Filozofická fakulta Univerzity Palackého

Perceptual adaptability and its generalization across the system of English vowel in native and non-native listeners

(Bakalářská práce)

2018 **Martin Zeidler**

Filozofická fakulta Univerzity Palackého
Katedra anglistiky a amerikanistiky

Perceptual adaptability and its generalization across the system of English vowel in native and non-native listeners
(Bakalářská práce)

Autor: **Martin Zeidler**
Studijní obor: Angličtina se zaměřením na komunitní tlumočení a překlad
Vedoucí práce: **Mgr. Václav Jonáš Podlipský, Ph.D.**
Počet stran: 72
Počet znaků: 115 304 (bez příloh)
Olomouc 2018

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a uvedl úplný seznam citované a použité literatury.

V Olomouci dne 17. 5. 2018 **Martin Zeidler**

Motto: “*Be the change you wish to see in the world.*”

– Mahatama Gandhi

Děkuji vedoucímu diplomové práce panu doktoru Podlipskému za obětavou pomoc s návrhem výzkumu, dále za zpracování všech použitých skriptů a samohláskového kontinua, vypracování grafů 3–6 a 8, za odbornou pomoc při analýze výsledků, doporučení první testované hypotézy a inspiraci k druhé hypotéze myšlenkou, že existuje spojitost mezi krátkodobou percepční adaptabilitou a diachronním vývojem jazyků, a konečně za cenné připomínky k vypracování diplomové práce. Rovněž děkuji doktoru Jeffrey K. Parrottovi za ochotnou asistenci při přípravě výzkumu a všem ochotným účastníkům experimentu.

V Olomouci dne 17. 5. 2018 **Martin Zeidler**

Abstract

Previous research has established the concept of short-term perceptual adaptability as a quick adjustment of listeners' vowel categories in accordance with recently perceived speech characteristics. For instance, exposition to existing words containing the sound /s/ manipulated in such way to resemble both [s] and [f] causes the listener's perceptual boundaries between /s/ and /f/ to shift. The present thesis studies perceptual adaptability in native and non-native listeners of English due to exposure to non-standard tokens of English close front vowel /i/, namely exposure to words with more close [i̝] or more open [i̞] tokens, produced via computer manipulation of formants. The goals are, first, to replicate the findings of adaptation for vowels, this time with native English listeners. Second, to answer the question of whether adaptation occurs for L2 listeners in the same way as for L1 listeners. Finally, it tests the hypothesis that the shift of perceptual boundaries between English /i/ and /ɪ/ triggered by exposure to manipulated tokens of /i/ also leads to a shift of the boundary between English /ɪ/ and /ɛ/; the study attempts to experimentally induce a push- or pull-chain adjustment of vocalic categories, such as those described for diachronic changes of vowel systems. In the exposure phase, respondents were presented with manipulated vowel tokens in the form of a lexical-decision task. In the following testing phase, their perceptual boundaries were measured in a vowel categorization task of tokens from a densely sampled [i]~[ɪ]~[ɛ] continuum. Pilot results are presented and discussed.

Keywords

perceptual adaptability, English vowels, front vowels, native listeners, non-native listeners, push-chain, pull-chain shift

Anotace

Předchozí výzkum doložil jev krátkodobé percepční adaptability, tedy okamžité úpravě hláskových kategorií posluchačem v závislosti na specifičnosti nedávno vnímaných promluv. Například, pokud je posluchač vystaven smysluplným slovům obsahujícím hlásku /s/, která je upravena tak, aby zněla mezi [s] a [f], dojde u něj k posunu percepční hranice mezi kategoriemi /s/ a /f/. V této práci je adaptabilita zkoumána u rodilých a nerodilých posluchačů angličtiny pro vnímání nestandardní výslovnosti zavřené přední anglické samohlásky /i/ (konkrétně s otevřenější [i̞] nebo zavřenější [i̝] kvalitou docílenou počítačovou manipulací samohláskových formantů). Cílem je v první řadě replikovat dosažené výsledky adaptace pro samohlásky u rodilých mluvčích a dále zodpovědět otázku, zdali k adaptaci dochází obdobným způsobem i u nerodilých mluvčích. Konečně je testována hypotéza, že posun percepční hranice mezi anglickými samohláskami /i/ a /ɪ/ vyvolaný expozicí manipulovaným exemplářům /i/ povede také k posunu hranice mezi /ɪ/ a /ɛ/. Záměrem experimentu je vyvolat řetězový posun hláskových kategorii (tzv. „push/pull-chain“ popisovaný pro diachronní změny samohláskových systémů). Pro expozici manipulovaným samohláskám slouží test lexikálního statutu slova (posluchač rozhoduje, zda jde o existující slovo, nebo ne) a percepční hranice je určena následným kategorizačním experimentem s izolovanými samohláskami z plynulého kontinua [i]~[ɪ]~[ɛ]. Práce popisuje výsledky pilotních testů.

Klíčová slova

percepční adaptabilita, anglické samohlásky, přední samohlásky, rodilí mluvčí, nerodilí mluvčí, posun samohláskových kategorií

OBSAH

1 Literature review

Language offers incredible diversity. Owing to the brilliance of this human communicative device, we can be sure that no two identical conversations were ever held and we may always encounter a new, unprecedented sentence. This is due to the fact that all language constructions are composed using only a limited set of meaningless units, phonemes, which can be, however, put into a virtually unlimited number of meaningful combinations. Furthermore, these combinations function as higher-level units and can also be combined into more elaborate constructions and used again in the same way, potentially *ad infinitum*. This is an intrinsic property of human language called double articulation (or also dual patterning) which ensures its productivity (Martinet 1984). However, the variety of linguistic choices is not the only one. In fact, even if we were to witness two presumably identical conversations—for example listening repeatedly to a play with scripted dialogues—there still would be differences, although some of those may be less noticeable since they would not (or would not be supposed to) affect the intended meaning.

In order to understand why there are differences in speech which are insignificant or even detrimental to mutual comprehension, we need to illustrate that, from an acoustic point of view, variability of speech can be seen as on a continuous scale, while from a linguistic standpoint it is rather viewed as a set of discrete categories (Repp and Libermant 1984, 31).

1.1 Linguistic vs. acoustic variability

The aforementioned higher-level units are in essence what we would call in everyday life "words". In linguistics, however, the definition of a word is not entirely unified. It is thus essential to bear in mind that, in the current study, the term "word" will be used as any meaningful sequence of phonemes: the higher-level units described above.

Words never take an ambiguous form per se because they are defined on the basis of phonemes. For example, every literate speaker of English can identify that the word *feel* is composed of three sounds (phonemes) /f/, /i/, and /l/, that all these sounds are also in the word *leaf*, and that some of them are also in the words *fake*, *me* and *low*. The only possible ambiguity is connected directly with the nature of the phonemes. We cannot reliably identify a word unless we are sure of what sounds it contains. This implies that we need to be somehow able to categorize what we hear into phonemes because these constitute the building material for all language constructions. That brings us to the question of what exactly defines a phoneme. For the sake of this illustration, it is sufficient to say that phonemes can be defined as having certain phonological features, such as voicing or nasality, which are derived from the possibilities of the human vocal tract. This featural makeup is what creates contrasts between individual phonemes which, in turn, distinguish individual words. Therefore, the **perceivable** identity of a phoneme is conditioned by having a particular feature or a cluster of features. This basic conception, however, is not reflected in measurable acoustic properties. It can be argued that the features are, like linguistic signs, arbitrary (Culler 1986, 28). It can be furthermore illustrated on cross-linguistic differences. Phonemic inventories of various languages differ. For instance, Czech but not English has the /ɲ/ phoneme, and English but not Czech has /θ/ and /ð/.

The essentiality of categorizing the acoustic spectrum can also be logically deduced when we look at the perceptual characteristics of a sound. Among them are pitch, duration, loudness, and timbre. Simply put, the quality of a sound, as that of a mechanical wave, is determined by a composite of sound waves of different frequencies and amplitudes varying over time. These variables are subject to many factors, which are both external (modulating the sound on its way through space) and internal (limiting and defining the sound source). Because of the interplay of these factors, if phonemes were to be identifiable by unique canonical acoustic signatures, we would most likely not be able to identify that exact phoneme ever again. Communication could then be successful only in perfect conditions where sound would remain undistorted. We can see that this is hardly ever the case since successful communication takes place routinely. The drawbacks of one-by-one matching between linguistic categories and acoustic profiles are also apparent in the typical struggles of voice recognition technology. Its sometimes amusing and certainly not useful creations may also be viewed as evidence of how ambiguous can the unprocessed acoustic signal be. But people are quite successful at mutual comprehension even under noisy conditions. We may thus infer that listeners possess a categorization tool which helps them sort out the characteristics of speech sounds and then map them onto the set of stored representations. This problem of finding a clear match between acoustic and phonemic categories, perhaps via a single hidden variable, is called lack of invariance (vide Liberman, Cooper, Shankweiler, and Kennedy 1967; Jusczyk 2000).

The internal and external factors mentioned above pertain to sound quality in general, but considering perception of human speech, we may speak of more specific ones.

First, the acoustic nature of a speech sound may be influenced by its phonetic context (Hillenbrand, Clark, and Nearey 2001; Lisker and Abramson 1967), that is, by its surrounding sounds and suprasegmental features, such as stress, intonation, speech-rate and degree of enunciation. The acoustic influence of these features can also be inferred from the fact that in some languages they differentiate meaning, e. g. tone in Mandarin Chinese (Ladefoged and Johnson 2011, 255). This variation in distinctiveness of features across languages further supports the claim that phonetic and acoustic categories do not correspond in a straightforward, unambiguous way and categorization has to be carried out by the listener.

Second, there is variability due to speaker characteristics (Peterson and Barney 1952) which comprise the shape and size of vocal tract as well as differences in dialects or pronunciation idiosyncrasies.

Any other factors, such as the communication channel (quality of a recording, telephone signal, etc.), the surrounding acoustics or position of the speaker, are considered external. Both internal and external factors with respect to how listeners compensate for their effects on sound, and especially speech, will be discussed in more detail later.

As stated above, if listeners were not capable of identifying and decoding relevant information, the interplay of all the mentioned factors together with the lack of invariance would make communication impossible. What is more, listeners can tell whether a voice is coming from a distance, a large hall, whether it is stationary or moving, whether it belongs to a child, a man, or a woman, and other details, even without other sensory information.

1.2 Phonetic categorization

The arguments stated above imply that listeners identify and sort out layers of relevant information in a speech signal, and thus are essentially performing a categorization task (Goudbeek, Smits, Cutler, and Swingley 2005; Goudbeek, Swingley, and Smits 2009; Holt and Lotto 2008; 2010). This process of speech perception will be here labeled **phonetic categorization**. Owing to phonetic categorization, listeners are able to hold daily conversations fairly effortlessly.

It is important to realize that the contrast between continuous variability of acoustic properties and discrete linguistic categories does not in itself necessarily pose challenges to an auditory system only capable of grouping sounds from a specific range of acoustic values and assigning them linguistic labels. It is the lack of invariance, the possibility of overlap between linguistic categories on acoustic dimensions that would prove such statically working system insufficient for communication.

Interestingly, categorization exceeds the scope of daily conversation and provides listeners with an ability to swiftly adapt to non-canonical speech as well; whether due to a foreign accent (e. g., Clarke and Garrett 2004; Bradlow and Bent 2008; Baese-Berk, Bradlow, and Wright 2013), dialectal variation (e. g., Sumner and Samuel 2009; Trude and Brown-Schmidt 2012), or distorted speech (e. g., Shannon, Zeng, Kamath, Wygonski, and Ekelid 1995; Davis, Johnsrude, Hervais-Adelman, Taylor, and McGettigan 2005). What mechanisms allow listeners to achieve this and their exact function has been a subject of a large body of research. What is certain is that these mechanisms are highly intricate because there were multiple similar processes identified, each acting upon perception in a slightly different manner. Among other, we may distinguish perceptual compensation, speaker normalization, or selective adaptation, all of which will be discussed later in this study. For current purposes, the ability to adjust the process of phonetic categorization according to the characteristics of unfamiliar speech will be referred to as **perceptual adaptability**. As with any scientific research, while many aspects of the subject matter are being studied and numerous hypotheses are creating further and further subfields, relevant terminology sporadically remains consistent. For that reason, many phenomena described on the following pages will be referred to by various terms with respect to individual authors' choice. All can be, however, also understood through the concept of perceptual adaptability.

1.3 Perceptual compensation

Referring back to the external factors affecting the quality of a speech signal mentioned earlier, we can distinguish research on so-called perceptual compensation. Watkins (1991) in his rigorous study focuses on what are the effects of changes in spectral envelope of a sound upon its perception. The term spectral envelope is understood as the spectrum of amplitudes for various frequencies in a sound signal. As a sound travels through an acoustic environment, its reverberations lead to certain frequencies being amplified while others get attenuated. Previous research states that listeners compensate for these changes and perceive sounds with subjective constancy (Risset and Wessel 1982), which enables them to recognize identical sounds coming from, or being perceived in, different acoustic environments despite the associated differences in soundwaves.

Watkins in his experiments (1991) presented listeners with a carrier phrase “the next word is” distorted by one of two spectral filters followed by a test sound drawn from an [ɪtʃ] to [ɛtʃ] continuum. The carrier phrase filters were designed to have a frequency response of the difference of the tested vowels’ spectral envelopes. This way, the amplitudes of frequencies these vowels share are flattened, while the differences get amplified. Consequently, an /ɛ/ vowel played through an /ɪ/ minus /ɛ/ filter takes on the spectral envelope, and consequently perceivable quality, of an /ɪ/ vowel and vice versa. The effects of a carrier phrase manipulated in this fashion on following testing words are similar to a phenomenon called the enhancement effect (Kiefte and Kluender 2005) described by Summerfield, Sidwell, and Nelson. “This effect demonstrates the existence of a mechanism or set of mechanisms, by which newly arriving acoustical energy receives a favored auditory representation in relation to pre-existing energy” (1987, 701). This essentially means that listeners perceive preferably changes in amplitude as opposed to its absolute values. Therefore, when a testing word is heard after a carrier phrase with particular frequencies accentuated, these frequencies are in the following testing words perceived less prominent than after an unaltered phrase. Due to this effect, perceptual compensation caused the listeners’ boundary to shift in a corresponding direction to the manipulation of the carrier phrase. In other words, when the carrier phrase was stripped of the /ɪ/-specific frequencies, some of the testing words that would without manipulation be identified as [ɛtʃ] were now identified as [ɪtʃ] since the listeners’ perception tuned to the carrier phrase and considered the manipulation a distorting element.

Based on further experiments in the same study (Watkins, 1991), varying in form of exposition to the carrier phrases and the nature of the phrases themselves, Watkins concluded that the mechanisms in question are susceptible only to low-rate variation changes in the sound signal, which suggests that they do not account for changes in sound source but only in channel. Their use in speech perception is, therefore, that of compensation for spectral envelope distortions caused by reverberations and surrounding noise. This is supported by his findings, for example, that perceptual compensation occurs even with non-speech precursors and its effects disappear when the carrier phrase and the testing word are presented binaurally, each to a different ear. This suggests that each stimulus is coming from a different place; therefore the sources have a different acoustic profile.

Apart from the findings mentioned above, Watkins' experiments did not find any significant contribution of several other perceptual phenomena identified by previous research, such as lateral inhibition (Crowder 1978; 1981), negative auditory after-image effect (Wilson 1970) and contrast effect (Fox 1985). The influence of these phenomena of auditory perception is generally too transient to be significant. The effects are, however, still measurable and their presence illustrates the multifaceted nature of the subject matter.

In a follow-up study, Watkins and Makin (1994) carried out another series of experiments using a similar methodology to study the relation of perceptual compensation for channel distortions to perceptual adaptation which is speaker-specific. Before proceeding to their findings, the concept of formant frequencies and their manipulation should be introduced.

Formant frequencies, labeled F1, F2, etc., are those resonant frequencies which are being amplified in the vocal tract during phonation. The range of their value is determined by the size of a person's vocal tract (Peterson and Barney 1952), while their relative values to the fundamental frequency (perceived as the pitch of the voice) change with the movements of the articulators, predominantly the tongue and lips. The first formant changes are determined mainly by the size of the pharyngeal cavity and the second formant by that of the mouth cavity. Because the tongue practically separates these two cavities, its movement affects the first two formants, as well as the other ones, at the same time. Especially the second and the third formant are also modulated by lip movement. The size and the physiological configuration as well as the movement and position of a speaker's articulators are thus reflected in relative formant values, which are considered to convey information both about linguistic units such as speech sound identity, as well as indexical information, such as the speaker's accent and personal characteristics (Ladefoged and Broadbent 1957, 103). Due to the correlation between the manner of articulation and formant values, formant frequencies are not only carriers of information about vowel quality, but they also define it, meaning that they cannot provide misleading information and their alteration actually changes the perceivable quality of the vowel. On this matter, it is notable that Ladefoged and Broadbent in their classic study (1957) showed that perceived vowel quality also depends on its preceding phonetic context and changes in average formant ranges can affect the perception of subsequent vowels. Since formants are not only descriptive but defining, this means that the contextual effects concern the way speech sounds are processed. In their experiment, Ladefoged and Broadbent presented listeners with a manipulated carrier phrase “Please, say what this word is” followed by a test word of the “b_(vowel)_t” structure. The range of F1, F2 or both formants was either raised or lowered across the whole phrase. Generally, men’s vocal tract is of larger proportions than that of women or children, and the absolute formant values of the speech of men, women, and children vary from the lowest to the highest, respectively. We can thus compare such change to reduction or increase in the size of the speaker’s vocal tract or a similar change affecting the speaker’s production of formants as a whole. As for the testing words, this study did not use a two-vowel continuum, but a set of vowels /ɪ/, /ɛ/, /æ/, and /ʌ/, therefore no perceptual boundaries were measured. They found that raising a formant’s frequency caused the testing word to be perceived as having lower formants and vice versa. That is to say that, one and the same sound was more likely perceived as /ɪ/ when presented after the carrier phrase with lowered frequencies and as /ɛ/ when presented after the carrier phrase with raised frequencies. Formants in a given vowel thus determine the perceived quality of that vowel and may as well affect the perceived quality of sounds in the following (Ladefoged and Broadbent 1957) and preceding (Johnson 1991) speech. The latter study observed differences in categorizing tokens from a synthetic [s]-[ʃ] continuum depending on the gender-associated average formant frequencies of the following naturally spoken vowel. Listeners were more likely to perceive the tokens as /s/, that is, with comparably lower frequencies, when heard together with a female voice and vice versa. This relationship can also be explained employing the abovementioned enhancement effect. Based on these two roles formant frequencies play in speech perception, we may distinguish between intrinsic and extrinsic factors (Ainsworth 1975, 103). Extrinsic factors are such perceivable sound qualities which are situated **outside** the segments whose perception they influence, while intrinsic factors lie **within**.

