FACULTY OF ARTS
Charles University

MASTER THESIS

Monika Kučerová

The impact of bidialectal input on children's perceptual adaptation of vowels in a non-native language

Vliv bidialektálního inputu na percepční adaptaci vokálů v nerodilém jazyce u dětí

Institute of Phonetics

Supervisor of the master thesis: Dr. Kateřina Chládková, M.A.

Study programme: Fonetika

Prague 2025

I declare that I carried out this master thesis on my own, and only with the cited sources, literature and other professional sources. I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

I would like to extend my appreciation to my supervisor, Dr. Kateřina Chládková, M.A., for her guidance and support throughout the entire process of writing this thesis. I am grateful to have benefited from her expertise, insight, and encouragement.

I also wish to thank Lucie Doležal Nováková and Kateřina Machová for their assistance with data collection, Marie Smejkalová for her help with segmentation, and Leona Straková for lending her voice to some of the task instructions.

This work was made possible by funding from the Czech Academy of Sciences, project LQ300252401, and the European Regional Development Fund, project "Beyond Security: Role of Conflict in Resilience-Building," reg. no.: CZ.02.01.01/00/22_008/0004595.

Title: The impact of bidialectal input on children's perceptual adaptation of vowels in a non-native language

Author: Monika Kučerová

Institute: Institute of Phonetics

Supervisor: Dr. Kateřina Chládková, M.A., Institute of Czech Language and Theory of Communication

Abstract: I investigate 5–6-year-old Czech L2 learners' adaptation to a difficult contrast. 51 children with non-immersive L2 experience were exposed either to native (General British English, GBE), non-native (Czech-accented English, CE), or multi-accent (Multi) input. All watched training videos and a test video with a word identification task featuring novel native talkers. The task included minimal pairs with the TRAP and DRESS vowels, homophones and dissimilar words. Multi-accent input was not confirmed to enhance novel accent adaptation. GBE and CE groups replayed minimal pairs more often than dissimilar words, the Multi groupwas not confirmed to replay minimal pairs more often. All groups identified dissimilar words above chance. Only the CE group identified minimal pairs above chance (produced as homophones, during training), suggesting reliance on multiple cues or exemplars facilitated by congruence in speaker and listener categories. No group identified minimal pairs at test above chance. Despite long-term experience with the contrast, children at this stage of L2 learning still use equivalence classification when processing minimal pairs with TRAP and DRESS.

Keywords: non-native language acquisition, vowel perception, bidialectal input, preschool children, accent adaptation

Název práce: Vliv bidialektálního inputu na percepční adaptaci vokálů v nerodilém jazyce u dětí

Autor: Monika Kučerová

Ústav: Fonetický ústav

Vedoucí bakalářské práce: Dr. Kateřina Chládková, M.A., Ústav českého jazyka a teorie komunikace

Abstrakt: Práce zkoumá adaptaci na obtížné anglické vokály českými 5–6letými dětmi s dlouhodobou zkušeností s angličtinou. 51 dětí sledovalo pohádky s rodilým přízvukem (britská angličtina, GBE), nerodilý přízvukem (angličtina s českým přízvukem, CE) nebo kombinací přízvuků (Multi). Všechny děti sledovaly 4 animovaná tréninková videa a poté testovací video s úkolem na identifikaci slov od nových rodilých mluvčích. Úkol zahrnoval minimální páry, minimální páry vyslovené jako homofony a výrazně odlišná slova. Analýza nepotvrdila efektivnější identifikaci slov u dětí vystavených kombinaci přízvuků. Skupiny GBE a CE přehrávaly minimální páry častěji něž výazně odlišná slova, u dětí ve skupině Multi se častější přehrávání minimálních párů nepotvrdilo. Všechny skupiny identifikovaly odlišná slova nad úrovní náhody. Přestože akustická analýza neodhalila rozdíly v produkci českých mluvčích, skupina CE identifikovala minimální páry produkované jako homofony nad úrovní náhody (během tréninku), což naznačuje využití exemplářů nebo kombinace vodítek, efektivnější zpracování těchto položek je možné vlivem shody mezi kategoriemi mluvčích a posluchačů. Žádná skupina neidentifikovala minimální páry při testu nad úrovní náhody. I přes dlouhodobou zkušenost s tímto kontrastem děti v této fázi učení L2 stále používají ekvivalenční klasifikaci při zpracování obtížných minimálních párů.

