses on allophones in the L2 which are, according to Hypothesis 1 (H1), related to the phonetic categories perceptually closest in the L1, he thus argues that this perceptual link is on the positional perceptual level, rather than on the more abstract phonemic. This particular hypothesis is based on evidence from studies that show better L2 learners' success perceiving or producing only some, but not all, allophones of certain phonemes, as is known for the native Japanese learners of English, who are more accurate in word-final compared to word-initial position (ibid, 239).



Postulates

P1 The mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning.

P2 Language-specific aspects of speech sounds are specified in long-term memory representations called *phonetic categories*.

P3 Phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.

P4 Bilinguals strive to maintain contrast between L1 and L2 phonetic categories, which exist in a common phonological space.

Table 1 presents us with the postulates on which Flege builds his seven hypotheses about L2 sound acquisition (adapted from Flege 1995, 239).

The second hypothesis deals with the formation of new phonetic categories. At the starting point it seems that L2 learners identify L2 sounds as those present in their L1, the learners identify them either as belonging to the same continua as those of L1, or they regard them as non-speech. While the first encounters with a foreign language may lead to L2 speech sounds being recognized as instances of L1 sound categories, it seems that later on L2 learners notice the cross-language phonetic differences and it is possible that they establish new phonetic categories for L2 sounds, and they will ultimately end up producing them according to their phonetic category representation. In case the established category is the same as that of the native speaker, accurate production is expected.

The third hypothesis mentions the fact that the more an L2 sound is perceptually different from the closest L1 sound, the more likely it is that the L2 learner recognizes such a difference, and as a result more accurate learning ensues. As support for this hypothesis, Flege mentions the evidence found for the native Japanese learners of English, more specifically, Japanese /r/ is perceptually closer to the English /l/ than /ɹ/, consequently more accurate learning of English /ɹ/ ensues.

Challenging the aforementioned neurological maturation hypothesis gave rise to SLM’s fourth hypothesis, advocating, as Age of Learning (AOL) increases, more continuous loss in the recognition of “phonetic differences between L1 and L2 sounds, and between L2 sounds that are non-contrastive in the L1” (ibid, 239). To support his argument against complete neurological maturation as a cause of the age-related effects in L2 speech learning, Flege (ibid, 234-235) presents data from a previous study, Flege et al. (1995), which investigated the degree of foreign accent of native Italian (NI) speakers of English, living in Canada for over 30 years on average and differing in their AOL. They found that "the later in life the NI subjects begin learning English, the more strongly foreign-accented their English sentences were judged to be," so there is gradual increase in likelihood of having a FA with increased AOL.

Flege bases his fifth hypothesis on language interference, which, according to him, is bi-directional, more specifically, not only has L1 an influence on L2, as would be suggested by the traditionalists, there is also the reversed influence, namely an L2 can change the way we produce L1 speech sounds, as "L1 and L2 sounds that are perceptually linked to one another (diaphones) come to resemble one another in production" (Flege 1995, 241).

The following hypothesis builds on the idea of a common phonological space for L1 and L2 categories. Flege finds evidence for his arguments in the historical sound change, more specifically the vocalic chain shifts which led to e.g. the Great Vowel Shift which occurred in the Middle English. Nevertheless, the evidence for interference between L1 and L2 can also be found in the production of voice onset time of bilinguals, in particular, Flege lists a case in which a 10-year-old boy maintained phonetic contrast between three different categories across two languages, he "spoke French at home and English elsewhere [and] produced /b d g/ with short-lag voice-onset time (VOT) values in both French and English. He produced /p t k/ with VOT values averaging 66 ms in French but [...] 106 ms in English" (ibid, 242). Even though, the child is able to maintain three phonetic contrasts it is at the cost of producing those contrasts with different values compared to the native monolingual. A deflection of categories in both L1 and L2 may thus take place when considering a speaker proficient in both languages.

It may also be the case that instead of deflection, the bilingual establishes a category which is based on different features or feature weights than that of a monolingual, and as such "*L2 sound might not be produced exactly as it is* [...] *by native speakers*" (ibid, 243). As a result, Flege argues that the mastery of a certain language is unattainable in the same way as monolingual attains it, due to the fact that there is unavoidable mixing of L1 and L2 as the bilingual's both languages are constantly engaged (ibid, 243).

3.2.2 The Revised Speech Learning Model (SLM-r)

By publishing the revised SLM, Flege and Bohn (2021) tried to clarify how early and late learners differ. In addition, they sought to capture the dynamicity of phonetic systems, that is, how they “reorganize over the life-span in response to the phonetic input received during naturalistic L2 learning” (Flege and Bohn 2021, 3). There are several aspects which define the SLM-r, some of them based on the SLM, others diverging.

Similar to SLM, SLM-r follows sequential bilinguals, persons who learn L1 and start learning their L2 only after reaching a certain level of experience in their L1, this is in contrast to simultaneous bilinguals, who are learning both L1 and L2 at the same time. SLM viewed early or late L2 speakers as end-state, while SLM-r recognizes that both L1 and L2 influence each other incessantly. Due to such dynamicity, the bilinguals’ speaking and listening abilities will never completely match those of the monolinguals. Moreover, the input received by the bilinguals, upon which they base their novel L2 phonetic categories, is not the same as for the native speakers.

For cross-language dissimilarity, SLM-r is in concordance with SLM, supporting the hypothesis that the L2 learners relate an L2 phone to L1 phonetic categories in a subconscious and automatic way. The more dissimilarity between the L2 and L1 sounds, the more likely it is for learners to form new L2 phonetic categories. Moreover, in some phonetic contexts, the speakers may perceive the phones as more similar to L1 category than in others.

Besides the similarity and dissimilarity of the L2 sound and L1 sound, the model also mentions quantity, as well as quality of the input which the learner obtains from meaningful conversation as a potential factor in the formation of new L2 category. The quality of L2 input is rarely taken into consideration in other studies. For example, its importance can be observed for native Spanish speakers of L2 English, who learned English in youth but were mostly exposed to Spanish-accented English, their English VOT values were short, resembling the native Spanish speakers who learned English in adult age (Flege and Hammond 1982).

Furthermore, the degree of precision of the closest L1 category to the L2 sound in question may also play a role. The authors expect L2 learners with more precise L1 phonetic categories to have an easier time discerning phonetic differences and thus they are more likely to form new L2 categories. As an effective procedure to obtain the data, the authors recommend gathering two judgments of the same stimulus from random tokens of L2 phonemes presented for classification in L1 categories, also labeling them for perceived dissimilarity. In addition, according to SLM-r, the rating of dissimilarity should be obtained at an early stage of L2 development in order to have a predictive value, because the L2 learners' perceptions are subject to change as their development progresses.

Being against the critical period (CP), which advocated neural maturation and the impossibility to learn L2 after certain age, SLM hypothesized that with increasing age of first L2 exposure, decreases the prospect of developing a new L2 category. SLM-r opposes the CP and the aforementioned SLM hypothesis, suggesting that first exposure is not significant, promoting the idea of a continuous development without there being a single point after which the L2 learners are unlikely to attain the L2. The authors view the formation of the L2 phonetic category as crucial for the learner's phonetic organization and mention that if such category is not formed, a blend of L1 and L2 phonetic category, exhibiting features from both languages, will come to rise.

Another aspect in which SLM and SLM-r differ is how they view native vs. non-native category formation as far as native features are concerned, the former advocates the feature hypothesis, according to which unused features in the L1 are inaccessible to the L2 learners due to lower perceptual acuity and when new L2 phonetic categories are formed, they are different from those formed by the L1 speakers. SLM-r, on the other hand, defends the full-access approach, on the bases of which "L2 learners can gain access to such non L1-features [... that is to say] that all processes and mechanisms used to develop L1 phonetic categories [...] remain intact and accessible for L2 learning" (Flege and Bohn, 65).

No reliable method exists, which would determine whether an L2 phonetic category was created. A promising method of research, however, is brain imaging. While testing the L2 speaker, new L2 category would respond to more activity in frontal speech regions and would provide evidence for/against the emergence of such a category. SLM-r proposes that most differences in L2 learners are based on the following aspects: the specification of L1 phonetic categories, as in cue weighting and category precision, mapping L2 sounds to L1 categories, perceived dissimilarity between L2 sounds and their closest L1 counterparts and the amount and quality of L2 input they received. The authors suggest that further research, based on SLM-r, needs to become more individualized, "treating each individual as a separate experiment" (Flege and Bohn 2021, 59).

SLM-r is relevant to the topic of L2 speech perception in noise before and after being immersed in the L2 environment, because it promotes continuous development of L2 learners without referring to first exposure to L2. The non-native learners are thus, according to this model, expected to develop their L2 perception in noise during their immersion even without being exposed to L2 early in their life. Moreover, the authors also hypothesize the full-access to non-L2 features which suggests further L2 learning and the possibility of having a more native-like perception in noise.

3.2.3 Perceptual Assimilation Model

In this subsection, we will briefly summarize the original PAM by Best (1995) aimed at naïve monolinguals. It predicts how naïve listeners perceive L2 sound contrasts by exploring the possible ways in which non-native phones can be categorized via L1 phonological categories. The listeners can only distinguish phonetic and phonological levels in their L1 "in which perceived differences at the phonetic level became systematically related to the functional linguistic categories of a phonological system during early lexical and grammatical development" (Best 2007, 18). Due to the fact that non-native naïve listeners cannot distinguish the difference between phonetic and phonological plane of unfamiliar language, Best suggests an attempt of the L1 system to constrict the unfamiliar phones by their phonological sieve, suggesting a bridge between the phonological and phonetic levels, in other words, upon hearing non-native phones, the naïve listeners tend to perceive them "according to their similarities to, and discrepancies from, the native segmental constellations that are in closest proximity to them in native phonological space" (Best 1995, 193).

According to PAM, there are three possibilities of how the L2 contrasts are categorized in the mind of the non-native listeners (ibid, 194-195). They are *categorized*, in case they hear the phones as part of their native phonemes, irrespective of whether they are good or poor examples of such phonemes. In addition, they can also be *uncategorized*, assuming the listener cannot place them as being part of their native phonemic inventory. Thirdly, the listener may find the speech sounds as not being part of any human language and thus perceives them as non-speech sound, much like a random noise, in this case we talk about n*on-assimilated* sound.

The *categorized* phones are further sorted according to whether and how the listener discriminates them. More specifically, if they perceive each of the sounds from the contrast as a good or excellent token of the native phoneme, a *two-category* assimilation is predicted, and hence good perceptual discrimination, both non-native phones are considered as being part of two distinct native phonemes. When two non-native phones are assimilated within the same native phoneme, it yields a *single category* assimilation, predicting poor discrimination between the contrasting sounds. The *category goodness* scenario occurs when two non-native phones are assimilated as part of one native phoneme, but they differ in how well they fit into that category. This predicts an immediate ability to discriminate the L2 sounds.

Another situation occurs if a non-native contrast, after being perceived by the non-native listener, remains uncategorized and as a result belong to the phonological space but not to any particular L1 phoneme. This assimilation pattern is called *uncategorized-uncategorized*. The discrimination of the contrast is predicted to be variable, depending on how close the two sounds are, perceptually. In case one of the non-native phones is perceived as part of a native phonological category and the other falls outside, we talk of *uncategorized-categorized* assimilation, with discrimination at a fairly good level, owing to the fact that the sounds are sufficiently separated from each other within the phonological space.

3.2.4 PAM-L2 (Best and Tyler 2007)

PAM-L2 is meant as an extension to PAM as proposed by Best (1995) in her paper *A direct realist view of cross-language speech perception*, adding to the nonnative speech perception the aspect of L2. As Best and Tyler (2007, 15) argue, non-native and L2 perception of speech are basically very different, even though many consider them equal. PAM is concerned with the perception of non-native language by naïve monolinguals, whereas PAM-L2 is similar to Flege's SLM, but, putting production aside, the focus is only on L2 perception. This model strives to bridge the evidence from research inspired by the SLM, concerned mainly with experienced bilingual speakers, and PAM, focusing on inexperienced, naïve listeners. PAM-L2 aims at explaining L2 perceptual learning and the interactions found between the native L1 and non-native L2 phonological systems, whether perception is influenced by the fact that the listener knows more languages and to what degree it affects them. Best and Tyler focus on how L2 learners acquire the higher order invariants of the L2 and, assuming a common L1-L2 inter-lingual system emerges, how it incorporates both phonetic and phonological levels, with the ultimate goal to outline how the system changes during L2 development (ibid, 24).

One of the issues presents, whether or not perceptual assimilation of L2 phone into L1 phonological category takes place. The equivalence is investigated on the phonetic level, similar to SLM, but also on a higher phonological or lexical-functional level, as it seeks to discover if the "phonological category has a similar contrastive relationship to surrounding categories in the phonological space" (ibid, 24). As an example, Best and Tyler compare English and French rhotic phones, the former is a liquid, the latter a voiceless uvular fricative. Phonetic differences notwithstanding, the two sounds are similar when it comes to higher linguistic levels, they reflect "similar patterning of rhotics [...] in terms of syllable structure, phonotactic regularities, allophonic and morphophonemic alternations" (ibid, 25).

3.2.4.1 L2 Minimal Contrasts Predicting (Un)Successful Perceptual Learning

Best and Tyler (ibid, 25-30) base their predictions on multiple possibilities of how a listener assimilates non-native sounds into his or her L1 phonological system.

The first considered case of L2 minimal contrast focuses on the situation when one L2 phonological category is perceived as equivalent to an L1 phonological category. In case only one member of the non-native phonetic contrast is perceived as a good example of the L1 phonological category, no further perceptual learning for the phone is expected, because the contrast with other categories leads to either two-category assimilation or an uncategorized-categorized assimilation and thus minimal contrastive word discrimination problems. When equivalence occurs on the phonological level, as well as on the phonetic level between L1 and L2 sounds, they are perceived as a common L1-L2 category with a possible shift towards the L2 values, for instance adjusting VOT values towards the L2 model. On the other hand, the L2 phone categorized in an L1 phonological category could also be perceived, compared to the L1 phone, deviant, as is the case e.g. for different rhotic sounds in distinct languages which are phonologically similar, but phonetically very different, leading to a fairly good dissimilation ability.

As a second case considered by Best and Tyler (ibid, 26-27), we take a look at two L2 phonological categories being assimilated into a single L1 phonological category of which one being somewhat inferior or deviant (what PAM lists as category-goodness assimilation). Moderate discrimination is expected, ultimately a new L2 phonetic and phonological category is expected to be formed for the deviant exemplar, whereas for the better exemplar of the lot, assimilation to the mutual L1-L2 phonetic category is more likely than the creation of a new separate L2 phonetic category.

When both L2 phonological categories are perceived as equally good or bad examples of the same L1 phonological category, little perceptual learning is expected. Nonetheless, there are conditions that increase the possibility of new phonological and phonetic category formation. For instance, when perceived as bad exemplars of L1 phoneme and being "high frequency words or com[ing] from two dense phonological neighborhoods, which contain many minimal contrasting words," it is more likely for a new category to arise, otherwise the L2 learners would not benefit from such phonetic and phonological differentiation (ibid, 28).

Supposing the contrasting L2 phones possess characteristics which the non-native listener perceives as similar to more than one L1 phonological category (those mentioned in PAM as uncategorized), the learning of a new L2 phonological category is expected. In case each of the uncategorized phones bears a resemblance to diverse sets of L1 phones, the features perceived are somewhat distant within the L1 phonological space, thus a perceptual learning of two new L2 phonological categories is predicted. In contrast, when both L2 phones share similarities to the same L1 phonemes, it is more likely that a single new phonological category will be established, incorporating the L2 phones.

The last remark about L2 phonetic contrasts that Best and Tyler (ibid, 29-30) make relates to how non-native listeners learn L2 phones which fall outside their phonological space. The case of Zulu clicks is considered which the American English listeners perceived as non-speech, yet were very able to discriminate the minimal contrasts, compared to non-native listeners of Zulu who contrast clicks in their native language. That being said, the authors predict either gradual incorporation to the phonological space as uncategorized speech sounds and eventual perceptual learning of new phonological categories or their permanence outside the phonological space. However, if the latter occurs, the listeners "will likely have difficulty co-articulating them properly with vowels," therefore if they appear in lexical items, errors in their production are to be expected (ibid, 30).

3.2.5 Similarities and Differences between PAM and SLM

This subsection reports considerations about Flege's SLM by Best and Tyler 2007 (19-23) reviewing his four postulates trying to find convergences and divergences between the SLM and PAM. Postulates from Flege's SLM are presented in Table 1.

The first postulate talks about the mechanisms and processes, such as forming phonological categories, involved in the L1 learning, which remain intact throughout the life and are possibly applicable to L2 learning. A principle that is basically compatible with the PAM, according to which "perceptual learning processes remain intact throughout life" (ibid, 19). Even though such processes remain intact, speech perception continues to develop, as the listener may come across a new dialectal variant or may have to adjust to changes within the dialect of his region. The need to adjust is central and according to Best and Tyler is very similar to learning a new language, nevertheless such learning does not occur for the adult in the same way as for the child, since the adult is a different being with distinct perceptual history for whom the environment changes "including the response of others to the individual's appearance and behavior as a physical, cognitive and social being, and particularly as a language-user" (ibid, 19). A crucial difference concerns how the two models approach the invariant structures of the speech signal (some are formed by articulators, some generated by the auditory system of humans). More specifically, PAM stresses the former, suggesting that "perceivers extract invariants about *articulatory gestures* from the speech signal, rather than forming categories from acoustic phonetic cues" (ibid, 20).

In the second postulate (Table 1), Flege stresses long-term memory which stores language-specific aspects of sounds, known as phonetic categories. Best and Tyler criticize such mental representation and, as suggested in the previous paragraph, view the listener as able to perceive the articulatory gestures directly and by perceptual learning to become attuned to them, in other words, being able to perceive "invariants across instances of specific types of objects and events (actions) [...] amplitudes and phasings of speakers' vocal tract gestures in the L1 and/or L2" (ibid, 20). Nevertheless, it is not only the phonetic level that is important for those perceptual objects or events, L2 listener's discrimination may also be influenced by gestural or phonological differences between the L1 and L2.

The third postulate involves phonetic categories for L1 sounds that supposedly keep evolving from the childhood times to encompass L1 and L2 phones which come to be identified as belonging to certain category. Now Best and Tyler consider this postulate as an extension of the second one and as such reject the theoretical concept of categories stored in the long-term memory. Nonetheless, they are in concordance with the ever-evolving concept of SLM, we can say that tuning to the perception of speech sounds is a task for the listener that is never quite finished.

As for the last postulate, bilinguals aim to preserve the contrast between L1 and L2 phonetic categories occupying the same phonological space. PAM agrees only with the SLM's concept of L1 and L2 phonological categories occupying the same phonological space while being against the idea of there being an interaction between phonetic and phonological levels in L2 speech learning (ibid, 22).

3.2.6 PAM-L2 Predictions for Foreign Language Acquisition

PAM-L2 investigated how an L2 learner actively acquires L2 categories in an immersion setting "with rich native-speaker input [who] would have ample opportunity for the sort of perceptual learning that is required" for successful L2 category acquisition (Tyler 2019, 616). Nevertheless, in a fairly recent study, the author focuses on L2 learners who acquire L2 during a formal instruction in classrooms, where the conditions are far from being ideal, focusing on three areas: spoken language input, written language input and foreign language exposure the learners received prior to the classroom foreign language acquisition (FLA). Furthermore, he seeks to afford predictions to FLA based on the observations from the PAM-L2.

Tyler argues that the problem of beneficial spoken language input does not necessarily consists in native vs. non-native, but more of whether the speech "maintains a phonological distinction between all L2 phonemes, and native speakers unambiguously perceive them as intended," thus even accented speech might facilitate the L2 learner's acquisition of novel L2 categories (ibid, 616). The problem is, however, when such speech does not maintain the coveted phonological distinction, the contrasting minimal pairs become homophones and the perception is strengthened by more vocabulary which leads to fossilization.

It is trickier with the written language input, as in the case of the former, they might be beneficial or detrimental, depending on what approach the learners choose. Generally, it can be beneficial to revise and consolidate the knowledge acquired in class, nevertheless it can also be detrimental for the acquisition of L2 phonological contrasts, as the learners might acquire too large vocabulary in too short a time, leading to phonetic learning fossilization (ibid, 613). When the L2 orthography offers signals of clear phonological difference, the L2 learners can tune in to it and written material might thus be beneficial, nevertheless when no such signals of clear difference are present in the orthography, it might lead the L2 learners to think that there is no difference between two L2 phonemes and it might strengthen their belief that the phonemes are equivalent, rather than different (ibid, 617).

Prior foreign language exposure is something that the original PAM-L2 does not consider at all, as the model aims at naïve L2 learners in an immersion setting without any prior L2 experience. The author mentions that there is a large variety of possible prior exposure to L2, amounting diverse types of classroom instruction, input from movies, TV, participating in a study abroad program, or having a family member capable of speaking in the L2. There might also be the orthography factor, the learners might learn the L2 by learning to read it. Something like that should be approached with caution, because initial exposure to spoken input early on is crucial for PAM-L2, it is the starting point for successful, native-like perceptual learning to occur and a way to avoid fossilization (ibid, 617-618). Another issue is that the L2 learners might not be exposed to a single variety or a single foreign accent and, indeed, typical for the modern world is to listen to L2 with many different accents. Foreign language exposure brings confusion to possible predictions in accord with PAM-L2, and as such it might be beneficial, as well as detrimental, depending on the particular history of exposure the particular L2 learner received.

The study offers predictions for FLA based on PAM-L2 (ibid, 618). According to the author, two-category assimilation follows the same path as in the immersion setting as the learners can recognize the distinction between the two phonemes which initially conceptually occupy the same phonological space as two distinct L1 phonemes. Regarding category-goodness assimilation, the phonetic difference is not as likely to be acquired as in the case of immersion, especially when there is little perceptual difference between the two phonemes in the perceived L2 speech or when the perception is learned by reading the written form without being exposed to the spoken input. Similarly to PAM-L2, rapid vocabulary acquisition is expected to hamper perceptual learning, which is even more so when the learner acquires them by written form. Very little or no phonological learning is predicted for single-category assimilation, fossilization is expected, as to learn to perceive the difference between those two phonemes is unlikely even for the immersion setting. For the acquisition of uncategorized dispersed L2 phonemes, those that are not perceived as being similar to any L1 category, the authors expect a course resembling the one present in the immersion setting, especially when L2 orthography provides beneficial information about the contrast and enough spoken input is supplied. In case of unrecognized focalized phonemes, very weakly perceived as one L1 category, and unrecognized clustered phonemes, very weakly perceived as multiple L1 categories, the predictions are rather unclear for the classroom setting as it is not clear how the phones are acquired in the immersion setting, on which the FLA is based.

Several ideas to create more hospitable conditions for successful L2 category acquisition during FLA were considered by the author. For instance, it is important to provide "rich and varying speech from native speakers" for L2 acquisition, nevertheless, due to the fact that the foreign language learners (FLL) normally spend quite a lot of time practicing between each other, they cannot completely avoid imperfect non-native speech input, it is, thus, even more important to provide the kind of input in which phonological contrasts between all phonemes of the target L2 are present (ibid, 622). The teacher should also explicitly stress how important contrasts the phonemes represent and how important it is to be able to distinguish them, providing good examples of the contrasts but only if he can satisfactorily produce the difference, otherwise it is more beneficial to provide audio-visual materials of native speech. Moreover, the author recommends using high variability training (also used in the studies we review in section 4.3), employing varied phonetic context from varied speakers, many studies consider such training crucial for improving the perception of L2 speakers, e.g. Bradlow et al. (1997), Lengeris and Hazan (2010) and Giannakopoulou et al. (2013). In addition, Tyler (2019) lays emphasis on the importance of feedback in high variability training and also on the probing of the student's background before tailoring the training specifically to the students' state. Regarding vocabulary, learning large amounts of words should be avoided, owing to the fact that it might cause fossilization, failing to acquire L2 in the desired manner. Early vocabulary should include primarily words that are predicted to be easily assimilable for the FLL or those that involve uncategorized phones. The words containing phonemes that could cause trouble assimilating, such as those expected to produce single-category assimilation, are to be introduced slowly, preferably by way of intensive perceptual training. The students should also have enough opportunity to compare several words containing certain phonemes and their minimal pair counterparts containing the contrasting phoneme. Another suggestion of Tyler (2019) is to delay presenting orthography to the students, especially for languages like English, where grapheme-phoneme correspondences are less frequent, this is especially useful for students who are from the L1 background in which such correspondences are common. The author also suggests teaching the FL students IPA symbols, which could be used to monitor how the FLLs progress in a better way than using orthography, it could also help in training exercises focusing mainly on the phonemes compared to using minimal pairs.