The study of Watkins and Makin mentioned earlier (1994) replicated their previous results and additionally used manipulation of F1 frequency as an extrinsic factor to study the relation of Ladefoged and Broadbent’s findings to the perceptual compensation described in their former study (Watkins 1991). Since both of the studies worked with static or low-rate variation changes, Watkins and Makin hypothesized that the critical source of information which listeners utilize to adapt to such changes is the long-term average spectrum (LTAS) of the speaker’s voice. They found that compensation for LTAS does not require the carrier phrase to follow the “constraints of natural utterances” (Watkins and Makin 1994, 1264) and significant shifts in perception occur even with the carriers played backward. It was concluded that knowledge of the LTAS of a given sound source does not wholly explain speaker compensation (i. e. adaptation to speaker-specific characteristics) and both the findings thereof and of Ladefoged and Broadbent (1957) were mainly due to distortions related to the transmission channel. Even though the latter study (1957) showed a relation between average formant values indicative of vocal tract size and speaker-specific adaptation, inter-speaker variability encompasses more variables than this; therefore other mechanisms must be involved.

Findings on perceptual compensation show presence of auditory mechanisms responsible for **immediate adaptation** effects that occur regardless of whether the acoustic input is perceived as speech or not. The likely role of these mechanisms is to provide for an identical interpretation of one and the same sound in different acoustic environments. This compensation for external factors was, to some extent, observed to account also for speaker-specific variability such as vocal tract size. Such speaker characteristics are presumably affected by the mechanisms in question due to the static influence they have on the sound quality, which is similar to that of external factors. Perceptual adaptation to more complex differences, for example, different accents, perhaps requires higher-level processing using also different kind of information than the direct acoustic input.

1.4 Speaker normalization and perceptual learning

The focus on speaker-specific differences in speech production introduces another subfield of research into perceptual adaptability termed speaker normalization. The subject of studies carried out around this concept may be divided into two categories based on the type of information used by listeners as well as the degree of its processing needed. One category is auditory-oriented, focusing on interspeaker differences resulting from anatomical variation. Normalization for these effects on vowel quality has been studied, for example, by Hillenbrand et al. (1995) and Nearey (1989). The second category covers mechanisms, sometimes called *phonetic*, which additionally employ other contextual information needed for adaptation to accent variation and idiosyncratic speaker characteristics. The term "normalization" is being used for both of these categories. The latter is also often referred to as **perceptual learning**, because adaptation to speech with systematic context-dependent deviations from a listener's speech (e. g. producing a different allophone only in open syllables), needs to be backed up by experience.

1.4.1 Auditory-oriented normalization

Auditory-oriented speaker normalization follows up on the concept of long-term average spectrum mentioned in the previous section. The perceptual mechanisms of adapting to the LTAS were later tested in 2011 by Sjerps, Mitterer, and McQueen who found that, although not speech exclusive, the adaptation driven by these mechanisms is determined by other factors than LTAS. They found that short-term adaptation may instead depend on variation in amplitude and fundamental frequency. When these characteristics are disturbed, effects are significantly smaller, for speech sounds, or absent, for non-speech sounds. Moreover, adaptation was observed independently of the listener's conscious view of the sound's "speechiness" which suggests that the mechanisms in question operate on an auditory level. Sjerps et al., in the same study (2011), suggest that the crucial factor for normalization is learning of the spectrotemporal characteristics of speech.

> Because adult listeners have had an abundance of exposure to speech from different speakers, they will gain experience with the fact that certain voice properties are stable within a speaker. They could therefore learn that it is beneficial to perceive vowels relative to those voice properties. This can be achieved if listeners learn to normalize for the LTAS properties of preceding sound sequences. (Sjerps, Mitterer, and McQueen 2011, 1196)

Sjerps et al. (2011) further argue that this mechanism is general and could be trained to apply for non-speech input as well. They also claim that those precursors (heard utterances preceding a target word) for which no adaptation effects were found sounded very unnatural and therefore listeners could not have been already accustomed to their characteristics. As the findings of Stilp, Kiefte, and Kluender (2010) show, LTAS normalization undoubtedly works on the basis of general auditory mechanisms and does also apply to non-speech sounds, since significant alterations in perception were also achieved by exposure to the sounds of musical instruments. Several questions remain open. First, what is the precise nature of the information listeners need to acquire to normalize for a particular sound source characteristics, and second, what is the relation to and possible interactions with a higher-level, not only auditory-oriented adaptation.

1.4.2 Phonetic normalization

The second part of this branch of research views speaker normalization as a process during which listeners filter out individual speaker-specific factors and "normalize" the speech signal to a universal form that is mapped onto linguistic units. Based on this assumption, several studies have proposed different vowel normalization algorithms proceeding from either intrinsic or extrinsic information, i.e., information contained within, or outside, the vocalic segment to be identified (Gerstman 1968; Lobanov 1971; Sundberg and Nordström 1976; Nearey 1978; Syrdal and Gopal 1986; Miller 1989). These algorithms are successful at removing the imprint of personal speaker-specific information, related mainly to vocal tract size, on the quality of a vowel. Yet, this imitation of vowel processing yields a normalized speech signal where dialectical differences are preserved, which is not the case in the presumed normalization listeners perform.

A likely explanation is that, unlike physiologically related speech characteristics, dialectical variation and individual idiosyncrasies cannot be inferred solely from sampled auditory information because they do not affect the phonemic system of an individual as a whole. This is perhaps related to the fact that, while some speakers may be physically predisposed to certain distinctive pronunciations (e. g. by a vertical overlap), the nature of dialectical and idiosyncratic speech variation stems from the acquired principles of each individual's speech production. It is a widely held view that speech development in childhood is governed by the child's social and linguistic environment (Brown 2000, 42; Goldstein, King, and West 2003; Hoff 2003) which makes characteristics of speech production at least partially independent of the particular individual. A given person's pronunciations thus differ considerably with the accent area he or she is born into. This calls for other contextual clues to assist speaker normalization.

An extensive line of research on consonants started in 2003 with a study by Norris, McQueen, and Cutler, aiming at the role of lexical knowledge during adaptation. In their broadly used experimental paradigm, Dutch listeners were exposed to an ambiguous fricative with sound qualities intermediate between the phonemes /s/ and /f/. Here, it will be represented as [s^f]. One group of listeners heard the [s^f] in words ending with /s/ together with words ending with /f/ sounding as a canonical [f]. The other group was presented with a canonical [s] sound in /s/-ending words and with the ambiguous [s^f] in /f/-ending words. Subsequent categorization of tokens from a [s]~[f] continuum showed that listeners adapted their perception of the ambiguous sound based on the context it was heard in. Those who heard the [s^f] sound in /f/-ending words were more likely to categorize ambiguous tokens from the continuum as /f/ and vice versa. Importantly, the ambiguous sounds in the continuum were not identical to the [s^f] from the exposure phase. Listeners, therefore, generalized their adaptation to other /s/ or /f/ sounds from that speaker. A likely explanation of these results, which would not require the use of lexical knowledge, is that listeners assumed the ambiguous fricative to be the one whose canonical form was missing. Each group of participants was, in fact, always exposed to one canonical form and one ambiguous. However, no adaptation occurred when the same experiment was carried out with non-words. This confirms that **listeners indeed use lexical clues to adapt to non-canonical speech sounds**. Note that these findings are not conclusive regarding the question whether lexical clues are the only possible source of necessary information.

1.4.2.1 Speaker (in)dependent normalization

The same experimental paradigm was later adopted by Kraljic and Samuel (2006) who also found that adaptation generalizes to other instances of the tested phoneme. Furthermore, their findings did not only show adaptation to an accent of a specific talker but generalization across different speakers as well. Such results were presented before in a study by Bradlow and Bent (2003) on Chinese accented English.

In this study, listeners were first exposed to a Chinese-accented set of English sentences either from a single speaker or a group of five different speakers. After this "training session," participants were asked to transcribe another set of simple English sentences mixed with white noise on a recording of the same or a different Chinese-accented talker. Importantly, listeners exposed to a different talker in the post-test showed better performance when trained on multiple talkers. Their comprehension was equal to that of the group exposed to the same talker in both the training and the post-test phase which implies that they were able to generalize the accent across different speakers. “Because of the nature of Bradlow and Bent’s task, however, the type of information extracted (e.g., featural, segmental, prosodic, or rhythmic) cannot be determined” (Eisner and McQueen 2005, 225).

While both Bradlow and Bent (2003) and Kraljic and Samuel (2006) found generalization across different speakers, Bradlow and Bent also observed a necessary condition of sufficient exposition to multiple talkers. Otherwise, listeners in their study did not achieve accent abstraction. Naturally, exposition to a single talker cannot trigger generalization since the non-canonical speech presented could stem from mere speaker idiosyncrasy. Still, even when a listener is assured of a collective accent, sheer talker-independent adaptation may prove detrimental. As there is no guarantee of encountering only talkers sharing common accent, *a priori* adaptation could result in unintelligibility. Evans and Iverson (2004) indeed found constraints on speech interpretation posed by learned accent characteristics.

For that reason, adaptation presumably does not always generalize for accent. This argument is favored by Eisner and McQueen’s study from 2005 which experimented with talker variation for both exposition and testing of the listeners. Subjects were again presented with the paradigm with fricatives and the [ɛf]~[ɛs] continuum that had been used by Norris, McQueen, and Cutler (2003), but this time they treated ambiguous sounds as an idiosyncratic characteristic and retained adaptation only for the talker heard in the exposure phase. Although the exposition here was limited to a single talker, adaptation occurred when that talker’s tokens were spliced with a vowel of a different one. This indicates that a mere presence of another voice does not inhibit generalization. Even so, note that Eisner and McQueen did not observe adaptation for other talker’s fricatives. An explanation for these seemingly contradictory results is offered by Kraljic and Samuel (2006). Their study featured exposition in the form of a lexical-decision task (categorization of words against non-words) followed by categorization of tokens from a [d]~[t] or [b]~[p] continuum. Their use of stops instead of fricatives is a crucial element here. After they found generalization of lexically induced adaptation to a different talker, they argued that the contrast of their results with those found in Eisner and McQueen (2005) points to perceptual learning mechanisms being flexible and variable with the ambiguous input. Notably, adaptation in their study generalized from the manipulated [d]~[t] to the unexposed stops [b]~[p] as well. According to their interpretation, perceptual learning generalizes at a featural level, rather than phonemic, and because the defining features for stop consonants are less variable between speakers than those of fricatives, generalization of the adapted perception of a voicing contrast among stops of a different manner of articulation is observable even after a single-talker exposure. In their earlier study, Kraljic and Samuel had made the same proposition but spoke of **acoustic** cues such as duration of aspiration in stops and spectral envelope in fricatives rather than abstract phonological features (2005, 171). The distinction between abstract phonological features and acoustic cues to these linguistic units carries implications for vowel perception and is discussed below in section 1.8.2 of this study.

Concerning perceptual adaptation to vowels, the 21st century's research on consonants is only weekly paralleled, even though "dialect differences are often carried by vowels" (McQueen and Mitterer 2005, 233). This imbalance might be caused by the fact that "vowel acoustic shape varies substantially with phonetic context, unlike the reasonably stable characteristics of fricatives and stops used previously" (McQueen and Mitterer 2005, 233). This difference probably stands behind the greater variety in experimental paradigms used in the research of adaptation in the perception of vowels compared to the research on consonants. Nevertheless, several findings are in accordance. For example, listeners' general tuning into a vowel-manipulated accent was achieved by Maye, Aslin, and Tanenhaus (2008).

Similarly to the studies of lexically guided adaptation in consonants, this study focused on medium-term adaptation effects measured in a post-test paradigm. In contrast to the research on consonants, listeners were here presented with a coherent piece of text rather than segmented exposure. The authors' intentions with this experiment were to explicitly address the difference between the dialectical variation affecting the whole speaker's formant production on one side, and more particular characteristics on the other (this comparison was discussed under 1.4.2). As the authors suggest:

> Although dialectal variations can involve wholesale shifts in the entire vowel space (i. e., all vowels are remapped to some extent, and the remapping is applied to all lexical items), this need not be the case. Adaptation could be applied only to specific regions of the vowel space, to a specific subset of vowels within or across different regions of the vowel space, or to specific lexical items that have been encountered by a listener exposed to the novel dialect. (Maye et al. 2008, 546)

On that account, the vowel-manipulation in this study concerned only front vowels. These were systematically changed for lower tokens in the vowels space so that words containing the vowel /ɪ/ were pronounced with [ɛ] and those with /ɛ/ as [æ]. Adaptation was tested in a following lexical-decision task. Note that use of this task in previous experimental paradigms was limited to the exposure phase. Systematic shifts in front vowel representations were indicated by listeners identifying of word forms with vowels altered according to the artificial accent, normally judged nonsensical, as existent words (e. g. *witch* [wɪtʃ] altered to *wetch* [wɛtʃ]). Interestingly, listeners' adaptation to the manipulated tokens did not replace the original phoneme representations as the unaccented word forms were still considered real words (*witch* [wɪtʃ] was not perceived as weech [witʃ]). This suggests that listeners adopted the artificial accent in the form of new allophones which were used side by side with the original ones according to the accent recognized. In a subsequent experiment with an opposite direction of the vowel manipulation, Maye et al. (2008) found evidence that "the adaptation effect was specific to the direction of the shift in the vowel space and not to a general relaxation of the criterion for what constitutes a good exemplar of the accented vowel category" (Maye et al. 2008, 543).

Other findings of the research on consonants were replicated for vowels by Skoruppa and Peperkamp (2011), who also found a correlation between adaptation and phonological features. They exposed listeners to three artificial accents of French, each of which was distinctive in the distribution of the feature lip rounding. In all three accents, this feature was treated as binary, i. e., either present or absent. The critical vowels in the exposure to the different accents all showed relation to the roundedness of its preceding vowels; having either the same or the opposite value. Critically, this relation was common to all of the tested vowels in the first two accents, and mixed in the third one; some vowels in the third accent were rounded after rounded vowels, some after unrounded. Skoruppa and Peperkamp (2011) hypothesized that if listeners perceived the characteristics of the accent for each vowel separately, learning of the accent would be equally demanding in all three cases. On the other hand, if learning were taking place on a featural level, the first two accents would be acquired more successfully than the one with mixed rounding distribution. This is precisely what the authors found when they tested their listeners for recognition of the pre-exposed accented words.

Similarly to Kraljic and Samuel (2005), acoustic nature of the properties guiding perception was also proposed for vowels by Llompart and Reinisch (2018). They found experimental evidence for this proposition in their recent study (Llompart, Reinisch 2018); see 1.8.2 for their findings.

To the author's best knowledge, there have been only two studies adopting the often mentioned paradigm proposed by Norris, McQueen, and Cutler (2003) and applying it also to vowel-related adaptation. The first one was carried out by McQueen and Mitterer in 2005, and its results, albeit confirming adaptation, left the question of generalization for other phonological categories somewhat unsettled. This problem will be dealt with in a later part of the current study in the context of more contemporary research. The second study was conducted by Chládková et al. (2017) and found other than lexically guided adaptation which is further discussed below in 1.6. Findings on generalization from the same study are discussed in 1.8.3.

A different experimental approach was presented in a study carried out by Evans and Iverson (2004), which found an influence of accented input on subsequent vowel perception as well as constraints on adaptation posed by the listeners' dialect. The subjects in this study were presented with synthetic vowels embedded in carrier phrases with either the same or contrasting accent to that of the listeners. Their task was to compare the vowel heard to another one, which was presented visually, and give a rating of similarity. Based on this rating, a four-dimensional acoustic algorithm adjusted the first three formants and duration to change the vowel's quality accordingly. The evaluation continued to the point when the algorithm had approximated the vowel suggested by the orthographic label as much as possible. Crucially, this exemplar was not presented in an auditory form, and the respondents thus had to rely on their own conception of that vowel's quality to zoom on into its prototypical sound. This implies the presence of abstract phoneme sound quality representations, to which newly arriving acoustic input is being compared. The fact that the subjects of Evans and Iverson's experiment reached different "perfect vowel examples" based on the accent in which these were embedded implies that their model phonemes had shifted. This argument suggests that there is a different approach to adaptation, which was, in fact, adopted in parallel to the concept of normalization among some related studies (e. g., McQueen, Cutler, and Norris 2006; Kraljic and Samuel 2006). The underlying mechanisms have been considered to behave as a computational unit, applying normalization algorithms to extract linguistic and other information (Johnson 2005). Adaptation to different speakers would be then achieved by virtue of creating different algorithms. The second approach explains normalization as dynamic adjustment of the abovementioned mental representations of phonemes based on acoustic and contextual information.