Klíčová slova: osvojování nerodilého jazyka, percepce vokálů, předškolní děti, bidialektální prostředí, adaptace na přízvuk

Contents

Preface

Worldwide, 1,453 million people speak English. Of those, more than a billion are non-native speakers (Eberhard et al., 2024). When using English for international communication, interacting with non-native speakers is inevitable. Czech non-immersive learners of English are typically exposed to General British English (GBE) or General American English (GAE) speech during classroom learning (e.g. Ministry of Education & Sports, 2024). However, communication in English as a lingua franca (ELF) often involves accents other than these, and to comprehend speech in those accents, experience with a wider range of varieties has been shown to be beneficial (e.g. van der Feest et al., 2022). Therefore, gaining experience with a variety of different L2 accents seems a desirable goal for learners who aim to increase their ability to adapt efficiently, and so comprehend speech in novel L2 accents faster with little difficulty.

Exposure to some, especially non-native, varieties is seen by some as possibly inhibiting "native-like" attainment, especially when it comes to pronunciation (Cook, 1999). Many studies focus on the attainment of L2 speech sounds, typically choosing GAE as the reference variety whose categories learners are assumed to be aiming to acquire (e.g. Lee & Iverson, 2012). For example, many studies focus on the acquisition of the /ɛ-æ/ contrast by listeners whose L1 lacks a counterpart to /æ/ (e.g. learners of Mandarin and Korean (S. Barrios & Hayes-Harb, 2021), Dutch (Escudero et al., 2012), German (Schlechtweg et al., 2023), Spanish (Escudero & Chládková, 2010), Czech (Šimáčková, 2003; Šimáčková & Podlipský, 2018; Šturm & Skarnitzl, 2011). Acquiring the contrast is beneficial in settings where the ambient variety uses it. However, the languages above, and by extension, often learners of English with these L1s, do not. It is not unlikely that L2 English speakers will meet other non-native speakers, who may not use the contrast. In this case, experience with non-native accents can lead to faster adaptation, and prevent misunderstanding.

Admittedly, when communicating with speakers whose varieties do contrast the sounds above, not perceiving relevant segments as members of different categories can lead to categorizing some words as homophones, which increases lexical competition during speech processing. A larger number of lexical entries being activated when perceiving speech does pose a challenge to successful word recognition. However, in everyday communication, because listeners are skilled in top-down processing, the advantage gained by having established accent-specific representations should override this supposed drawback. This is not to say that acquiring a contrast (in production or perception) is futile. Rather, being aware that some speakers may make use of it, while others may not, promotes flexible speech processing, helping listeners understand speech more easily and more quickly.

When it comes to the native language, frequent multi-accent input is common, especially in multi-cultural urban areas like London, Hong Kong, or Miami. How soon could exposure to multi-accent input show benefits in adapting to novel L2 accents? Many Czech children exposed to English experience multi-accent input, typically Czech-accented English along with General American or British English. In this study, I explore the effect of multi-accent input on adaptation to vowel

categories of two novel accents of English. Five- to six-year-old children were exposed to four animated videos narrated by two talkers: they heard two Czech talkers speaking English (CE), or two talkers of General British English (GBE), or one GBE talker and one CE talker. The focus is on the vowels /ɛ/ and /æ/ (as transcribed for GBE) in minimal pairs. The CE talkers who provided the input merged these vowels, while the GBE talkers produced them as different in quality. After the four training videos, children watch a test video, featuring a talker of Singapore English (SE), who merges the two vowels into /ɛ/, similarly to CE talkers; and a White South African English talker (WSAE), who, similarly to GBE talkers, contrasts /ɛ/ and /æ/. Adaptation is assessed through a word identification task which allows the participant to choose one of two sounds to match a displayed picture. Participants also completed an executive function task, and some completed a categorization task using words with synthesized vowels from a continuum between /ɛ/ and /æ/.

1 Literature review

1.1 Dealing with variability in speech

There is no fixed or straightforward mapping between acoustic properties of speech and speech sound categories. Phonologically identical utterances can be produced using acoustically different signals, and the same signal can be mapped onto different categories in perception depending on a wide variety of factors, including immediate phonetic context (Broadbent et al., 1956), f0 of the segment (Miller, 1953), or perceived gender or sex of talker (e.g. Strand & Johnson, 1996). This has been termed the lack of invariance problem (Appelbaum, 1996; F. S. Cooper et al., 1952). Differences in accent can also be seen as contributing to this lack of invariance. On the segmental level, accent-based variability between talkers involves both subphonemic and phonemic differences. To be able to comprehend speech, listeners need to be able to map cues onto optimal categories, by which they adapt to novel talkers, or, by extension, accents.