This subsection summarized the ideas of Tyler (2019) about how the PAM-L2, formerly aiming at naïve L2 learners in an immersion setting, could apply to FLA taking place in classrooms. The author considered spoken input, written input and prior foreign language exposure specific for the FL learners in such environment, offered predictions according to PAM-L2 and ways to counteract the negative effects that could stem from the nature of FLA.

3.2.7 Summary

The current section tried to present the models that describe ideas about how L2 speech is acquired, mentioning taking into account perception in case of PAM and both perception and production in case of the SLM. It offers a comparison between the two crucial models and presents the updated versions of the studies, PAM-L2 and PAM-L2 for FLA, in case of the former and SLM-r, in case of the latter, informing the reader on the novelty in which it differs from the original study.

It is important to note that the models mentioned previously might come in handy when noise comes into play and perceptual assimilation is hampered. Due to the fact that in everyday, non-laboratory, language contact, there are several adverse conditions which can influence how the listener perceives speech sounds either separately or in combination. For instance, we can talk of reverberation, as a fair amount of the verbal communication takes place in rooms which are not anechoic, whether they be classrooms, lecture theatres, meeting halls or cafés. Moreover, the interaction could take place in a busy street or at an outside restaurant, where environmental noise may have an effect, as well as the noise caused by other people's speech. The prediction would be that the difficulty of certain L2 contrasts would be even more accentuated when adverse conditions are taken into account when trying to figure out L2 learning.

The study by Tyler (2019), focusing on acquisition of L2 phones in foreign language classrooms and how it can be complemented by PAM-L2, provide interesting ideas about how the L2 perception develops for L2 learners that acquire it not through immersion in L2 environment but in classroom setting. The ideas are potentially very useful for L2 learners of widespread languages such as English. The majority of L2 learners of such languages acquire them in classroom settings, it is thus essential that we understand how the L2 learner comes to categorize the L2 contrasts. The native Czech learners of L2 English are no different, most of their acquisition is also carried out in classrooms rather than in immersion setting, therefore using the ideas of Tyler (2019), when conducting a research with Czech speakers of English, seems paramount.

The models of phonetic category formation are relevant to our investigation, because they predict improvement for L2 perception provided there is good quality and large amount of native input, high level of L2 use and low level of L1 use. Immersion in an L2 environment should, thus, be crucial, because it usually includes all the aforementioned aspects and should, therefore, lead to improved perception in noise.

3.3 Factors Affecting L2 Learning

The following paragraph aims to summarize the main factors which may affect L2 learning. We decided to include this section, because it sheds light on how L2 perception can be improved, with better quality and larger portions of input and increased L2 use playing the key role. As all this is usually present when being immersed in L2 speaking country, perceptual learning of L2 sounds is expected, leading to improved L2 speech perception in noise.

3.3.1 Age

Age is relevant when discussing second language acquisition, because "both naturalistic evidence and formal instructional evidence can be interpreted as being consistent with what may be termed the 'younger = better in the long run' view," even though such view with respect to SLA "needs to be seen in the perspective of a general tendency and not as an absolute, immutable law" (Singleton and Lengyel 1995, 3-4). Such view is consistent with Flege (1995, 234), who reported a study by him, Munro and Mackay where English sentences produced by native Italian speakers were judged by native speakers for foreign accent and found to be correlated with the AoA. The earlier the arrival, the more native-like the L2 speaker sounded, which suggests a more gradual view not unlike the one proposed by Singleton and Lengyel.

Khasinah (2014, 259-261) asserts that the difference between child and adult learning consists not in age per se but in motivation (which we will consider below in 3.3.5) and the way the learners are able to apply linguistic rules. There also exists a certain trade-off, the older L2 language learners are more able to actively use linguistic rules, but often lack children's motivation yearning to earn acceptance from their social group.

SLM-r does not place much emphasis on age, because, according to this model, all mechanisms used in FLA remain available in SLA. The reason why, the apparent age-related effects on L2 speech perception and production arise, is in reality due to factors confounded with age, such as L1/L2 use, the amount and quality of input, motivation, full-time equivalents, featuring years of learning multiplied by L2 use proportion.

3.3.2 Native Input

Native input is arguably the most influential factor with impact on second language acquisition. This factor is very important to mention when considering immersion setting, as frequent access to native input of good quality and high variability positively influences L2 speech perception and subsequently have a positive influence even on L2 speech perception in noise. We turn primarily to Gass and Selinker (2008, 304-310) who investigate how the theories viewed the role of native input in SLA as time progressed.

First concepts of second language learning were based on behaviorist theories of Pavlov, Watson and Skinner. The advocates of this theory saw input as "the major driving force of language learning," nevertheless, they focused mainly on children and saw the whole process as unavoidable imitation of behavior that surrounds them, much like an animal who learns to eat or hunt (ibid, 304). However, as the popularity of the behaviorist theories fell down, so did the interest for native input. The focus was shifted to internal mechanisms and innateness of the learner. The theory looked upon the language learners as some sort of gods, creating their language systems as they developed and native input had, at least for the children, no significant importance.

A distinction between input and intake was made by Corder in 1967, the former basically refers to what the learner hears, but it is not necessarily what he is able to comprehend, take in (the latter), as it "may not even be possible to separate the stream of speech into words [...], because it 'goes in one ear and out the other'; it is not integrated into the current learner-language system" (ibid, 305). This theory thus supports only the role of useful input or as Corder calls it "intake," which the listener can absorb, any kind of input is similar to no input at all.

Ferguson (1968) focused on the issue of linguistic simplicity, which the non-native speakers might come across when talking to native speakers, more specifically, a native speaker might make adjustments to his speech, deliberately simplifying it, in order to facilitate comprehension for the non-natives. This might be beneficial, if the non-native is at a lower stage in the acquisition process and the pure native speech, with its speech rate, reductions, co-articulations, might, similarly to what suggested Corder, hamper any further learning, as the speech could be incomprehensible for the non-native. On the other hand, it might impair the non-native's further learning as well as being perceived as offensive, depending on how far the non-native in the acquisition process is. The adjustments (simplifications) the native makes can be in the area of pronunciation, grammar or lexicon, as well as in the speaking style, repetitions or restructuring of the discourse.

An important theory for the SLA is the Input Hypothesis crafted by Stephen Krashen (1985). Based on the natural order of acquisition, it focused on comprehensible input, something Krashen defined as a "bit of language that is heard/read and that is slightly ahead of the learner's current state of grammatical knowledge. Language containing structures a learner already knows essentially serves no purpose in acquisition" (Gass and Selinker 2008, 309). Krashen's theory is based on innateness, the Language Acquisition Device, but stresses the specific input the non-native must gain access to, if *i* is seen as the current state, only *i+1* provides the right input. The main problem of this theory, according to Gass and Selinker, resides in being too theoretical and questions, such as, how to define the current level of knowledge, or what is sufficient amount of appropriate input, arise.

The role of input was investigated by Flege and Liu (2001) in order to find out how it affects the natural acquisition of adults' L2. The authors assessed Chinese participants on three experiments, namely word-final identification of English consonants, grammatical judgment and listening comprehension test. A division into four groups of participants was made according to their length of residence (LoR) in the L2 speaking country: relatively long vs. relatively short, and their occupation (students vs. non-students). LoR serves as an index of how much native input the non-native perceives. The authors found a significant difference between students with long LoR, who received higher scores, compared to the ones with shorter LoR. Nevertheless, no significant difference was found between the non-student groups and as a result Flege and Liu challenge the LoR as a good index of "rich input that is needed for successful L2 learning" for some non-native learners (ibid, 550). The reason might be that the learners who study at school "interact frequently with native speakers of the L2, whereas adult immigrants often enter the workplace, where they interact frequently with fellow native speakers of their L1," they may thus receive far less input and of a far worse quality than those who go through the educational system in the country of residence (ibid, 528-529).

3.3.3 L2 Use

It is important to mention the difference between speaking in L2 while being immersed in L2 environment (such as living in the country where L2 predominates), and using L2 in an L1-speaking environment. The difference is immense, the former, that is, using L2 in an L2 environment, is predicted to trigger much more L2 learning as it usually combins with high amount of input of better quality.

When considering the notion of L2 use, basically we can draw data from the studies investigating immersion of non-native speakers in an L2 environment, such as Aoyama et al. (2004), McCarthy et al. (2014) and Kim et al. (2018), even though they did not aim at L2 use particularly, they were investigating cross-sectionally or within certain social group consonantal, vocalic or voicing contrasts in an immersion setting. A naïve expectation would be that those, who voluntarily immigrate into a country, in which the community speaks in a different language, would logically strive to achieve proficiency in the target language as soon as they possibly can, in order to achieve higher socio-economic position, as well as socialize with the native speakers. There might, however, be circumstances, which prevent the immigrants to use L2 or listen to L2 input, we can turn to the study of Flege and Liu (2001) already mentioned in the previous paragraph. The authors talk of social constrains for some non-natives, usually adults who begin working right after their arrival, they are more likely to use and listen to their L1, as their peers are frequently of the same language background. On the other hand, those, who begin studying in the country, usually children and adolescents, have more opportunity to use and be exposed to L2, as they are eager to fit in the social group.

Flege and MacKay (2004) employed in their methodology the amount of L1 use. In experiments 3 and 4, they investigated vocalic contrasts of native Italian (NI) speakers living in Canada for at least 10 years and compared them to native English speakers. The NIs were divided into early (age of acquisition

(AoA) 2-13 years) and late learners (AoA 15-26) and also subdivided according to their L1 use into low (1-15 %) vs. high (25-100 %). The study found in both experiment 3 and 4 low-L1-use NIs (i.e. those, who used L2 more frequently than L1) to obtain higher scores than the high-L1-use NIs, in other words "the low-L1-use participants in this study perceived English vowels more accurately than the high-L1-use participants, because they used English more often and had thus heard English vowels more frequently" (ibid, 27).

Casillas (2020) also takes into account L2 use as an important factor. Unlike Flege and MacKay (2004), the design of this study is semi-longitudinal, taking place within the 7-week time-span, rather than taking participants who live in a non-native country for several years. The participants in the Casillas study are made up of adult non-native Spanish learners (having English as their L1) without any previous experience with other languages and Spanish-English bilinguals as the control group. The reason for using bilinguals rather than monolinguals as control is because they "tend to differ from monolinguals in production, perception and lexical processing," and non-native learners "undertake the endeavor of language learning with the goal of becoming bilingual and not to replace their native language" (Casillas 2020, 13). A weekly-progress questionnaire was always administered before their participation on the experimental tasks. The immersion took place in an innovative design, as part of the Spanish domestic immersion program, taking place in the USA, the country of origin of the participants. The participants lived in residence halls with other students and professors, they attended 4 hours of classes in the morning and other activities in the afternoon. Interestingly, the participants signed a language pledge, an agreement in which they "promised to use only the target language (in this case Spanish) for 7 weeks," if they failed to comply, it could result in expulsion from the school (ibid, 14). The authors found phonetic learning in both production and perception over the program's course, which might have been caused by maximized L2 use.

As there is a tight connection between native input, LoR and L2 use, Flege and Bohn (2021, 7) advocate a different concept of measuring native input, they speak about something called "full-time equivalent" (FTE), a figure that is "calculated by multiplying years of residence [...] by proportion of English use," it thus represents “a better estimate of quantity of L2 input that LOR alone,” even though the quality of the input is still not completely allowed for.

3.3.4 Classroom vs. Naturalistic Settings

Gass and Selinken (2008, 368-372) argued that in home environment, especially if the language learning occurs via being taught in classroom settings through classmate input, there are differences from learning the language in the L2 environment. More specifically, they mention distinctions in quality and quantity of input, i.e. “there is not only limited input, but a large part of the input comes from classmates whose knowledge of the foreign language is restricted” (ibid, 368). In addition, there are also limitations on the learners’ opportunities to interact.

One aspect of the classroom setting can constitute the modification of language when talking to language learners, which would mean that the input they receive via naturalistic learning and the one from classroom setting, is not equal. Whether or not such difference is beneficial is however unclear. We can find evidence in Gass and Selinker (2008) who report on a study which investigated the speech of eight teacher trainees towards four groups of ESL students differing in proficiency. The study found the proficiency level as a “statistically significant predictor of the syntactic complexity of these teachers’ speech,” which means the teachers did modify their language according to learners’ proficiency (ibid, 369). Language learning in classroom setting thus occurs via three sources, the teacher, who may simplify his output, classmates, who are far from producing faultless speech, and study materials. Two of the three sources might be unreliable to produce natural output, compared to what the non-native would receive through naturalistic learning. Studies, such as Stevens (2001), Nagle et al. (2016) or Casillas (2020), suggest that learning L2 in a more naturalistic setting, such as taking part in a study abroad, is more beneficial as the learners are more able to produce native-like sounding speech, compared to those who learn L2 in the “traditional” way, that is learning solely via classroom instruction.

Stevens (2001) is one of those studies devised to investigate the difference between more naturalistic setting (abroad study in the L2 environment) vs. more formal setting (studying at home institution, meaning an American university), seeking to solve the crucial second language acquisition (SLA) issue, namely, whether the “learners acquire an L2 better by formal instruction or comprehensible input alone” (ibid, 137). Native speakers of English took part in the study without ever receiving formal instruction in Spanish pronunciation, the participants were divided into two groups, one being part of Spanish language course at the home university, the other taking Spanish course abroad in Madrid (subdivided into seven-week intermediate group and sixteen-week advance group with more comprehensive program). The author aimed at VOT, hypothesizing VOT reduction for the participants in the abroad study programs and, at the same time, no reduction for those who remained at the home institution learning solely in formal environments. As expected, the author found support for his hypothesis, as the “study abroad learners did make significant improvement in acquiring the Spanish voiceless stops in terms of reduced VOT, whereas the home learners did not,” even though there was a slight but non-significant improvement for the latter group (ibid, 147).

The effect of a short-term six-week intensive study abroad program was investigated by Nagle et al. (2016). Advanced learners took part in the study, in particular, those who had received over 6 years of Spanish classes in a private university in the USA, had not studied abroad before and had been fluent only in their L1 English. The study abroad program took place in Barcelona and prior to the program, the students signed a pledge, committing themselves to use only their L2 Spanish during the course of the program, if they failed to meet such requirements, they could face reduction of their final mark or even expulsion from the program. The program itself consisted of various activities, the students had to enroll three courses taught in Spanish on visual arts, history and politics; moreover, they had to attend fieldwork led by the faculty, visitations to regional sites and conversations with native Spanish speakers were included, providing “ample opportunities for exposure to input that was varied and rich in information” (ibid, 683). The research focused on the participants’ L2 Spanish production of stop consonants (namely /p/, /t/, /k/, /b/, /d/, /g/) and tested the participants pre- and post-stay. In general, the authors found them to produce “forms that increasingly aligned with bilingual norms, both in terms of lenition and duration,” and especially for the tokens with voiceless stops whose VOTs were found to shorten significantly as the result of the stay (ibid, 690). Nevertheless, in this particular study, no control group which would attend solely traditional classroom lectures, was employed, so no comparison similar to that of Stevens (2001) could be made.

Casillas (2020) employed a similar design to that of Stevens (2001) and Nagle et al. (2016), nevertheless, it was unique in the way the immersion was created, the environment for the study abroad experience was created under the home institution, an American university. The author focused on both perception and production, namely he explored phonetic category development of Spanish stop voicing contrast, and used native English beginning learners of Spanish in an environment in which both L2 use and L2 input were maximized. As was already mentioned, the study had a unique design of a domestic immersion program occurring at the home institution in an American university. This is probably possible due to the fact that there exists a large Hispanic community in the USA and, as such, one can experience an immersion in a foreign language community, as if he were taking part in an actual study abroad program. The participants had to sign a language pledge of a similar design to that of Nagle et al. (2016) with the similar repercussions to the students violating it, promising to use only the target L2 Spanish for the duration of the program, which covered 7 weeks. The students resided in residence halls and attended four hours of lessons in the morning focusing on communication and being taught in Spanish, together with "co-curricular activities in the afternoon [...] with the intention of creating an experience comparable to living abroad" (Casillas 2020, 14). The results suggest that the L2 phonetic categories can form up abruptly for the beginning learners of L2, their formation is "perceptually driven [...] and especially susceptible to cross-linguistic interference during the initial stages of learning" (ibid, 42-43).

It is important to mention that there is a crucial difference between being taught L2 in an L1 environment, e.g. through formal education, and being exposed to it in a naturalistic, immersion setting, where one has access to more input of a better quality and is forced to use the L2 much more often than in the classroom setting, and, on that account, much more L2 learning occurs for those in the naturalistic immersion setting.

3.3.5 Motivation

As the previous paragraph suggested, one of the factors that could have an influence on how the non-native speaker learns an L2 is motivation. It is defined as the aspiration towards certain goal or orientation (Khasinah 2014, 258). According to Carrió-Pastor and Mestre (2014, 240-241), motivation of non-native learners consists of three components, namely effort (the amount of time the L2 speaker spends learning and the drive he has towards the language), desire (the eagerness the L2 speaker has to become a proficient user of the language) and affect (whether or not and to what degree emotions are involved when the L2 speaker is studying).

We can speak of two basic types of motivation which need to be differentiated, namely integrative and instrumental. While the former focuses on the L2 speaker's interest in the culture and the population speaking the L2 and the desire to speak their language while communicating with them, the latter addresses functionality and pragmatics, the L2 speakers wants to improve in order to pass a test, get a promotion, be able to apply for a better position, or also to be able to follow foreign news, deal with the authorities and municipal or state offices. While both strategies are beneficial, each type seems to be more effective in different situations. Khasinah (2014, 258) argues that integrative "motivation plays a major role where L2 is learned as a 'foreign language'," meaning if the non-native speaker studies the L2 in a country where the language isn't dominant and as such is not immersed in it. On the other hand, if the non-native speaker studies the L2 in a country where it is widely used, instrumental motivation has more importance (ibid, 258).

We can also differentiate motivation into intrinsic, in which case, the learner does not achieve any particular reward for learning the L2, he is rewarded solely by the activity of learning itself. As Deci et al. (1975, 82) state, when "a person is intrinsically motivated the locus of causality is within himself. [...] People are intrinsically motivated to perform activities which make them feel competent and self-determining." On the contrary, if we talk about extrinsic motivation, some sort of apparent reward is expected, for instance financial (as is also the case when e.g. people get a promotion), a praise or other type of positive feedback. Even though extrinsic motivation would, at first sight, seem superior, Khasinah (2014, 258-259) argues that "intrinsic motivation leads to greater success in learning a foreign language, especially in the long run."

3.3.6 Summary

As it was suggested, age seems to be a relevant factor, nevertheless, the ability to acquire more native-like L2 when younger rather than older should be seen as a general tendency not an absolute as suggested by the CP defined by Lenneberg, owing to the fact that some L2 learners are still able to reach proficiency, even though they start learning L2 after the CP.

Another factor that receives a lot of attention is native input. We considered the beginnings when researchers started thinking native input, a distinction between input and intake (useful input) was made by Corder, advocating only the input that the listener can comprehend, which was followed up by Krashen, who talks about the right input for the listener (i+1). LoR was considered as a possible measuring device for the amount of native input received by the L2 learners. L2 use is a concept closely connected to native input and the length of residence, but the focus is on the speech production side of the L2 learner and the studies discussed in 3.3.3 indicate that more L2 use leads to more accurate perception of L2 sounds and also more communication opportunities which also prompts more L2 input. The connection between native input, LoR and L2 use is, indeed, very tight and as such Flege and Bohn consider using other concept, full-time equivalent, which measures the amount of native input by combining LoR and L2 use and provides more accurate quantity of native input the L2 learner receives.

It is important to consider in which setting the L2 acquisition takes place, when it comes to classroom setting, less L2 learning is expected, unless we consider a special kind of immersion in an environment similar to the one employed in Casillas (2020), which resembles immersion in the L2-speaking country, even though it takes place at the home institution.

When it comes to motivation, it is rather intangible to understand how it influences L2 learning, mainly because there is no easy way to measure how high highly motivated L2 learners really are and as such this factor receives only marginal attention.

3.4 Speech Perception in Adverse Listening Conditions

The current thesis is concerned with native and non-native perception in adverse conditions and, as such, it is necessary to present the particularities which define speech in such conditions. The current section presents preliminary information about noise and reverberation, the reader will also find additional information about adverse conditions in sections 4.1.3 and 4.1.4, providing evidence from experimental studies.

Any kind of unwanted sound masking the speech signal can be thought of as potentially harmful to the speech perception. Such sounds are present in everyday speech, sounds from other sources than the speech source or the sounds reflected from surfaces around us (i.e. reverberation). Lecumberri et al. (2010, 865) focus on artificially added noise (and hence precisely controlled by the experimenter) which is most commonly used in studies on non-native speech perception, namely additive noise making it harder to recognize target speech sounds, at the level of auditory periphery, known as energetic masking, or at a higher level, namely informational masking.

According to Febo (2003, 4) noise functions as a masker especially to those portions of the signal that are less intense, leading to the fact that consonants, containing less acoustic energy, tend to be masked more than vowels. Noise present in the environment logically leads to “reduction in the redundancy of the acoustic and linguistic cues characteristic of speech” (ibid, 4). Speaking of noise, we also have to mention the importance of relation between the overall intensity of the speech signal and the overall intensity of the noise, known as signal-to-noise ratio (SNR) and as Febo (ibid, 4-5) mentions, a high rate of successful speech perception is expected when SNR is favorable, such as +10 dB, with a proportional decrease in success when the SNR lowers.

3.4.1 Energetic Masking

We talk about energetic masking (EM) when the energy intruding from other sound sources interferes with the target sound, neural representation is occupied by the unwanted sound, which renders the target sound inaudible to the auditory nerves; in case of EM such intrusion occurs at the auditory periphery, overwhelming the neural representation by masking sounds (Wang and Xu 2021, 110). Lecumberri et al. (2010, 872) speak about EM in a similar way, but stressing the unavailability of “potential cues to the identity of segments and their boundaries as well as interfering with access to prosodic cues,” even though complete blocking of prosody is unlikely, as it is of long-term nature. Generally, dental fricatives are, due to their quality, most liable to succumb to EM, as they are inherently of low energy (ibid, 872).

The degree of EM depends on the SNR as well as the property of the masker itself. Stationary maskers, such as speech-shaped noise[1], are more detrimental than speech-modulated noise when the SNR is held constant, the release from masking depends on the modulation of the target speech, only “those parts of the speech signal which are intrinsically energetic [...] [such as] formants and strong frication [...] are likely to survive SNRs below 0 dB” (ibid, 872). On the other hand, in maskers which are not stationary, such as competing talker or multi-talker babble, “weaker target signal components may well be audible,” some may thus be released from masking, depending on the interaction between the modulation of the target and the masker (ibid, 872).

Considering EM, sound source distance is a critical factor, as “even relatively quiet noise source close to the listener can make a more distant target source more difficult to comprehend” (ibid, 871). Ezzatian et al. (2010, 927) found such effect of spatial separation when speech perception in competing-talker noise was tested, the listeners were released from masking regardless of being native or non-native speakers of the target language, suggesting that “the degree of release from masking due to spatial separation appears to have more to do with the ability to segregate the speech target from the masking background than [...] fluency in the language,” hence the release from spatial separation seems to be connected with EM rather than informational masking.

3.4.2 Informational Masking

Information masking (IM) is a different kind of masking and applies in cases when the information contained in the noise interferes with the target speech signal, but rather than blocking the target at a peripheral level, making it unable to enter into cognition, the target is indistinguishable from the masker at a higher level of processing, making “both the target and masker [...] audible but not distinguishable and [...] the masker information is [thus] misattributed to the target” (Wang and Xu 2021, 111). In other words, when the noise is able to initiate linguistic and semantic processing, it may interfere with the linguistic and semantic processing of the target speech signal, affecting the processing of the

[1] Represents a noise which is tailored to a particular utterance.

target speech at a more central level compared to EM (Ezzatian et al. 2010, 920-921). Lecumberri et al. (2010, 873) stress the fact that the most effective IM is regarded competing speech or multi-talker babble with small number of talkers, even though they admit that "all maskers have some occasional IM potential."

Brungart et al. (2001) investigated whether the distinction between sexes and whether the fact that the listeners were exposed to the target speech and noise that came from the same speaker or a different one, would come into play in perception in noise. They found that IM is weaker when there is a difference between the quality of the target and the masker, namely, "[d]ifferent-sex maskers degrade performance less than same-sex maskers, and same-sex maskers degrade performance less than same-talker maskers," such effect applies especially when the sound to noise ratio is above 0 dB (ibid, 2537).

3.4.3 *Reverberation*

Reverberation is commonly known as the persistence of a sound after being produced in an enclosed room, it is reflected off the walls and gradually being absorbed by the objects which are located in such room. Lecumberri et al. (2010, 873) mention the fact that the sound reaches the listener from indirect paths contributing reverberation to the direct signal being received at the ear of the listener, leading to masking of speech components in a different manner than additive noise. The effect of reverberation also depends on the qualities of the room in which the speech sounds are produced.