1.5 Normalization vs. representation account of speaker-specific adaptation

These two approaches to the process of speaker-specific adaptation were confronted in a study carried out by Dahan, Drucker and Scarborough (2008), which explored whether adaptation systematically affects also perception of sounds unaffected by the speaker's accent. Note that this is not the same concept as generalization, because here knowledge acquired by adaptation is used in recognition of familiar sounds, although their perception is not being adjusted. For this purpose, they imitated a feature found in some American dialects, in which the vowel /æ/ is pronounced with a raised quality [æ̝], approaching the vowel /ɛ/, before the consonant /g/ but not before the voiceless counterpart /k/ (Dahan et al. 2008, 711). In these dialects, the distinction between words such as *bag* and *back* is enhanced, which should make them easier to distinguish. This hypothesis builds on two premises. First, words are initially perceived in a bottom-up fashion and later are confronted with stored information from the mental lexicon, following principles of the cohort model. This model describes a concept of "earliness of spoken word recognition" (Marslen-Wilson 1987, 73–76). That is, the selection of words in the mental lexicon narrows down with the perception of consecutive sounds. For example, after hearing the sounds /fi/, the range of possible forms for that given word in a listener's mental lexicon narrows down to *fee feel*, *feature*, *feedback* and such, excluding all words differing in the first two sounds. The predictions of this model into practice have been observed by an abundant research; see Marslen-Wilson (1987) for a review. The second premise is that listeners can learn the characteristics of an unfamiliar accent and, as illustrated above, apply them on new speech input.

Listeners were divided into two groups, each of which was performing a forced-choice task, pairing an auditory word input with four written words on a computer screen. One group was hearing the standard vowel /æ/ in both contexts, while the other had the vowel in /g/-final words replaced with the raised [æ̝] variant. Their decisions were monitored by tracing the movement of the cursor, which they used to select their option, and by tracking of eye fixations during perceptual processing. This information put on a time scale served to assess the word-recognition time. The results showed that the dialect-induced enhancement of the *bag/back* contrast not only increased the overall success rate of word recognition but also facilitated the recognition of the /k/-final words. Since the vowel before /k/ in those words corresponded to the listeners' perceptual standards, this facilitation gives evidence for systematic use of perceptual adaptation and contrary to the algorithmic nature associated with normalization. Normalization algorithms would not affect /k/-final words since no adjustment for their perception was needed. The representation account for perceptual adaptation (Dahan et al. 2008), however, offers a convincing explanation. After the initial exposition to the raised vowel token, listeners adjusted their phonemic representations according to the lexical clues. As part of the accent abstraction process, they used additional contextual knowledge to form a raised [æ̝] allophone of the phoneme /æ/ and stored it in the mental lexicon together with its phonological context; before the phoneme /g/. Correspondingly to the cohort model of word recognition, perceiving the raised allophone ruled out the /k/-final words and led to a faster distinction between the proposed options. This suggests that perceptual adaptation does not work through normalization algorithms and involves higher-level cognitive processing, which enables listeners to alter their categorization even for **sounds directly unaffected by the exposed accent**.

1.6 Role of the lexicon in perceptual adaptation

All the studies discussed so far were dealing either with a solely auditory-based or a lexically guided kind of perceptual adaptation. One may get the impression of the dichotomy between adaptation to personal speaker information or environment distortions on one side and adaptation to accent or speech idiosyncrasies on the other, with lexical information being the decisive factor between these two. This might be the case because of the popularity of the experimental paradigm relying on lexical guidance introduced by Norris, McQueen, and Cutler (2003). Or it may simply come natural to perceive this importance of lexicon since there are more or less non-linguistic characteristics on the one side and linguistic ones on the other. And what makes a better distinction between linguistic and non-linguistic sounds than meaning. The influence of lexical information, however, does not seem so prominent at all, as numerous studies found effects of adaptation to meaningless sounds as well.

For example, Mitterer (2006) in his study focusing on perception of native vowels by Dutch listeners showed significant effects of perceptual adaptation with words as well as non-words. For his experiments, he used the carrier phrase followed by a testing word model. In the first experiment, listeners were exposed to an across the board manipulation of F2 range in a meaningful or a meaningless carrier phrase containing vowels with a wide range of frequencies, /u, i, ɑ, ɛ/. The testing words/non-words contained tokens from a continuum of front to back mid vowels. Importantly, all three testing vowels had similar F1 values and differed mainly in F2. Corresponding to perceptual adaptation effects, listeners were more likely to identify a test vowel as front (i. e. with high F2) after exposure to the lowered F2 carriers.

Mitterer concluded that short-term perceptual adaptation was independent of lexical processing because non-word carriers had the same effect as the real-word ones. However, it should be noted that manipulation of a formant across all the vowels in the exposure phase may be perceptually comparable to a long-term average spectrum change, which is associated with perceptual compensation, or in other words, auditory-based adaptation (Mitterer 2011). As he also noted, "manipulation of F2 range in the current experiment more or less resembles anatomically grounded speaker variation" (Mitterer 2006, 225).

A different experiment in the same study was conducted to test whether the observed adaptation effects can be explained by purely acoustic principles; a concept of "acoustic history" (Holt 2005) similar to the effects of LTAS. Such findings would in a sense unify the auditory and phonetic categories under a single mechanism possibly encompassing all the results of previous research. To test for the acoustic history effect, Mitterer used LTA spectra and amplitude envelopes of the carriers used in his first experiment applied on stretches of white noise. Otherwise, the same experimental design was used. This experiment failed to find significant perceptual boundary shifts in listeners. These results imply that short-term adaptation is not based on auditory processes alone and is probably limited by a certain threshold of similarity to speech. This threshold may be connected to the aspect of learning (Mitterer 2011) discussed above in the section 1.4.1. However, one major drawback of the experiments in Mitterer (2006) is the small number of participants (8 and 15 for the mentioned experiments, respectively) which makes the statistical significance of the results questionable.

Studies on consonants have shown that lexical status of words influences perception changes (e.g., Norris et al. 2003), which indicates that the mechanisms in question may operate on multiple cognitive levels; this time, the decisive aspect was not only auditory information but also lexical knowledge. The fact that this higher-level cognitive processing in adaptation is not exclusive to the mental lexicon can be illustrated by replaceability of the lexicon's influence. Several studies have given evidence of perceptual adaptation acting on a multisensory level (e.g., Vroomen, van Linden, Keetels, de Gelder, and Bertelson 2004). In fact, the term itself is in cognitive psychology also used more generally to refer to an adaptation of senses. We can thus speak of, for example, visual adaptation, which has also been a subject of research (Harris 1965, 419). Auditory perceptual adaptation has been found to be influenced by not only lexical, but also visual and even conceptual clues (Bertelson, Vroomen, and de Gelder 2003; Johnson, Strand, and D'Imperio 1999, respectively).

Studies on visually-induced auditory adaptation work with the so-called McGurk effect (Bertelson et al. 2003; Vroomen et al. 2004). McGurk effect is a relatively widely known concept of a lip-reading contribution to auditory perception. Its principle lies in the interference between auditory and visual input typically illustrated on an unambiguous sound /b/ superimposed on a silent video-recording of a speaker pronouncing the sound /g/. This kind of stimulus is ultimately perceived as the sound /d/; intermediate between /b/ and /g/ in terms of the manner of articulation. The studies mentioned at the beginning of this paragraph found that McGurk effect may also provide guidance for perceptual adaptation. When a video recording of a speaker pronouncing the syllable /aba/ was played together with the **ambiguous** vowel-stop-vowel auditory sequence [a?a], with the consonant intermediate between /b/ and /d/, listeners perceived the ambiguous sound correspondingly to the gesture pronounced in the video recording. As with lexically guided adaptation, the shift in interpretation of the exposed ambiguity persisted to a unimodal listening categorization task. Note that these studies, among others, use the term "recalibration" to describe perceptual adaptation effects. This term is being used in the same fashion as perceptual learning or normalization.

Other studies have shown that conceptual clues can have the same effects as formant manipulation. Johnson et al. (1999) observed a shift in listeners' perceptual boundary between the vowels [ʌ] and [ʊ] after instructing them to imagine the vowels being spoken by either a male or a female speaker. These results also illustrate that although adaptation to a speaker's vocal tract is considered auditory, there is also a connection to higher level cognitive processing.

Lastly, recent research has brought evidence that effects similar to those of lexically guided perceptual adaptation can be achieved solely on the basis of auditory input. Based on the findings of perceptual adaptation research, we may assume a connection between phonetic categorization and first as well as second language acquisition (Norris et al. 2003, 205; Chládková, Podlipský, Chionidou 2017, 423, respectively). Adaptation can be described as adjustment of the phonemic boundaries for categorization of acoustic characteristics. Language acquisition, on the other hand, requires these phonemic categories to be created. Infants are born with the ability to recognize even those sound contrasts that are not phonemic (meaning distinguishing) in their environmental language (Maye and Gerken 2000, 522). This ability is lost during their first year of life, presumably between the sixth and tenth month (Werker and Tees 1999), as they learn to distinguish only those sounds that differentiate meaning in their native language (Werker and Tees 1984) and ignore variability which is not linguistically significant. As Maye and Gerken put forward in their study (2000), infants learn phonetic categories without the aid of lexical information (i. e. meaning) and retain this ability into adulthood. Experimental results indicate that information about the statistical distribution of different phonemes is sufficient. This is because, despite the great acoustic variability between individual phoneme realizations, "exemplars of a particular phoneme cluster together along one or more acoustic dimensions" (Maye and Gerken 2000, 523). Remarkably, it seems that listeners make use of statistical distribution also during perceptual adaptation to an unfamiliar sound (Chládková et al. 2017, 423).

There are several differences between lexically guided adaptation and the use of statistical distribution that should be accounted for. Remember that Norris et al. (2003) found no effects with non-words as well as no contribution of statistical distribution to adaptation specifically. Even when omitting the fact that their study focused on consonants, there is an explanation for the difference in results. The discussion in Norris et al. mentions the following:

> Learning took place when exposure to those fricatives was limited to 20 words, spread over a list of 100 words and 100 nonwords. The rapid learning seen in these studies stands in contrast to the very much slower process of learning to perceive the speech either of one's native language, or of a second language. (Norris et al. 2003, 227)

As the wording implies, their study possibly did not comprise a sufficient number of trials and linguistic variability in the exposure phase so that listeners could make use of statistical distribution of the ambiguous sound. Lexically guided perceptual adaptation thus may operate on the same basis as that using statistical distribution, only at much faster pace thanks to the amount of information available.

Numerous studies on perceptual adaptation and the findings thereof point to the interpretation that its purpose is to facilitate, or in fact enable, communication despite the lack of invariance. There is, however, also a different kind of perceptual phenomenon, selective adaptation, which appears with a striking similarity while producing inverse perceptual adjustments. Scientific publication on this subject has been initiated by a study of Eimas and Corbit (1973) who coined the term selective adaptation. The following section will address its seeming contradiction to the mechanisms and effects described in this paper.

1.7 Selective adaptation

Since being described for the first time in the year 1973, selective adaptation has enjoyed great attention. This perceptual phenomenon can be described as acting on the established phonetic categories in a fatigue fashion. In other words, exposition to a token from a certain contrasting pair of sounds "fatigues" the receptors of that particular sound which leads to reduction of its perception on a one-to-one continuum. In practice, listeners exposed repetitively to a voiced stop [b] will perceive more voiceless [p] tokens on a subsequent categorization from a [b]~[p] continuum and vice versa (Eimas and Corbit 1973). These results were attributed to effects on "linguistic feature detectors" (Eimas and Corbit 1973, 99).

Later studies, however, proposed that selective adaptation is entirely auditory-based, suggesting that there is no relation to the speech-specific higher-level cognitive processing. Roberts and Summerfield (1981) in their study based on the abovementioned McGurk effect created a conflict of auditory and visual information by superimposing a pronunciation of the syllable [be] on a video recording of a speaker pronouncing [ge]. They followed up on this exposition with a categorization task from a [be]~[de] continuum. Since the audio-visual conflict changed the perceptual quality of the consonant from [b] to [d], the exposition should lead to a predominant reporting of /b/. If, however, selective adaptation operated merely on an auditory basis, the perceptual adjustment due to the McGurk effect should affect it and, in fact, fewer [b] tokens should be reported. This is exactly what Roberts and Summerfield found.

Nevertheless, considering the multi-level operations of perceptual adaptation (i. e. perceptual learning), one may argue about how interdependent are the individual kinds of information (auditory, lexical, visual) as regards their processing. If we assumed that there was no hierarchical distinction between individual perceptual cues and auditory and visual information was processed on the same level, we would have to consider that the audio-visual contrast could create an impression of ambiguity. In this case, the ultimate percept of the two opposing cues, perceived with equal importance, would result in an input halfway between the phoneme /b/ and /g/, resembling /d/ in quality. Due to this resemblance, listeners might **learn** to assign this percept to the phoneme /d/ and categorize it accordingly even with one of the initial cues missing. In such case, it would also be possible to explain the observed results by the operation of perceptual learning. Without prior conclusion about how information from different sources is processed, it cannot be determined which mechanisms (selective adaptation or perceptual learning) are accountable. Therefore, no conclusions about the nature of selective adaptation can be inferred with confidence.

The difference between selective adaptation and perceptual learning is partially reconciled in a further study by Vroomen et al. (2004) which found both selective adaptation and perceptual learning effects, concluding that the (un)ambiguity of the stimulus is the distinguishing variable. While perceptual learning occurs after exposition to an **ambiguous** stimulus and leads to increase in its identification, selective adaptation acts on **unambiguous** stimuli in an inverse manner. Even though auditory nature of selective adaptation was not confirmed, different experiments still suggest it (Simon and Studdert-Kennedy 1978, 100).

Kraljic and Samuel (2005, 168–170) provide a comprehensive discussion of the specifics of selective adaptation as compared to perceptual learning. Among major differences between the processes of these two phenomena is stated the engagement period. Perceptual changes due to learning are relatively abrupt, while selective adaptation requires longer exposition. Notably, despite the slower onset of selective adaptation, its effects with longer exposure time exceed the degree of influence on perception that perceptual learning has. However, each of these phenomena requires a different kind of input; ambiguous vs. unambiguous. A peculiar combination is proposed by Vroomen et al. (2004). They suggest that a new representation of a phoneme, which is created due to perceptual learning, may thereafter function as a canonical representation and be subject to the effects of selective adaptation. This account corresponds to the later dominance of the changes induced by selective adaptation. Lastly, the same study also observed that the effects of selective adaptation are longer-lasting than those of perceptual learning.

From the characteristics above an account of the mechanisms of selective adaptation can be inferred by relating it to the enhancement effect discussed in 1.3 (Kiefte and Kluender 2005) and the concept of the long-term average spectrum and statistical distribution. If particular sounds are categorized on the basis of clustering of their statistical distribution, repetitive exposition to a specific stimulus decreases the variance and makes the average more robust. As a result, the cluster of the exposed phoneme tightens, and the surrounding categories expand. Consequently, the perceived relative contrast to the tokens from the subsequently presented continuum increases which makes these more resembling the other endpoint. This presumed nature of the mechanisms behind selective adaptation would not require it to be speech-specific.

Overall, we can say that "selective adaptation paradigm can be a powerful tool for investigating the perception of complex acoustic stimuli like speech" (Samuel 1986, 452) but the possibility of its interference with the effects of perceptual adaptation should not be omitted.

1.8 Generalization of perceptual adaptation across the phonemic inventory

As was implied in the section 1.4.2.1 above, generalization of perceptual adaptation does not concern solely the speaker level but also the phonological level as perceptual adjustments can spread across other sounds. This kind of phonemic generalization will be discussed hereon.

1.8.1 Generalization of perceptual adaptation for consonants

When Kraljic and Samuel (2006) found that the shift in categorization of [t]~[d] induced by an ambiguous [t^{d}] consonant also affected the categorization of [p]~[b], they assumed that generalization occurs for phonologically related contrasts; voiced/voiceless distinction is characterised by the difference in the voice onset time (VOT) feature. This conclusion can also be drawn from the results of Reinisch, Wozny, Mitterer, and Holt. (2014) and Reinisch and Mitterer (2016). These studies carried out on consonants showed that generalization presumably concerns only phonologically related contrasts and is specific to intrinsic features, as no effects were found, respectively, for manner of articulation or place of articulation, where the surrounding vowel formant transitions are the guiding features.

1.8.2 Contrasting features of phoneme recognition

The nature of the contrasting features guiding generalization was not, however, further examined. These can be approached simply as acoustic cues or as more abstract phonological units. Perception based on purely acoustic information is foundational for so-called episodic or exemplar models of perception; proposed in Goldinger (1998). In brief terms, these state that every instance of perceived words is stored in memory in its whole auditory quality. Phonetic categorization then requires no abstract units such as phonemes because individual speech sounds are recognized simply by mapping of the signal onto the stored information and finding the closest approximation. Each sound's representation is thus virtually an average of all its instances perceived so far. This episodic view differs from the representational account in that it assumes greater demand on storage than on computation during language processing. This opposition has been a subject of discussion not only on the field of phonetics but also among theoreticians of morphology and lexicology. There is, nonetheless, empirical evidence against purely episodic models, for example, the mere fact that perceptual adaptation generalizes to other sounds, and the nature of an applicable speech perception model is thus often proposed to be hybrid at most (Mitterer 2006, 227). One possibility for a hybrid model further proposed by Goldinger (1998) actually admits speaker normalization. The necessity of some degree of processing, as opposed to direct storage, is implied in that various characteristics inferable from a speech signal are all encoded by the same variables, i. e., formants. This hybrid model is proposed to operate by storing procedural information of speaker normalization together with information this processes yield.