L1 accent adaptation, i.e. altering cue-to-category mappings to recognize and understand speech in a novel accent, has been widely studied in adults (e.g. Bradlow & Bent, 2008; A. Cooper & Bradlow, 2018; Kleinschmidt, 2020; Kraljic & Samuel, 2005; Maye et al., 2008). Some have demonstrated that when adults first encounter a talker with a novel accent, their speech is processed more slowly and with lower accuracy, but in a matter of minutes, they adapt (Bradlow & Bent, 2008). Adaptation brought on by perceptual learning has been shown to last, for example, in Eisner and McQueen (2006), participants showed the same extent of adaptation when tested immediately after exposure and 12 hours after exposure.

When it comes to children, already during the first year of life, they learn to dynamically adjust cue-to-category mappings for the L1 input that they are routinely exposed to, perceptually normalizing incoming speech (see Kuhl, 2004; McLeod & Crowe, 2018, for reviews). Regarding unfamiliar L1 accent adaptation, the ability to adapt is reported around the age of two: infants start recognizing familiar words presented in an unfamiliar accent around 19 months of age (Best et al., 2009). Some studies report familiar word recognition in an unfamiliar accent even before that for children that are at least briefly exposed to a familiar story in the novel accent before test (15 month-olds in van Heugten & Johnson, 2014). Two-year-olds' adaptation is already similar to that demonstrated by adults: they adapt quickly (e.g. two minutes in White & Aslin, 2011) and generalize onto untrained words (White & Aslin, 2011).

1.2 Targeted shifts and general criterion relaxing

White and Aslin (2011) provide findings that support toddlers employing an adaptation mechanism which allows them to shift cue-to-category mappings in an evidence-based way to achieve adaptation to novel stimuli. White and Aslin (2011) used manipulated stimuli with a simple shift in one vowel (GAE /ɑ/ was

shifted to /æ/). The control group was exposed to unmanipulated GAE stimuli. Participants exposed to the manipulated stimuli accepted shifted pronunciations at test as familiar words. They also generalized to words not heard during familiarization that contained the same shift, recognizing them as words. The control group did not demonstrate adaptation, indicating shifted pronunciations to be mispronunciations at test. The experimental group did not demonstrate simple widening of categories to recognize words with vowel tokens shifted in any direction. Rather, they adapted only to the shift signalled by the input, which they demonstrated by reacting to test vowel shifts that they did not experience during exposure as to mispronunciations. Studies on adults also find L1 adaptation specific to shifts observed in input (e.g. Maye et al., 2008).

Though the targeted shifts mechanism has been observed in both children and adults, it may not be used for processing natural unfamiliar accents, which typically deviate from the listener's L1 accent in multiple ways. The targeted shifts mechanism requires that shifts present in the input lead the listener to adjust cue-to-category mappings to categorize words as intended by the talker. There is a competing mechanism, which does not require any evidence to be observed. The general expansion strategy involves changing cue-to-category mappings in a uniform way: the listener increases the range of cues that map onto a given category to increase the likelihood that word-like forms map onto already established lexical entries, rather than consider them novel words. There is evidence for both targeted shifts and general expansion in toddlers' L1 adaptation. According to Schmale et al. (2012), when confronted with unfamiliar words in a novel accent, the general expansion strategy may be more likely to be used by toddlers, because it does not require fine-tuning to the input. It involves simply relaxing the criteria for matching acoustic signal with lexical form, so that cues that deviate more than was previously tolerated are now accepted as matches to lexical entries. This does result in increased word recognition, but it also causes confusion about word pairs that are very similar (see an illustration of cue-to-category mapping resulting from targeted shifts adaptation and general criterion relaxing adaptation in Figure 1.1).