In a spectrographic representation of such sounds, we notice smearing of energy in time, "enhancing 'horizontal' structures such as static formants and blurring 'vertical' structures such as bursts and transitions" both within- and across-segments are affected (ibid, 873). Sounds containing lower energy, such as fricatives or closures of plosives, are more affected by reverberation, together with formant transitions of diphthongs, unlike static vowels, which remain fairly robust. The robustness of static vowels was proven by Nabelek (1988), who investigated vowel identification in quiet, noise and reverberation across 4 groups of subjects who varied in age and hearing levels and found formants of vowels robust in reverberation and thus affected only slightly by such phenomenon, only the group with the lowest hearing abilities performed differently compared to the three other groups, also in comparison with the noisy condition (12-talker babble), in which all four groups differed significantly. Similar to Lecumberri et al. (2010), Masuda (2016, 74) argues that reverberation can cause "unintelligibility due to temporal and spectral distortion of the target sound," expressed by reverberation time, the period the signal decays to 60 dB. She also stresses the fact that "speech perception in such conditions can be difficult for all listeners, but the challenge is greater for non-native listeners" even those with high proficiency (ibid, 74). Both prosodic and segmental information is affected by reverberation, reducing "the availability of cues to duration and rhythm," moreover, this phenomenon has a greater effect on children, the hearing impaired and older listeners (Lecumberri et al. 2010, 873-874).

Some studies, such as Febo (2003) and Masuda (2016) focus on the fact that in real-life conditions, especially in enclosed environments, noise is not the only issue for the listeners, often enough, it is also reverberation and particularly when these adverse conditions occur simultaneously, there is an increased, combined, detrimental effect compared to being exposed only to one of the adverse conditions at a time. The participants of Masuda's study showed concordance with such hypothesis, among the conditions to which they were exposed, they showed lowest identification rates when they were exposed to multi-talker babble combined with reverberation. Nevertheless, the results of Febo (2003), unexpectedly, did not show poorer performance in simultaneous noisy and reverberant condition, but in the noisy, anechoic one. The author attributes such result to the order effect, owing to the fact that the noisy reverberant tests were always administered as final and the participants may have gotten used to the task at the end of testing (Febo 2003, 23).

Reverberation is an important factor for L2 learning, because one of the most common environments in which L2 is learned, the classroom, is reverberant, and as such reverberation is a relevant issue to the majority of students, also considering its increased detrimental effects in combination with noise.

4. LITERATURE REVIEW OF EXPERIMENTAL EVIDENCE ON L2 PERCEPTION

As mentioned above, longitudinal research on L2 speech perception before and after an immersion experience is virtually nonexistent. The empirical findings reviewed in this section are therefore gathered from three different areas of research, which are connected but are not without their methodological differences. The subsection 4.1 reviewed studies in noise, which do not take a longitudinal perspective, nor any sort of training, they focus simply on the performance of the L2 speakers in noise at a single point in time, comparing them with native speakers and/or between each other according to their L2 proficiency. Studies with a longitudinal immersion design, which do not however use noise, are the center of attention in the subsection 4.2. The third group is represented by studies with a training paradigm, i.e. observing participants at multiple points in time though not really longitudinally, that do not employ noise (with the exception of Lengeris and Hazan, 2010, and Cooke and Lecumberri, 2018).

4.1 Non-Longitudinal Studies of L2 Speech in Noise

The current subsection aims to review, with a few exceptions commenting also upon production, perception of speech in noise, and the factors that might be beneficial or detrimental for the non-native speaker as he perceives degraded speech.

4.1.1 Age of Acquisition (AoA) – Early vs. Late Bilinguals

Many studies considered carefully, whether to employ early or late bilinguals as participants for their tests. Most of those studies suggest that learning L2 earlier is more beneficial for understanding L2 in noise, for instance, Mayo et al. (1997, 690) found a higher tolerance for background noise for early bilinguals compared to late bilinguals who learned English after puberty.

Some studies focused, in line with Mayo et al. (1997), solely on early bilinguals, those were, for instance, Febo (2003), Rosenhouse et al. (2006) and Tabri et al. (2010). Febo (2003) stressed the fact that in the USA, there are many individuals, who learn more than one language very early in their life, specifically, Febo's study focuses on the large Hispanic community in the US centering on Spanish-English bilinguals, who learn English prior to the age of six. The communities where more languages are learned very early do exist and there is, therefore, a need for studies focusing on the effects of adverse listening conditions on early bilinguals which will have "important implications for speech perception in educational, occupational and rehabilitative settings for this population" (Febo, 2003, 4). The pool from which Febo recruited his participants consisted of students, instructors, staff and the surrounding community of the University of South Florida, we can thus say that the author tried to choose the participants with similar language background in order to create as homogenous group as possible. It is rather problematic to place the participants of Rosenhouse et al. (2006), the authors themselves do not mention whether their participants are early or late bilinguals, nevertheless, they do mention the fact that their participants began learning their L2 Hebrew in the 3rd grade, have been exposed to Hebrew for more than 10 years and resided in Israel, where the dominant language is Hebrew, we thus assume that their performance resembled those of early bilinguals (Rosenhouse et al. 2006, 121-122).

In a similar spirit to that of Febo (2003), Tabri et al. (2010, 411) speak about the number of multilingual children increasing world-wide, many of whom "are being taught in their non-native language under poor classroom conditions." Moreover, they speak about the need to add more evidence to studies such as Mayo et al. (1997), who found early L2 learners to perform better than late ones, it also expands the evidence by employing trilinguals to their field of study. Tabri et al. (2010) used a complex strategy to select their participants, a speech pathologist, native speaker of English, assessed their L2/L3 according to their fluency and proficiency in a face-to-face interview, so it seems that Tabri et al. (2010) used not only early bilingualism or trilingualism as the decisive factor but also proficiency to select their participants. A further selection took place according to AoA: English had to be acquired prior to 6 years of age; the participants had to receive more than 5 years of formal English education, rate themselves fluent in written and spoken English, spent more than half of their time reading or listening in English and 25 % of their communication had to be done in English.

On the other hand, e.g. Wijngaarden et al. (2002) focused only on late non-native learners of Dutch, unlike other studies; however, they chose to investigate not perception, but the intelligibility of L2 production under adverse conditions. Their L2 speakers were all late learners (AoA ranging from 19 to 28) and they came from a different L1 background (American English, Chinese, German and Polish). Even though the L2 speakers were late bilinguals, it does not really enter into much consideration, the authors center more on L2 experience, L2 speakers' accent self ratings and foreign accent ratings obtained from native Dutch language judges. The aim of the study was to investigate the possibility of establishing L2 speech intelligibility from accent strength and self accent ratings and also to investigate the effect of phoneme vs. sentence level of L2 speech production on intelligibility. The authors found experience with L2 as an important factor for L2 speech intelligibility, as well as the received overall foreign accent and self-ratings of L2 proficiency (Wijngaarden et al. 2002, 3012).

Some studies concentrated on both early and late bilinguals, e.g. Mayo et al. (1997), Ezzatian et al. (2010) and Rogers et al. (2010). The aim of Mayo et al. (1997) was to establish how AoA influences the perception of L2 speech in noise. The authors were pioneers to focus in detail on the difference between early vs. late bilinguals, apart from monolinguals (MON), the study consisted of 3 other groups of participants: bilinguals since infancy (BSI), who could not recall which language they learned first and who spoke Spanish with one and English with the other parent; bilinguals since toddle (BST), learning L2 English prior to the age of 6 and bilinguals post puberty (BPP), learning L2 English after the age of 14. The results for BSI indicated that the group was different from MON while being very similar to BST and as such was merged with the latter into the group of early bilinguals (EB). The authors used Speech Perception in Noise Test (SPIN), employing sentences with controlled predictability (context aids the listener in 50 % of the sentences) in babble noise, consisting of 8 lists of 50 sentences with the target, in form of a noun, falling on the end of every sentence (Mayo et al. 1997, 687). As was already mentioned before, EB tolerated more noise than BPP, even though, they did not quite reach the same level of tolerance as the MON group.

The participants of Ezzatian et al. (2010) also consisted of 4 groups as in the previously mentioned research, but they were of a different linguistic background, more specifically, the native group consisted of L1 speakers of English, born and raised in English-speaking countries; the second group arrived in Canada between the age of 7-14, at which point they judged themselves fluent in their L2; the third group was from a non-English speaking country being raised there for at least 15 years prior to their arrival in Canada, receiving only some formal English foreign language education; the forth group consisted of English-educated participants but in a non-English-L1 environment, arriving in Canada when they were older than 10 years. Similar to Mayo et al. (1997), the authors found that the earlier the L2 was acquired, the better the participants performed in all conditions, in other words, they found that "the later the age of language acquisition, the higher the threshold for speech reception under all conditions" (Ezzatian et al. 2010, 919).

Surprisingly, many studies do not mention whether their participants are early or late bilinguals, it is the case of e.g. Cutler et al. (2007), Brouwer et al. (2012), Ishida and Arai (2015), Marchegiani and Fafoutis (2015), Cooke and Lecumberri (2016), Masuda (2016), it is unclear why the researchers would avoid such an important factor, considering the effect it could have, as suggested by e.g. Mayo et al. (1997). Brouwer et al. (2012) in their experiment 2 do not mention whether their participants are early or late bilinguals either, L2 proficiency seems to be a more deciding factor when selecting their participants. Nevertheless, the authors talk about the fact that the participants' English lessons started at the age of 11, from which we can gather that, rather than being early, they represent late bilinguals, moreover, the authors mention typically high quantity and quality of English exposure in Holland hence they consider their participants as highly proficient non-native English listeners (Brouwer et al. 2012, 1455). Marchegiani and Fafoutis (2015) do not take into account bilingualism either, their focus is however somewhat different from the other studies, more specifically, they investigate L1 speech perception in the presence of L2 (English) noise, which, as Lecumberri and Cooke (2006) suggested, might be more about L1 robustness than L2 abilities. More specifically, the authors found small native-English-speaker benefit when exposed to unfamiliar (Spanish) noise compared to L1 (English) noise, "perhaps due to a reduced attentional engagement" (ibid, 2453-2454). Cooke and Lecumberri's participants were either Spanish monolinguals or Spanish-Basque bilinguals, but since they were second-year students of English

Philology at the University of Basque Country, we expect their proficiency to be rather high and might even consider them late bilinguals. Masuda (2016) used a group of heterogeneous L1 Japanese speakers, who differed in age (18 to 36), onset of L2 (English) learning (3 to 13) and length of residence in L2 speaking country (0 to 60 months). Unlike Mayo et al. (1997), she did not find effect of age of learning, which she contributed to incomparable participants' background, as some of her participants were not bilinguals and although they started learning L2 early, the quality of input was put into question, some of them did not even experience residency in an English-speaking country, while the participants in Mayo et al. (1997) were bilinguals with a constant native input being immersed in L2 environment.

Despite the fact that in some studies the consideration of bilingualism was neglected or ignored, it should not be taken as a rule, especially when we have a strong precedence in form of the results of Mayo et al. (1997), who showed us that bilingualism can be a very important factor affecting the perception of L2 speakers of English when listening to speech in the presence of background noise.

4.1.2 L2 Proficiency

Some studies did not take into account AoA and relied on proficiency instead, the listeners had to undergo a test or fill in a questionnaire and in that way the studies satisfied their desire to separate them into groups, according to their proficiency as in Van Engen (2010) and Calandruccio and Buss (2017). Such way of sorting is rather problematic, as e.g. Febo (2003, 2) argued that even the highly experienced L2 learners exhibit significantly poorer recognition of L2 sounds compared to their monolingual counterparts.

In Van Engen's research (2010), 20 native Mandarin Chinese listeners with an average age of 24.5 participated. All of them were first-year-graduate students of a US university who were attending English language and acculturation program. Even though the author admits imperfect uniformity of English proficiency, all the participants had attained the TOEFL scores the university required them to attain in order to be admitted, their participation took place within three months of their arrival. They also had to complete an additional lab-internal Language Experience and Proficiency Questionnaire (LEAP-Q) in order to further characterize their proficiency in Mandarin and English. The authors found that the non-native speakers required a significantly more favorable SNR in order to identify English sentences in noise, nevertheless, no significant effect of English experience and proficiency was found (Van Engen 2010, 950).

Calandruccio and Buss (2017) also focused on proficiency, but unlike Van Engen (2010), they did not employ noise, but manipulated frequency regions instead, in order to create adverse conditions, employing low frequency band centered at 500 Hz, high frequency band centered at 2500 Hz and the combination of the two. The non-native participants were also from Mandarin Chinese cohort, 4 males and 15 females of a mean age of 26 participated in the research. Before taking part in the experiment, all non-native participants had to complete a linguistic and demographic questionnaire developed by the Linguistics Department at Northwestern University assessing areas of language status, language stability, language competency, language history and demand for language use (Calandruccio and Buss 2017, 1647). In addition, they had to complete the Versant English test, an automated speech recognition test over the phone, usually used by businesses hiring non-native employees, or by universities testing the level of English language skills of their potential graduates and English teaching candidates, testing sentence mastery, fluency, vocabulary and pronunciation, providing English assessment score on a scale from 20 to 80 points, in this case taking place in a double-walled, sound-isolated room. According to the authors, the "Versant scores are positively correlated with non-native English speakers' ability to understand English speech in noise," which explains why the authors do not take into account early or late bilingualism, as they have a different way to assess the ability to understand English speech in noise (ibid, 1647). Nevertheless, the authors mention the way the participants learned their L2, they talk about little "diversity in the age of acquisition, [and that] all of these listeners learned English in a similar manner (predominately for higher education), and all were either in their late teens or adults when they emigrated to the U.S.," suggesting that we are dealing with late, rather than early bilinguals. Similar to Van Engen (2010), Calandruccio and Buss (2017) found the non-native speakers to perform significantly worse than the native ones, more specifically, they needed a significantly wider bandwidth for both low and high frequency bands, they were also less adept to combine speech cues across the two bands in order to support sentence recognition (ibid, 1651). The authors expected to find greater L2 experience to allow recognizing "speech based on more spectrally limited cues," compared to the less experienced participants (ibid, 1651). Nevertheless, the findings were mixed, the overall Versant score was non-predictive and the AoA was only predictive for high frequency band, but not for the low band (ibid, 1651). The authors argued that greater English L2 experience leads to better utilization of high frequency band for the non-native listeners, but also pointed out that the participants formed a rather homogeneous group for language experience to show any significant distinction between them (ibid, 1651).

Judging from the two studies, it seems that proficiency should not be the main decisive factor when sorting participants into groups, or we should not employ such factor as the sole device in such sorting, instead, we should focus on or add the early/late bilingual distinction to create a sensible division between the participants.

4.1.3 Type of Noise

A fair amount of types of noise was investigated by studies with a predominance of multi-talker babble, e.g. Mayo et al. (1997), Rosenhouse et al. (2006), Van Engen (2010) and Tabri et al. (2010) and speech-shaped noise, for instance, Cutler et al. (2007), Golestani et al. (2009), Marchegiani and Fafoutis (2015), Cooke and Lecumberri (2016); but there are also other types of adverse conditions investigated, i.e. reverberation, studied by e.g. Febo (2003) or Masuda (2016) and competing talkers, e.g. Cutler et al. (2007), Ezzatian et al. (2010). The reason why the multi-talker babble is a commonly used masker noise is the fact that together with competing speech, with small amount of talkers present in the background noise, they are "regarded as the most effective form of informational masker [...] [even though] all maskers have some occasional IM potential" Lecumberri et al. (2010, 873). Similarly to Lecumberri et al. (2010), for Silbert et al. (2014, 2227) multi-talker babble is also an excellent masker for speech in perceptual experiments, making arguments for its ecological validity compared to other maskers, such as white noise or speech-shaped noise, especially due to its similarity to speech noises frequently present in everyday's life (ibid, 2227). It thus seems that experiments, which aim at creating conditions that are close to those encountered in real life, should employ multi-talker babble. The reason why multi-talker babble and similar types of noise constructed from speech are so successful maskers is the fact that they are comprised of “acoustic properties similar to that of the target signal,” hence they have the biggest potential to obscure it (ibid, 2227).

Studies which employ multi-talker babble vary in the use of talkers, as was already mentioned in the former paragraph, according to Lecumberri et al. (2010) a smaller number of talkers in the babble seems to serve as a more effective informational masker compared to cases of babble with a higher number of talkers, a fact that is supported e.g. by Freyman et al. (2004), investigating the effect of the quantity of talkers in the babble masking. In particular, the authors found 2-talker babble the “most effective in creating informational masking” (Freyman et al. 2004, 2250). For Marchegiani and Fafoutis (2015, 2208), a “boundary for natural babble between conditions in which words are still detectable and conditions in which they are not,” was found to be three talkers, which suggests most detrimental effect of informational masking when the number of talkers in the noise equals three. Their noise condition included speech-shaped noise and varied number of speakers in natural babble and babble-modulated noise. Nevertheless, unlike expected, more talkers, up to 64, in the natural babble actually worsened the identification rate of their listeners, as we can see in the Figure 2. The authors used /ɑ/-CONSONANT-/ɑ/ pattern to test for the consonant-identification rate, studying phonemes rather than words, offering a context-free environment leading to more detrimental effect of higher number of speakers in the babble, which could mean that energetic masking has more effect in such cases, preventing the speech sounds to enter into cognition of the listeners. Some studies followed the same reasoning as Lecumberri et al. (2010), employing only small number of talkers in the babble, such as Van Engen (2010) and Brouwer et al. (2012) who used two talkers in their noise condition. Nevertheless, some studies chose a different path, using rather high number of talkers in the babble, e.g. 8-women talkers in Rosenhouse et al. (2006), 12-talker in Rogers et al. (2010) and Tabri et al. (2010), 20-talker in Rogers et al. (2004) and if what Marchegiani and Fafoutis (2015) suggests is true, meaning that three talkers mark a threshold for informational masking to take place, it would mean that their listeners were exposed basically to energetic masking. Other studies do not mention the number of talkers in their babble-noise condition, such as Mayo et al.

(1997) and Masuda (2016), which is surprising, considering the effect it might have on the results, if we consider Lecumberri et al. (2010).

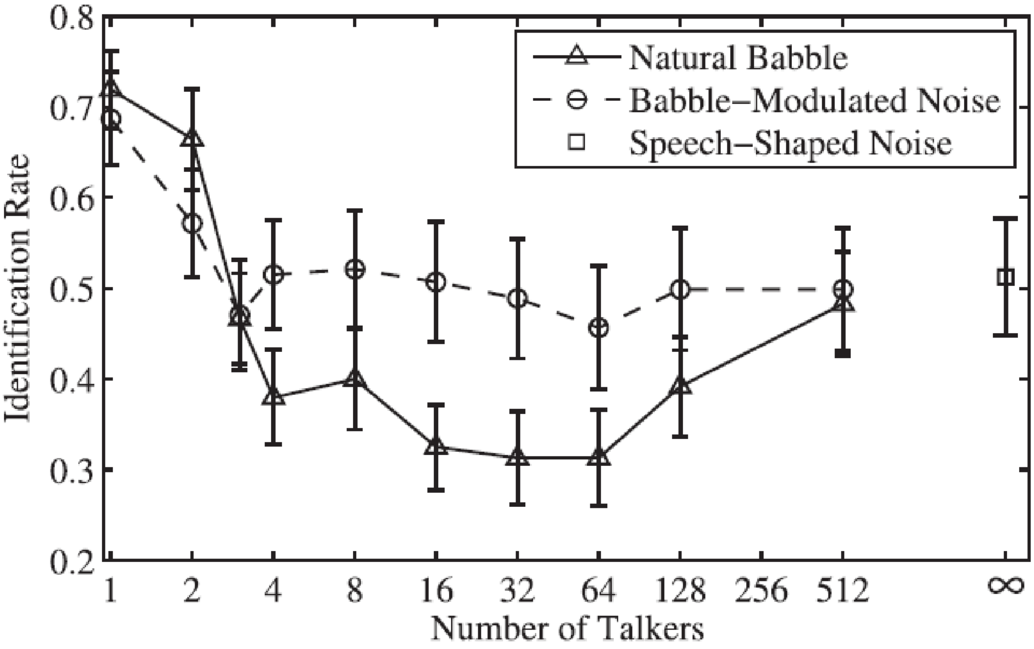


Figure 2 represents identification rates when the stimuli were presented in conditions of natural babble, babble-modulated noise with different number of talkers in the masker. Condition with speech-shaped noise was also presented, indicated in the figure with ∞ symbol (adapted from Marchegiani and Fafoutis 2015, 2208).

We can also note somewhat different case of Ishida and Arai (2015) who examined perceptual accuracy and phonemic restoration of L2 participants. Their native Japanese speakers listened to a pair of words, one of which consisted of English word or non-word with either phoneme covered by or replaced with signal-correlated noise, the second of which consisted of an English word or non-word without noise, after hearing the pair, they evaluated their similarity on an 8-point scale. They found out that the words (or non-words) containing phonemes with added noise were more similar to those without noise compared to the ones whose phonemes were replaced by noise (ibid, 3410-3411).

4.1.4 Background Noise (Native/Non-Native/Unfamiliar Language)

When considering background noise, some researchers also focused on the effect of language of the masker on the performance of native and non-native listeners.

For example Lecumberri and Cooke (2006) investigated the effects of speech-shaped noise, 8-talker babble and most importantly English vs. Spanish competing talker. They focused on the perception of English intervocalic consonants of native monolingual speakers of English studying at the University of Sheffield and the non-native group of L1 Spanish speakers, students of the University of Basque Country studying English as a foreign language. Even though the level of English competence of the non-native group was not uniform, they were all advanced L2 speakers of English, as all of them passed the Cambridge Advanced Exam. When comparing the results of the native English and the non-native Spanish speakers, the authors found "a significant interaction of language and nativeness [...] due to a small but significant difference for the native group," suggesting that the English listeners were better in tuning out the masker when it consisted of unknown Spanish competing talker, as opposed to Spanish speakers for whom no release from masking in either Spanish or English competing talker was found, as we can see in Figure 3 (ibid, 2448-2449).

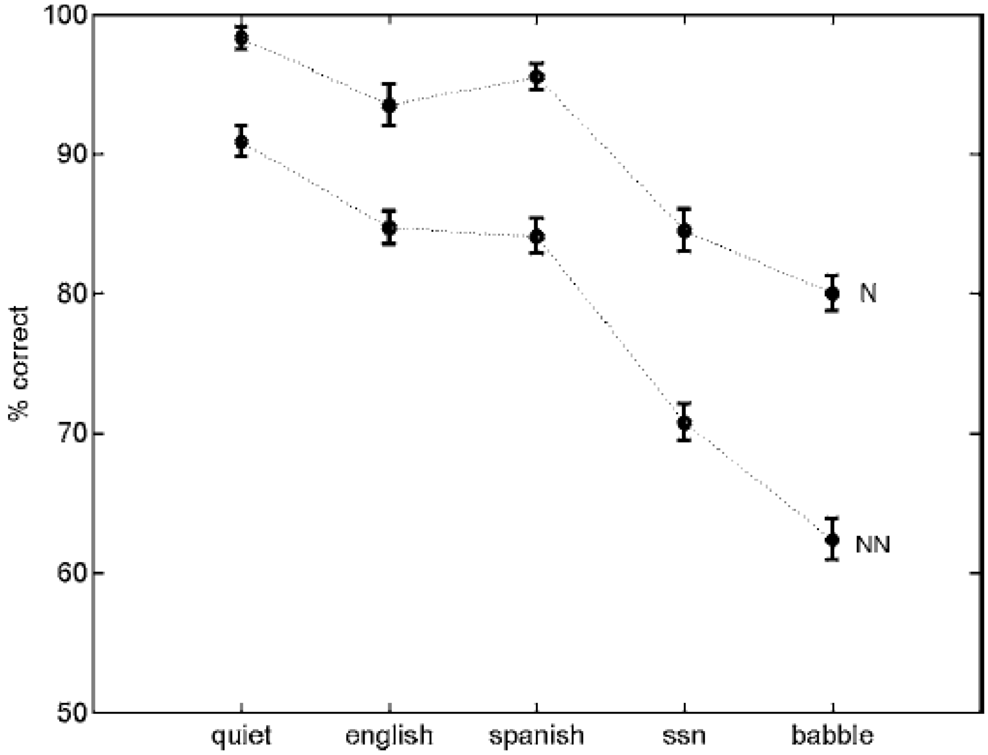


Figure 3 shows consonant identification scores for native (N) and non-native (L1 Spanish) (NN) speakers of English in quiet, competing English and Spanish talkers, speech-shaped noise and 8-talker babble. Important to notice is the performance of native speakers who receive release in competing Spanish speech (adapted from Lecumberri and Cooke 2006, 2448).