On the other hand, representation models also do not require an intermediate abstract form of features. In fact, phonetic categorization could act as direct match-making between acoustic cues and phonemes. On the other hand, results of Chládková, Boersma and Benders (2015) suggest that the abstract units guiding perception are rather individual features than their combinations represented in phonemes. In their study on Czech, categorization of vowels (differing in F1 values) was performed on two artificial listeners. The two listening simulations assessed vowel quality either on the basis of F1 or F1 **and** F2. Their categorizations were plotted on a vowel space and compared to results of human listeners of Czech with Moravian dialect (for discussion of Moravian and Bohemian dialect, see p. 35). What they found was that as far as the feature vowel height is concerned, human listeners very much ignored the values of F2 and categorized vowel tokens on a featural level rather than on the basis of phonemes. We may speculate that analogical results would be achieved with F2 values for the feature vowel position as well, assuming that “these two dimensions are linked to one perceptual cue each” (Llompart and Reinisch 2018, 19). Is this to be interpreted as that there are in fact no phonemes and the only abstract prelexical units are phonological features? No, not exactly.

First, the concept of abstract features in itself is questionable, taken into account that, for vowels, these features often correlate with single acoustic cues and their assumption may seem redundant. However, there are combined features which make up phonemic contrasts between vowels and are at the same time composed of multiple acoustic cues. If these were treated during perception in the same fashion as the single-cued ones, that would actually favour the assumption of abstract features. This was exactly proposed in a study on German, where the combined feature tenseness establishes phonemic contrast, carried out by Llompart and Reinisch (2018). Tense vowels are generally more peripheral and are associated with more tension in the articulators, hence the term tense vs. lax. The contrast thus lies in spectral difference as well as difference in duration. Moreover, front tense vowel /i/ has a higher F2 than the lax /ɪ/, while /u/ and /o/ have lower F2 frequencies than their lax counterparts, /ʊ/ and /ɔ/ (Llompart and Reinisch 2018, 15). This makes the featural contrast incongruent with the acoustic one. The paradigm of selective adaptation was used, whose effects manifest in a lower identification rate of the sounds heard. If abstract features were the basis for perception, exposition to tense vowels should in a subsequent categorization task result in a larger number of lax vowel responses and vice versa. If, however, perception functions on a basis of acoustic clues, the categorization bias will respect individual formant patterns. The results confirmed the latter, following the patterns of adaptation to exposition to individual acoustic cues as opposed to a compound feature tenseness. However, one cannot deem these results conclusive until the existence of the feature tenseness in German is confirmed in a first place

Second, phonetic categorization is, of course, not limited to perception of a single acoustic cue. Different sounds are characterized by various numbers of cues. Lisker (1986), for example, compiled no less than 16 individual acoustic features defining the voicing contrast. These cues are “weighted”, meaning listeners give them different values based on their reliability for phonetic categorization (Toscano and McMurray

2010, 434). The cue-weighting strategies are adopted by listeners in course of learning from statistical distribution and may change as a function of this distribution (Holt and Lotto 2006, 3059).

According to the propositions above, features are more likely than phonemes to govern perception, and the nature of these features is purely acoustic. Yet, some degree of abstraction is necessary for the adaptation to generalize across sounds. It can be thus hypothesized on whether features are in fact abstracted but only for those acoustic cues which are given the most value, or whether the abstract units are actually phonemes, and information about cue-weighting is stored as their part; only the most weighted cues are playing a role in generalization.

1.8.3 Generalization of perceptual adaptation for vowels

The inconclusiveness of this concept is also reflected in the research on inter-phonemic vowel generalization. For instance, lexically guided adaptation of the /i/ vs. /e/ boundary in a study by McQueen and Mitterer (2005) showed little to no generalization for neighbouring contrast /ɪ/ vs. /ɛ/ and a slightly bigger boundary shift for spectrally dissimilar and farther contrast /ɑ/ vs. /ɔ/. This was an unexpected result since formant frequencies in vowels constitute common and intrinsic features which is the proposed requirement for generalization in consonants (Reinisch and Mitterer 2016). Mitterer's study from 2006 focusing on **immediate** perceptual adjustments found perceptual adaptation effects independent of lexical information; presumably due to the statistical distribution learning or perceptual compensation (see 1.6 above). Results in this study further indicated a correlation between generalization and the variety of exposed vowels. When the exposure phase contained only mid to high front vowels, adaptation concerned only this part of vowel space. Mitterer concluded that adaptation was vowel-dependent. Maye et al. also failed to observe generalization of adaptation effects from front to back vowels.

A later study by Chládková et al. (2017) focused on the same front-back generalization but considered phonological relation as the prerequisite; namely symmetry in vowel space, that is contrasts limited to the difference in only one formant frequency. Significant shifts in perceptual boundaries for the pre-exposed /i/-/e/ contrast as well as generalization to the /u/-/o/ contrast were observed on the symmetrical system of Greek vowels. These results support the hypothesis that generalization of perceptual adaptation, as perception itself, are guided by acoustic cues, and weighting of these cues may play a key role as well.

1.9 Perceptual adaptation in second language listeners

Research on perceptual adaptation in non-native listeners has been so far rather scarce. However, it seems that the same principles as with native language adaptation are followed even though second language perception is much complicated by the interference with the established native language system.

The proposed models of second language perception suggest that novel phonetic categories and contrasts are being assimilated to the established ones (see Flege 1995; Escudero Neyra 2005; Best and Tyler 2007 for a review of Speech Learning Model, Perceptual Assimilation Model [PAM] and Second Language Linguistic Perception model [L2LP], respectively). The resulting phonemic inventory is then likely to resemble a hybrid between the two languages. Apart from number and demarcation of categories, there also may be differences in what acoustic cues listeners rely on during phonetic categorization. Ineffective cue-weighting may then cause difficulties in understanding the foreign language (Holt and Lotto 2006, 3061). One example of such difference in cue-weighting is the perception of the /i/-/ɪ/ contrast by Bohemian and Moravian Czech listeners. While Moravian listeners rely mostly on duration, Bohemians put weight also, if not more, on the quality (i. e. formant frequencies) (Podlipský, Skarnitzl, Volín 2009). This difference carries implications for second language perception. A study by Chládková and Podlipský (2011) identified differences between Bohemian and Moravian listeners in the assimilation of Dutch vowel categories non-existent in Czech which implies different learning paths for speakers of the two dialects. Note, however, that a different study carried out by Iverson and Evans found that there is “a surprising degree of uniformity in the ways that individuals with different language backgrounds perceive second language vowels” (Iverson and Evans 2007, 2842). It is also crucial for understanding second language perception to consider listeners’ experience with the particular non-native language (Flege, Bohn, and Jang 1997). In this regard, the literature distinguishes “naïve” and “experienced” listeners.

As for the efforts to replicate first language perceptual adaptation effects in L2 listeners, notable is the achievement of lexically guided adaptation by Reinisch, Weber, and Mitterer (2013). Their experiment essentially replicated the adaptation for a [s^{f}] ambiguous fricative in /s/-final or /f/-final Dutch words. Germane to L2 perception research, this experiment found significant category retuning also for German listeners. Moreover, the adaptation in Dutch listeners generalized also for the speaker's non-native English, which favors the hybrid nature, or merging, of the two language systems in a listener. However, this study focused on a contrast which is more or less corresponding between the two languages concerned. It is the contrasts which are represented only in one of the two languages that are, according to PAM, assumed to pose the most difficulties to second language learners.

To further inquire into this issue, Drozdova, van Hout, and Scharenborg (2014) followed up on the study on Dutch, this time comparing the contrast /l/-/r/ with English. The important distinction here lies in the discrepancy of /r/ sound representations in the two languages. In Dutch, speakers pronounce the /r/ sound in several varieties, or allophones, depending on the phonological context as well as the speaker's dialect. Critically, the realization most approaching the quality of the English bunched approximant [ɹ], which in Dutch occurs only at the end of a syllable, was not included in the exposition since non-rhotic[1] English dialect was used. Listeners were therefore exposed exclusively to the non-native /r/ category. The exposure phase was designed as a set of testing words containing ambiguous [r^{l}] phonemes embedded in a short story. This paradigm has yielded positive adaptation effects in previous research on vowels (Maye, Aslin, and Tanenhaus 2008). A significant shift in the contrast boundary was observed, and it was therefore concluded that lexically guided perceptual adaptation could affect non-native categories as well.

Lastly, a recent study found evidence that L2 listeners are capable of actively using statistical distribution information to adjust their perceptual strategies. Particularly, Schertz, Cho, Lotto, and Warner (2016) presented English and Korean listeners with sentences containing stop-vowel sequences for voicing categorization on a [p]~[b] scale. Importantly, the stop sounds were manipulated in such way that the primary cue for English speakers, that is VOT, provided conflicting information to the secondary cue: fundamental frequency (F0). As indicated by the results of the categorization task, listeners adjusted the use of their secondary cue accordingly. These results were obtained for non-native speakers as well with the choice of cue for adjustment being determined by their initial cue-weighting strategies.

[1] In non-rhotic English dialects, such as the Received pronunciation, the /r/ sound is not pronounced at the end of a syllable and is replaced by the central vowel /ə/ or /ɜ/ (Ladefoged and Johnson 2011, 94).

In general, the research on L2 listeners has so far found no limitations or dissimilarity related to perceptual adaptation. Introducing a new language system to the picture, of course, brings many new variables to be accounted for and consequently opens many further questions. For example, there is no information on what principles govern generalization of cross-language adjustments within a speaker and what are the possible interactions of adaptations in different language systems. To deal with such questions, a thorough and incontestable model of speech perception might be necessary in a first place.

2 The aim of the present study

This thesis defines three main goals. The first goal was to replicate previous findings of perceptual adaptation for vowels in native English listeners. Secondly, we aimed to fill the gap between research on perceptual adaptation in consonants and vowels which, after attended to in the field of native language, remains unfulfilled in the domain of non-native language perception. For this purpose, we sought to replicate effects of lexically guided perceptual adaptation with Czech listeners using the paradigm of Norris et al. (2003); that is, exposition in the form of a lexical-decision task followed by a categorization task. Note that similar experiment was conducted with native speakers of Dutch in 2005 by McQueen and Mitterer who found significant effects even though inconsistent. We conducted a series of no-supervision individual computer-based offline experiments with Czech listeners with equal experience with English. Participants were divided into two groups whose categorization choices were compared. As the ambiguous vowel, we used manipulated [i] tokens in words pronounced by a native speaker of American English. The manipulation constituted changes in formant frequencies to achieve a more open [i̞] quality (higher F1, lower F2) for one group of participants and a more close [i̝] quality (lower F1, higher F2) for the other to maximize the contrast and consequently the potential boundary shift. For the categorization task, an [i]~[ɪ]~[ɛ] continuum was used to assess both the /i/-/ɪ/ and the /ɪ/-/ɛ/ boundary in the sake of our second hypothesis.

Our second aim was to elucidate the issue of inter-phonemic generalization of vowels in native speakers where there is only little unequivocal evidence (see 1.8.3). Building on the hypothesis that short-term perceptual adjustments contribute to or even drive long-term changes in the whole language system we offer a different explanation for generalization of perceptual adjustments as well as its principles. We suggest that vowel adjustments may generalize based on the principle of contrast maintenance and that under favorable conditions these generalizations follow the path of chain shifts of phoneme representations.

2.1 Diachronic changes in the system of English vowel and connection to perceptual adaptation

Cruttenden (2001, 64–65) identifies four types of sound change based on its nature and origin: internal isolative, internal combinative, external, and changes of length and accentual pattern. For current purposes, only the **internal isolative** changes are of particular interest. It is because, unlike the other mentioned types, these "tend to affect a phoneme in all its occurrences" (Cruttenden 2001, 64) and, by extension, also the whole phonemic system of a language. Furthermore, there is no external factor such as growth in borrowed terms, which could influence both number and distribution of phonemes (2001, 65). This means that an internal mechanism is present. One such is the already mentioned phenomenon of contrast maintenance which for one thing prevents loss of phonemic contrast and for the other keeps the inter-vowel distances uniform. For example, the contrast between the words *sit/set/sat* in some English dialects demonstrates both aspects of this phenomenon. In cases where

> the vowel of *sat* has a closer articulation […] that of *set* must be raised, too […] Alternatively, if the vowel phoneme of *sat* is realized as a front open vowel, as in many English dialects, the vocalic area in which the phoneme of *set* can be realized becomes more extensive; in fact, in those kinds of English where this occurs, the vowel in *set* tends to be open-mid variety. (Cruttenden 2001, 67).

The same principle of contrast maintenance might be in effect during perceptual adaptation. It would at least favor its function of helping successful communication. Research shows that the bigger the vowel space in general and, by extension, the distances between vowels, the more intelligible the speaker is (Bradlow, Torretta, and Pisoni 1996). That implies the motivation for a listener to adjust his vowel space accordingly to the contrast maintenance principle.

Keeping with the mentioned example of contrast maintenance between the front vowels /i/, /ɪ/, /ɛ/, and /æ/, our experiment also aimed to test the push-chain shift hypothesis on these vowels. The decision to choose front vowels was motivated by the fact that they are often subject to diachronic changes of this fashion in different English dialects. Simply put, a chain shift occurs when multiple neighboring vowel categories change their demarcation in a vowel space in the same direction. This process is initiated by change at one of the end-points of the chain. For example, in the London working-class vernacular Cockney, diphthongization of /i/, i. e. becoming /ɪi/ or /əi/, is projected into a push-chain shift across the vowel space as in **Figure 1** (Cruttenden 2001, 87).

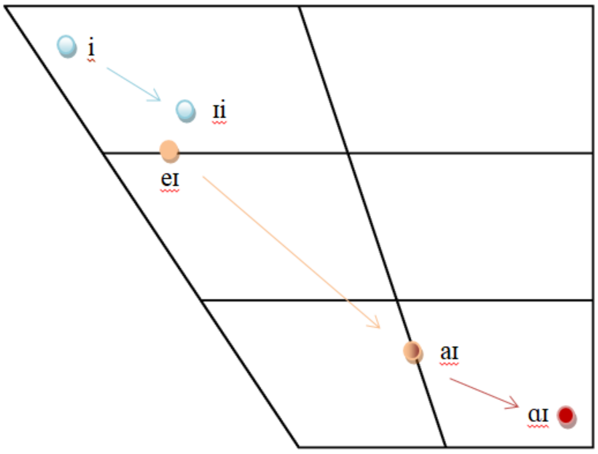

Figure 1. Front vowel shift in Cockney; colored dots represent starting points of individual diphthongs, arrows represent the direction of their shift.

Other cases are the Canadian and Californian English dialects, where a similarly looking pull-chain shift of front vowels is arguably triggered by the loss of contrast between the sounds /ɒ/ and /ɔ/ (Eckert n.d.; Boberg 2005; Clarke, Elms, and Youssef 1995).

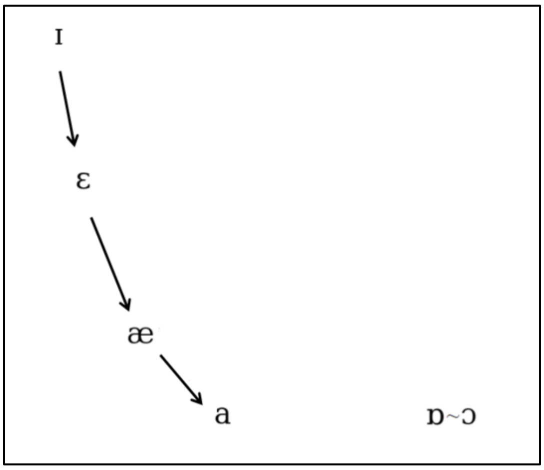

Figure 2 Canadian front vowel shift as described by Clarke et al. (1995).

Such vowel space alterations show that English vowel system may be subject to a chain-like kind of change when certain conditions are fulfilled; in this case, it is the cot/caught merger and /i/ diphthongization. This may not be the case in other languages, such as Arabic, where the inter-vowel space is large enough so that a shift would not reduce the contrast with other vowels substantially[2].

To the author's best knowledge, previous research has so far been looking at inter-phonemic generalization only from the perspective of phonological or acoustic similarity. Our approach seeks higher-level motivation stemming from the systematic operation of the whole language system. Indirect effects of lexically guided adaptation on sound representations have been observed in Dahan et al. (2008). In order to achieve effects of the systematic operation of perceptual adaptation, we provided listeners with ample information about our speaker's vowel space. However, we put emphasis on not exposing the area of vowels liable to chain shifts. Given that perceptual adaptation causes a shift in the /i/-/ɪ/ boundary and reduces the contrast between these two vowels, we believe that the principle of contrast maintenance should cause a shift in the neighboring /ɪ/-/ɛ/ boundary as well. Another reason to present our listeners with multiple vowels was to make explicit that the manipulation concerns only specific vowel(s) and is not a mere characteristic of the speaker's vocal tract shape.

We tested both our hypotheses by means of a single experiment not only for practical reasons but also to be able to test our second hypothesis on non-native listeners in case we found positive evidence within native speakers.

[2] Note that a chain shift in Arabic has been, in fact, observed (Kirchner 1996). Yet, unlike diachronic shifts, this synchronic shift concerns only certain phonological contexts and is presumably driven by different mechanisms than those studied here.