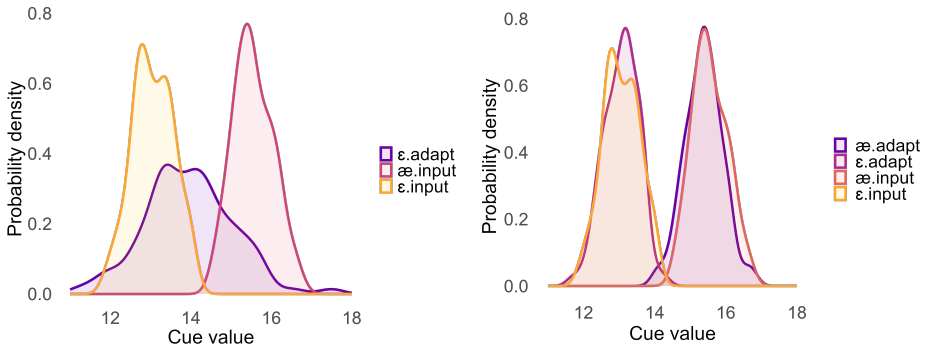

Figure 1.1 A simulated illustration of adaptation outcomes, simplified to a single cue. The x axis shows the values of the cue (e.g. F1 in ERB) that signal category identity. The two lighter distributions (with means of 13 and 15.5) are simulated cues to two categories that a listener may encounter. The y axis represents the kernel density estimates. **Left:** The purple distribution (mean = 14) illustrates the outcome of general criterion relaxing adaptation. Notice that a single category is used to process two intended categories, potentially giving rise to confusion about minimal pairs. **Right:** The purple distributions (with means of 13 and 15.5) illustrate the outcome of targeted shifts adaptation.

Creel (2012) constructed an artificial accent to assess whether three- to five-year-old's adaptation to atypically produced familiar words indicates the use of general criterion relaxing. The artificial accent differed from the children's accent only by a quality shift in front vowels to more open or more closed realizations. Eye-tracking data revealed that when presented with an atypical pronunciation, participants looked to the target more slowly, and spent less time looking at it than when presented with a canonical pronunciation. Yet, the children overwhelmingly selected the target picture from four alternatives, even in the presence of one unfamiliar alternative. The more deviant the atypical form was (expressed by the number of phonological features changed), the more frequent were non-target object responses, indicating greater uncertainty about the lexical identity of the auditory form. However, the target picture always prevailed as the most frequent response, even when atypical pronunciations included multiple feature changes. Preschoolers demonstrated the use of general criterion relaxing in adaptation. The minority of novel image responses to atypical pronunciations indicate that some aspects of unfamiliar accents are not accommodated by general criterion relaxing as automatically as others, and children may require exposure to be able to connect them with established lexical items.

A. Cooper et al. (2023) suggest that a general expansion strategy increases processing difficulty, as the observed input can be matched with more representations. So, maybe this strategy is only resorted to when the situation does not provide sufficient evidence that would help disambiguate the unit, or when the child has increased uncertainty about their representation during novel word learning.

Among the studies that examine L2 accent adaptation is A. Cooper and Bradlow (2018), who focused on phonetic adjustment to L2 English vowels by adult Dutch learners. The L2 learners endorsed more forms in a lexical decision task and showed more successful word identification than controls who did not receive training, indicating adaptation. L2 learners demonstrated higher tolerance for

deviant productions, endorsing them as words. The tolerance increase was constrained to trained shifts, items that were non-words with respect to the artificial dialect were rejected. The authors suggest that training brought on both "general increase in tolerance for atypical speech input" and "targeted adjustments to specific categories" (A. Cooper & Bradlow, 2018, pp. 84).

A. Cooper et al. (2023) hypothesized, based on Schmale et al. (2015), that because children's linguistic representations are less robust than those of adults, children are more likely to increase category variability to adapt to accented input than utilize targeted shifts. They note that this strategy can be used without the need of top-down linguistic information like lexical knowledge, as it relies on across-the-board loosening of criteria for determining if the input constitutes an acceptable match to the stored representation. Further, A. Cooper et al. (2023) posited that the more dissimilar the accent was to the native accent of the children, the more likely it was that they would use the general expansion strategy.

The targeted shifts and general criterion relaxing mechanisms differ in more than just the predicted outcomes. The targeted shifts mechanism has been connected with abstractionist approaches of speech sound representation. It has been argued that changes in the cue-to-category mappings are adjusted on the pre-lexical level, leaving lexical representations intact, which requires some level of abstraction (E. K. Johnson et al., 2022). General criterion relaxing has been said to be supported by both abstractionist and strictly exemplar theories of representation, because accepting a wider range of cues as signalling a given category can include changes in lexical representations, as well as adjustment on the pre-lexical level. While it is true that generalization to unfamiliarized words, as seen in White and Aslin (2011), could not be explained with a strictly exemplar model of representation, there has been a move towards hybrid accounts of representation, which integrate properties of abstractionist and strictly exemplar frameworks (E. K. Johnson et al., 2022). Using a hybrid account removes the conflict in the argued representational implications of finding general criterion relaxing versus targeted shifts, i.e. one does not have to assume strict abstractionist representation from observing targeted shifts.