Van Engen (2010) investigated first and second language sentence recognition of native English and non-native Mandarin Chinese speakers in English and Chinese 2-talker babble. The author found similarities to that of the previous study by

Lecumberri and Cooke (2006), namely, both studies found a significant release from masking when L1 English speakers listened to target stimuli in the presence of masker in an unknown language. Nevertheless unlike the previously mentioned research (Lecumberri and Cooke 2006) which found no effect of language for the non-native listeners, Van Engen's results suggested a release from masking even for the L2 listeners when they listened to L2 target sentences in L1 Mandarin Chinese babble, the L2 listeners thus benefited from target-noise mismatch. Importantly, the native English speakers received a greater release from masking compared to the non-native Mandarin Chinese when listening to English target sentences in Mandarin Chinese babble, suggesting that "acoustic and/or linguistic similarity between the speech signal and the noise may be the most critical factor driving noise language effects" (Van Engen 2010, 951). The release from masking in English and Mandarin Chinese 2-talker babble for the two groups is presented in the Figure 4.

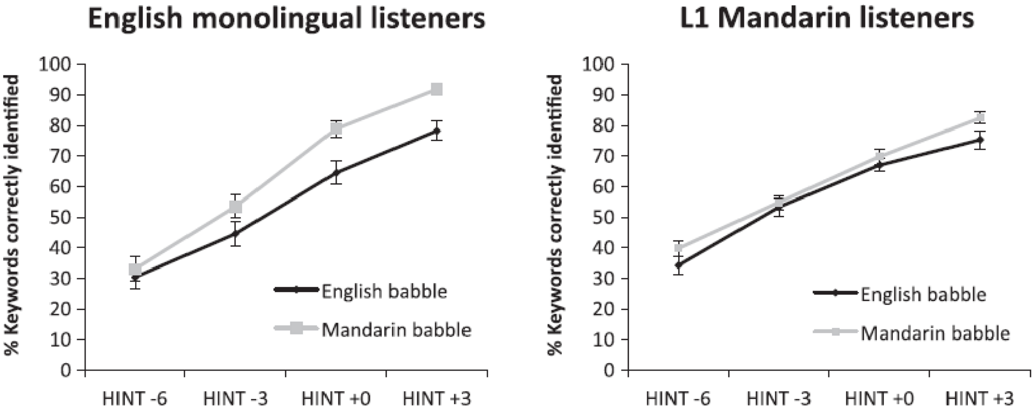


Figure 4 shows percentage of correctly identified keywords for L1 speakers and L2 speakers of English (L1 Mandarin Chinese) when exposed to target English sentences in English and Mandarin babble (adapted from Van Engen 2010, 950).

The language of the background noise was also part of Brouwer et al. (2012), who employed three experiments to investigate how speech recognition varies in relation to target and masker languages. The first experiment examined native English listeners presented with meaningful L1 English target sentences in L1-English/unfamiliar-Dutch 2-talker babble with meaningful or anomalous sentences. The second investigated Dutch-English bilinguals highly proficient in English presented with the same material as those in the first experiment, but this time the language of the Dutch 2-talker babble was the same as their L1. The third experiment involved Dutch-English bilinguals from similar linguistic background as those taking part in Exp 2 using the same background speech material as in Exp 1 and 2 but with L1 Dutch target sentences. Results showed similarly to Lecumberri and Cooke (2006) and Van Engen (2010) release from masking for monolingual English listeners when the targets were L1 sentences and background noise was unfamiliar, 2-talker Dutch babble (Brouwer et al. 2012, 1460). The second experiment found concordance with Van Engen (2010), finding release from masking for Dutch-English bilinguals for L2 English targets with L1 Dutch background noise (the case of target-language-noise-language mismatch) but diverged from Lecumberri and Cooke (2006), for whom no such effect was found for their Spanish-English bilinguals (Brouwer et al. 2012, 1460).

Research seems to suggest that the language of the background might offer release from masking when there is target-masker mismatch as far as native speakers are concerned but also for the bilingual speakers, the reason why Lecumberri and Cooke (2006) did not find effect of the target-masker mismatch for their Spanish-English bilinguals might be due to differences in methodology, as their stimuli involved consonants in VCV pattern, whereas Van Engen (2010) and Brouwer et al. (2012) focused on target words in sentences, it might thus be more difficult to tune out the bilinguals' L1 when the targets are small elements such as individual phonemes, compared to bigger elements, such as words. Moreover, the discrepancy between the results in the discussed studies could be due to the difference in the masker type as Lecumberri and Cooke (2006) employed a competing speech, whereas both Van Engen (2010) and Brouwer et al. (2012) employed 2-talker babble.

4.1.5 Native Input and Length of Residence (LoR)

When native input of the non-native speakers is considered, such factor is not viewed consistently, in many studies, there is no information provided about the level of native input, for instance, Cutler et al. (2007), Golestani et al. (2009), Brouwer et al. (2012), Ishida and Arai (2015), Marchegiani and Fafoutis (2015). Some studies do not provide information as such, nevertheless, we can infer something from the description of the participants, such as those in Cooke and Lecumberri (2016), who talk about second-year students of English philology, suggesting a higher degree of input, as they are possibly in contact with native speakers on a daily basis. Moreover, Van Engen (2010) talks about her participants as having recently arrived in the USA and taking part in the English language acculturation program, which would suggest a previous lower degree of native input, even though, we cannot be sure as the participants' past is unknown, they could have received more native input than expected. There are studies that consider such input, but include a varied group with the input ranging from half a year to 28 years, as in case of Wijngaarden et al. (2002)[2]; from 8 months to 6 years, as in the case of Rogers et al. (2004)[3]. Some studies are fairly specific, e.g. Mayo et al. (1997) and Febo (2003), both of whom are concerned with Hispano-Americans with a high degree of native input and length of residence; moreover, a very specific situation involves the participants in Rosenhouse et al. (2006): they are native Arabic speakers immersed for more than 10 years in L2 Hebrew environment; in addition, Ezzatian et al. (2010) also mention a high degree of native input, their participants emigrated to Canada and were immersed in Canadian English for at least three years, nonetheless their situation is somewhat different than that of the other studies with designated high degree of input, as their participants are from various L1 backgrounds. Some studies are very specific in their description, such as Tabri et al. (2010), selecting to their bilingual group only those spending "more than 50 % of their time reading, listening to music, or watching television and films in English, and [... communicating] in English more than 25 % of the time" (Tabri et al. 2010, 413). Masuda (2016) in her study does not mention native input as such, nevertheless, the length of residence is observed, a subgroup of her participants include those who took part in a one-month study abroad program resulting in higher TOEIC scores and higher identification rates in all the conditions they were tested, suggesting that even such short study abroad program as one month "may play a role in improving learners' TOEIC® scores as well as higher accuracy in identifying non-native speech in degraded listening environments" (Masuda 2016, 82).

As we have seen, studies differ in their approach to native input and the length of residence in L2-speaking countries as there is little consistency in the provided information. The finding of Masuda (2016) is interesting and very

[2] Concerned L2 speech production.

[3] Concerned L2 speech production.

relevant to our topic as it suggests that even a relatively short length of residence in L2 environment might lead to better results.

4.1.6 Contextual Cues

Some studies also investigated, whether predictability of the target words may provide release for the listeners and how it is reflected in the perception of native and non-native speakers. For instance, Mayo et al. (1997) administered SPIN test to their listeners in which predictability from the context of the sentence for the target final words was provided for half of the sentences the listeners were exposed to. Sentences in which the final word was contextually predictable included e.g. "The boat sailed across the bay," as opposed to the non-predictable "John was thinking about the bay" (ibid, 687). As mentioned in chapter 4.1.1, Mayo et al. (1997) employed four groups of listeners: English monolinguals (MON), Spanish-English bilinguals since infancy (BSI), since toddler (BST) and those learning their L2 after puberty (BPP). The results for the BPP listener group were important as different effect of context was revealed: the slopes for correctly identified high and low predictability target words in noise were almost equal, an interesting finding if we consider that all the other groups showed "considerable advantage from hearing parts of the words in sentences in which the target word can be deduced from the carrier phrase," indicating that late L2 learning hampers the BPP's ability to use across-word contextual information when trying to identify the target word if the sentences are masked by noise (ibid, 691). As we can see in the Figure 5, the averages of slopes for high and low predictability for the BPP group are very close to those of the low-predictability for the BSI and BST group, suggesting almost no effect of context for the BPP group.

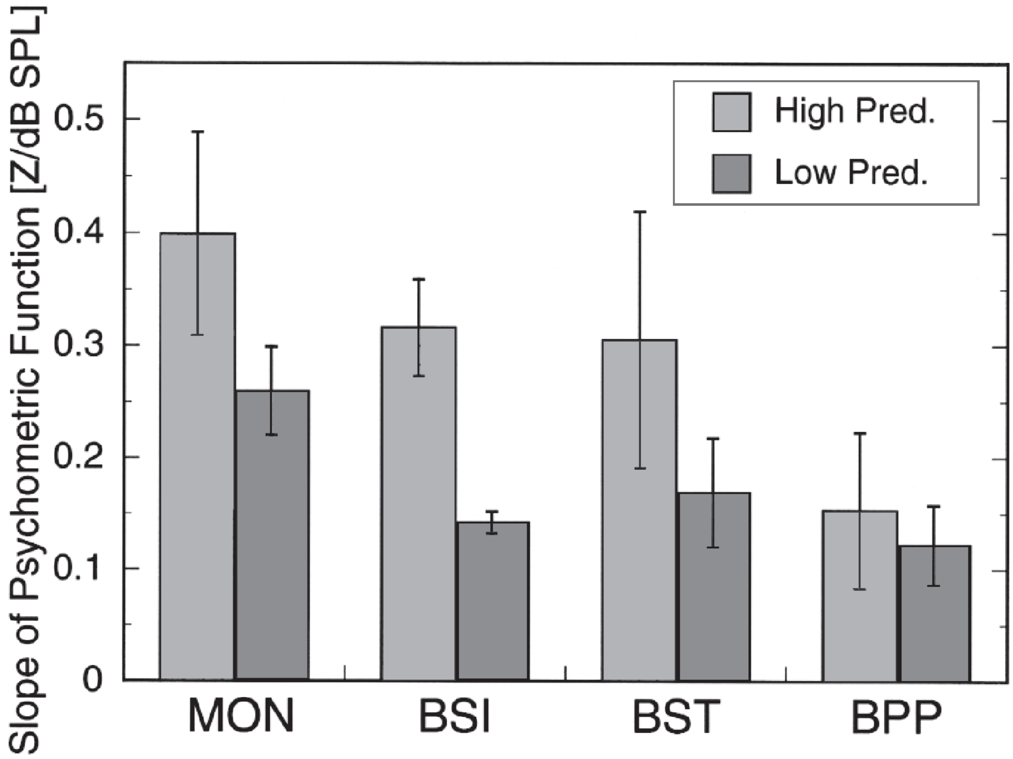


Figure 5 shows the average slopes of psychometric functions taken from the z scores transformed from the raw data for the four groups of listeners (adapted from Mayo et al. 1997).

Ishida and Arai (2015) investigated the perception of English liquids and nasals by Japanese speakers of English and compared their performance to native speakers of English in Mattys et al. (2014). The participants listened to English words or non-words with noise which was added to or replaced the phoneme. After hearing the modified word or non-word, they were exposed to an English word/non-word without noise in a row. Their objective was to evaluate the similarity of the pairs on a scale from 1, not similar and 8, very similar. The authors argued that lexical contextual knowledge should help the listeners restore the missing phonemes: restoration should be better in words, compared to non-words (Ishida and Arai 2015, 3408). As seen in Figure 2, the listeners, however, awarded no different score for non-words, compared to words, suggesting that context did not help the non-native listeners in recognizing the original speech behind noise (ibid, 3409). It is difficult to compare the results of Ishida and Arai (2015) with the previous study, as it does not offer much information about the proficiency or bilingualism of their native Japanese participants; nevertheless, they seem to perform similarly to the BPP group in

Mayo et al. (1997), for whom the authors did not find significant contextual effects.

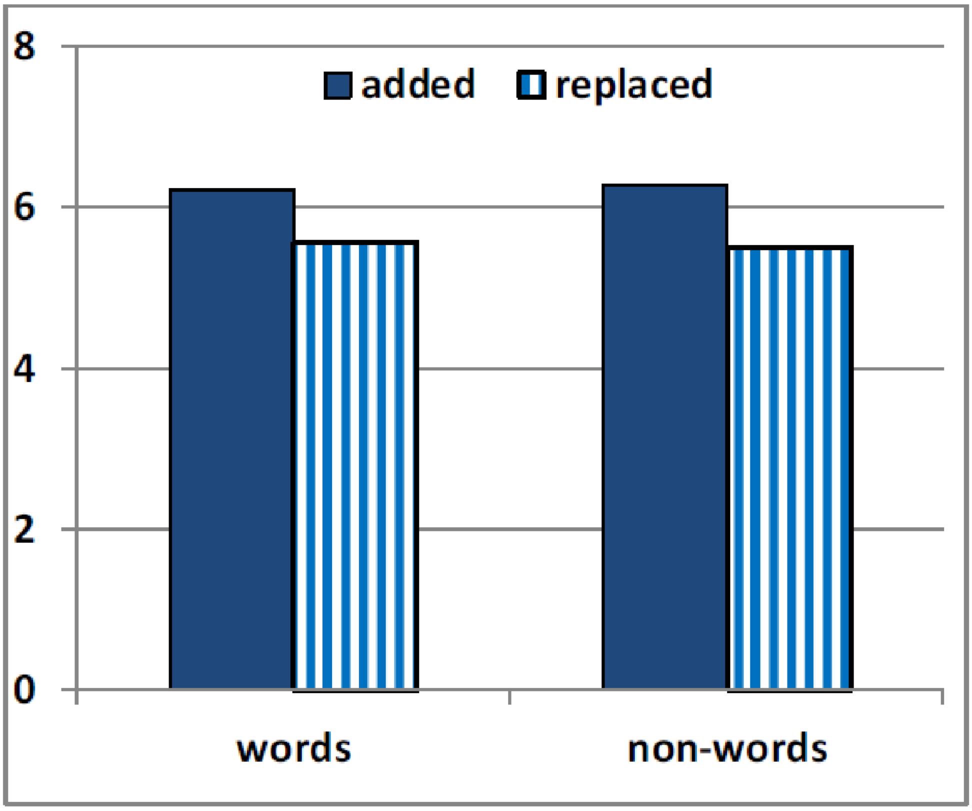


Figure 6 shows the mean similarity scores on 1-8 scale awarded by non-native listeners of English when exposed to pairs of words/non-words with added/replaced noise and words without noise (adapted from Ishida and Arai 2015, 3410).

In a similar manner to Mayo et al. (1997), the results of Golestani et al. (2009) suggested semantic level of language to contribute to the native language advantage for speech in noise. The authors used a different design, rather than using sentence recognition, they aimed to isolate solely the semantic level, investigating whether "semantically related target word facilitates the perception of a previously presented degraded prime word" for native and non-native languages (ibid, 385). Their native French participants were exposed to native French and non-native English pairs of words, the first of the pair called prime, was degraded with noise in different SNR levels (from -7 dB to -4 dB), it was also presented without noise. Moreover, half of the degraded prime words were semantically related and half unrelated to the second target word (the former included e.g. *parrot* as prime and *bird* as target, the latter e.g. *parrot* as prime and *cake* as target). After hearing the prime and the target, two words were displayed, the first one the prime and the second one semantically related foil and the listeners were to decide which of the two corresponded to the prime, by pressing a button. As was already mentioned, the authors found semantic advantage for native but not for non-native language. Furthermore, a surprising finding was discovered: the authors found semantic detriment when the participants listened to non-native language pairs, as seen in Figure 7, suggesting that "in the participants' less fluent language (English), hearing degraded words followed by semantically related words results in semantic interference" (ibid, 390). As the participants of Golestani et al. (2009) were similar to that of the BPP group in Mayo et al. (1997), even though their proficiency was probably lower, at least according to the description provided, the results seem to confirm that late learners do not benefit from semantic relatedness when the task concerns their non-native language, in case their proficiency is rather low, it might even cause harm to their perception of speech in noise.

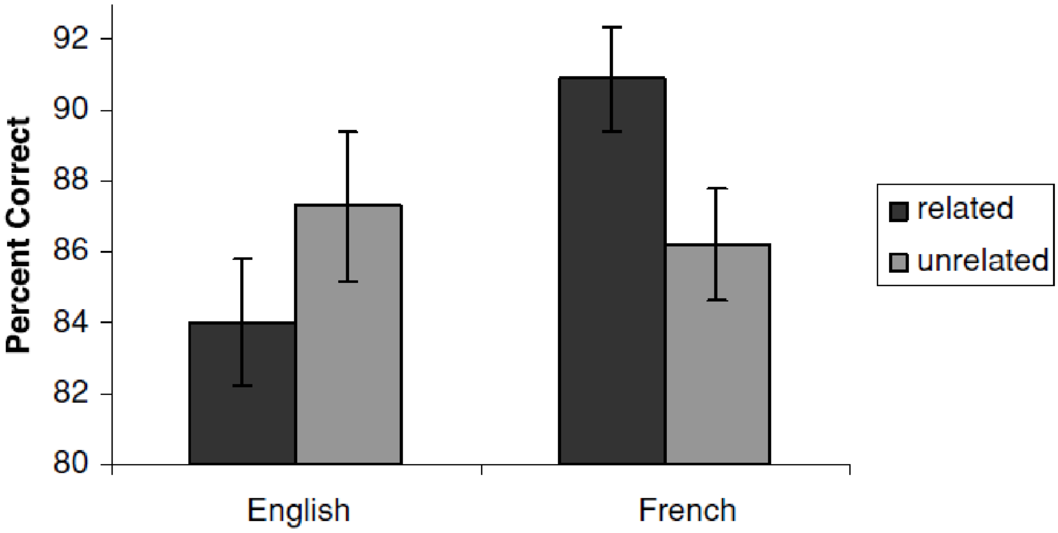


Figure 7 shows language-by-relatedness interaction for the native French and non-native English language of the participants (adapted from Golestani et al. 2009, 389).

As the outlined studies suggest, context might be beneficial for native listeners, we should, however, be careful when focusing on non-native listeners, it seems that early bilinguals also receive a release from masking when the target is contextually related, as found in Mayo et al. (1997), nevertheless for late learners, as the BPP group in Mayo et al. (1997), context is not beneficial for their performance, or may even be harmful, as Golestani et al. (2009) suggested.

4.1.7 *Speech Rate*

Not many studies focused on speech rate as a possible factor in perceptual studies in noise. From the more recent ones, we can mention Rosenhouse et al. (2006) and Shi and Farooq (2012). The effect of speech rate on L1 and L2 speech perception in optimal condition and in background noise was investigated by the former. The authors based their stimuli on 64 CHABA sentences, adapted half of them to the participants' L1 colloquial Arabic and the other half to their L2 Hebrew, content of both languages was similar, involving daily objects and actions, with special effort made to use the "same target words and as much as possible the same number of syllables in the parallel sentences in the two languages," with differences stemming solely from linguistic necessity, when it was not possible to maintain such structural similarity due to the quality of one of the languages (Rosenhouse et al. 2006, 122). Sentences were recorded in two conditions, one "normal" (at about 3 syllables per second) and the other "fast" speech rate (at about 4 syllables per second). The noisy condition included 8-women-talker babble at an SNR +6 dB, creating 4 conditions: in quiet with normal speaking rate, in quiet with fast speaking rate, in background noise with normal speaking rate and in background noise with fast speaking rate. As we can see in Figure 8, no significant difference was found for bilingual perception in L1 and L2 under optimal condition in quiet with regular speech rate, the authors, nevertheless, found deterioration of speech perception in quiet condition with fast speech rate, with L2 target words being perceived with more difficulty compared to the L1 words (ibid, 127). Moreover, the noisy condition proved to be more difficult than fast speaking rate, with the combined condition of noise and fast speaking rate creating the most deteriorating environment with a tendency to hamper the listeners' ability to perceive words in their L2 considerably more than in their L1 (ibid, 126-127).

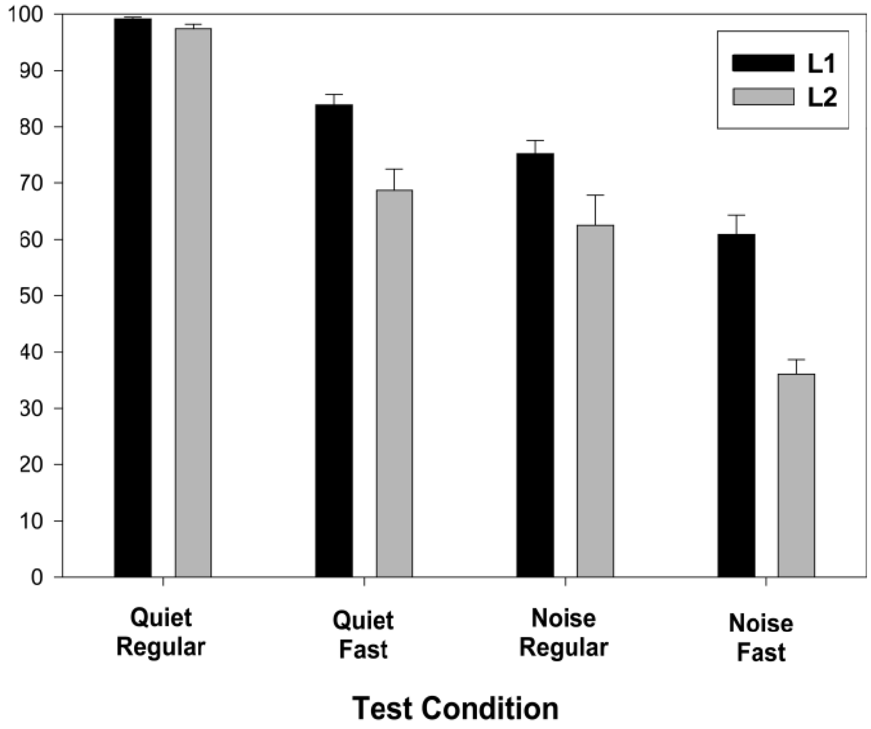


Figure 8 shows group means of correct words in percentage for L1 and L2 of the Arabic-Hebrew bilinguals in the four conditions investigated (adapted from Rosenhouse et al. 2006, 125).

Shi and Farooq (2012) used English monolinguals, English-dominant bilinguals (EDB) and non-English-dominant bilinguals (NDB) in Connected Speech Test (CST) with passages of 9-11 sentences and a total of 25 key words per passage, read by female native speaker of English. In the noisy condition, 6-talker babble originally present in the CST recordings was used. The authors employed five speech rate conditions, the normal average speech rate was counted at 4.38 syllables per second (s/s). Apart from the normal speech rate, CST passages were temporally manipulated to create two fast speech rate conditions compressing time by 15 % and 30 %, creating average speech rate of 5.13 s/s and 5.63 s/s. For the two slower speech rate conditions, the duration of the stimuli was lengthened also by 15 % and 30 %, resulting in average speech rate of 3.86 s/s and 3.38 s/s, the listeners were thus exposed to the total of ten conditions (quiet and noisy multiplied by the five speech rate conditions). Similar to the previously discussed study, the perception of target words deteriorated with faster speech rates for both monolinguals and bilinguals, but with the presence of noise, the EDB was affected significantly more than the other groups, as their performance in noise drew near to that of the NDB (ibid, 133). The authors found “the amount of decline in performance [... to be] rate dependent," with the bilingual listeners failing to reach the monolingual level even at slower speech rates, suggesting a rather limited benefit even for the EDB, compared to the monolingual speakers.

We can see similarity between Shi and Farooq (2012) and Rosenhouse et al. (2006) as the participants' performance declines the most in noise and faster speech rate, even though Rosenhouse et al. (2006) seems to suggest a more drastic decline in performance for L2 targets in combined noise and fast speech rate viz Figure 8. Such contrast might be explained by differences in methodology as Shi and Farooq (2012) used adaptive noise procedure to establish 50 % correct performance level and avoid ceiling and floor effects for the three listener groups, reaching the average SNR of -3.10 dB for the English monolinguals, -1.50 dB for the EDB and 0.35 dB for the NDB, trying to maintain the noise at the level at which the groups would reach comparable results and thus investigating mainly the effect of slowed and accelerated speech rate. Such method was not employed by Rosenhouse et al. (2006), who used only SNR +6 dB in both L1 and L2 of the participants. The differences in SNR, thus, seem to explain why noise might have been more detrimental in the combined condition of Rosenhouse et al. (2006), compared to Shi and Farooq (2012).

The studies suggested a possible effect of speech rate on the perception of speech in quiet, with intensified detrimental effect when noise was present and when the listener was exposed to L2 rather than L1 targets. It might be, therefore, beneficial to keep track of the average speech rate of the stimuli that the listeners are exposed to, in order to avoid the creation of a more detrimental effect than originally intended.