3 Methodology

3.1 Participants

Our first hypothesis was tested on a group of 18 Czech students from the Department of English and American Studies at Palacký University in Olomouc. All participants had English proficiency level C1 as defined by the Common European Framework of Reference for Languages (CEFRL). Except for three individuals, all come from Moravia and speak Moravian dialect. The other three, who come from Bohemia, formed a separate group for within subject analysis.

To test our second hypothesis, a group of 22 native speakers of English was contacted and agreed to take part in the experiment. Data from only 15 listeners were eventually collected due to insufficient time for analysis or technical problems on the side of the participants. Results used in this study were obtained from 14 native speakers of English from different parts of the United States and one from Great Britain. The majority has resided in the Czech Republic but was not bilingual. The mean age of the group was 36, ranging from 19 to 59 and participants were equally distributed. There were five female respondents who were in the end distributed in a four to one ratio since the lack of results from other participants disrupted the former equal distribution. Two male participants provided results for both versions of the experiment and were thus analysed in both inter subject and within subject design. The results are presented separately.

3.2 Experiment design and preparation

The experiment was designed to identify and measure perceptual boundaries between the vowels /i/, /ɪ/, and /ɛ/ in two groups of participants exposed to differently manipulated stimuli. Based on a comparison of the two groups' results, the influence of the sounds' manipulation on the subjects' phonetic perception was determined. To test for a chain like adaptation, tokens containing the vowel [i] were manipulated in such way to either increase or reduce the contrast with its neighboring more open vowel /ɪ/ and to potentially trigger a compensatory shift between /ɪ/ and /ɛ/ without prior exposition to its tokens. Manipulation in the opposite direction, that is, towards a greater contrast was intended for one thing to increase the qualitative difference between the two groups' stimuli so as to result in more significant potential shift and for the other to determine whether this shift occurs to preserve sufficient contrast or whether the inter-vowel space merely has a tendency to remain uniform. In the latter case, the shift of the /ɪ/-/ɛ/ boundary would also be triggered by increasing the contrast between /i/ and /ɪ/.

For both groups of participants, the experiment consisted of two phases: an exposure phase and a testing phase. For the exposure phase, a lexical-decision task was used. Respondents were asked to classify presented stimuli into word and non-word categories. Among the real word stimuli were also the critical words, which were expected to trigger perceptual boundary shift. For the following testing phase, another two alternative forced choice task was used. This time, listeners were asked to categorize isolated vowel tokens from an [i]~[ɪ]~[ɛ] continuum. The probabilities of the answers /i/, /ɪ/ or /ɛ/ for each presented token served to determine each respondent's perceptual vowel boundaries.

Respondents were allowed several short breaks throughout the two phases and were instructed to not engage in any verbal activity.

Most of the preparatory work put into the experiment was done in a freeware tool for scientific purposes in the phonetic field called Praat[3].

3.2.1 Design and preparation of the exposure phase

The exposure phase was devised to present the listeners with three categories of stimuli in random order. These categories were: critical words with manipulated [i]-tokens (30×3), word fillers (30×2), and non-word fillers (44×2). All together formed 208 classification trials performed by each participant. While the fillers and non-words were the same for both participant groups, the critical words were prepared in two sets of 30, each set presented to one group, differing in the direction of manipulation.

3.2.1.1 Preparation of the critical words

Originally 33 critical words, list of which can be found in **Table 2**, were extracted from the English corpus of the InterCorp v9 parallel corpora of the Czech National Corpus[4]. The critical words were manually selected from a lemmatized list of concordance search results displaying words with an "obstruent-/i/-obstruent" structure. To obtain this list,

[3] Praat has been developed by Paul Boersma and David Weenink, from the Institute of Phonetic Sciences in the University of Amsterdam.

[4] Currently run by the Czech National Corpus Institute under the auspices of Charles University's Faculty of Arts. This corpus was chosen for its free access after registration and the option of conditioned search. As a parallel corpus, that is, consisting of translations from Czech, it is not advisable for use in corpus-driven research. However, in this experiment the corpus was used for searching common words only.

an *ad hoc* KWIC search command was used, which utilized multiple English allographs of the vowel /i/ and of different obstruents. The criteria for each critical word were: to contain the vowel /i/ between obstruents, to not form a minimal pair with /ɪ/-words, and finally to not contain any other front vowels. In other words, the critical words were selected freely to contain any vowel, including diphthongs, except for /i/, /ɪ/, /ɛ/, and /æ/. Obstruents were chosen as the phonological context to avoid vowel formant transitions in the /i/-tokens which would get distorted by the planned manipulation. The second criterion was set to ensure lexical guidance of adaptation and not a simple alteration of the word perceived. Front vowels had to be omitted from the exposure phase to target their representations indirectly. One criterion was not kept in the case of the word *coffee* with /i/ in an open syllable. This, however, had no negative effect on the manipulation outcome and the word was consequently kept in the set of stimuli due to general lack of words that would satisfy all of the criteria. In two other cases, the limit imposed by the possibility of words forming minimal pairs with the vowel /ɪ/ had to be bypassed by incorporating the periphrastic to-infinitive. This marked the words as verbs and eliminated the possibility of a minimal pair; compare the existing pair *tease – tizz* and *to tease – *to tizz*[5].

3.2.1.2 Preparation of the filling words

Choosing of the filling words was far less conditioned, and therefore aimed more at making the lexical-decision task engaging for the participants. Existing words of moderate frequency and length were chosen to form a preliminary list of 40 specimens with more or less proportionate representation of English obstruents and vowels. As with the critical words, the vowel inventory for the filling words was reduced by four front vowels. Otherwise, there were no limitations. This list was eventually reduced to 30 words, which was the initially planned number. The excluded words were either too frequent or too prone to be perceived as containing a front vowel.

3.2.1.3 Preparation of the non-words

A list of 47 non-words was assembled following the same principles as with the filling words. This selection was intended to resemble the real-word stimuli in order to draw more of the participants' attention to the lexical-decision task.

[5] An asterisk stands for ungrammaticality or no meaning.

As an added value, making the participants pay more attention and be wary of the slight phonetic differences in order to successfully recognize real words from the non-sense ones may reduce potential effects of top-down perception and phonemic restoration; this is but a mere hypothesis, though. The term phonemic restoration describes a perceptual phenomenon which takes place when listeners are exposed to words with certain sounds replaced by noise. In these experiments, Samuel (2001) tested the human ability to mentally restore missing phonemes by means of existing knowledge about the lexical unit; in other words, by recognizing the word as a whole. Furthermore, listeners in this experiment were even unable to reliably distinguish words with sounds replaced by noise from words with noise only superimposed on the same sounds. This is an example of a top-down perception, which may lead to incorrect labeling of a non-word as its similar existing counterpart or to the perception of a front vowel where it is actually not pronounced. The risk of the phonemic restoration occurring in the critical words is, however, minimized anyway by their very choice.

Raw data for the list of non-words was gathered from Soybomb[6] online generator, which uses a frequency list of individual phonemes to form phonotactically correct nonsensical English words. Particular words were then selected manually. It was eventually concluded that four of the selected non-words were too similar to already existing words. Also, one word was included twice (*borthoog* and *borthug*) due to our speaker's dual pronunciation during the recording, so the final total was 44.

3.2.1.4 Recording of the stimuli

After the selection of stimuli, a session with a native speaker of American English was arranged to take place in a recording cabin of the audio-visual department in Zbrojnice library in Olomouc. A sound-proof booth with acoustic panels was used as the setting to minimize any reverberations that might cause distortions. Such distortion would jeopardize the accuracy of boundary shift measurements by possible perceptual compensation effects. For the recording purposes, all three categories of stimuli (critical words, filler words, and non-words) were put together in a single file and fed into a Praat script, which served as a presenter for the speaker. This presenter displayed the stimuli one by one at a fixed pace and random order to be read by the speaker and

[6] Nonsense word generator, accessed March 7, 2018
www.soybomb.com/tricks/words/

recorded. The order of the words was randomized to prevent coarticulation; that is, adapting the pronunciation of one sound to the preceding or following one, which is a natural way of facilitating articulation during speech. Two cycles of the script were performed for a total of 246 (2×123) words recorded.

3.2.1.5 Pre-manipulatory editing

The raw recordings were afterward edited for further processing. This included setting the word-boundaries and splitting the recording into individual files accordingly. Each word file was then converted to a monaural signal, and all words were scaled to the same intensity. Afterward, the quality of the sound was assessed, and better version from the paired recordings was chosen for each word. In case of the i-words, suitability for manipulation was the key factor. That is, how apparent were individual formants and presence of any distortions, such as creaky voicing or nasalization. Notably, in the majority of the words, there was perceivable pre-voicing of voiced plosives in syllable-initial position, which is atypical for English pronunciation (Ladefoged and Johnson 2011, 57). This may be either the speaker's idiosyncrasy or an acquired accent from Czech speakers. Because this experiment does not focus on consonants but vowels, and also for reasons related to the technicalities of formant manipulation, no editing was needed here.

From this stage, the filling words and non-words were effectively ready for use in the experiment. Nonetheless, the vowel inventory of both the filling words and the non-words had to be reconsidered based on the pronunciation differences between the Received Pronunciation (i. e. British English) and General American English. Because the list was formerly designed for a speaker of British English, General American pronunciation put a number of the words out of lines with the inclusion criteria. In particular, words pronounced with the open back unrounded vowel /ɑ/ in RP are in some contexts pronounced by a speaker of GA with the open front vowel /æ/ (Cruttenden 2001, 86). In consequence, these words had to be excluded. In the word-list in **Table 2**, the excluded words are highlighted.

Further preparation of the critical words was necessary with regards to the manipulation, which was to be carried out by means of a pair of Praat scripts. These were designed to resynthesize each word once its formants had been measured, extracted and manipulated using information from tables and text grids. Text grid files served to indicate which part of each word was to be manipulated; since only the tokens of the vowel /i/ needed manipulation and the rest of each word was to remain intact. For this task, it was prudent to adopt methodology and conventions suggested in the handbook *Principles of Phonetic Segmentation* by Skarnitzl and Machač as "both inter-labeller and intra-labeller consistency is an issue in manual segmentation" (2001, 13) which causes the precise boundaries of phonetic units to be often disputable. Besides creating a text grid file for each critical word, it was also necessary to write down a table of correct settings for the Praat's built-in algorithm for tracing formants (see **Table 5**). These settings included the number of formants for measurement and the ceiling value of frequency to set the bandwidth for formant-detection. By this step, the input for the first Praat script was completed.

The first resynthesis script was adopted from a similar study (see Chládková et al. 2017) in its original form; therefore, its output was used only partially. Namely formant grids, separate files showing precise tracing of each formant in each critical word's [i]-interval, were collected for hand correction. By correcting inaccuracies of the software formant measurements, the stimuli were readied for the second script.

3.2.1.6 Manipulation of the critical words

Before the manipulation itself, the degree of shift for each formant was calculated. For this purpose, a psychoacoustic unit of measurement called equivalent rectangular bandwidth (ERB) was used. This unit works on the basis of approximation of frequency-filters of human hearing. As a result, the range between two points on an ERB scale corresponds to the same range perceived by the human ear.

Because the testing phase tokens were recorded later in a separate session, there were no tokens of the speaker's vowels available but those of /i/. For this reason, formant frequencies for the target vowels needed to be approximated. Average values were taken from a table published in J. Hillenbrand's article "Acoustic characteristics of American English vowels" (1995, 3103). These values are provided in **Table 3**. The deviation of our speaker's /i/ from Hillenbrand's averages was calculated as a ratio, which was then applied to find values of /ɪ/. This method assumes that deviation from average does not vary with different vowels, which might not have been the case. On that account, our assumptions were verified by later measurements. Our speaker's /ɪ/ actually corresponded to the stated average, and the deviation of /i/ was most likely due to hyper-articulation. The degree of our manipulation was adjusted accordingly.

Based on the approximated values, the degree of the shift was calculated; first, to achieve [i̞] tokens with more open (i. e. lower) quality, approximately halfway between our speaker's /i/ and the calculated /ɪ/, and second, to achieve hyper-articulated [i̝]. After the actual formant values of our speaker's /ɪ/ were obtained, we reconsidered the degree of manipulation and used the **whole** distance between measured /i/ and the **calculated** /ɪ/ instead. The measured values for /ɪ/ were not used for technical reasons. Raised quality tokens were attained by using half the distance of the lowering manipulation and an inverse vector; flipping the ± sings for each formant value. For the manipulation in numbers, see **Table 1**.

Manipulation		***Vowel***	***Formant values in ERB***			***Formant values in Hz***		
			F1	F2	F3	F1	F2	F3
Hillenbrand	Average	[i]	7.99	22.17	24.29	342	2322	3000
		[ɪ]	9.29	21.07	23.37	427	2034	2684
Speaker	Measured	[i]	7.06	22.42	24.08	287.72	2392.19	2923.09
		[ɪ]	9.55	21.07	23.92	445	2033.5	2867
	Calculated	[ɪ]	8.21	21.30	23.17	356	2092	2618
Degree of shift		[i̞]	1.15	-1.11	-0.91	30.16	29.16	23.55
		[i̝]	-0.575	0.557	0.455	-15.081	-14.582	-11.777

Table 1. Formant values measured to be used in the exposition stimuli manipulation. Values for our speaker's /i/ were attained as an average of the tokens from the exposure stimuli. Measurements of the vowel /ɪ/ come from isolated tokens recorded for the testing phase continuum.

The critical stimuli were then finalized. This process comprised manual correction of manipulation artifacts and assessing quality of the input. During the finalization work, the number of critical words was reduced to the initially planned 30, excluding the words *beekeeper*, *beep*, and *secret*, either because of poor outcome of the manipulation or to avoid top-down perception of a front vowel by the participants. The list of stimuli with the excluded words marked by color is provided in **Table 2**. Also, to ensure there were indeed no front vowel tokens among the stimuli and hence no other than the critical words could influence respondents' perception of the vowels /ɪ/ and /ɛ/, formant values for all vowels in the stimuli were measured and mapped onto a vowel space. For all the manipulated [i̝] and [i̞] tokens placed inside a vowel space, see **Figure 5**, other vowels, from critical words inclusive, are displayed in **Figure 3**.

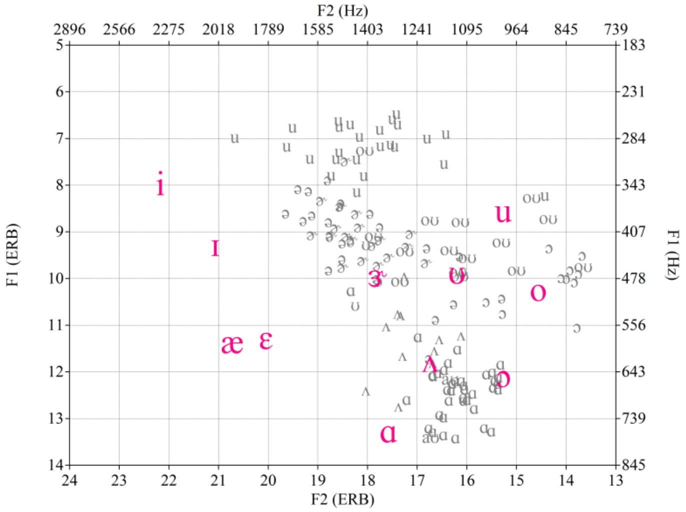

Figure 3. All of our speaker's vowel tokens that were used in the exposure phase but were not manipulated plotted onto an F1 by F2 field together with a mapped average of American English vowel space adopted from Hillenbrand (1995).

3.2.2 Design and preparation of the testing phase

The three front vowels /i/, /ɪ/, and /ɛ/ were each put into a phrase of the following structure: "*V*, *obstruent-V-obstruent* has the best *V*." For example, "[i], *speed* has the best [i]." The other two carrier words were *kid* and *Ted*. These three phrases were then presented to our speaker to read out loud and were recorded. The carrier phrases were used to make the pronunciation more natural for the speaker as well as to have a phonological context to check for a possibly unusual pronunciation of the isolated tokens.

3.2.2.1 Preparation of the vowel continuum

The phrases were recorded by the same speaker as the stimuli for the exposure phase but in a separate session. Same recording cabin was used, but the recording equipment was slightly altered. In particular, a different microphone was used to minimize background noise. To ensure that the timbre of the speaker's voice is preserved, a test recording was made and compared to one made with the previous microphone. Three rounds of the three phrases were recorded for a total of 18 vowel tokens, each vowel appearing twice in each round. All recordings were again converted to mono and edited into a form of individual cropped, scaled, and sorted vowel tokens. Another formant measurement was carried out to form a set of reference points for the continuum to trace. All of our speaker's front vowels recorded for the experiment are given in **Figure 4**.

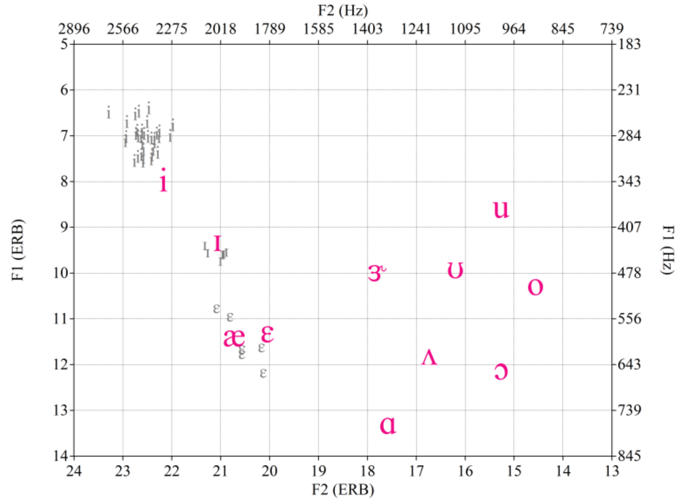

Figure 4. Tokens of our speaker's front vowels recorded for the lexical decision task (/i/ vowels from words) and for preparation of the continuum (/ɪ/ and /ɛ/ isolated) plotted on an F1 by F2 field. Vowels in pink are the average vowel referents provided in Hillenbrand (1995, 3103)

An [i] across [ɪ] to [ɛ] continuum of 85 uniform steps of equal duration was then created via a Praat script from a selected /ɛ/ token. Precise tracing of the points on the continuum together with the manipulated tokens of /i/ and Hillenbrand's reference values are given in **Figure 5**.