4.1.8 Languages Investigated

The target language of the vast majority of studies in adverse conditions was English, e.g. Mayo et al. (1997), Febo (2003), Lecumberri and Cooke (2006), Cutler et al. (2007), Ezzatian et al. (2010), Rogers et al. (2010). Some studies also focused on other languages as their target, for instance, Rosenhouse et al. (2006) employed as their targets 64 CHABA sentences, translated and adapted half of them to Arabic (L1) and half to Hebrew (L2); Golestani et al. (2009) used, apart from English (L2), also French as their target language in the task of retroactive word-priming. Furthermore, Brouwer et al. (2012) also focused their attention on Dutch (L1) target words in sentences in the presence of English or Dutch noise.

The native language of the participants usually included English, Spanish speakers were also often involved, e.g. in Mayo et al. (1997), Febo (2003), Cutler et al. (2007), Rogers et al. (2010), Cooke and Lecumberri (2016). Furthermore, native Arabic listeners were included by Rosenhouse et al. (2006) and Tabri et al. (2010); native Japanese listeners participated in the study of Ishida and Arai (2015) and Masuda (2016); native Dutch listeners took part in Cutler et al. (2007) and Brouwer et al. (2012). A different approach was chosen by a small amount of studies, for instance, Wijngaarden et al. (2002)[4], Ezzatian et al. (2010) and Marchegiani and Fafoutis (2015), in those studies, participants from a larger spectrum of L1 languages were selected.

The studies investigated include a wide range of languages, the author of this thesis, however, knows of no perceptual studies in adverse conditions which would include native Czech speakers. There is, nevertheless, a study in adverse conditions which concerned native Czech speakers, namely Volín and Skarnitzl (2010), the authors, however, focus mainly on the effect of signal degradation on foreign accent rating. The study investigated Czech accented English of three groups (near-native, strongly accented and averagely accented speech) by recording fluent reading of BBC news bulletins containing pronunciation features typical for Czech-accented English in the presence of brown noise, coffeeshop noise and filtered speech. The listeners were also recruited from the native-Czech-speaker populace, they were "three groups of undergraduate students of English phonetics and phonology [listening to the stimuli] in a sound-treated room via high-quality loudspeakers" and they "represented young educated Czech adults with very good command of English" (Volín and Skarnitzl 2010, 1014). We can thus say that there is a gap, concerning perceptual studies in adverse conditions as L2 listeners of English with L1 Czech background are practically absent, therefore future research should aim at including speakers from such L1 background and investigating their performance compared to speakers from other L1 backgrounds.

[4] Concerned L2 speech production.

4.2 Longitudinal Studies of L2 Perception

As there is lack of evidence of longitudinal perceptual studies where a noisy element would be employed[5], we review research papers which do not use such condition, hoping to shed light on the methods used in longitudinal studies. There is a fair amount of longitudinal L2 production studies, for example Munro et al. (2003), Tsukada et al. (2004), Oh et al. (2011), Holliday (2015), Chang (2019), Kartushina and Martin (2019) and Nagle (2019). Nevertheless, as we primarily focus on perception, our interest should steer in that direction, unfortunately, studies of L2 perception spread in a longer time span are rather scarce, which is surprising, considering the fact that their evaluation should take less effort than that of the production studies. From the studies that consider only perception, we can list Kim et al. (2018) and Sun et al. (2021). Furthermore, there are also studies that carried out both perceptual and production experiments, such as Aoyama et al. (2004), Tsukada et al. (2005), Aoyama et al. (2008) and McCarthy et al. (2014). This section will be concerned primarily with longitudinal studies of L2 perception, commenting upon production only marginally, those studies will be principally mentioned to show that production receives a disproportional interest compared to perception.

4.2.1 Research Aims and Results

Most studies were cross-sectional, employed both non-native (NN) children and adults, investigating whether either of the groups show improvement over the time span and comparing the performance of the two groups, e.g. Aoyama et al. (2004), Aoyama et al. (2008) and Kim et al. (2018). Even though NN adults had a starting advantage compared to NN children in all three studies, children improved greatly in most observed aspects between T1 to T2 compared to adults who usually exhibited either stagnation or marginal improvement and, as such, the "adult-child differences disappeared or became smaller over a one-year time frame" (Aoyama et al 2008, 85). Apart from the cross-sectional approach, Aoyama et al. (2004) also investigated whether their Native Japanese (NJ) participants would have

[5] This means to say that there are very few, if any, studies investigating L2 perception in noise in a longer time span, some studies employing perceptual training in noise were, however, carried out, usually taking place within shorter time span, sometimes with an element of formal instruction, we will focus such studies in chapter 4.3.

more success acquiring English /r/ than /l/. The Speech Learning Module (SLM) predicts more learnability for English /r/ due to its higher level of dissimilarity from Japanese /r/ compared to English /l/. The results revealed support for the SLM, especially for the NJ children, who exhibited in T2 greater learning for [ɹ] than [l] which was proved in the perceptual, as well as the production experiment (ibid, 246). Furthermore, when "'the room for improvement' is considered, the relative improvement for [ɹ] was larger than [l] for both the NJ adults and children in production (ibid, 246).

In addition to cross-sectionality, Kim et al. (2018) also wanted to find out if spectral and durational cue weighting change over time and across varied contrasts for non-native listeners, their results were mostly in conformity with the developmental stages for acquisition of /i/-/ɪ/ contrast found for Spanish learners of English, the stages were proposed by Escudero (2000) and included "(0) inability to distinguish the vowel contrast, (1) use of the duration cue to distinguish between the two vowels, (2) use of both duration and spectral cues but with main reliance on duration, (3) use of spectral cues to distinguish between the two vowels," some participants were, however, already in a more advanced stages of development even at the beginning of the testing (Kim et al. 2018, 15). Unlike the aforementioned studies,

McCarthy et al. (2014) focused solely on children, investigating how L1 interferes with L2 acquisition of early child learners, the authors discovered that their English plosives were initially determined by their L1 but after being exposed to a lot of L2 speech "changed to match that of their monolingual peers" (ibid, 1965). Sun et al. (2021) carried in a similar direction as McCarthy et al. (2014) in that it did not incorporate cross-sectionality, more precisely, it focused only on adults and investigated explicit and implicit auditory processing with their influence on L2 segmental and prosody acquisition in a form of sound discrimination threshold and music memory testing in case of the explicit processing and neural encoding of sound testing in case of the implicit processing. The authors found out that most L2 gains were associated with explicit auditory processing, especially remembering and reproducing music segments (Sun et al. 2021, 567).

4.2.2 Methods

4.2.2.1 Participants

All the reviewed longitudinal perception studies focus on English as the target language and from the participants' native languages predominate Asian ones: native Japanese children and adults took part in Aoyama et al. (2004) and Aoyama et al. (2008), at the time of the first test, the non-native listeners (NNL) were living in the US for about 5 months; in case of Tsukada et al. (2005) and Kim et al. (2018), we find native Korean speakers as participants of their studies, the location of the former is again the US, being exposed to English for 3 or 5 years on average; while in the latter the participants resided in Canada for 2 months at the onset of the study. McCarthy et al. (2014) also investigated native speakers of an Asian language, as native speakers of Sylheti, an Indo-Aryan language, living in the London-Bengali community, took part in their study. Their child participants were initially tested in nursery and a year later, when they finally started to receive massive English input, before that they basically stack to their Bengali community. Furthermore, Sun et al. 2021 followed the same course, investigating native Chinese listeners at an early phase of immersion in the UK, who had received considerable formal foreign language education in China, ranging from 10-19 years.

As seen, all the considered longitudinal studies of perception employed non-native speakers whose language had origins in Asia and, as such, there is a gap to be filled in the form of employing other NNS and consider their performance compared to those already investigated.

4.2.2.2 Time Span

In the methodology of longitudinal studies, researchers usually use time span around 1 year between the initial and the final testing, such as is the case of Aoyama et al. (2004) and Aoyama et al. (2008), who begin testing their participants when they are in their 5th month of immersion in the US, making the retest after 1.1 years, in their 1.6th year of immersion. With almost the same time between the tests, McCarthy et al. (2014) retested their child participants after 12 months and, even though they were exposed to English for a couple of years before, as they attended nursery, the defining language was that of their Bengali community. The authors found out that after the children had begun attending the school, joining increasingly the native English community, their perception and production of English plosives started changing dramatically towards their monolingual peers (ibid, 1965). Unlike the aforementioned studies, Kim et al. (2018) registered progress continually, the authors tested their participants initially within 2 months of their residence in Canada and retested them subsequently after 4 months, 8 months and 12 months. Such design might prove useful, as it could shed more light at which point, within the observed year-long period, most changes occur. A special case represented Tsukada et al. (2005), first aiming at longitudinality, retesting their participants 1.2 years after their first test, the authors, however, did not find significant effect of time nor did they find any significant interaction of time with other factors and, as a result, averaged the scores over T1 and T2, basically discarding the longitudinal aspect, without offering an explanation as to why the longitudinal aspect did not bare any significant effects (ibid, 273). Contrary to all the other observed studies, the participants of Sun et al. (2021) were retested after only 5 months from their initial test, it is, however, important to note that the researchers focused on considerably different aspects than the previously mentioned ones.

Most studies used analogous design of longitudinality, keeping the time between the initial test and the final test at about 1 year. Future research of longitudinal L2 perception might focus on a longer period of immersion, or, similarly to Kim et al. (2018), test the participants at more points within the immersion to see how they were progressing continually.

4.2.2.3 Stimuli

Use of target stimuli differed among studies, consonantal contrasts were investigated by Aoyama et al. (2004) and Aoyama et al. (2008), in the case of the former, the perceptual part focused on categorical discrimination of the following contrasts: /l/-/r/, /r/-/w/, /s/-/θ/ and /b/-/s/ as a control, the logic behind analyzing those particular contrasts is the fact that they pose difficulty for Japanese L2 speakers of English (Aoyama et al. 2004, 236). As far as the latter study is concerned, it focuses more closely only on the /s/-/θ/ while employing again the /b/-/s/ control contrast. The consonants for both Aoyama et al. (2004) and Aoyama et al. (2008) were taken from naturally produced tokens of adult male speaker of American English producing /Ca/ syllables in a carrier phrase: "Then I saw /Ca/ there" (Aoyama et al. 2004, 237).

Tsukada et al. (2005) and Kim et al. (2018), on the other hand, focused on vowels, the former, according to the results of their experiment 1 mapping problematic vowel contrasts for native Korean speakers, centered their attention on the discrimination of four vocalic contrasts: /i/-/ɪ/, /e$^{\mathrm{I}}$/-/ɛ/, /ɛ/-/æ/, /ɑ/-/ʌ/ in /bVb/ context. With regard to the latter, Kim et al. (2018), the aim was on the cue weighting, investigating spectral vs. durational cues. The authors used the forced-choice identification task, rather than focusing on discrimination, creating a continuum "ranging in vowel duration from 70 to 230 ms for *bit-beat* and from 100 to 260 ms for *bet-bat*," with duration varying by 40 ms for each of the 5 steps of the vowel-spectral continuum, generating 25 tokens for each of the vowel contrasts (Kim et al. 2018, 5).

4.2.2.4 Procedure

This subsection investigates the procedures adopted in the reviewed studies.

Aoyama et al. (2004) and Aoyama et al. (2008) used the same design, i.e. a categorial discrimination testing, during which each of the consonant contrasts was investigated by "eight triadic change and eight triadic no-change trials with an inter-stimulus-interval of 0.5 s," the participants responded by pressing the "1", "2" or "3" or the forth button indicating a no change. The responses were successful if they indicated the correct odd item in case of the change trials and in the latter case, with no change of the category, if they pressed the forth button (Aoyama et al. 2004, 237). In both aforementioned studies, the participants were able to set the volume to a comfortable level before the onset of testing, the volume was non-adjustable thereafter. A session practicing the testing procedure was included as the participants had to respond correctly in at least 9 out of 10 /wa/ and /sa/ practice items. The testing took place in a quiet room with the stimuli being presented over headphones. Very similar design to that of Aoyama et al. (2004) and Aoyama et al. (2008) was employed by Tsukada et al. (2005), based on change and no-change trials motivated by the fact that "establishing a phonetic category will increase sensitivity to differences between […] newly formed category and other L1 and L2 categories, and will also reduce sensitivity to token-to-token variation *within* the newly formed category" (Tsukada et al.

2005, 272). The differences consisted in using vowel contrasts instead of consonant contrasts and also by using a longer, 0.8 s, inter-stimulus interval.

Quite distinctive procedure was used by McCarthy et al. (2014) and Kim et al. (2018). Due to the fact that the former had only child participants, the procedure was adapted to suit the young listeners: it had the form of a computer game presented in a quiet room in a familiar environment of a nursery (in T1) or a school (in T2). A two-alternative forced choice task was employed, the participants identified the word by pointing at the screen, indicating one of the two possibilities with a subsequent reward appearing on the screen after each trial. In order to appeal to children, as already mentioned, the instructions were delivered in form of a game saying that '"Panda is learning to say new words, and because you already know these words, you're the best person to help him. Listen to Panda and point to what he says"' (McCarthy et al. 2014, 1970). A familiarization block with a feedback after each trial was also used prior to the actual test. Unlike other studies, an adaptive procedure was used presenting the stimuli in a way that reflected how each of the listeners identified each stimulus. According to the authors, in this way the study made an efficient use of a small number of presentations concentrating on the most crucial regions: phoneme boundary and slope of the function (ibid, 1970). The latter study, Kim et al. (2018) focused on cue weighting stimuli employing, similarly to McCarthy et al. (2014), a two-alternative forced choice identification task, the response words appeared in form of pictures and the participants made their choice by pressing either left (←) or right arrow key (→), pictures were used so as to avert a possible orthographic bias. The participants could only listen to each trial once, nevertheless, they determined how fast the task progressed, in other words, it was self-paced without a time limit, contrary to e.g. Aoyama et al. (2008) and Tsukada et al. (2005), who employed a 1s interval between a response and the subsequent trial. Prior to testing, in order to avoid misinterpretation of the pictures, the participants were asked to provide words for each picture making sure each of them understands what each picture represents. At T1, the authors also employed a familiarization practice session based on naturally produced *sheep* and *ship* tokens amounting up to 10 trials, ensuring the participants understood how to proceed during the test.

4.3 Intensive L2 Perceptual Training – a Semi-Longitudinal Design

Perceptual training features a similar design to that of the longitudinal one, mentioned in section 4.2 with the difference of a shorter time span and the fact that it does not employ an immersion setting, with an exception of Nielsen et al. (2015) and Casillas (2020). The former used participants that took part in an army training program, an "intensive language learning [that] in this respect entails language learning as a full-time job" (Nielsen et al. 2015, 6). Spanning over 19 months, it could possibly be classified as longitudinal study, same as in the previous section (4.2), however, the design resembled that of a university course, which made us categorize it as semi-longitudinal. In the latter study (Casillas 2020), it was not the length of the training (7 weeks) by which the study differed from others, but the way it was devised. More specifically, it constituted an immersion but in a domestic US university campus environment, as the students signed a pledge to use only the target language for the duration of the training (Casillas 2020, 14). Such immersion was possible due to the fact that the target language was Spanish and the number of native Spanish speakers in the USA.

Early study of Strange and Dittmann (1984) has proven that L2 speech perception training may have beneficial effects. The authors used a same-different discrimination task and focused on synthesized continuum between the word-initial consonants /r/-/l/ in a training task with immediate feedback. Improvement was found in the identification and discrimination of perceived synthesized stimuli, as well as generalization to a novel minimal pair. Nevertheless, no improvement was found in the natural word-initial /r/-/l/ contrasts, that is to say, no generalization from the trained synthetic contrasts to natural contrasts was found (Strange and Dittmann 1984, 131).

Training settings differ as to their variability, we talk either about low-variability e.g. Strange and Dittmann (1984), or high-variability training, e.g. Bradlow et al. (1997), Lengeris and Hazan (2010), Giannakopoulou et al. (2013), Shinohara and Iverson (2015), depending on the number of speakers and phonetic contexts, it will be presented more closely in section 4.3.1. Some studies also used both low and high variability training, focusing on which one is more effective, e.g. Perrachione et al. (2011), Giannakopoulou et al. (2017).

The source of inter-study variation include the choice of the participants' native language, of which the most common include Greek, e.g. Lengeris and Hazan (2010), Giannakopoulou et al. (2013), Gianakopoulou et al. (2017); Spanish, e.g. Iverson and Evans (2009), Kondaurova and Francis (2010), Wanrooij et al. (2013), Escudero and Williams (2014), Cooke and Lecumberri (2018); Japanese e.g. Strange and Dittmann (1984), Bradlow et al. (1997), Shinohara and Iverson (2013), Shinohara and Iverson (2015); but also less common languages such as Danish (Nielsen et al. (2015)); English (Casillas (2020)); German (Iverson and Evans (2009)); Salento Italian (Sisinni and Grimaldi (2011)) and Finnish (Ylinen et al. (2010)). With the exception of Dutch (Escudero and Williams (2014)); Arabic and Dari (Nielsen et al. (2015)) and Spanish (Casillas (2020)); English predominates as the target language, e.g. in studies by Strange and Dittmann (1984), Bradlow et al. (1997), Kondaurova and Francis (2010), Lengeris and Hazan (2010), Ylinen et al. (2010).

A training using identification, e.g. Lengeris and Hazan (2010), Shinohara and Iverson (2015), Cooke and Lecumberri (2018); discrimination, e.g. Strange and Dittmann (1984), Escudero and Williams (2014), Shinohara and Iverson (2015), Giannakopoulou et al. (2017) and word-learning (Giannakopoulou et al. (2017)) was employed.

Some studies focused on perceptual cue learning in consonants, e.g. Francis et al. (2008), as well as vowels, e.g. Kondaurova and Francis (2010), Ylinen et al. (2010), Giannakopoulou et al. (2013).

Few studies also employed noise in their training, multi-talker babble (without specified number of talkers) in case of Lengeris and Hazan (2010), 8-talker babble and speech-shaped noise in case of Cooke and Lecumberri (2018), those are perhaps the most important to our thesis, as they investigate noise in a study of a longer time span.

The studies usually include young adults as participants, e.g. Bradlow et al. (1997), Kondaurova and Francis (2010), Sisinni and Grimaldi (2011), Cooke and Lecumberri (2018), scarcely they also include older participants, as in the case of Wanrooij et al. (2013) and Escudero and Williams (2014). The former exhibits age range from 19 to 60, the latter from 24 to 63. Few studies try to investigate the effect of training cross-sectionally, comparing children and adults, e.g. Giannakopoulou et al. (2013) and Giannakopoulou et al. (2017), other studies focus solely on children, such as Shinohara and Iverson (2013), focusing on children of 6-8 years of age, and Shinohara and Iverson (2015), aiming at children of 6-12 years age.

4.3.1 Low vs. High Variability Training

Perceptual training involves low and high variability training, as was already mentioned in 4.3, which, in case of the former, is explained as using a single speaker and single phonetic context during the training phase, while in the latter case it involves multiple speakers speaking in multiple contexts.

Strange and Dittmann (1984, 133) is a perfect example of a low-variability training (LVT), as we can talk about a single speaker (or in this case synthesized speech created from the recordings of a single native talker) and a single phonetic context, end points made up by "rock" and "lock" series of stimuli. Nevertheless since their pioneer work, high variability training (HVT) was developed using stimuli from varied speakers and also varied phonetic contexts which, according to Giannakopoulou et al. (2013), turned out to be critical for such training to have success in improving the perception of L2 speakers. For this reason HVT tends to be considered by the majority of research, e.g. Bradlow et al. (1997), Francis et al. (2008), Lengeris and Hazan (2010), Ylinen et al (2010), as more beneficial to listeners compared to LVT. Some studies, nevertheless, challenge such view, e.g. Perrachione et al. (2011) and Giannakopoulou et al. (2017).

4.3.1.1 High Variability Training and Its Benefits

After the results of Strange and Dittmann (1984) were revealed, researchers tried to find different ways to train non-native listeners that would also help them generalize the gained knowledge to new words and novel talkers not present in the training.

Lively et al. (1993) employed HVT, as well as LVT, and compared how they affected his listeners. It comprised of similar participants as those in Strange and Dittmann (1984) that is Japanese learners of English. Moreover, they also chose the same consonantal contrast, specifically /r/-/l/, which is known to cause difficulty to Japanese learners of English. Unlike the former study, the authors trained the listeners on different stimuli in more phonetic contexts using multiple talkers and focusing on consonant identification, rather than discrimination. The results of the two experiments in Lively et al. (1993) showed that, even though both listeners in HVT and LVT improved in the post-test relative to pre-test, only the listeners who underwent HVT also "generalized to new words produced by a familiar talker and novel words produced by an unfamiliar talker" (ibid, 1242). Logan et al. (1991), focusing on the same consonantal contrast as the aforementioned studies, used participants from the same L1 background. In the training phase, they employed multiple natural examples of words in different phonetic contexts spoken by six different talkers, finding "the new procedure to be more robust than earlier training techniques [... and obtaining] reliable differences in performance" when comparing the post-test with the pre-test (ibid, 874).

The previous studies focus only on consonant contrasts, nevertheless, studies also found that HVT improved cue weighting in vowel perception, approximating the perception to that of the native listeners, e.g. Kondaurova and Francis (2010), Ylinen et al. (2010) and Giannakopoulou et al. (2013). Kondaurova and Francis (2010) focused on the cue weighting of American English /i/-/ɪ/ by native Spanish listeners. The training phase included cue enhancement, cue inhibition and natural distribution. The English lax and tense vowel contrast used by Kondaurova and Francis (2010) is distinguished by two acoustic dimensions, those are the spectrum (specifically the first three formants) and the duration, the way native and non-native speakers of English weight those cues, however, differs, as native English speakers use spectrum as the primary cue with duration playing only a secondary role (ibid, 570). Native Spanish speakers, on the other hand, almost exclusively, make use of duration, which is one of the reasons, why it prevents them from reaching native-like perception (ibid, 570). Native Spanish listeners received one of the three training methods mentioned above (cue enhancement, cue inhibition and natural distribution) and the results showed that reliance on spectrum increased in all three training conditions, nevertheless, cue enhancement and especially cue inhibition training were more efficient than natural distribution, suggesting that cue-specific training seems to be "more effective for the acquisition of second-language speech contrasts" (ibid, 569). Similar to Kondaurova and Francis (2010), Giannakopoulou et al. (2013) also used the English contrast /i/-/ɪ/ trying to find out whether perceptual training would improve the perception of native Greek adults and children. The authors employed both natural and modified stimuli, the latter consisting of tokens with modified vowel duration to prevent the participants from making use of the durational cue. HVT was found to improve the performance of both native Greek adults and children, even though for the latter group the improvement was more pronounced, as we can see in Figure 9 below (ibid, 201).

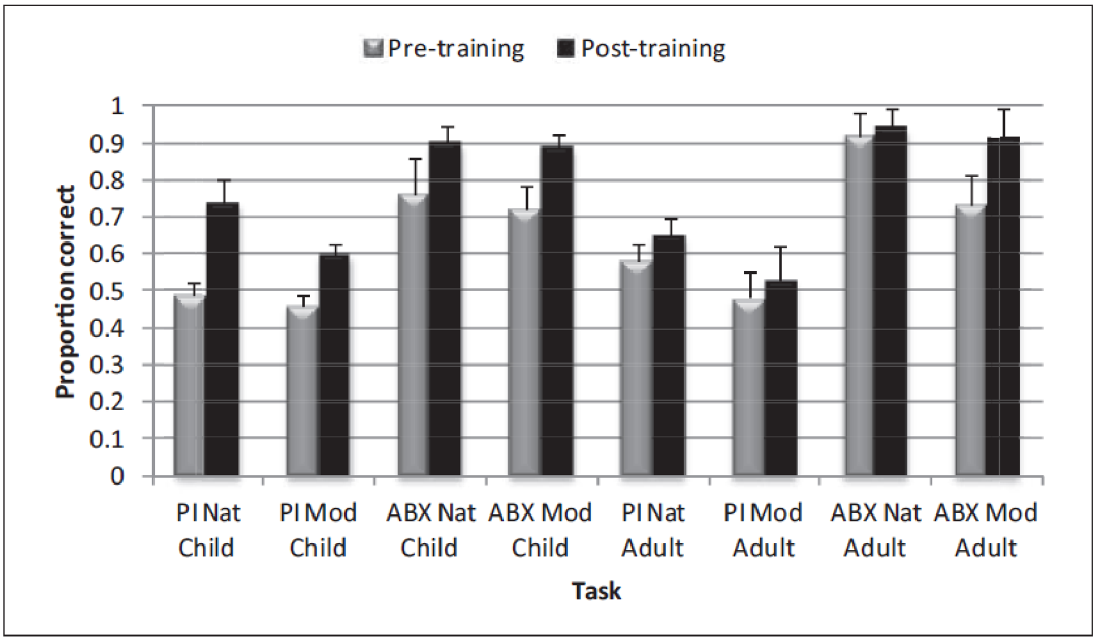


Figure 9 shows how the training of native Greek adults and children affected their performance in perceptual identification (PI) and auditory discrimination (ABX) tasks for both natural (Nat) and modified (Mod) condition (adapted from Giannakopoulou et al. 2013, 209).

Additionally, HVT in L2 perception was also found to have an effect on L2 production as judged by native speakers in Bradlow et al. (1997). The authors investigated how /r/-/l/ perceptual identification training affected /r/-/l/ production. Native Japanese speakers were recorded pre- and post-HVT, producing English words that contrasted /r/ and /l/. The results showed significant perceptual learning as a result of the training program, which also generalized to novel items (ibid, 2299). Using native English listeners as judges, two separate tests were carried out to determine the improvement in the production. The first directly compared pre- and post-training recordings and showed significant improvement of perceived rating for the latter ones, while the second employed two-alternative minimal pair identification and the post-training tokens were identified more accurately, the results thus suggest the transfer of L2 perceptual learning of /r/ and /l/ to the production domain (ibid, 2299).

4.3.1.2 High Variability Training Challenged

Even though HVT received the most attention as a training method, some studies put it under criticism, e.g. Perrachione et al. (2011) and Giannakopoulou et al. (2017).

Perrachione et al. (2011) tried to discover how pre-training abilities, such as the ability to learn foreign languages, are influenced by various training patterns. Their participants learned to recognize words of simulated foreign language, having to "learn to use a novel phonological contrast called 'lexical tone' to distinguish the words" which involves pitch contours not present in the native English language of the participants (ibid, 462). Their criticism was aimed at the fact that HVT was seen as being beneficial in any circumstances, the authors, however, found that only some listeners benefited from HVT, more specifically, those individuals who had strong perceptual abilities exhibited enhanced word recognition (ibid, 461-462). In addition, HVT was found to disrupt the perceptual abilities of those with weak perceptual abilities, for whom LVT was more beneficial (ibid, 461-462). The authors thus stressed the need to consider the pre-training perceptual abilities of the listeners to match them with the appropriate type of training.

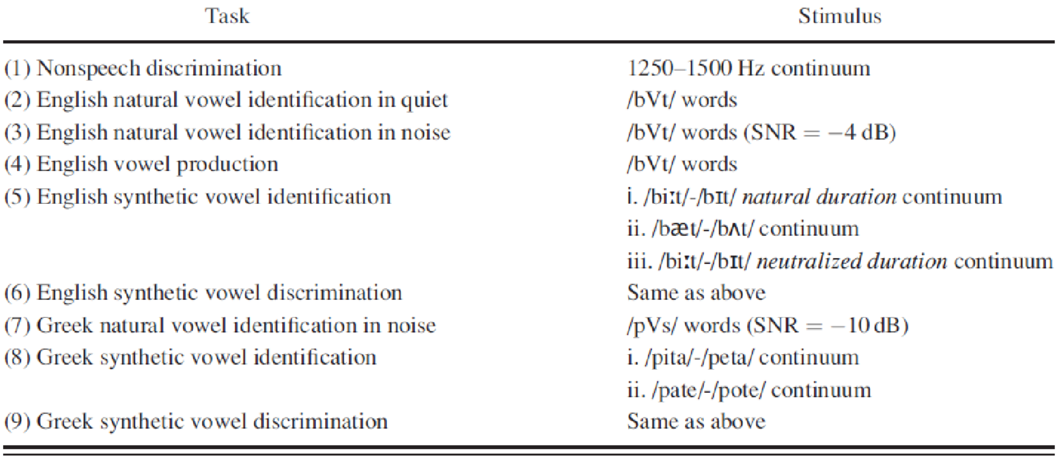
The primary aim of Giannakopoulou et al. (2017) was to investigate how HVT and LVT affect children's phonetic discrimination and word-learning. Moreover, the authors expected that, compared to adults, children show more brain plasticity which means they should also show greater improvement from pre- to post-training testing, the authors thus, reviewing the literature, expected children to perform better than adults in both discrimination and word-learning. Furthermore, as HVT was expected to trigger more learning, the children in the HVT group were to exhibit better results, as well as more generalization to novel talkers and novel words in the post-training test. The authors found improvement for both groups, but unexpectedly, after being trained with a single talker (LVT), this LVT benefit was greater for the child group and extended even for generalization of novel items (ibid, 30). Comparing the HVT of adults and children, the former group showed advantage that did not extend to novel items (ibid, 30). The condition of training variability did not affect lexical learning, nor did brain plasticity of children prove more phonetic learning for this group (ibid, 30). Giannakopoulou et al. (2017), together with Perrachione et al. (2011) thus provide an interesting contra-evidence against using HVT as the established, more beneficial training method for all the participants and in the research to come, it could be favorable to also include LVT and screen the participants carefully as to their pre-training perceptual abilities matching the right training method to the right participant, in order to gain maximum benefit from the perceptual training. As far as children and perceptual training are concerned, very little research was carried out and as such there is a gap that needs to be filled.

4.3.2 Perceptual Training in a Noisy Condition

Studies making use of noise in their training design are very scarce, out of those we can note Lengeris and Hazan (2010) and Cooke and Lecumberri (2018). Moreover some very recent works were published by Cooke and Lecumberri (2019), Mi et al. (2021) and Gong et al. (2021) focusing on perceptual training in noise, the studies are, however, inaccessible to the author of this essay and as such won't be discussed in this subsection.

4.3.2.1 Lengeris and Hazan (2010)

Five sessions of HVT, trying to find out, if successful L2 vowel learning is related to native language vowel processing (as per L1-phonetic hypothesis) or to their general frequency discrimination acuity (as per auditory-processing hypothesis). A group of native Greek (NG) speakers received training, other completed pre/post tests without it. Different aspects of L2 and L1 vowel processing and frequency acuity was accessed through variety of tests, assessing 1) natural English (L2) vowel identification both in quiet and in mt-babble, 2) natural Greek (L1) vowel identification in mt-babble, 3) categorization of synthetic English vowel continua, 4) categorization of synthetic Greek vowel continua, 5) discrimination of synthetic English vowel continua and 6) discrimination of synthetic Greek vowel continua. The tasks completed in the pre/post testing are presented in Table 2. As we can see, the authors also investigated English vowel production.



Task	Stimulus
(1) Nonspeech discrimination	1250–1500 Hz continuum
(2) English natural vowel identification in quiet	/bVt/ words
(3) English natural vowel identification in noise	/bVt/ words (SNR = −4 dB)
(4) English vowel production	/bVt/ words
(5) English synthetic vowel identification	i. /biːt/-/bɪt/ *natural duration* continuum ii. /bæt/-/bʌt/ continuum iii. /biːt/-/bɪt/ *neutralized duration* continuum
(6) English synthetic vowel discrimination	Same as above
(7) Greek natural vowel identification in noise	/pVs/ words (SNR = −10 dB)
(8) Greek synthetic vowel identification	i. /pita/-/peta/ continuum ii. /pate/-/pote/ continuum
(9) Greek synthetic vowel discrimination	Same as above

Table 2 represents the tasks in the order of their presentation completed by the Greek participants, whether those participating in the research or the controls (adapted from Lengeris and Hazan 2010, 3760).

A total of 28 NG participated in the study, 18 took part in the vowel training, 10 served as control, being tested by the same pre/post test without receiving any training. Even though the participants had received in the past 10 to 12 years of formal English instruction, they had very little to none native input interaction and did not spent more than one month in an English-speaking country. Nevertheless, their proficiency was moderately high, having obtained FCE to CAE language certificates.

Stimuli used in the natural vowel task both in quiet and in mt-babble included Greek vowels /i, e, a, o, u/ in /pVs/ words, while the English vowels comprised of /i, ɪ, e, ɜ, æ, ʌ, ɑ, ɒ, ɔ, u/ in /bVt/ words. The stimuli were recorded by two NG speakers, one male and one female, in the former and by two native British English (BE) speakers, also a male and a female, in the latter case with another BE speaker being recorded for the generalization test.

In order to assess vocalic categorization in L2 and how well the participants categorize L1 vowels, the authors also included the tasks of identification and discrimination of English and Greek synthetic vowel continua. The range of Greek vowel continua comprised of /i/-/e/ and /a/-/o/, while the English utilized /i/-/ɪ/ and /æ/-/ʌ/, which, according to the authors, “cover similar areas in the acoustic/perceptual space across languages,” offering a possible comparison between the two languages (ibid, 3759). Previous study made by one of the authors of Lengeris and Hazan (2010) discovered that the /i/-/ɪ/ was a more difficult contrast for Greek speakers of English than /æ/-/ʌ/ and thus they expected similar patterns also for the synthetic vowel continua. Parallel to the natural vowels, the synthetic continua had to be placed in a different carrier word, two phonetic contexts had to be used, due to the fact that Greek does not contrast all used vowels in a single context, Greek /i/-/e/ was placed in /pVta/ word, /a/-/o/ in /pVte/, both English continua /i/-/ɪ/, /æ/-/ʌ/ were placed in /bVt/ word.

As the authors also aimed at discovering whether L2 vowel learning is related to L1 vowel processing or the participants' general frequency discrimination acuity, they made use of non-speech continuum consisting of a single formant of a varied 1250 and 1500 Hz frequency, parallel to natural vowel F2. Moreover, the continuum had a constant F0 of 120 Hz, resembling fundamental frequency of a male speaker.

The perceptual training phase included 14 English vowels (10 monophthong vowels and 4 diphthongs) arranged in 4 minimal-pair groups: 1) /i, ɪ, aɪ, eɪ/; 2) /u, aʊ, ɜ/; 3) /ɒ, əʊ, ɔ/ and 4) /e, ɑ, æ, ʌ/ with the first three groups being expected to cause the Greek speakers problems, 10 words with minimal pairs for each vowel were used, a total of 140 tokens. The participants partook in five training sessions consisting of vowel identification with provided feedback on their choices. Each of the sessions stretched for 45 minutes with 225 tokens presented, of which "70 tokens were five random repetitions of the 14 English vowels, the next 85 were based on the participants' errors" and the last 70 were again random repetitions (ibid, 3760).

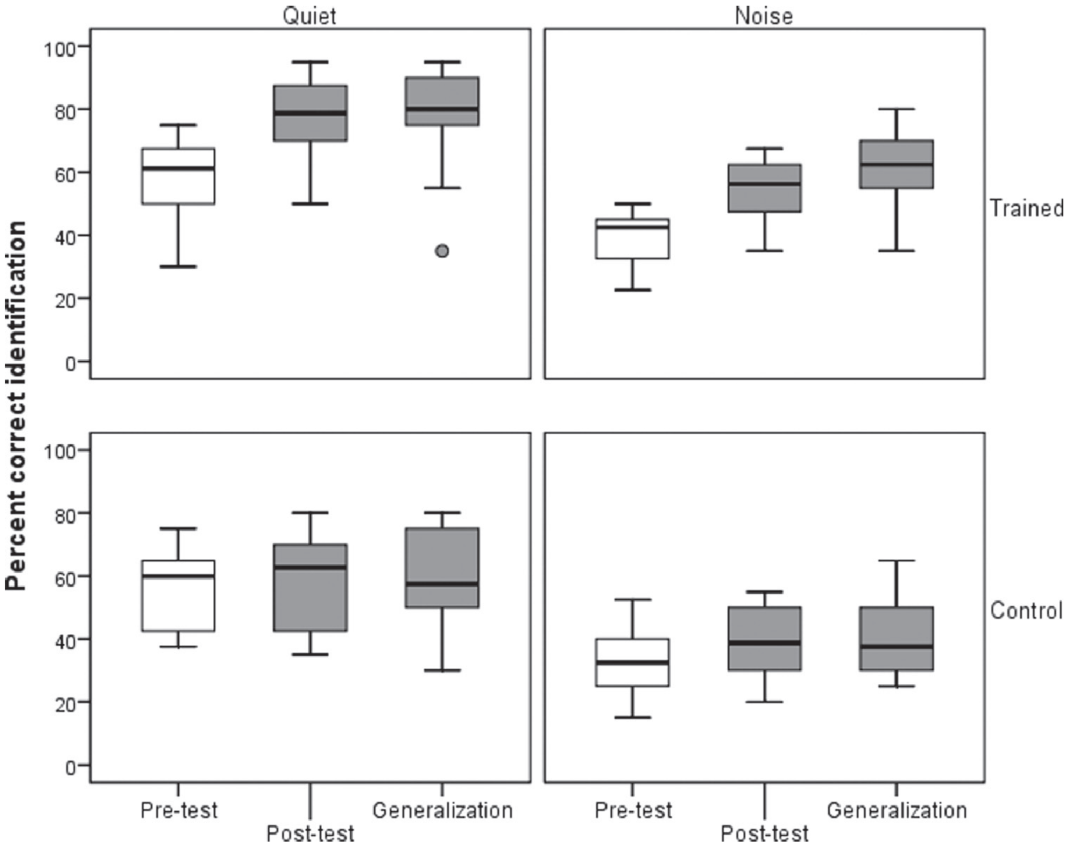


Figure 10 shows English vowel identification accuracy together with generalization to novel stimuli in normal and noisy condition for trained NG participants and for the NG controls (adapted from Lengeris and Hazan 2010, 3762).

Lengeris and Hazan (2010, 3761-3767) investigated how L2 vowel learning varies among learners and whether it is more related to their L1 vowel processing or their general frequency discrimination. Similar to e.g. Logan et al. (1991) and Lively et al. (1993), the results provided evidence for perceptual HVT which significantly improved L2 English vowel identification as visible in the Figure 10. Additionally, the results also replicated the findings of Bradlow et al. (1997), specifically the L2 perceptual training significantly improved the L2 production, as we can clearly see in Figure 11, if we compare the trained participants and the controls.

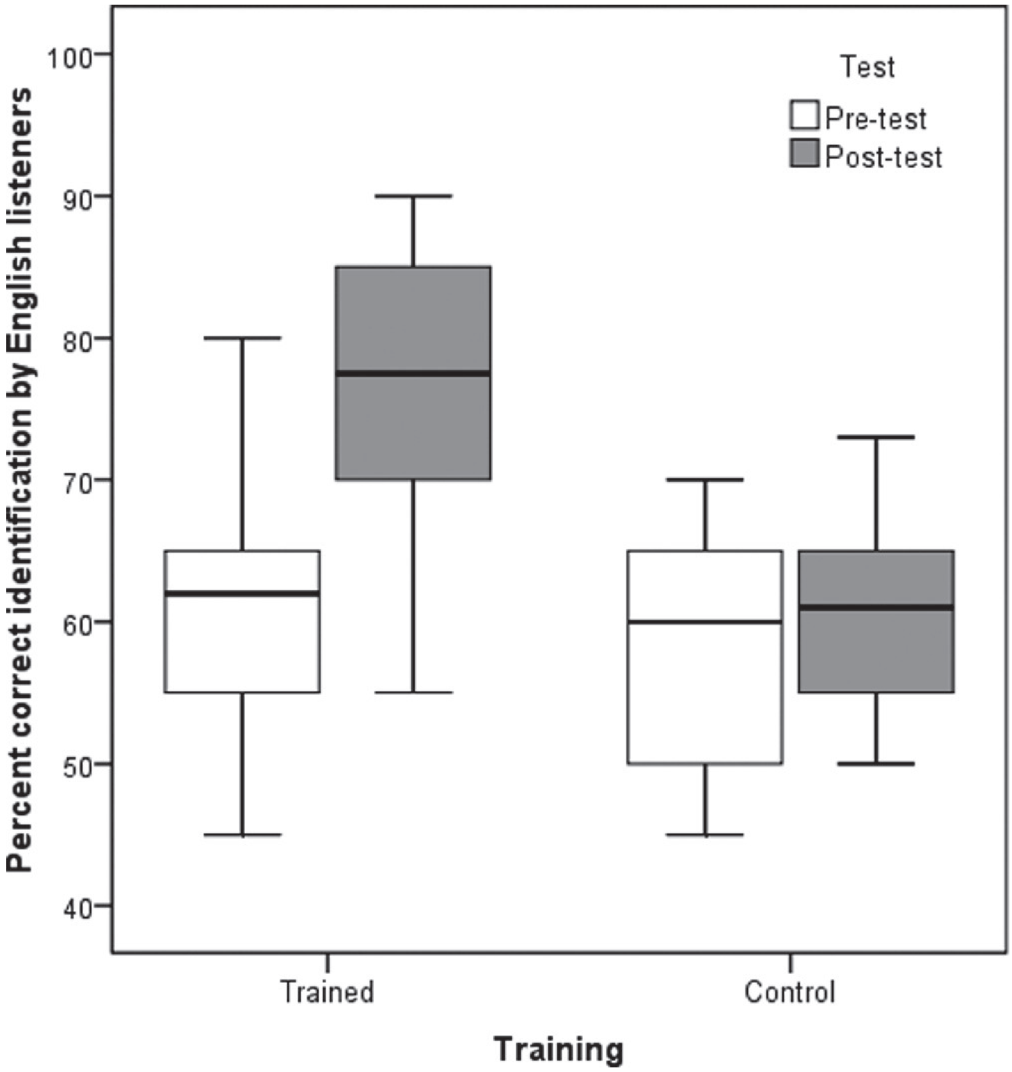


Figure 11 presents the percentage of correctly identified English vowels which were produced by trained NG speakers and NG controls in the pre/post test and judged by native English listeners (adapted from Lengeris and Hazan 2010, 3764).

In real-life conditions speech perception is hampered by noise and, as such, the novelty implemented in this study consists in assessing how phonetic training affects L2 perception in noise. As was expected from the results of other studies which investigated the effect of noise, e.g. Mayo et al. (1997) and Rogers et al. (2004), both pre/post tests showed significantly poorer performance for L2 vowel perception in noise, compared with the quiet condition and even though the perceptual training of Lengeris and Hazan (2010) did not include noise, the pre/post test battery did. Importantly, although no noise condition was present in the perceptual training, "training in quiet significantly improved perception in noise by about 15% points," as can be seen in the Figure 10 when comparing the trained and control cohorts (ibid, 3766).

The aforementioned evidence provided by Lengeris and Hazan (2010) is important for the purposes of the present thesis, notwithstanding, the main goal of the study was to discover, whether training was related more to the ability to discriminate frequencies generally or to native vowel processing. The authors found that "individuals with better frequency discrimination acuity for synthetic vowels in L1 and L2 and nonspeech stimuli were better at identifying natural L2 vowels both before and after training" (ibid, 3766). The results thus support the auditory-processing hypothesis playing the crucial role in L2 vowel learning.

Even though it was not the primary aim of the study, the authors provide evidence to the studies concerning noise, especially the novel finding that training in quiet increases the perception in noise which makes sense considering how L1 speakers perform better in noise, as was mentioned earlier. More experience with the L2 seems to be crucial and as intensive training provides a condensed form of experience, it seems to positively affect L2 speakers in that they improve their L2 perception in noise.

4.3.2.2 Cooke and Lecumberri (2018)

The study was based on the premise that people acquire their native language perfectly even if they face adverse listening conditions, such as background noise. Basing their investigation on the L1 learning, the authors are asking "whether the presence of masking noise during consonant training is a barrier to improvement," or whether noise can prove beneficial for L2 listeners (Cooke and Lecumberri 2018, 2602).

There were four homogenous groups of L1 Spanish learners of L2 English, each group being exposed to different kind of training but a common pre- and post-test involving forced-choice consonant identification in three conditions: quiet, speech-shaped noise and multi-talker babble, the final condition constituting an untrained masker, investigating whether training in specific noise offers benefits when listening to a different kind of noise, that is to say, whether generalization to novel noise can be observed. The training spanned for ten sessions with forced-choice identification task, two of the four groups underwent consonant identification training in VCV tokens, the other two served as controls, undergoing vowel identification training in CVC tokens. Both the vowel and consonant groups were further divided into subgroups trained either in quiet or in speech-shaped noise, thus forming a total of four separate groups.

The participants constituted 88 L1 Spanish learners of English in their second-year study of the English Philology program at the University of Basque Country, which suggests moderate to high command of English possibly with native English input. The participants were assigned to one of the four groups randomly with 22 of them being assigned pseudo-randomly "following a group score balancing procedure [... so] that the four group mean scores were within one percentage point of each other" in the pre-test conditions (ibid, 2604). One participant in the vowel-quiet training group did not complete the training sessions and another, also from the same group, exhibited a 25% drop in one of the test conditions, comparing post- vs. pre-test data, both were, therefore, excluded from the study. Along with taking part in the training tasks, the participants were also learning to transcribe and analyze English phonemes using IPA symbols as they were obligated to participate in the course of English phonetics as part of their studies.

The stimuli for both the training and the pre/post test were taken from the Consonant Challenge Corpus, the authors selected part of the corpus which consisted of non-sense VCV tokens spoken by 12 male and 12 female speakers and containing all 24 British English consonants (/p, b, t, d, k, g, tʃ, dʒ, f, v, h, ð, s, z, ʃ, ʒ, h, m, n, ŋ, l, r, j, w/) in nine vowel combinations (/i, u, æ/) in front, as well as end stress. Different talkers were used in the pre/post test and the training, the former was comprised of four male and four female talkers, while the remaining 8 male and 8 female speakers were used in the latter. Duration of the tokens ranging from 290 to 1002 ms with mean of 602 ms. The tokens for the control vowel groups consisted of monosyllabic CVC tokens containing all 11 English vowels /i, ɪ, e, æ, ʌ, ɑ, ɒ, ɔ, ɜ, ʊ, u/ spoken by 7 British English speakers.

The training included 10 sessions each made up of five blocks of equal length spread over a five-week period. Unlike the pre/post test, the procedure also included feedback on the wrong responses and, as a result, the participants listened to the token once more. During the training, the listeners of the consonant-quiet (CONS-Q) and the consonant-noise (CONS-N) groups "identified four VCV tokens for each of the 24 English consonants in each block, i.e., 20 exemplars per consonant per session" (ibid, 2604). The authors used speech-shaped noise (SSN) in CONS-N, the listeners had tokens presented in each of the training block with a different SNR (+2, 0, -2, -4, -6 dB). The reason for training the listeners in multiple SNRs was to provide diverse noise conditions similar to those in the real life. As was already mentioned, the listeners had 10 sessions of training which means they responded to 4800 tokens, 200 per each of the 24 consonants. As for the vowel-quite (VOW-Q) and vowel-noise (VOW-N) control groups, the participants were also presented with five blocks, each spoken by a single talker. Moreover, feedback was also present as in the consonant groups. Nevertheless, the VOW-N group differed in that the participants listened to SSN in only one SNR of -6 dB, matching the one used in the pre/post tests.

During the pre/post tests, the participants completed three computer-based tasks, in each of them they were forced to identify the stimuli by choosing one of the 24 consonantal alternatives, they selected their choice by clicking on an IPA symbol on an onscreen keyboard. The first task consisted of consonant identification in VCV tokens in quiet. In the second one, they identified stimuli in SSN with SNR of -6 dB, which was also the most difficult noisy condition in the training phase. The third task consisted in identifying stimuli in a novel 8-talker babble noise with SNR of -2, not present in the training phase. Each of the three test tasks were comprised of 16 examples of each of the 24 consonants, that is to say, the participants were exposed to every talker out of the eight twice, once for the front-stressed and once for the end-stressed words, making up 384 tokens per each training block, a total of 1152 tokens per each pre/post test. It is also important to note that for the two test blocks containing maskers, the participants went through a practice session, 16 stimuli with tokens in noise were heard before listening to the test tokens.

Unlike the previously mentioned study (Lengeris and Hazan (2010)), the one by Cooke and Lecumberri (2018) provides a more detailed account of the noise. Firstly, it lists the number of talkers in the novel babble, namely eight are present in this noise. According to the authors the tokens with noise were "generated by mixing speech with randomly chosen masker fragment of 1.2 s duration [... and] was scaled [...] in the region containing the speech signal, i.e., discounting the leading and lagging noise-only section of the waveform" (ibid, 2604).

Cooke and Lecumberri (2018) considered the absence of noise in second language acquisition and aimed to establish whether or not noise functions as an obstruction to non-native language acquisition, taking into consideration the fact that, due to reduced accessibility of speech cues, noise might have a detrimental effect. Nevertheless, they hypothesized that perceptual training with exposure to tokens with background noise would lead to the use of more robust perceptual cues, more native-like cue weighting and, as such, would be a useful strategy for non-native acquisition of consonants.