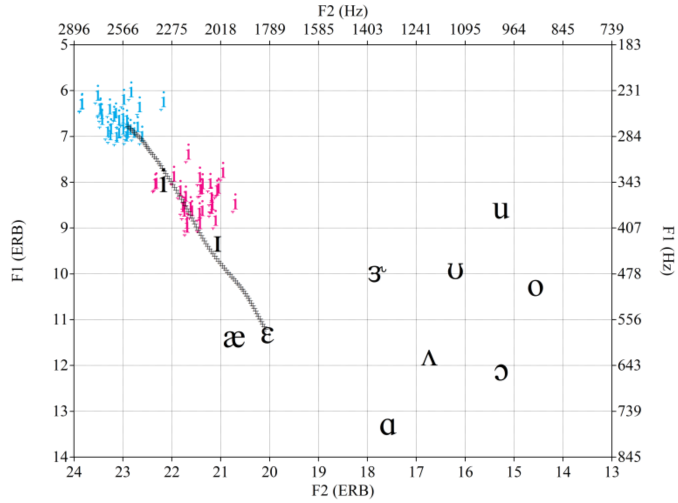

Figure 5. Manipulated [i̝] and [i̞] tokens (in color) and tracing of the 85-stepped [i]~[ɪ]~[ɛ] continuum on an F1 by F2 field. Hillenbrand's averages are in black.

3.2.3 Finalization of the experiment

The whole experiment was encoded into a single Praat script which ensured that each participant was provided with necessary instructions and that stimuli were presented in a given number and random order. This script together with all the recordings and multiple versions of Praat for different operating systems were made into an experiment package.

3.3 Procedure

The experiment package was distributed among the participants via electronic mail. Participants were assigned numbers which determined which version of the experiment would be played to them. They were advised on how to run the experiment and instructed to use headphones, perform the experiment in a quiet environment and to not engage in any verbal activity before the end of the experiment. After the participants confirmed they had understood what their task was as well as all the related information, they could fit the experiment into their schedules freely. The results were collected from each participant individually. The within subject group was given the other version of the experiment with a minimal inter-period of one day.

4 Results

All responses from the lexical-decision task and the categorization task were recorded for each participant separately. The number of correct responses as well as the reaction times were assessed. One American respondent was excluded due to a high percentage of incorrect answers in the lexical-decision task. Another three Czech respondents, two from the between subject and one from the within subject group, had to be excluded because no /i/-/ɪ/ boundary could be determined from their categorization choices for reasons presumably related to their cue-weighting strategy (see section 1.9). Responses in the categorization task for which reaction time exceeded 10 seconds were also excluded from the analysis.

Using the software Statistica, logistic regression was carried out for each participant with /i/ choice as the binary dependent variable (1 = stimulus categorized as /i/, 0 = categorized otherwise). The same procedure was then repeated with /ɛ/ as the dependent variable to assess both of the perceptual boundaries. As for the independent variable, we chose the first formant frequency since the stimuli and the continuum tokens varied predominantly in vowel height, for which F1 value is the associated acoustic clue (Llompart 2018, 19). Note that the exposition stimuli were manipulated for each word individually so as to preserve the natural voice timbre. For this reason, they varied in F2 and F3 as well.

Perceptual boundaries were calculated as the F1 value for which there is a 50% chance of either categorization; i. e. 1 = /i/, 0 = other or 1 = /ɛ/, 0 = other. **Figure 6** displays logistic curves from the between subject measurements divided according to participant group (American/Czech), the direction of manipulation (lowered/raised quality) and the vowel boundary (/i/-/ɪ/ or /ɪ/-/ɛ/). Every line corresponds to a single type of experiment (e. g. American participants exposed to lowered [i̞] tokens]. A t-test of the measured boundaries showed no significant shifts between any of the participant groups ($p > 0.05$). The direction of shifts also does not correspond to our expectations. Under our hypothesis, boundaries should lower, meaning higher F1 values. This difference was observed only with the directly unstimulated /ɪ/-/ɛ/ boundary in the Czech lowered group.

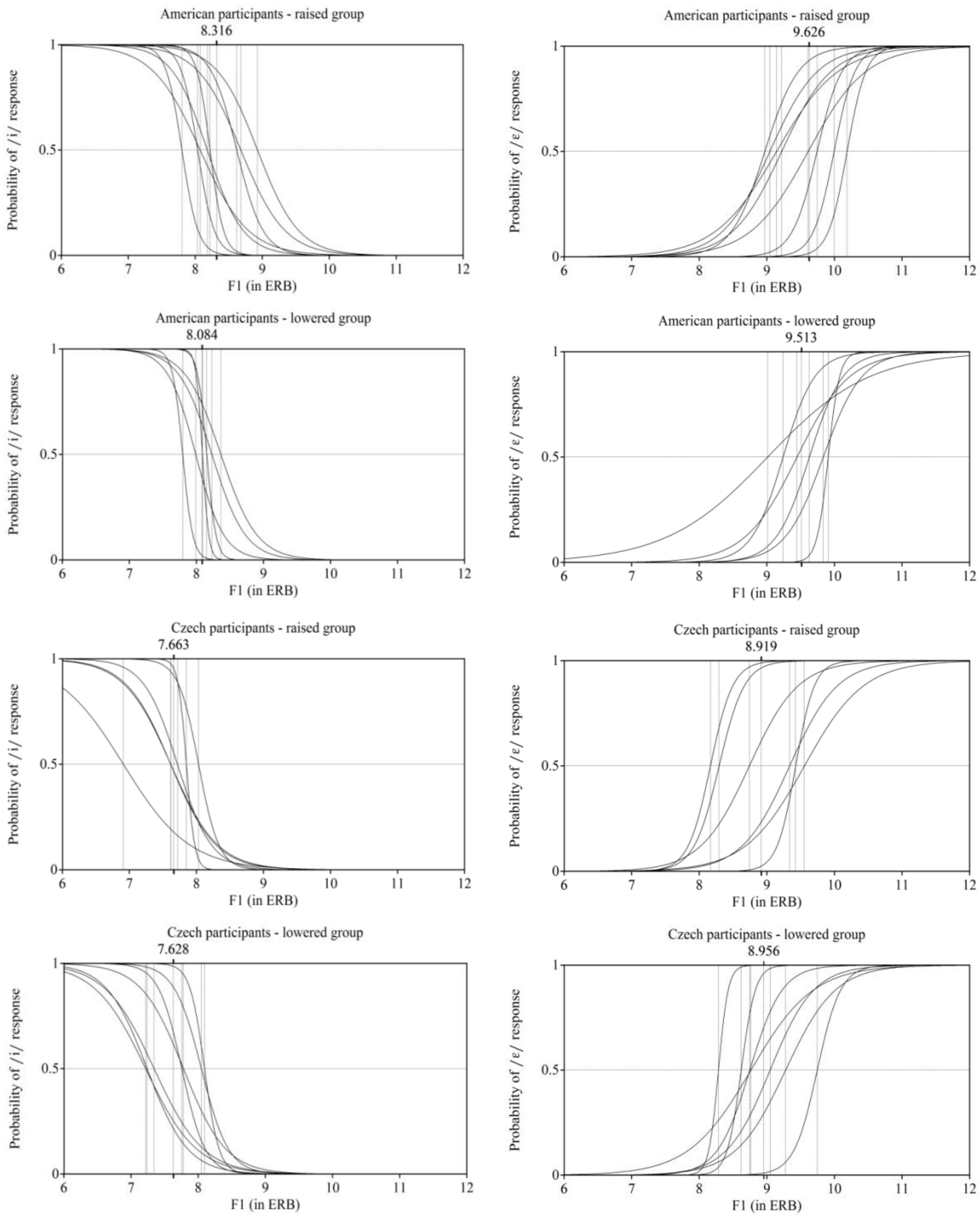

Figure 6 Logistic regression curves mapping the categorization of the [i] and [ε] tokens (left-hand and right-hand side) on the F1 and probability dimensions. Each curve represents one participant, its center, marked by a vertical line, corresponds to the perceptual boundary measured. The average boundary for each group is marked on the horizontal scale with a numeric value above it.

A significant difference in perceptual boundaries was found between American and Czech listeners. As can be seen from **Figure 7**, both groups had their boundaries fairly high compared to Hillenbrand's averages. Perception of American listeners was a bit more evenly distributed with the F1 value of the higher boundary amounting to 8.25 ERB (327 Hz) and the lower at 9.56 ERB (411 Hz). Czech listeners had their boundaries at 7.65 ERB (292 Hz) and 8.9 ERB (367 Hz). Most of the participants' vowel space seems to be occupied by the vowel /ɛ/. As Daniel Jones stated in his *Outline of English Phonetics*, “the vowel /ɛ/ varies a good deal with different speakers” (1956, quoted in Ladefoged and Broadbent 1957, 101). Categorization of such variable vowel thus requires broad phonetic demarcation. All averaged boundaries together with manipulated stimuli and averaged vowel placements from Hillenbrand (1995) plotted as if on a vowel space are provided in **Figure 7**.

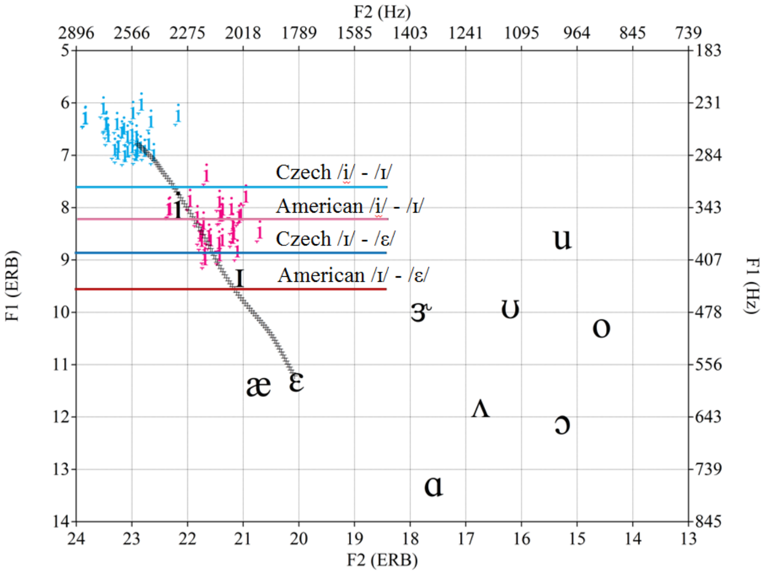

Figure 7. Average perceptual boundaries of the /i/-/ɪ/ and /ɪ/-/ɛ/ contrast measured in the categorization task (cf. the overall above-average placement as well as the difference between the Czech and the American group).

The within subject design regarded two American and two Czech participants. Analogous with the between subject analysis, logistic regression was performed with /i/ and /ɛ/ picks taken as the independent variable and F1 as the dependent one. Logistic curves in **Figure 8** show that a larger shift was achieved in all cases. However, an ensuing t-test proved the degree of shift still insignificant ($p > 0.05$).

In one of the four cases, the observed shift took a direction opposite to the one expected (American participant n. 2). This case was also unique in that the indirectly stimulated boundary /ɪ/-/ɛ/ shifted to a greater extent than the one whose shift was primarily targeted. In the other three cases, the indirectly induced shift was systematically smaller as expected or did not occur at all.

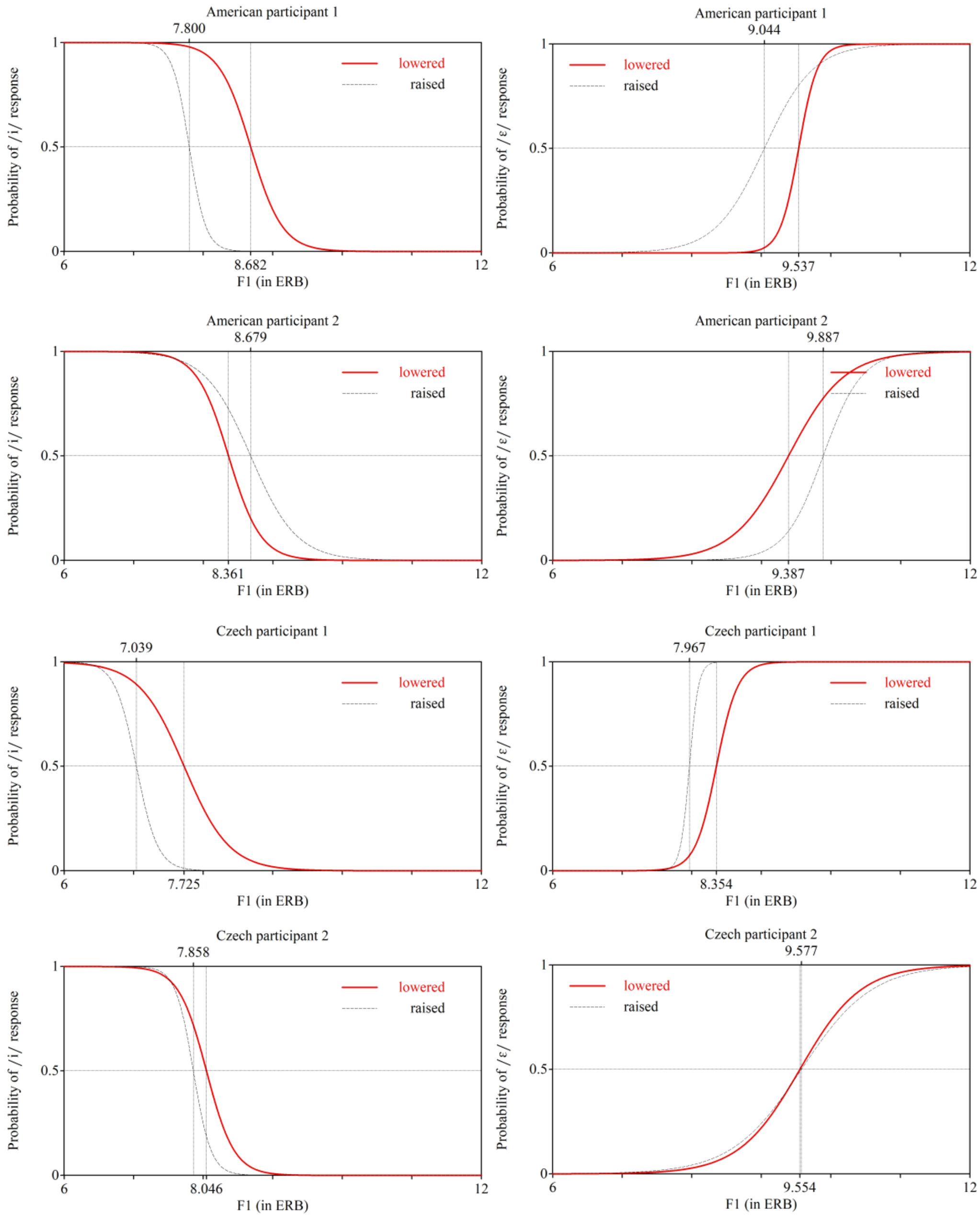

Figure 8. Logistic curves for within subject analysis of results from two American and two Czech listeners. Perceptual boundaries correspond to the centers of the curves and are marked by the horizontal lines with numeric values on the top or bottom side.

5 Discussion

The experiment conducted in this study failed to replicate previous results on perceptual adaptation; our second hypothesis, therefore, could not be tested. The lack of effects is probably attributable mainly to practical factors such as the small and varied sample of participants and relatively low control over the experimental conditions. Since the experiment was performed by each participant individually and without supervision, we cannot be certain that all necessary conditions were met. Variation in type of the headphones used might have had a disrupting effect as well; using speakers instead would certainly have. We also did not measure participants' boundaries without prior exposition. There is a chance that there was a significant difference between the boundaries of the raised and lowered groups and adaptation to our stimuli made these come together and eliminated the difference. This coincidence would correlate with the fact that all of the measured boundaries were above Hillenbrand's average. Elimination of inter subject variability has also proven useful as larger shifts were observed. On that account, within speaker design seems a promising strategy for a new round of the experiment herein with a larger sample of participants.

The direction of shifts was prevalently opposite to that assumed by the principles of perceptual adaptation. This would suggest, were such results to be also detected in a larger scale study, the presence of interfering effects of selective adaptation. It is the case that our speaker tended to hyper-articulate during the stimuli recording. Since the degree of manipulation was equal for all /i/ tokens, some overly pronounced words might not have come out as ambiguous for certain participants. Another possibility is that exposition to the vowel tokens from the categorization task itself caused selective adaptation which overruled the perceptual adaptation induced shifts. Due to the architecture of our experiment script, it is, unfortunately, impossible to retrieve intermediate (e. g. from halfway to the categorization task) participants' boundaries from the collected data. In case interference of selective adaptation effects proves significant, supplementing the experiment with approximately 30 minutes' nonverbal task is advisable. This time period is allegedly sufficient for the selective adaptation effects to fade out (Harris 1980, quoted in Kraljic and Samuel 2005). Perceptual adaptation, on the other hand, stabilizes and possibly even strengthens with time (Kraljic and Samuel 2005).

Despite the fact that the measured perceptual boundaries indicate that the lowered [i̞] tokens were perceived by our listeners as /ɪ/ (see **Figure 7**), the results from the lexical-decision task show the opposite. This implies that, for this group, perceptual adaptation might have occurred but did not generalize to other instances of the vowel. Our participants were overall exposed to 90 tokens of the manipulated vowels which should be a sufficient amount. Previous studies have found effects even with a lower number of trials (Chládková et al. 2017, 418). Possible explanation is that during the lexical-decision task listeners utilized additional information about the identity of the vowels in the critical words, namely the spectral profile of the preceeding consonants. Particularly the aspiration part of voiceless plosives retained its original quality, which might provide a sufficient acoustic cue so that no adaptation for the vowel itself was necessary. On the other hand, in the process of manipulation of the critical vowel tokens we provided for progressive transitions at the start and at the end of each vowel. This, however, caused a slight diphthongizing effect, which may have also inhibited generalization of adaptation to other instances of the tested vowel.

5.1 Findings from the lexical-decision task

One respondent's results had to be excluded from the study since 75% of the manipulated stimuli were not recognized as existing words. Presumably, the degree of manipulation was for this individual excessive, and the vowel was identified as /ɪ/. Another possible explanation is that the lexical-decision as such made it less likely to treat manipulation as pronunciation idiosyncrasy. In fact, some non-word stimuli formed minimal pairs with existing words (e. g., *cardo* vs. *cargo*). Listeners could be thus motivated to be cautious of such oddities and not accept them as phonetic deviations. In a different experimental paradigm which would encourage participants to perceive words as meaningful, the adaptation might take a stronger effect. Such paradigm was used for example by Maye et al. (2008).

American participants were quite successful at categorizing, with an average of 5% incorrect answers per participant, including approximately 4.7% from the critical word set, 7% from the non-word set and negligible 1.3% from the filler word set. The prevalence of incorrectly categorized non-words (approx. 6 per participant) is attributable to the ambiguity of some of the stimuli. For example, the non-word *phonle* is actually homonymous to an existing, albeit infrequent, word *phonal*. Other frequently mislabeled non-words were *affold*, confusable with *a fold*, *toove*, and *gaaz*; the last two probably mistaken for *tooth* and *gas*. On the other hand, for American listeners, the vowel in the word *gaaz* should be a sufficient cue even with the final consonant devoiced, since in GA accent the vowel /ɑ/ before fricatives and nasals is realized as /æ/ (Cruttenden 2001, 86).

In comparison, Czech listeners were naturally less accurate with slightly over 13% of incorrect answers per participant. This included 12% of the critical words, 17.5% of the non-words, and 5% of the existing words. In both the critical word and the non-word categories, Czech listeners were with remarkable coincidence two-and-a-half-times more likely to choose an incorrect answer than native-listeners for both critical words and non-words. With unmanipulated existing words, the ratio was greater, which logically corresponds with the smaller vocabulary of non-native listeners.

In both participants groups, the words *to tease* and *to speak* could be interpreted as a single non-word. Nonetheless, regarding the number of incorrect choices, these were not outstanding. By contrast, the word *feedstock* was particularly often mislabeled even though it is not as rare compared to other stimuli. Merriam-Webster[7] online dictionary categorizes its frequency as of 28 April 2018 among the bottom 30% of English words.

5.2 Remarks on the chain shift hypothesis

Because our study did not replicate perceptual adaptation induced shifts, we were not able to test our second hypothesis regarding inter-phonemic generalization of the mentioned effects. However, several things should be noted for possible future studies on this topic.

In case that shifts in the targeted vowel were observed, it would be possible for the push-chain shift of the indirectly stimulated boundaries to not occur due to a lack of inter-vowel space. In dialects where front vowel push chain does occur, such as Californian English, there is this space created by reduction in the number of phonemes (Ladefoged and Johnson 2011, 96). In fact, pull-chain manipulation may prove more suitable to induce shifts as the front part of vowel space is more open, i. e., less restricted.

---

[7] Merriam-Webster, s.v. “feedstock,” accessed May 12, 2018, https://www.merriam-webster.com/dictionary/feedstock.

On the other hand, the results in the present study suggest that listeners are capable of successful categorization of words with ambiguous sounds even when no shifts manifest in categorization of isolated vowels. The question of what degree of manipulation is necessary or, by contrast, excessive regarding perceptual adaptation has not been so far widely addressed. This matter is closely related to the hypothesis that contrast maintenance principle plays a role in perceptual boundary shifts.

If perceptual adaptation was indeed able to generalize in a chain fashion, there would still be a missing fragment to the connection of short-term adaptation and diachronic change; that is whether perceptual adaptation is somehow reflected in production. This issue has been touched upon by Samuel and Kraljic (2009) in their comprehensive overview of the subject matter. In an experiment mentioned therein, Kraljic, Brennan, and Samuel (2008) had their subjects to fill in missing words in a narrative before and after being exposed to perceptual learning stimuli. Although a high degree of adaptation was reached, “the production system did not make a corresponding change” (Samuel and Kraljic 2009, 1214). However, the authors themselves identified several reasons why their experiment did not yield any changes in production. This question thus remains open to further research.

6 Conclusion

This thesis presents the topic of auditory perceptual adaptation from its basic underlying acoustic and phonetic principles to the findings of current research. The theoretical part also outlines several related topics of human perception and the relevant scientific work. Authors attempt to replicate previous findings and propose two follow-up hypotheses regarding perceptual adaptation in non-native speakers and generalization of adaptation effects across the system of English vowel in context of the diachronic development of language, respectively.

The practical part provides an experimental design as well as pilot results, their analysis, and discussion. Insofar as no significant effects were found, it is proposed for the experiment to be carried out on a larger scale and under more controlled conditions in line with further suggestions provided herein.

The main goal of this study was to contribute to the research on human speech perception as a deeper understanding of this matter has practical implications for various fields. For example, findings concerning language processing may help to deal with language impairments or in development of cochlear implants, while knowledge about adaptation to non-canonical speech and accents could find use in second language pedagogy and interpreter training.

7 Resumé

Lidská řeč je unikátní prostředek komunikace, který v přírodě nemá obdoby. S nástupem pokročilé informační technologie vzrůstají snahy o napodobení řečového vnímání a s nimi také o porozumění jeho zákonitostem. Jedním z klíčových aspektů této oblasti je lidská schopnost kategorizace rysů akustického profilu řeči do abstraktních lingvistických jednotek. Jelikož mezi akustickou a lingvistickou charakteristikou řeči neexistuje přesná shoda, musí se činnost tohoto kategorizačního systému dynamicky přizpůsobovat vnějším okolnostem. Tento jev se nazývá percepční adaptabilita.

Tato bakalářská práce představuje problematiku percepční adaptability od základních fonetických a akustických principů až po poznatky současného výzkumu. V teoretické části jsou rovněž zmíněny relevantní vědecké publikace na související témata ohledně lidského vnímání. Autoři navazují na předchozí studie s cílem replikovat jejich výsledky a formulují dvě hypotézy. Předmětem výzkumu je jednak percepční adaptabilita anglických samohlásek u nerodilých mluvčích, jednak principy její generalizace. První testovaná hypotéza tvrdí, že k percepční adaptaci anglických samohlásek dochází u nerodilých mluvčích obdobně jako u rodilých. Druhá hypotéza navrhuje vztah mezi krátkodobou adaptací a diachronním vývojem jazyků, konkrétně, že posun jedné mezisamohláskové kontrastní hranice vede k řetězovému posunu sousedících samohláskových reprezentací (tzv. „push/pull-chain“, jev popisován v historickém vývoji jazyků).

V praktické části je představen metodický návrh experimentu společně s pilotními výsledky, jejich analýzou a diskuzí. Jelikož se experiment nedopracoval k žádným statisticky významným výsledkům, navrhuje se jej zopakovat s obsáhlejším vzorkem respondentů, v přísněji kontrolovaných podmínkách a s přihlédnutím k dalším zde předloženým návrhům.

Hlavním cílem této studie bylo přispět k bádání v oblasti vnímání lidské řeči. Hlubší porozumění této problematice může s velkou pravděpodobností najít praktické využití v mnoha oblastech. Například studium zpracovávání řeči v lidském mozku může přispět řešením v oblasti poruch řeči či kochleárních implantátů, zatímco znalosti o percepci neznámých a nových přízvuků lze využít při výuce druhého jazyka anebo tlumočnickém výcviku.

8 References

Ainsworth, W. A. 1975. "Intrinsic and extrinsic factors in vowel judgments." *Auditory analysis and perception of speech*: 103–113. doi: 10.1016/B978-0-12-248550-3.50011-8.

Baese-Berk, Melissa M., Ann R. Bradlow, and Beverly A. Wright. 2013. "Accent-independent adaptation to foreign-accented speech." *The Journal of the Acoustical Society of America* 133(3): EL174–EL180. doi: 10.1121/1.4789864.

Bertelson, Paul, Jean Vroomen, and Béatrice De Gelder. 2003. "Visual recalibration of auditory speech identification: a McGurk aftereffect." *Psychological Science* 14(6): 592–597. doi: 10.1046/j.0956-7976.2003.psci_1470.x.

Best, Catherine T., and Michael D. Tyler. 2007. "Nonnative and Second-Language Speech Perception." *Language Experience in Second Language Speech Learning Language Learning and Language Teaching*: 13–34. doi: 10.1075/lllt.17.07bes.

Boberg, Charles. 2005. "The Canadian shift in Montreal." *Language Variation and Change* 17(2): 133–154. doi: 10.1017/S0954394505050064.

Bradlow, Ann R., and Tessa Bent. 2008. "Perceptual adaptation to non-native speech." *Cognition* 106: 707–729. doi: 10.1016/j.cognition.2007.04.005.

Bradlow, Ann R., Gina M. Torretta, and David B. Pisoni. 1996. "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics." *Speech communication* 20(3–4): 255–272. doi: 10.1016/S0167-6393(96)00063-5.

Brown, H. Douglas. 2000. *Principles of Language Learning and Teaching*. White Plains, NY: Longman. ISBN-13: 9780131991286.

Chládková, Kateřina, and Václav Jonáš Podlipský. 2011. "Native dialect matters: Perceptual assimilation of Dutch vowels by Czech listeners." *The Journal of the Acoustical Society of America* 130(4): EL186–EL192. doi: 10.1121/1.3629135.

Chládková, Kateřina, Paul Boersma, and Titia Benders. 2015. "The perceptual basis of the feature vowel height." In *Proceedings of the 18th International Congress of Phonetic Sciences Glasgow*: The University of Glasgow.

Chládková, Kateřina, Václav Jonáš Podlipský, and Anastasia Chionidou. 2017. "Perceptual adaptation of vowels generalizes across the phonology and does not require local context." *Journal of Experimental Psychology: Human Perception and Performance* 43(2): 414. doi: 10.1037/xhp0000333.

Clarke, Constance M., and Merrill F. Garrett. 2004. "Rapid adaptation to foreign-accented English." *Journal of the Acoustical Society of America* 116: 3647–3658. doi: 10.1121/1.1815131.

Clarke, Sandra, Ford Elms, and Amani Youssef. 1995. "The Third Dialect of English: Some Canadian Evidence." *Language Variation and Change* 7(02): 209–28. doi: 10.1017/s0954394500000995.

Crowder, Robert G. 1978. “Mechanisms of auditory backward masking in the stimulus suffix effect.” *Psychological Review* 85(6): 502. doi: 10.1037/0033-295X.85.6.502.

Crowder, Robert G. 1981. “The role of auditory memory in speech perception and discrimination.” *Advances in Psychology* 7: 167–179. doi: 10.1016/S0166-4115(08)60191-0.

Cruttenden, Alan. 2001. *Gimson’s pronunciation of English*. Edward Arnold Publishers Limited, London, England. ISBN-13: 9780340806685.

Culler, Jonathan D. 1986. *Ferdinand de Saussure*. Cornell University Press. ISBN-13: 9780801493898.

Dahan, Delphine, Sarah J. Drucker, and Rebecca A. Scarborough. 2008. “Talker adaptation in speech perception: Adjusting the signal or the representations?” *Cognition* 108(3): 710–718.
doi: 10.1016/j.cognition.2008.06.003.

Davis, Matthew H., Ingrid S. Johnsrude, Alexis Hervais-Adelman, Karen Taylor, and Carolyn McGettigan. 2005. “Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences.” *Journal of Experimental Psychology: General* 134(2): 222–241.
doi: 10.1037/0096-3445.134.2.222.

Drozdova, Polina, Roeland van Hout, and Odette Scharenborg. 2014. “Phoneme category retuning in a non-native language.” *Fifteenth Annual Conference of the International Speech Communication Association*.

Eckert, Penelope. n.d. “Penny Eckert’s Web Page: Vowel Shifts in Northern California and the Detroit Suburbs’ Accessed May 5, 2018. https://web.stanford.edu/~eckert/vowels.html

Eimas, Peter D., and John D. Corbit. 1973. “Selective adaptation of linguistic feature detectors.” *Cognitive Psychology* 4(1): 99–109.doi: 10.1016/0010-0285(73)90006-6.

Eisner, Frank, and James M. McQueen. 2005. “The specificity of perceptual learning in speech processing.” *Perception and psychophysics* 67(2): 224–238.
doi: 10.3758/BF03206487.

Escudero Neyra, P. R. 2005. *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. Utrecht: LOT. ISBN: 90-76864-80-2.

Evans, Bronwen G., and Paul Iverson. 2004. “Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences.” *The Journal of the Acoustical Society of America* 115(1): 352–361. doi: 10.1121/1.1635413.

Flege, James E. 1995. “Second language speech learning: Theory, findings, and problems.” *Speech perception and linguistic experience*: 229–273.

Flege, James E., Ocke-Schwen Bohn, and Sunyoung Jang. 1997. “Effects of experience on non-native speakers” production and perception of English vowels.’ *Journal of Phonetics* 25: 437–470. doi: 10.1006/jpho.1997.0052.

Fox, Robert Allen. 1985. “Auditory contrast and speaker quality variation in vowel perception.” *The Journal of the Acoustical Society of America* 77(4): 1552–1559. doi: 10.1121/1.391998.

Gerstman, Louis. 1968. “Classification of self-normalized vowels.” *IEEE transactions on audio and electroacoustics* 16(1): 78–80. doi: 10.1109/TAU.1968.1161953.
Goldinger, Stephen D. 1998. “Echoes of echoes? An episodic theory of lexical access.” *Psychological review* 105(2): 251. doi: 10.1037/0033-295X.105.2.251.
Goldstein, Michael H., Andrew P. King, and Meredith J. West. “Social interaction shapes babbling: Testing parallels between birdsong and speech.” *Proceedings of the National Academy of Sciences* 100(13): 8030–8035. doi: 10.1073/pnas.1332441100
Goudbeek, Martijn, Daniel Swingley, and Roel Smits. 2009. “Supervised and unsupervised learning of multidimensional acoustic categories.” *Journal of Experimental Psychology: Human Perception and Performance* 35(6): 1913–1933. doi: 10.1037/a0015781.
Goudbeek, Martijn, Roel Smits, Anne Cutler, and Daniel Swingley. 2005. “Acquiring auditory and phonetic categories.” In H. Cohen and C. Lefebvre (Eds.), Handbook of categorization in cognitive science. Amsterdam: Elsevier: 497–513. doi: 10.1016/B978-008044612-7/50077-9.
Harris, Charles S. 1965. “Perceptual adaptation to inverted, reversed, and displaced vision.” *Psychological review* 72(6): 419. doi: 10.1037/h0022616.
Hillenbrand, James M., Michael J. Clark, and Terrance M. Nearey. 2001. “Effects of consonant environment on vowel formant patterns.” *Journal of the Acoustical Society of America* 109(2): 748–763. doi: 10.1121/1.1337959.
Hillenbrand, James, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. 1995. “Acoustic characteristics of American English vowels.” *The Journal of the Acoustical society of America* 97(5): 3099–3111. doi: 10.1121/1.411872.
Hoff, Erika. 2003. “The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech.” *Child development* 74(5): 1368–1378. doi: 10.1111/1467-8624.00612.
Holt, Lori L. 2005. “Temporally nonadjacent nonlinguistic sounds affect speech categorization.” *Psychological Science* 16(4): 305–312. doi: 10.1111/j.0956-7976.2005.01532.x.
Holt, Lori L., and Andrew J. Lotto. 2006. “Cue weighting in auditory categorization: Implications for first and second language acquisition.” *The Journal of the Acoustical Society of America* 119(5): 3059–3071. doi: 10.1121/1.2188377.
Holt, Lori L., and Andrew J. Lotto. 2008. “Speech perception within an auditory cognitive neuroscience framework.” *Current Directions in Psychological Science* 17(1): 42–46. doi: 10.1111/j.1467-8721.2008.00545.x.
Holt, Lori L., and Andrew J. Lotto. 2010. “Speech perception as categorization.” *Attention, Perception, and Psychophysics* 72(5): 1218–1227. doi: 10.3758/APP.72.5.1218.
Iverson, Paul, and Bronwen G. Evans. 2007. “Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration.” *The Journal of the Acoustical Society of America* 122(5): 2842–2854. doi: 10.1121/1.2783198.

Johnson, Keith, Elizabeth A. Strand, and Mariapaola D'Imperio. 1999. "Auditory–visual integration of talker gender in vowel perception." *Journal of Phonetics* 27: 359–384. doi: 10.1006/jpho.1999.0100.