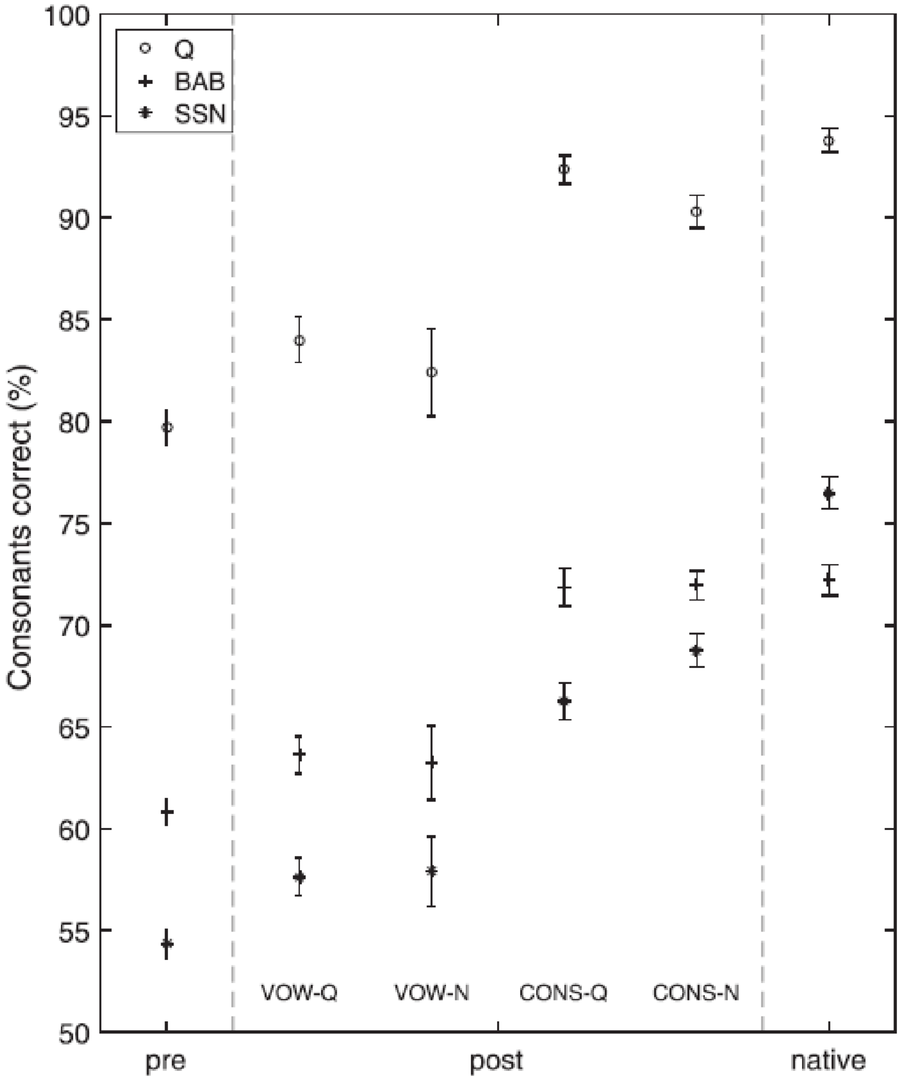


Figure 12 shows consonant identification rate for the four groups of participants in the three conditions (quiet, with SSN, with 8-talker babble). The left-most column represents mean scores from pre-tests from across all the four groups, the right-most productions of native speakers and the middle columns represent the post-tests (adapted from Cooke and Lecumberri 2018, 2605).

As we can see from Figure 12, all participants showed improvement in consonant identification if we compare the pre-tests and the post-tests, the consonant-trained groups, however, showed far more improvement compared to the ones who were trained on vowels, closing in on the identification rates of the native speakers. Even though all participants showed improvement, the authors argued that the improvements for the vowel-trained groups from “pre- to post-test are quite similar to the rapid gains observed between the pre-test and the first training session for the consonant-trained groups” and as such suggests an in-task accommodation, a kind of procedural learning, rather than real improvement, illustrated by 2-4% improvement vs. 10-14% for the consonant group (ibid, 2608). We can see the improvement of the consonant-trained groups in Figure 13. As we can see, the most defining were the first six training sessions after which the participants of both groups exhibited either very little or no progress in the consonant identification rate.

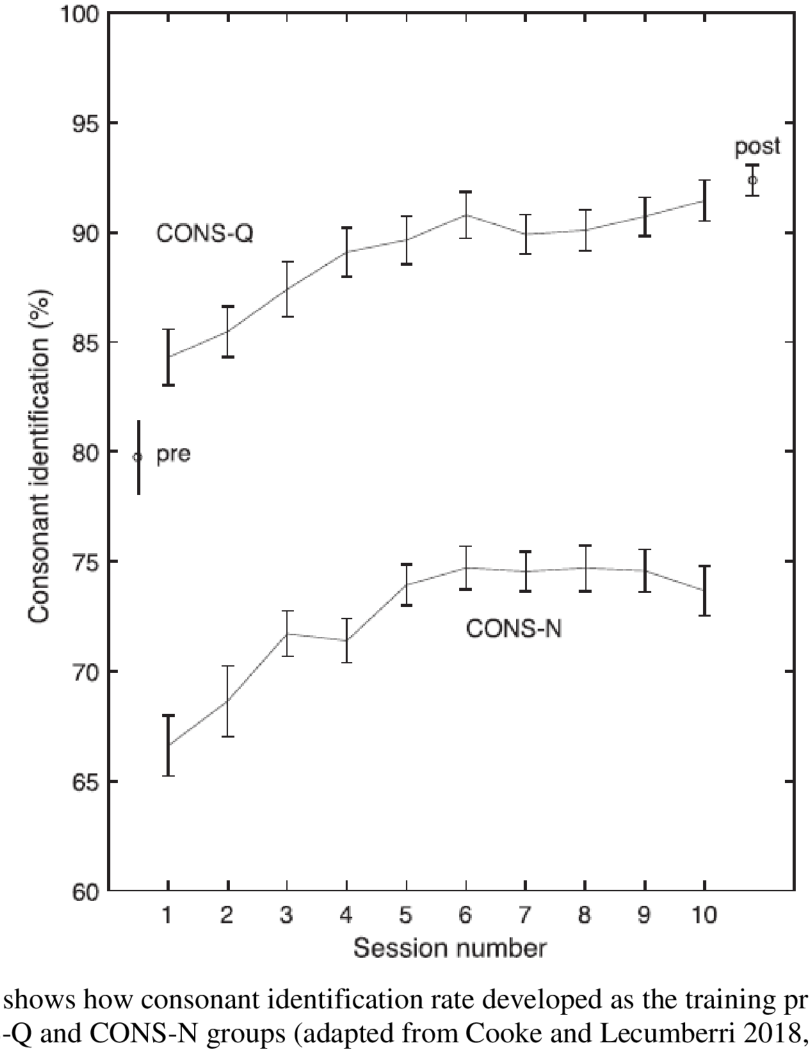


Figure 13 shows how consonant identification rate developed as the training progressed for CONS-Q and CONS-N groups (adapted from Cooke and Lecumberri 2018, 2606).

The study also found a small benefit of 2-3 % for participants in the consonant groups, when the training condition was matched to the post-test task, that is the CONS-Q exhibited higher identification rates when being tested on the consonants in quiet compared to the CONS-N and the CONS-N received higher scores when identifying consonants in SSN than their CONS-Q counterparts.

As was mentioned previously, the authors aimed predominantly on discovering whether the exposure to tokens with noise would lead the non-native speakers of English to acquire non-native consonants with more ease. Moreover, the formation of more robust perceptual cues and more native-like cue weighting habits were investigated. According to the authors, if such course were to occur, the participants exposed to the SSN would have to show larger gains in the task involving the novel 8-talker babble. No such gains for the SSN-trained group were registered in the babble condition, in fact, they were "almost identical to those from the group trained on consonants in quiet" (ibid, 2609). Strategy-wise, the study equated the training in noise with the one in quiet, even though both reached their success rates on different grounds. In the former, masking caused information loss and as a result, the CONS-N group received "incomplete spectro-temporal data," they, nevertheless, compensated for it by learning which information was relevant when they listened to tokens with background noise (ibid, 2609). In the latter, the CONS-Q group, received complete spectro-temporal data, but were less able to determine the relevant information when listening to tokens in noise, compared to the CONS-N group, thus Cooke and Lecumberri (2018) seem to provide evidence that training in noise is neither beneficial, nor detrimental for the acquisition of non-native consonants, as the CONS-N reached similar success rates in quiet as well as in novel babble noise compared to the CONS-Q.

4.3.2.3 Summary

As was already mentioned, the studies that include training with background noise are rather scarce, the presented studies, Lengeris and Hazan (2010) and Cooke and Lecumberri (2018), are very different from each other, the main differences being that the former focuses on vowel perception of native Greek learners of English, involves five training sessions and does not include noise in the training sessions but only in the pre/post test battery, it is thus an investigation of whether L2 speakers of English improve their perception in noise by being trained in quiet.

Noise is not the main aim of their study, their interest lies in finding out how the participants' L2 is affected by their L1 and whether the effect of L1 on their L2 learning is more significant than their general frequency acuity. The authors found out that even training in quiet significantly increased the L2 perception in noise by about 15 %. Cooke and Lecumberri (2018), on the other hand, are more noise-centred. The study, involving 10 training sessions, focuses on consonant perception of native Spanish learners of English including vowel-trained group as control, aiming to discover benefits of the speech-in-noise L2 training and potential formation of more robust cues or more native-like cue weighting. We would witness such fact if the noise-trained group exhibited higher identification rates when perceiving consonants in a novel babble noise compared to the group trained in quiet. However, the results did not support such hypothesis, as both consonant groups exhibited similar gains in the babble noise condition. The authors speculated that the noise trained group might have formed more robust cues and more native-like consonant cue-weighting but the gains coming from such benefit were cancelled by the fact that the noise present in their training provided incomplete spectro-temporal data, compared to the participants exposed to consonants in quiet.

5. FUTURE RESEARCH

The following paragraphs will be devoted to steer possible future perceptual research in background noise in the right direction with the focus on native Czech speakers of English, making use of the sources investigated in the literature review section. Most of the evidence come from perceptual studies in noise which are neither longitudinal nor do they employ a training program, which means they are basically carried out by a single perceptual test, comparing monolinguals with early and/or late bilingual speakers. Nevertheless, there is a fair amount of perceptual studies employing training paradigm, most of those, however, do not employ noise, with the exception of Lengeris and Hazan (2010) and Cooke and Lecumberri (2018). There are also several studies with a fully developed longitudinal immersion program, typically using immigrants in an English speaking country, such as the USA, Canada or the UK, such studies do not consider noise in their research design and focus solely on the improvement of their participants and/or the cross-sectional comparison between the improvement of adults and that of children. Due to the lack of studies which would best serve our purpose, the evidence is obtained from three different areas of research which are connected but are not without their methodological differences, more specifically, non-longitudinal, non-training perceptual studies in noise; longitudinal immersion perceptual studies which do not employ noise and studies with a training paradigm of which only two studies focus on noise.

The original intention was to perform a perceptual experiment on Czech students intending to travel and live abroad for the duration of their Erasmus study program in an English-speaking country, testing them on non-native English perception in noise before leaving the country with the intention to gain pre-immersion data, serving as the pre-tests similar to those present in the aforementioned research. The study abroad program should have served as kind of a naturalistic perceptual training for the participants, during which they were expected to immerse themselves in the foreign linguistic environment, communicating predominantly in their non-native language. After one semester, we planned to assess how the immersion influenced the perception of the non-native speakers of English in noise, providing the post-test data. Nevertheless, due to COVID-19 restrictions, we chose to abandon the longitudinal research and focus more closely on the literature providing evidence on the topic we aimed to investigate.

Considering non-native perceptual research, it seems necessary to assess the level of participant's L2 proficiency, if we refer to the literature provided in this thesis, several options arise, such as the way it was handle by Van Engen (2010, 947), the participants had to attain the required TOEFL scores and, in addition, they had to complete a lab-internal language history questionnaire and a Language Experience and Proficiency Questionnaire, assessing different aspects of their L2 performance, such as their age of initial L2 learning, years of formal education and self-reported proficiency. Moreover, a very comprehensive and useful form of participant sorting was presented by Febo (2003) or Calandruccio and Buss (2017) probing various parts of language, such as language status, language stability, language competency, language history and demand for language use. In accordance with Krashen (1985), the learner is only really able to learn efficiently if his knowledge is at a level which allows it and in this way we could avoid the floor and the ceiling effects, possibly impeding their learning because they simply do not comprehend or because they are at a level which is very high, near-native, therefore very little or no learning would occur. It also seems beneficial to use the method of Tabri et al. (2010) to assess weekly the ratio of native and non-native language use in detail, e.g. the ratio of watching television, reading, conversing in L2 as opposed to L1. We could later compare whether there was a significant difference between those who used L2 more frequently and those who used it to a lesser extent.

Even though many studies, e.g. Cutler et al. (2007), Golestani et al. (2009), Brouwer et al. (2012), Ishida and Arai (2015), Marchegiani and Fafoutis (2015), provide little information about native input, we should not discard such factor when attaining L2 and we should focus carefully on the L2 history of the participants when establishing their L2 proficiency. The fact that the performance of L2 speakers might be influenced by native input is supported by Masuda (2016), who found a subgroup of her participants, who had taken part in a one-month study abroad program, to receive higher TOEIC scores and identification rates in all conditions, in which they were investigated. Tabri et al. (2010) also took native input very seriously, including in her study only bilinguals who were passively perceiving L2 input 50 % of the time, such as reading and listening to the non-native language. Moreover, they also had to actively communicate in the L2 25 % of the time, basically including only those bilinguals who had very high level of native input. In case of longitudinal studies, e.g. Aoyama et al. (2004), McCarthy et al. (2014) and Kim et al. (2018), all the participants were immersed in the L2 environment. Future research could also employ a questionnaire which would periodically assess the native input of the L2 learners and compare it to their performance.

We turn to the type of bilingualism used in perceptual studies with background noise. Mayo et al. (1997) constitutes a crucial research, as it found a significantly higher tolerance for background noise in early bilinguals compared to late bilinguals, closing in on the performance of the native speakers, a finding which, rightfully, influenced other studies to use only early bilinguals, e.g. Febo (2003), Rosenhouse et al. (2006) and Tabri et al. (2010). Some studies, e.g. Ezzatian et al. (2010) and Rogers et al. (2010) used both early and late bilinguals and replicated the findings of Mayo et al. (1997). Future research should, therefore, take into account that early bilinguals in perceptual studies with noise perform much better than late bilinguals and should accommodate its task accordingly to avoid floor and ceiling effects. Future studies, which would consider using native Czech speakers of English should, therefore, choose to include both early and late bilinguals which would yield more interesting results than if only late bilinguals were included. Nevertheless, it is true that such early bilinguals are expected to be scarce, as most Czechs acquire English in imperfect classroom settings with not enough native input. On the other hand, there could be space for improving the imperfect non-native English of native Czech speakers in an immersion setting.

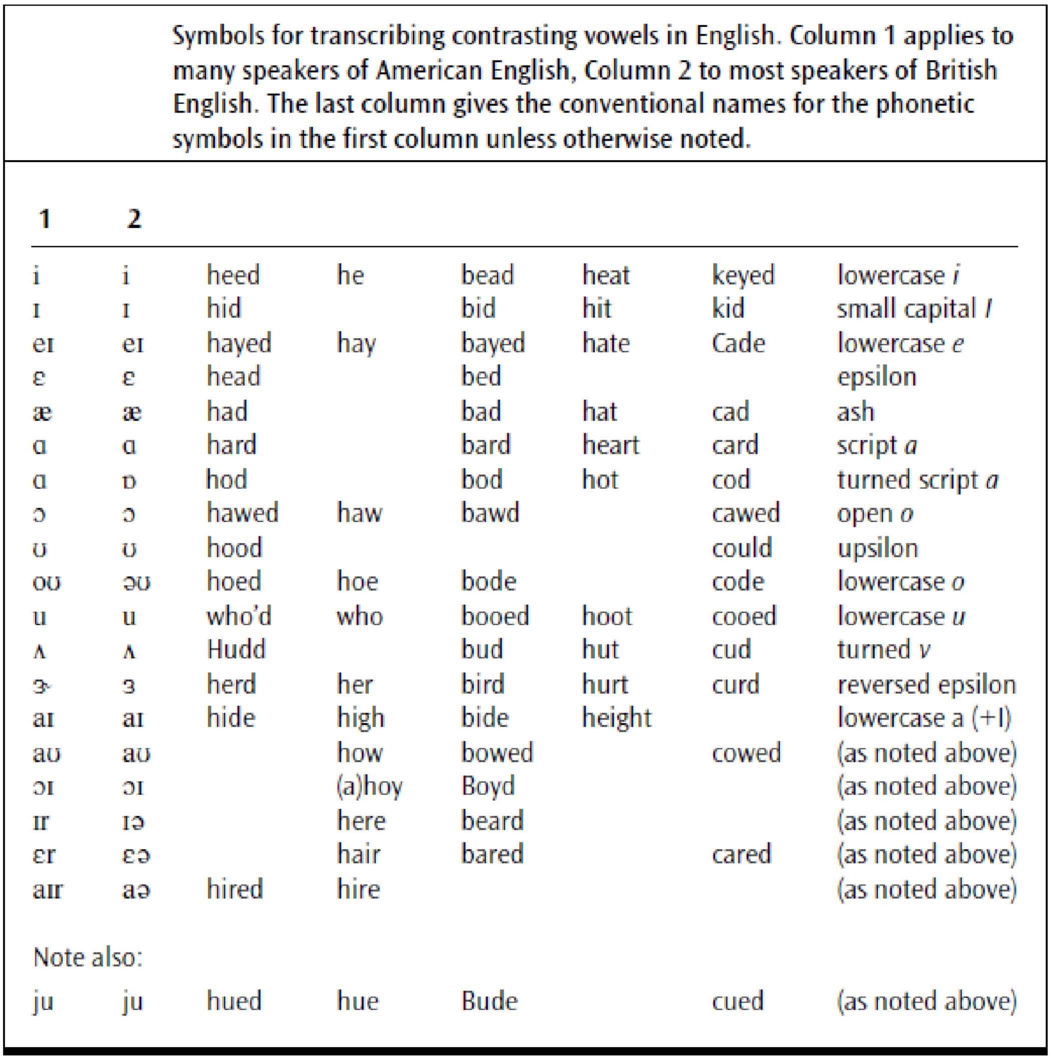
Concerning semantics, Mayo et al. (1997) used SPIN test with keywords always at the end of perceived sentences with controlled predictability, that is to say, the keywords of half of the sentences were inferable from the beginning of the sentences and half were not. The authors found that while the perception of predictable tokens in noise was similar for monolinguals and early bilinguals, unpredictable tokens posed difficulty even for the early bilinguals if we compared their performance with monolinguals and, even though the late bilinguals' performance was way below the performance of the monolingual or early bilingual groups, they were still influenced by the difference based on semantics, which we can see in Figure 5. Such findings are similar to those found by Golestani et al. (2009) who discovered semantic benefit for the native but not their late bilingual L2 speakers of English. Taking into consideration the impact of such mechanism, it should be included in the future research, the late learners might improve substantially after the longitudinal intensive exposure especially as far as the predictable keywords are concerned.

There is a large amount of noise types which can be used in perceptual research depending on its aim, the noise types are summarized in chapter 4.1.3. Due to the fact that the tests of the suggested longitudinal research aimed at establishing conditions which are close to everyday life, probing whether the participants' immersion improved their perception or not, it is highly recommended to use multi-talker babble because it is very near to noise encountered in common spaces, as suggested by Silbert et al. (2014), who mentions the ecological validity of the multi-talker babble compared to other kinds of maskers. In addition, the number of talkers is also the source of variation in the babble, nevertheless, if we want to achieve the highest level of informational masking, in which the voices in the babble are still recognizable, we should use smaller number of talkers, as suggested by Lecumberri et al. (2010). Such line of reasoning is also present in Freyman et al. (2004), investigating the effect of number of talkers in the babble on the perception of the stimuli and finding 2-talker babble to be the most effective masker. When focusing on everyday communication and adverse conditions, we should also consider implementing reverberation in future research and investigating the effects of such adverse condition also in combination with multi-talker babble, as it is a common condition in enclosed public spaces such as restaurants, hallways, cafés and office settings. Reverberation is also very common in formal settings of classrooms where most people learn English.

Another essential factor to control for in perceptual studies in noise is establishing the noise level in which the speech tokens are presented to the participants. If the researchers use too low SNR, the floor effect might arise, in other words, the test might be too difficult for them to take. On the other hand, if the SNR is too high, the researcher might observe a ceiling effect, the case when the participants are close to reaching the highest possible score. Mayo et al.

(1997) used adaptive psychophysical procedure in order to determine at what level of noise the non-native listeners were able to correctly repeat target words 50 % of the time. Initially using noise at 55 dB sound pressure level (SPL) with the increase of 5 dB until the participant made an error, after which the SPL was reduced by 2 dB, until responding correctly, at which point the SPL increased again by 2 dB, when reaching 8 reversals in accuracy, the procedure was halted and the author averaged the noise levels of the last six reversals to establish the level of noise at which the listener could reach 50% accuracy (ibid, 687). Van Engen (2010, 946) also used an adaptive procedure to establish the level of noise at which the participants could repeat full sentences correctly at 50 % of the time, thus establishing four levels of SNR, according to difficulty to avoid ceiling and floor effects from easy SNR +3 dB, 0 dB and harder SNR -3 dB and -6 dB. The author discovered that the L2 speakers needed significantly more favorable SNR (of about 8 dB) to identify English sentences in noise, compared to the native speakers. Due to such finding, it seems convenient for future research to establish some kind of handicap, especially for the late learners, as did e.g. Rogers et al. (2010), who established by his pilot testing an SNR of -8 dB for the monolinguals and early bilinguals and SNR -4 dB for the late learners, thus reducing the difficulty for the late learners, avoiding the floor effect.

The stimuli that the studies focused on were diverse, investigating both consonants and vowels, typically aiming at areas which cause trouble for the participants of the particular study due to the specificities of their native language. We can, e.g. mention the studies which employed Japanese speakers of English, e.g. Strange and Dittmann (1984), Lively et al. (1993), Aoyama et al. (2004) and Shinohara and Iverson (2013), who used almost exclusively the consonants /r/ and /l/ as the distinction is known to cause trouble and the authors investigated, whether they would improve their perception as the result of immersion or training period. Similarly, both Wanrooij et al. (2013) and Escudero and Williams (2014) focused on typical difficulties of Spanish speakers of Dutch, more specifically, the fact that the Spanish speakers rely more on duration and not spectral quality when distinguishing /ɑ/ and /a/. By manipulating the duration of the stimuli, keeping it equal for both vowels, the authors trained the participants to weight the F1, F2 and F3 of the spectrum more importantly and brought them closer to how the vowels are perceived by the native speakers. In a similar manner, the pre- and post-tests in a study involving Czech learners should reflect their deficiencies stemming from their L1 and the characteristics of the L2 English. Turning to vowels, one main difference between Czech and English is the fact that the former has vowel duration as a distinctive phonological feature and even though the ratio between the short and long vowels differs, as well as the fact that there might be differences in quality, the duration functionally distinguishes between Czech short /ɪ ɛ a o u/ and long /iː ɛː aː oː uː/ vowels, creating minimal pairs (Skarnitzl et al. 2016, 100-101). As we can see from the Table 3, the English vocalic system is, quality-wise, much richer and thus the vowels not present in Czech might cause confusions for the non-native speakers, especially distinguishing minimal pairs contrasting /i/-/ɪ/ as in *beat-bit*; /ɪ/-/ɛ/ as in *bid-bed*, /ɛ/-/æ/ as in *bed-bad*, /u/-/ʊ/ as in *pool-pull*, /ɔ/-/ɒ/ as in *cawed-cod*, /ɒ/-/ɑ/ as in *cod-card*.

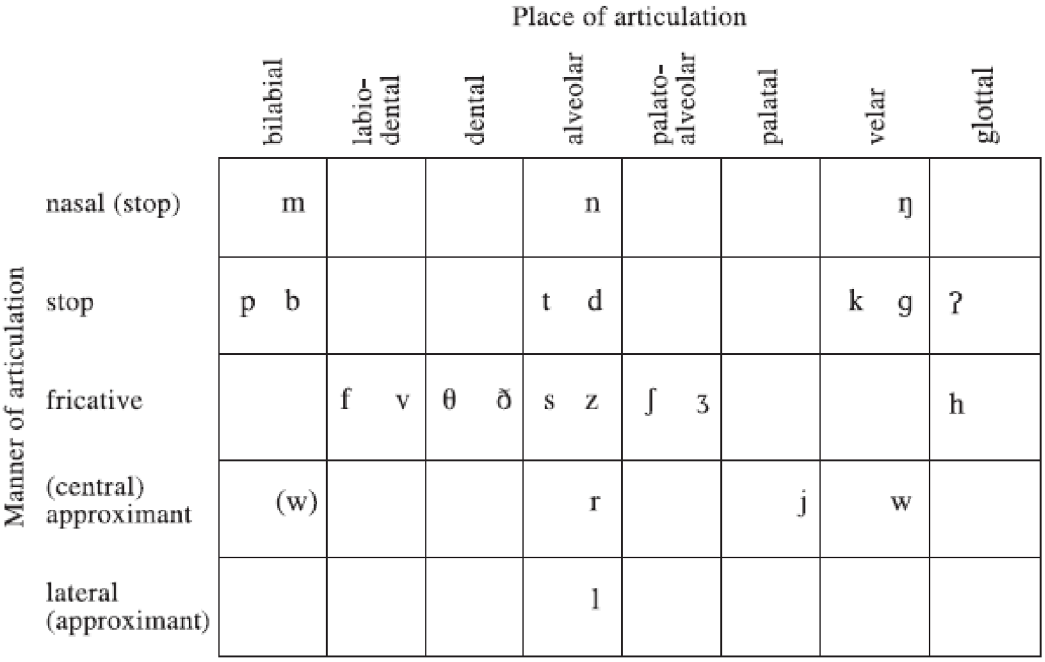


Symbols for transcribing contrasting vowels in English. Column 1 applies to many speakers of American English, Column 2 to most speakers of British English. The last column gives the conventional names for the phonetic symbols in the first column unless otherwise noted.