Johnson, Keith. 1991 "Differential effects of speaker and vowel variability on fricative perception." *Language and speech* 34(3): 265–279. doi: 10.1177/002383099103400304.

Johnson, Keith. 2005. "Speaker normalization in speech perception." *The handbook of speech perception*: 363–389. doi: 10.1002/9780470757024.ch15.

Jusczyk, Peter W. 2000. *The Discovery of Spoken Language.* Cambridge (Massachusetts): MIT. ISBN-13: 9780262600361.

Kiefte, Michael, and Keith R. Kluender. 2005. "The relative importance of spectral tilt in monophthongs and diphthongs." *Journal of the Acoustical Society of America* 17: 1395–1404. doi: 10.1121/1.1861158.

Kirchner, Robert. 1996. "Synchronic chain shifts in Optimality Theory." *Linguistic Inquiry* 27(2): 341–350. doi: 10.1002/9780470756171.ch21.

Kraljic, Tanya, and Arthur G. Samuel. 2005. "Perceptual learning for speech: Is there a return to normal?" *Cognitive psychology* 51(2): 141–178. doi: 10.1016/j.cogpsych.2005.05.001.

Kraljic, Tanya, and Arthur G. Samuel. 2006. "Generalization in perceptual learning for speech." *Psychonomic Bulletin and Review* 13: 262–268. doi: 10.3758/BF03193841.

Kraljic, Tanya, Susan E. Brennan, and Arthur G. Samuel. 2008. "Accommodating variation: Dialects, idiolects, and speech processing." *Cognition* 107(1): 54–81. doi: 10.1016/j.cognition.2007.07.013.

Ladefoged, Peter, and Donald Eric Broadbent. 1957. "Information conveyed by vowels." *The Journal of the Acoustical Society of America* 29(1): 98–104. doi: 10.1121/1.1908694.

Ladefoged, Peter, and Keith Johnson. 2011. *A Course in Phonetics.* Wadsworth, Cengage Learning. ISBN-13: 9781428231269.

Liberman, Alvin M., F. S. Cooper, D. P. Shankweiler, M. Kennedy. 1967. "Perception of the speech code." *Psychological review* 74(6): 431. doi: 10.1037/h0020279.

Lisker, Leigh, and Arthur S. Abramson. 1967. "Some effects of context on voice onset time in English plosives". *Language and Speech* 10(1): 1–28. doi: 10.1177/002383096701000101.

Lisker, Leigh. 1986. " 'Voicing' in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees." *Language and speech* 29(1): 3–11. doi: 10.1177/002383098602900102.

Llompart, Miquel, and Eva Reinisch. 2018. "Acoustic cues, not phonological features, drive vowel perception: Evidence from height, position and tenseness contrasts in German vowels." *Journal of Phonetics* 67: 34–48. doi: 10.1016/j.wocn.2017.12.001.

Lobanov, Boris M. 1971. "Classification of Russian vowels spoken by different speakers." *The Journal of the Acoustical Society of America* 49(2)B: 606–608. doi: 10.1121/1.1912396.

Machač, Pavel, and Radek Skarnitzl. 2009. *Principles of Phonetic Segmentation*. Prague: Epocha. ISBN: 978-80-7425-032-3.

Marslen-Wilson, William D. 1987. “Functional parallelism in spoken word-recognition.” *Cognition* 25(1–2): 71–102. doi: 10.1016/0010-0277(87)90005-9.

Martinet, André. 1984. “Double articulation as a criterion of linguisticity.” *Language Sciences* 6(1): 31–38. doi: 10.1016/S0388-0001(84)80003-0.

Maye, J. and Gerken, L. 2000. “Learning phonemes without minimal pairs.” *In Proceedings of the 24th annual Boston university conference on language development. Vol. 2*. Somerville, MA: Cascadilla Press: 522–533.

Maye, Jessica, Richard N. Aslin, and Michael K. Tanenhaus. 2008. “The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent.” *Cognitive Science* 32(3): 543–562. doi: 10.1080/03640210802035357.

McQueen, James M., and Holger Mitterer. 2005. “Lexically-driven perceptual adjustments of vowel categories.” At the *ISCA Workshop on Plasticity in Speech Perception*, London, UK.

McQueen, James M., Anne Cutler, and Dennis Norris. 2006. “Phonological abstraction in the mental lexicon.” *Cognitive Science* 30(6): 1113–1126. doi: 10.1207/s15516709cog0000_79.

Miller, James D. 1989. “Auditory-perceptual interpretation of the vowel.” *The journal of the Acoustical society of America* 85(5): 2114–2134. doi: 10.1121/1.397862.

Mitterer, Holger. 2006. “Is vowel normalization independent of lexical processing?” *Phonetica* 63(4): 209–229. doi: 10.1159/000097306.

Nearey, Terrance M. 1978. “Phonetic feature systems for vowels.” Indiana University Linguistics Club, Indiana.

Nearey, Terrance Michael. 1989. “Static, dynamic, and relational properties in vowel perception.” *The Journal of the Acoustical Society of America* 85(5): 2088–2113. doi: 10.1121/1.397861.

Norris, Dennis, James M. McQueen, and Anne Cutler. 2003. “Perceptual learning in speech.” *Cognitive psychology* 47(2): 204–238. doi: 10.1016/S0010-0285(03)00006-9.

Peterson, Gordon E., and Harold L. Barney. 1952. “Control methods used in a study of the vowels.” *The Journal of the acoustical society of America* 24(2): 175–184. doi: 10.1121/1.1906875.

Podlipský, Václav Jonáš, Radek Skarnitzl, and Jan Volín. 2009. “High front vowels in Czech: A contrast in quantity or quality?” *Proceedings of the Tenth Annual Conference of the International Speech Communication Association*. 132–135.

Reinisch, Eva, and Holger Mitterer. 2016. “Exposure modality, input variability and the categories of perceptual recalibration.” *Journal of Phonetics* 55: 96–108. doi: 10.1016/j.wocn.2015.12.004.

Reinisch, Eva, Andrea Weber, and Holger Mitterer. 2013. “Listeners retune phoneme categories across languages.” *Journal of Experimental Psychology: Human Perception and Performance* 39(1): 75–86. doi: 10.1037/a0027979.

Reinisch, Eva, David R. Wozny, Holger Mitterer, and Lori L. Holt. 2014. “Phonetic category recalibration: What are the categories?” *Journal of Phonetics* 45: 91–105. doi: 10.1016/j.wocn.2014.04.002.

Repp, Bruno H., and Alvin M. Libermant. 1984. “Phonetic category boundaries are flexible.” Haskins Laboratories Status Report on Speech Research (SR-77/78): 31–53.

Risset, Jean-Claude, and David L. Wessel. 1982. “Exploration of timbre by analysis and synthesis.” *The psychology of music*: 26–58. doi: 10.1016/B978-0-12-213562-0.50006-1.

Roberts, Martin, and Quentin Summerfield. 1981. “Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory.” *Perception & Psychophysics* 30(4): 309–314.

Samuel, Arthur G. 1986. “Red herring detectors and speech perception: In defense of selective adaptation.” *Cognitive Psychology* 18: 452–499. doi: 10.1016/0010-0285(86)90007-1.

Samuel, Arthur G. 2001. “Knowing a word affects the fundamental perception of the sounds within it.” *Psychological Science* 12(4): 348–351. doi: 10.1111/1467-9280.00364.

Samuel, Arthur G., and Tanya Kraljic. 2009. “Perceptual learning for speech.” *Attention, Perception, and Psychophysics* 71: 1207–1218. doi: 10.3758/APP.71.6.1207.

Schertz, Jessamyn, Taehong Cho, Andrew J. Lotto, Natasha Warner. 2016. “Individual differences in perceptual adaptability of foreign sound categories.” *Attention, Perception, & Psychophysics* 78(1): 355–367. doi: 10.3758/s13414-015-0987-1.

Shannon, Robert V., Fan-Gang Zeng, Vivek Kamath, John Wygonski, and Michael Ekelid. 1995. “Speech recognition with primarily temporal cues.” *Science* 270: 303–304. doi: 10.1126/science.270.5234.303.

Simon, Helen J., and Michael Studdert-Kennedy. 1978. “Selective anchoring and adaptation of phonetic and nonphonetic continua.” *The Journal of the Acoustical Society of America* 64(5): 1338–1357. doi: 10.1121/1.382101.

Sjerps, Matthias J., Holger Mitterer, and James M. McQueen. 2011. “Constraints on the processes responsible for the extrinsic normalization of vowels.” *Attention, Perception, and Psychophysics* 73(4): 1195–1215. doi: 10.3758/s13414-011-0096-8.

Skoruppa, Katrin, and Sharon Peperkamp. 2011. “Adaptation to novel accents: Feature-based learning of context-sensitive phonological regularities.” *Cognitive Science* 35: 348–366. doi: 10.1111/j.1551-6709.2010.01152.x.

Stilp, Christian E., Joshua M. Alexander, Michael Kiefte, and Keith R. Kluender. 2010. “Auditory color constancy: Calibration to reliable spectral properties across nonspeech context and targets.” *Attention, Perception, and Psychophysics* 72: 470–480. doi: 10.3758/APP.72.2.470.

Summerfield, Quentin, Andrew Sidwell, and Tony Nelson. 1987. “Auditory enhancement of changes in spectral amplitude.” *The Journal of the Acoustical Society of America* 81(3): 700–708. doi: 10.1121/1.394838.

Sumner, Meghan, and Arthur G. Samuel. 2009. “The effect of experience on the perception and representation of dialect variants.” *Journal of Memory and Language* 60(4): 487–501. doi: 10.1016/j.jml.2009.01.001.

Sundberg, Johan, Per-Erik Nordström. 1976. “Raised and lowered larynx-the effect on vowel formant frequencies.” STL-QPSR 17(2–3): 035–039.

Syrdal, Ann K., and Hundrai S. Gopal. 1986. “A perceptual model of vowel recognition based on the auditory representation of American English vowels.” *The Journal of the Acoustical Society of America* 79(4): 1086–1100. doi: 10.1121/1.393381.

TIBCO Software Inc. 2017. Statistica (data analysis software system), version 13. http://statistica.io.t

Toscano, Joseph C., and Bob McMurray. 2010. "Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics." *Cognitive science* 34(3): 434–464. doi: 10.1111/j.1551-6709.2009.01077.x.

Trude, Alison M., and Sarah Brown-Schmidt. 2012. "Talker-specific perceptual adaptation during online speech perception." *Language and Cognitive Processes* 27(7–8): 979–1001. doi: 10.1080/01690965.2011.597153.

Vroomen, Jean, Sabine van Linden, Mirjam Keetels, Béatrice de Gelder, and Paul Bertelson. 2004. "Selective adaptation and recalibration of auditory speech by lipread information: dissipation." *Speech Communication* 44(1–4): 55–61. doi: 10.1016/j.specom.2004.03.009.

Watkins, Anthony J. 1991. "Central, auditory mechanisms of perceptual compensation for spectral-envelope distortion." *The Journal of the Acoustical Society of America* 90(6): 2942–2955. doi: 10.1121/1.401769.

Watkins, Anthony J., and Simon J. Makin. 1994. "Perceptual compensation for speaker differences and for spectral-envelope distortion." *The Journal of the Acoustical Society of America* 96(3): 1263–1282. doi: 10.1121/1.410275.

Werker, Janet F., and Richard C. Tees 1984 "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life." *Infant Behavior and Development* 7(1): 49–63. doi: 10.1016/S0163-6383(84)80022-3.

Werker, Janet F., and Richard C. Tees. 1999. "Influences on infant speech processing: Toward a new synthesis." *Annual review of psychology* 50(1): 509–535. doi: 10.1146/annurev.psych.50.1.509.

Wilson, J. P. 1970. "An auditory after-image." *Frequency analysis and periodicity detection in hearing* 304.

9 Attachments

Table 2. List of the stimuli for the exposure phase of the presented experiment. Highlighted words were excluded during preparation for reasons which are here color-coded.

Green = Distorted by the manipulation process Turquoise = Risk of perception of a front vowel

Yellow = Similarity to an existing word Gray = Redundant filler words

/i/-words	filler words	non-words
beast	aloft	affold
beaver	approach	affong
beefeater	argument	akoot
beekeeper	assault	bomble
beep	barn	borthug
coffee	boost	borthoog
deepen	borrow	bourced
feature	broadcast	cardo
feeble	bun	chuzzer
feed	carbon	clawm
feedstock	code	conser
fever	comfortable	contawn
cheese	consumption	coopera
chieftain		donfous
Jesus	cotton	doodge
peacock	counter	dottar
peekaboo	document	dudawsh
peoples	donor	fonquer
seafood	follow	gaaz
season	formal	gerzoon
secret	frozen	gloods
seed	functional	gludge
siege	harmful	guvle
speaker	horoscope	harshoo
speed	jewel	haatch
steed	judgement	kerpong

steepness	law	larl
teacher	moral	maphondle
teaspoon	master	looga
teeth-marks	mushroom	marser
thief	normal	maufer
to speak	other	modust
to tease	overcharge	morole
	overview	nonpot
	perform	oxoloor
	popular	pulpoon
	snuggle	phonle
	spokesperson	protruff
	tolerant	snutter
	topple	sodalls
	unlawful	spotal
		stocue
		stotch
		tarkle
		thoof
		toove
		vook
		verhootch

Table 3. Average formant values of American English vowels used in this study. This table was adopted in its whole state from a study of James M. Hillenbrand "Acoustic characteristics of American English Vowels" (1995, 3103).

		/i/	/ɪ/	/e/	/ɛ/	/æ/	/ɑ/	/ɔ/	/o/	/ʊ/	/u/	/ʌ/	/ɝ/
Dur	M	243	192	267	189	278	267	283	265	192	237	188	263
	W	306	237	320	254	332	323	353	326	249	303	226	321
	C	297	248	314	235	322	311	319	310	247	278	234	307
*F*0	M	138	135	129	127	123	123	121	129	133	143	133	130
	W	227	224	219	214	215	215	210	217	230	235	218	217
	C	246	241	237	230	228	229	225	236	243	249	236	237
*F*1	M	342	427	476	580	588	768	652	497	469	378	623	474
	W	437	483	536	731	669	936	781	555	519	459	753	523
	C	452	511	564	749	717	1002	803	597	568	494	749	586
*F*2	M	2322	2034	2089	1799	1952	1333	997	910	1122	997	1200	1379
	W	2761	2365	2530	2058	2349	1551	1136	1035	1225	1105	1426	1588
	C	3081	2552	2656	2267	2501	1688	1210	1137	1490	1345	1546	1719
*F*3	M	3000	2684	2691	2605	2601	2522	2538	2459	2434	2343	2550	1710
	W	3372	3053	3047	2979	2972	2815	2824	2828	2827	2735	2933	1929
	C	3702	3403	3323	3310	3289	2950	2982	2987	3072	2988	3145	2143
*F*4	M	3657	3618	3649	3677	3624	3687	3486	3384	3400	3357	3557	3334
	W	4352	4334	4319	4294	4290	4299	3923	3927	4052	4115	4092	3914
	C	4572	4575	4422	4671	4409	4307	3919	4167	4328	4276	4320	3788

Table 4. Measured formant frequencies of /i/-tokens in different stimuli. The numbers in red indicate recordings which were used together with the rest to calculate our speaker's average values, but were not included in the experiment itself due to distortion caused by the manipulation process.

Stimulus /i/	**Formant frequency in Hz**		
	F1	F2	F3
Beast	278	2462	3003
Beaver	273	2266	2874
Beefeater	287	2306	2836
Beekeeper	280	2441	3133
Beep	259	2439	2949
Cheese	286	2352	2893
Chieftain	310	2380	2847
Coffee	301	2297	2851
Deepen	293	2323	2854
Feature	298	2407	2966
Feeble	283	2355	2999
Feed	263	2587	3180
Feedstock	303	2302	2916
Fever	283	2258	2839
Jesus	292	2203	2874
Peacock	280	2473	2915
Peekaboo	310	2398	2863
Peoples	310	2442	2984
Seafood	283	2324	2801
Season	298	2285	2880
Seed	276	2477	3038
Siege	277	2443	2997
Speaker	285	2370	2845
Speed	272	2482	2993
Steed	262	2578	2990
Steepness	309	2376	2970
Teacher	286	2417	2886
Teaspoon	293	2377	2870
Teeth-marks	310	2436	2857
Thief	286	2420	2892
To speak	284	2391	2836
To tease	297	2483	2908
Average	287.7	2392.2	2923.1

Table 5. A list of settings for each /i/-stimulus to accurately measure its formant values in Praat.

Stimulus /i/	Ceiling frequency	nF
Beast	4 800	4
Beaver	4 450	4.5
Beefeater	4 600	4.5
Beekeeper	4 300	4.5
Beep	4 320	4
Cheese	4 400	4.5
Chieftain	5 500	5
Coffee	4 000	4.5
Deepen	4 200	4.5
Feature	4 800	4.5
Feeble	4 615	4
Feed	4 950	4.5
Feedstock	5 000	4.5
Fever	5 600	5
Jesus	5 500	5
Peacock	5 500	5
Peekaboo	4 400	4
Peoples	4 500	4.5
Seafood	5 500	5.1
Season	4 900	4.5
Seed	4 320	4.5
Siege	4 500	4
Speaker	4 600	4.5
Speed	4 400	4.5
Steed	4 500	4.5
Steepness	5 500	5
Teacher	5 500	5
Teaspoon	5 000	5
Teeth-marks	4 500	4.5
Thief	4 700	4.5
To speak	4 400	4.5
To tease	4 300	4.5