1	2						
i	i	heed	he	bead	heat	keyed	lowercase *i*
ɪ	ɪ	hid		bid	hit	kid	small capital *I*
eɪ	eɪ	hayed	hay	bayed	hate	Cade	lowercase *e*
ɛ	ɛ	head		bed			epsilon
æ	æ	had		bad	hat	cad	ash
ɑ	ɑ	hard		bard	heart	card	script *a*
ɑ	ɒ	hod		bod	hot	cod	turned script *a*
ɔ	ɔ	hawed	haw	bawd		cawed	open *o*
ʊ	ʊ	hood				could	upsilon
oʊ	əʊ	hoed	hoe	bode		code	lowercase *o*
u	u	who'd	who	booed	hoot	cooed	lowercase *u*
ʌ	ʌ	Hudd		bud	hut	cud	turned *v*
ɝ	ɜ	herd	her	bird	hurt	curd	reversed epsilon
aɪ	aɪ	hide	high	bide	height		lowercase a (+I)
aʊ	aʊ		how	bowed		cowed	(as noted above)
ɔɪ	ɔɪ		(a)hoy	Boyd			(as noted above)
ɪr	ɪə		here	beard			(as noted above)
ɛr	ɛə		hair	bared		cared	(as noted above)
aɪr	aə	hired	hire				(as noted above)
Note also:							
ju	ju	hued	hue	Bude		cued	(as noted above)

Table 3 shows the symbols for contrasting vowels of British English in column 1 and of American English in column 2 (adapted from Ladefoged 2011, 39).

Unlike the Czech vocalic system, the one featuring consonants is fairly rich, Skarnitzl et al. (2016, 102-105) lists 26 consonants in Czech (/m n ɲ l r j r̝ p b t d c ɟ k g t͡s t͡ʃ d͡ʒ f v s z ʃ ʒ x ɦ /) classified by several distinctive features, such as voicing, place of articulation and manner of articulation. Looking at the Table 4, the main apparent distinction between Czech and English consonant system is the lack of dental fricatives /θ ð/ in the former, leading to errors in pronunciation and confusions in perception. A less visible but not less confusing aspect of Czech, as compared to English, constitutes the way it distinguishes plosives /p t k/ from their phonologically voiced counterparts /b d g/ by means of VOT. In case of Czech the phonologically voiced plosives /b d g/ are pre-voiced, as the voicing starts before the release of the plosion with the VOT going to negative values. On the other hand, the phonologically voiceless plosives /p t k/ are distinctively short-lagged, as the voicing for the succeeding vowel starts immediately or shortly after the release of the plosion. In English, the case is different as the phonologically voiced /b d g/ are realized as short-lagged and usually phonetically voiceless and the phonologically voiceless /p t k/ are realized as long-lagged or aspirated with a longer period of voiceless noise after the release of the plosion and before the onset of the following vowel (Ladefoged and Johnson 2011, 198-200). Future researchers should surely aim at one or more of the mentioned difficulties, investigating their development before and after the immersion. Generally, whether the language is Czech or any other, the researchers should always control for the specificity of L1 when performing a perceptual experiment involving non-native speakers, owing to the fact that it could have significant effect on the results of the study and bring unaccounted for influences.



	Place of articulation							
Manner of articulation	bilabial	labio-dental	dental	alveolar	palato-alveolar	palatal	velar	glottal
nasal (stop)	m			n			ŋ	
stop	p b			t d			k g	ʔ
fricative		f v	θ ð	s z	ʃ ʒ			h
(central) approximant	(w)			r		j	w	
lateral (approximant)				l				

Table 4 represents the inventory of English consonants, when the consonant is located on the right side of the cell, it indicates a voiced sound. It is also important to mention that /tʃ/ and /dʒ/ do not appear in the table, because they are considered by Ladefoged and Johnson (2011) as sequences of sounds rather than independent elements (adapted from Ladefoged and Johnson 2011, 43).

The current section offered ideas on future research based on the evidence gathered from three different types of studies. In addition, it offered observations about several aspects one should consider when carrying out a perceptual study with special focus on Czech speakers of English. It focused on L2 proficiency, native input, the type of bilingualism of the participants and how semantics might help the participants to perceive the correct words. Moreover, it also considered a suitable noise type for the pre- and post-test to evaluate the perception in noise one might encounter in everyday situations together with the level of such noise. Lastly, it offered a paragraph concerning suitable stimuli for the future perceptual study, mentioning some of the difficulties a Czech speaker might have when perceiving English speech.

6. SUMMARY

The current thesis offers a review of literature divided into two sections, the first of which provides information about recognition of speech sounds, models of phonetic category formation, L2 learning factors and speech-in-noise specificity. The second one analyzes experiments, which focus on the L2 speech perception with/without noise. The last content chapter, based on the acquired knowledge, gives advice towards future perceptual research in noise with a special focus on Czech speakers of English.

In the first part of literature review, we talk about recognition of speech sounds, the idealized communication model, the importance of redundancy and how it is generated. Models of phonetic category formation are considered, such as Flege's Speech Learning Model (SLM), focusing on perception and production, and the Perceptual Assimilation Model for L2 (PAM-L2) by Best and Tyler, aiming at perception and offering predictions about L2 sound assimilation, trying to predict the L2 sounds that will cause most trouble. The models are also supplemented by their revised versions, out of which the PAM-L2 for FLL by Tyler (2019) provides very useful information. This section also compares and contrasts such crucial studies. The literature review then moves on to investigating the factors that might affect L2 learning in general, analyzing native input, L2 use, learning setting, age distinctions and motivation. The following subsection introduces the topic of noise and comments upon different types of adverse conditions, distinguishing between energetic and informational masking, and also mentioning reverberation.

The second part of the literature review analyzes three types of experiments which are thematically connected but differ substantially in their methodology. Firstly, we investigated research in L2 perception in noise which does not employ longitudinal model, nor any sort of training, simply focusing on the performance of the L2 listeners in noise at a single point, comparing them with native speakers and/or between each other, according to their age of acquisition. Secondly, we focus our attention on longitudinal immersion studies, which do not use noise in their paradigm. The third group is represented by studies with a training paradigm which do not employ noise, with the exception of Lengeris and Hazan (2010) and Cooke and Lecumberri (2018).

The final content section takes into account the investigated studies to offer guidance for future research, speaking about the nature of evidence and considering several aspects of non-native perceptual research in noise. The future research is designed with the special focus on Czech speakers of English as there is a gap with respect to perceptual studies in noise that would include native Czech speakers and this section should work as a basic template for potential future research.

7. WORKS CITED

Aoyama Katsura, James Emil Flege, Susan G. Guion, Reiko Akahane-Yamada, and Tsuneo Yamada. 2004. "Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/." *Journal of Phonetics*, 32(2): 233-250.

Aoyama, Katsura, Susan G. Guion, James Emil Flege, Tsuneo Yamada, and Reiko Akahane-Yamada. 2008. "The First Years in an L2-Speaking Environment: A Comparison of Japanese Children and Adults Learning American English." *International Review of Applied Linguistics in Language Teaching (IRAL)*, 46(1): 61-90.

Best, Catherine T. 1995. "A direct realist view of cross-language speech perception." *Speech perception and linguistic experience*: 171-206.

Best, Catherine T., and Michael D. Tyler. 2007. "Nonnative and second-language speech perception: Commonalities and complementarities." In *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins.

Bongaerts, Theo. 1999. "Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners." *Second language acquisition and the critical period hypothesis*: 133-159.

Bradlow, A. R., D. B. Pisoni, R. Akahane-Yamada, and Y. I. Tohkura. 1997. "Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production." *The Journal of the Acoustical Society of America*, 101(4): 2299-2310.

Brouwer, S., K. J. Van Engen, L. Calandruccio, and A. R. Bradlow. 2012. "Linguistic contributions to speech-on-speech masking for native and non-native listeners: Language familiarity and semantic content." *The Journal of the Acoustical Society of America*, 131(2): 1449-1464.

Brungart, D. S., Brian D. Simpson, Mark A. Ericson, and Kimberly R. Scott. 2001. "Informational and energetic masking effects in the perception of multiple simultaneous talkers." *The Journal of the Acoustical Society of America*, 110(5): 2527-2538.

Calandruccio, Lauren, Emily Buss, and Kristina Bowdrie. 2017. "Effectiveness of two-talker maskers that differ in talker congruity and perceptual similarity to the target speech." *Trends in Hearing*, 21: 1646–1654.

Casillas, J. V. 2020. "Phonetic category formation is perceptually driven during the early stages of adult L2 development." *Language and Speech*, 63(3): 550-581.

Carrió-Pastor, María Luisa, and Eva M. Mestre Mestre. 2014. "Motivation in second language acquisition." *Procedia-Social and Behavioral Sciences*, 116: 240-244.

Chang, C. B. 2019. "Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research." *Journal of Phonetics*, 74: 96-113.

Corder, S. P. 1967. "The significance of learners' errors." *International Review of Applied Linguistics*, 5: 161–170.

Cooke, Martin, and Maria Luisa Garcia Lecumberri. 2016. “The Effects of Modified Speech Styles on Intelligibility for Non-Native Listeners.” *Interspeech 2016*: 868-872.

Cooke, Martin, and Maria Luisa Garcia Lecumberri. 2018. “Effects of exposure to noise during perceptual training of non-native language sounds.” *The Journal of the Acoustical Society of America*, 143(5): 2602-2610.

Cooke, Martin, Vincent Aubanel, and María Luisa García Lecumberri. 2019. “Combining spectral and temporal modification techniques for speech intelligibility enhancement.” *Computer Speech & Language*, 55: 26-39.

Cutler, Anna 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge: The MIT Press.

Cutler, Anna, Martin Cooke, Maria Louisa Garcia Lecumberri, and Dennis Pasveer. 2007. “L2 consonant identification in noise: Cross-language comparisons.” *Interspeech 2007*: 1585-1588.

Deci, Edward L., Wayne F. Cascio, and Judith Krusell. 1975. “Cognitive evaluation theory and some comments on the Calder and Staw critique.” *Journal of Personality and Social Psychology*, 31(1): 81–85.

Escudero, Paola. 2000. *Developmental patterns in the adult L2 acquisition of new contrasts: The acoustic cue weighting in the perception of Scottish tense/lax vowels by Spanish speakers*. (Doctoral dissertation).

Escudero, Paola, and Daniel Williams. 2014. “Distributional learning has immediate and long-lasting effects.” *Cognition*, 133(2): 408-413.

Ezzatian, Payam, Meital Avivi, and Bruce A. Schneider. 2010. “Do nonnative listeners benefit as much as native listeners from spatial cues that release speech from masking?” *Speech Communication*, 52(11-12): 919-929.

Febo, Dashielle M. 2003. *Effects of bilingualism, noise, and reverberation on speech perception by listeners with normal hearing*. Tampa: University of South Florida. (Doctoral dissertation).

Ferguson, Charles A. 1968. “Absence of Copula and the Notion of Simplicity: A Study of Normal Speech, Baby Talk, Foreigner Talk and Pidgins.” *Conference on Pidginization and Creolization of Languages*.

Flege, James Emil. 1995. “Second language speech learning: Theory, findings, and problems.” In *Speech perception and linguistic experience: Issues in cross-language research*, 92: 233-277.

Flege, James Emil, and Ian R. A. MacKay. 2004. “Perceiving vowels in a second language.” *Studies in second language acquisition*, 26(1): 1-34.

Flege, James Emil, and Ocke-Schwen Bohn. 2021. “The revised speech learning model (SLM-r).” In *Second language speech learning: Theoretical and empirical progress*: 3-83. Cambridge: Cambridge University Press.

Flege, James Emil, and Robert M. Hammond. 1982. “Mimicry of non-distinctive phonetic differences between language varieties.” *Studies in Second Language Acquisition*, 5(1): 1-17.

Flege, James Emil, and Serena Liu. 2001. “THE EFFECT OF EXPERIENCE ON ADULTS’ ACQUISITION OF A SECOND LANGUAGE.” *Studies in second language acquisition*, 23(4): 527-552.

Flege, James Emil, Murray J. Munro, and Ian RA MacKay. 1995. “Factors affecting strength of perceived foreign accent in a second language.” *The Journal of the Acoustical Society of America*, 97(5): 3125-3134.

Francis, Alexander L., Natalya Kaganovich, and Courtney Driscoll-Huber. 2008. “Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English.” *The Journal of the Acoustical Society of America*, 124(2): 1234-1251.

Freyman, Richard L., Uma Balakrishnan, and Karen S. Helfer. 2004. “Effect of number of masking talkers and auditory priming on informational masking in speech recognition.” *The Journal of the Acoustical Society of America*, 115(5): 2246-2256.

Giannakopoulou, Anastasia, Maria Uther, and Sari Ylinen. 2013. “Enhanced plasticity in spoken language acquisition for child learners: Evidence from phonetic training studies in child and adult learners of English.” *Child Language Teaching and Therapy*, 29(2): 201-218.

Giannakopoulou, A., H. Brown, M. Clayards, E. Wonnacott. 2017. “High or low? Comparing high and low-variability phonetic training in adult and child second language learners.” *PeerJ*, 5.

Golestani, Narly, Stuart Rosen, and Sophie K. Scott. 2009. “Native-language benefit for understanding speech-in-noise: The contribution of semantics.” *Bilingualism: Language and Cognition*, 12(3): 385-392.

Gong, J., Y. Yu, W. Bellamy, F. Wang, and X. Ji. 2021. “Effect of Perceptual Training with Noise on Chinese Learners' English Consonant Reception Thresholds.” *2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*: 1087-1091.

Hermansky, H. 2019. “Coding and decoding of messages in human speech communication: Implications for machine recognition of speech.” *Speech Communication*, 106: 112-117.

Holliday, J. J. 2015. “A longitudinal study of the second language acquisition of a three-way stop contrast.” *Journal of Phonetics*, 50: 1-14.

Ishida, Mako, and Takayuki Arai. 2015. “Perception of an existing and non-existing L2 English phoneme behind noise by Japanese native speakers.” *Interspeech 2015*: 3408-3411.

Iverson, Paul, and Bronwen G. Evans. 2009. “Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers.” *The Journal of the Acoustical Society of America*, 126(2): 866-877.

Kartushina, Natalia, and Clara D. Martin. 2019. “Third-language learning affects bilinguals’ production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production.” *Journal of Phonetics*, 77.

Khasinah, S. 2014. “Factors influencing second language acquisition.” *Englisia: Journal of Language, Education, and Humanities*, 1(2): 256-269.

Kim, Donghyun, Meghan Clayards, and Heather Goad. 2018. "A longitudinal study of individual differences in the acquisition of new vowel contrasts." *Journal of Phonetics*, 67: 1-20.

Kondaurova, Maria V., and Alexander L. Francis. 2010. "The role of selective attention in the acquisition of English tense and lax vowels by native Spanish listeners: Comparison of three training methods." *Journal of phonetics*, 38(4): 569-587.

Krashen, S. 1985. *The Input Hypothesis: Issues and Implications.* New York: Longman.

Ladefoged, Peter, and Keith Johnson. 2011. *A Course in Phonetics 6th Edition*. Boston: Wadsworth, Cengage Learning.

Lecumberri, M. L. Garcia, and Martin Cooke. 2006. "Effect of masker type on native and non-native consonant perception in noise." *The Journal of the Acoustical Society of America*, 119(4): 2445-2454.

Lecumberri, Maria Luisa Garcia, Martin Cooke, and Anne Cutler. 2010. "Non-native speech perception in adverse conditions: A review." *Speech communication*, 52(11-12): 864-886.

Lenneberg, Eric H. 1967. "The biological foundations of language." *Hospital Practice*, 2(12): 59-67.

Lengeris, Angelos, and Valerie Hazan. 2010. "The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek." *The Journal of the Acoustical Society of America*, 128(6): 3757-3768.

Lively, Scott E., John S. Logan, and David B. Pisoni. 1993. "Training Japanese listeners to identify English /r/and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories." *The Journal of the Acoustical Society of America*, 94(3): 1242-1255.

Logan, John S., Scott E. Lively, and David B. Pisoni. 1991. "Training Japanese listeners to identify English /r/ and /l/: A first report." *The Journal of the Acoustical Society of America*, 89(2): 874-886.

Marchegiani, Letizia, and Xenofon Fafoutis. 2015. "On cross-language consonant identification in second language noise." *The Journal of the Acoustical Society of America*, 138(4): 2206-2209.

Masuda, Hinako. 2016 "Misperception patterns of American English consonants by Japanese listeners in reverberant and noisy environments." *Speech Communication* 79: 74-87.

Mattys, Sven L., Katharine Barden, and Arthur G. Samuel. 2014. "Extrinsic cognitive load impairs low-level speech perception." *Psychonomic bulletin & review*, 21(3): 748-754.

Mayo, L. H., Mary Florentine, and Søren Buus. 1997. "Age of second-language acquisition and perception of speech in noise." *Journal of speech, language, and hearing research*, 40(3): 686-693.

McCarthy, Kathleen M., Merle Mahon, Stuart Rosen, and Bronwen G. Evans. 2014. "Speech perception and production by sequential bilingual children: A longitudinal study of voice onset time acquisition." *Child development* 85(5): 1965-1980.

Mi, L., S. Tao, W. Wang, Q. Dong, B. Dong, M. Li, and C. Liu. 2021. "Training non-native vowel perception: In quiet or noise." *The Journal of the Acoustical Society of America*, 149(6): 4607-4619.

Miller, G. A. 1951. *Language and communication*. New York: McGraw-Hill Book Company, Inc.

Munro, M., T. Derwing, and R. Thomson. 2003. "A longitudinal examination of English vowel learning by Mandarin speakers." *Canadian Acoustics*, 31(3): 32-33.

Nábělek, Igor V., Hsiao-Chuan Chen, and Sumalai Maroonroge. 1988. "Spectrographic comparison of whispered voiced and voiceless stop consonants in various vowel environments." *The Journal of the Acoustical Society of America*, 83(S68).

Nagle, Charles L., Colleen Moorman, and Cristina Sanz. 2016. "Disentangling research on study abroad and pronunciation: Methodological and programmatic considerations." In *Handbook of research on study abroad programs and outbound mobility*: 673-695. Hershey: IGI Global.

Nagle, C. L. 2019. "A longitudinal study of voice onset time development in L2 Spanish stops." *Applied Linguistics*, 40(1): 86-107.

Nielsen, Andreas Højlund, Nynne Thorup Horn, Stine Derdau Sørensen, William B. McGregor, Mikkel Wallentin. 2015. "Intensive foreign language learning reveals effects on categorical perception of sibilant voicing after only 3 weeks." *i-Perception*, 6(6): 1-26.

Oh, G. E., Susan Guion-Anderson, Katsura Aoyama, James E. Flege, Reiko Akahane-Yamada, Tsuneo Yamada. 2011. "A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting." *Journal of phonetics*, 39(2): 156-167.

Perrachione, T. K., J. Lee, L. Y. Ha, and P. C. Wong. 2011. "Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design." *The Journal of the Acoustical Society of America*, 130(1): 461-472.

Rogers, Catherine L., Teresa M. DeMasi, and Jean C. Krause. 2010. "Conversational and clear speech intelligibility of /bVd/ syllables produced by native and non-native English speakers." *The Journal of the Acoustical Society of America*, 128(1): 410-423.

Rosenhouse, Judith, Lubna Haik, and Liat Kishon-Rabin. 2006. "Speech perception in adverse listening conditions in Arabic-Hebrew bilinguals." *International Journal of Bilingualism*, 10(2): 119-135.

Selinken, L., and Susan M. Gass. 2008. *Second Language Acquisition. An introduction course*. New York: Routledge.

Shia, Lu-Feng, and Nadia Farooqa. 2012. "Bilingual Listeners' Perception of Temporally Manipulated English Passages." *Journal of Speech, Language, and Hearing Research*, 55: 125-138.

Shinohara, Yasuaki, and Paul Iverson. 2013. "Computer-based English /r/-/l/ perceptual training for Japanese children." *Proceedings of Meetings on Acoustics ICA2013*, 19(1).

Shinohara, Yasuaki, and Paul Iverson. 2015. "Effects of English /r/-/l/ perceptual training on Japanese children's production." *The 18th International Congress of Phonetic Sciences*.

Silbert, N. H., Kenneth de Jong, Kirsten Regier, Aaron Albin, and Yen-Chen Hao. 2014. "Acoustic properties of multi-talker babble." *The Journal of the Acoustical Society of America*, 135(4).

Singleton, David Michael, and Zsolt Lengyel. 1995. *The age factor in second language acquisition: A critical look at the critical period hypothesis*. Bristol: Multilingual Matters.

Sisinni, Bianca, and Mirko Grimaldi. 2011. "Validating a Second Language Perception Model for Classroom Context. A Longitudinal Study within the Perceptual Assimilation Model." *Twelfth Annual Conference of the International Speech Communication Association*.

Skarnitzl, Radek, Pavel Šturm, and Jan Volín. 2016. *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči*. Prague: Karolinum Press.

Strange, Winifred, and Sibylla Dittmann. 1984. "Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English." *Perception & psychophysics*, 36(2): 131-145.

Stevens, J. 2001. "Study abroad learners' acquisition of the Spanish voiceless stops." *MIFLC Review*, 10: 137-151.

Sun, Hui, Kazuya Saito, and Adam Tierney. 2021. "A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition." *Studies in Second Language Acquisition*, 43(3): 551-573.

Tabri, Dollen, Kim Michelle Smith Abou Chacra, and Tim Pring. 2010. "Speech perception in noise by monolingual, bilingual and trilingual listeners." *International Journal of Language & Communication Disorders*: 1-12.

Tsukada, K., D. Birdsong, E. Bialystok, M. Mack, H. Sung, and J. Flege. 2005. "A developmental study of English vowel production and perception by native Korean adults and children." *Journal of Phonetics*, 33(3): 263-290.

Tsukada, K., D. Birdsong, M. Mack, H. Sung, E. Bialystok, and J. Flege. 2004. "Release bursts in English word-final voiceless stops produced by native English and Korean adults and children." *Phonetica*, 61(2-3): 67-83.

Tyler, Michael D. 2019. "PAM-L2 and phonological category acquisition in the foreign language classroom." In *A sound approach to language matters–In honor of Ocke-Schwen Bohn*: 607-630. Arhus: Arhus University.

Ura, Masako, Kathleen S. J. Preston, and Jack Mearns. 2015. "A measure of prejudice against accented English (MPAAE) scale development and validation." *Journal of Language and Social Psychology*, *34*(5): 539-563.

Van Engen, K. J. 2010. "Similarity and familiarity: Second language sentence recognition in first- and second-language multi-talker babble." *Speech communication*, 52(11-12): 943-953.

Van Wijngaarden, Sander J., Herman J. M. Steeneken, and Tammo Houtgast. 2002. "Quantifying the intelligibility of speech in noise for non-native listeners." *The Journal of the Acoustical Society of America*, 111(4): 1906-1916.

Volín, Jan, and Radek Skarnitzl. 2010. "The strength of foreign accent in Czech English under adverse listening conditions." *Speech Communication*, 52(11-12): 1010-1021.

Wang, Xianhui, and Li Xu. 2021. "Speech perception in noise: Masking and unmasking." *Journal of Otology*, 16(2): 109-119.

Wanrooij, Karin, Paola Escudero, and Maartje E. J. Raijmakers. 2013. “What do listeners learn from exposure to a vowel distribution? An analysis of listening strategies in distributional learning.” *Journal of Phonetics*, 41(5): 307-319.

Ylinen, S., M. Uther, A. Latvala, S. Vepsäläinen, P. Iverson, R. Akahane-Yamada, and R. Näätänen. 2010. “Training the brain to weight speech cues differently: A study of Finnish second-language users of English.” *Journal of Cognitive Neuroscience*, 22(6): 1319-1332.