E. K. Johnson et al. (2022) formulate the *hybrid flexibility hypothesis* that combines general criterion relaxing and targeted shifts. It is posited that children use general expansion when encountering an unfamiliar accent for the first time, and with sufficient experience that allows them to form more evidence-based expectations, they turn to using targeted shifts. At least some level of abstraction is needed to support this account, because it is assumed that children use experience with between-accent differences to adjust cue-to-category mappings.

1.3 The ideal adaptor framework

The ideal adaptor framework (IAF) proposed by Kleinschmidt and Jaeger (2015) is a theoretical account of L1 adaptation, which can, like the hybrid flexibility hypothesis, account for both the discussed adaptation mechanisms. IAF builds on the finding that linguistic categories can be inferred from speech by making use of statistical properties of the input (Kraljic & Samuel, 2005), including their distribution, variability, frequency, and probability of co-occurrence (Thiessen et al., 2013). Kleinschmidt and Jaeger (2015) aim to describe how listeners adapt

to the changing statistical properties of speech in various perceptual tasks using distributional learning. This is a process by which infants have been proposed to, for example, infer the categories of their native language (Maye et al., 2002). The use of distributional learning connects L1 adaptation to L1 category acquisition and also to L2 category acquisition (Pajak et al., 2016). Both adaptation and L2 sound learning can be treated as a problem of inference under uncertainty: the listener tries to infer the category intended by the talker upon observing a given cue value in the input (Kleinschmidt & Jaeger, 2015; Pajak et al., 2016).

Kleinschmidt and Jaeger (2015) propose that listeners hold beliefs about underlying distributions that generate the speech sound instances observed in the input. As listeners start gaining experience with a novel talker, they adapt, i.e. adjust the relevant cue distributions according to the newly observed evidence about the characteristics of the underlying cue distribution (which is believed to have generated the observed cue values). Kleinschmidt and Jaeger (2015) focus on describing adaptation at the level of segments, i.e. changes in category means and variances. It is easiest to demonstrate adaptation on this level, but adaptation as a *problem of inference under uncertainty* is not constrained to the segmental level, it can be applied to linguistic units of any complexity (Kleinschmidt & Jaeger, 2015, p. 1).

In their IAF-based account of L2 sound learning, Pajak et al. (2016, p. 903) claim that "distributional knowledge of the covariance between linguistic and socioindexical structure" is crucial for L2 comprehension. Linguistic categories are described as probability distributions, describing the probability of each possible cue value to signal the relevant category. Listeners recognize phonological categories indicated in input by using and updating their knowledge about these cue distributions. The cue-to-category inference is formalized using Bayes' rule, taking into account how likely a category is apriori, and how well it predicts the observed cue value (the likelihood). Experience with values from specific cue distributions for a given linguistic unit guides the inference. Adaptation is achieved by iteratively updating knowledge about cue distributions to incorporate recently observed evidence from input.

To understand speech, listeners must take into account information from multiple sources. Adaptation is aided by context. For example, visual context in Vroomen et al. (2007) helped participants infer the intended meaning despite deviant form. Another source of information is top-down linguistic knowledge. This knowledge includes lexical information (Kraljic & Samuel, 2007). Other types of knowledge are also used, e.g. having encountered speech in a similar situation or setting previously, and thus being able to use that experience to process speech in the current situation (e.g. being familiarized with an accent in White and Aslin (2011)). Experience-based knowledge acquired throughout life is typically exploitable in everyday interaction, but often not exploitable in lab settings.

1.3.1 Integrating prior beliefs with observed evidence

When encountering a novel talker, adaptation depends on the selection of the initial cue distribution to be updated (the prior distribution). In other words, the beliefs held about a talker before perceiving their speech influence, at least to some degree, how linguistic units produced by them are categorized. Studies on

perceptual normalization serve as illustrations for this. For example, exposing two groups of listeners to androgynous-sounding speech while making one group think that the talker is female, and the other that the talker is male, has been found to affect which speech sounds listeners perceive (K. Johnson et al., 1999). The very initial cue distribution chosen to model the speech of the novel talker influences how speech is perceived. This initial choice is done based on context and top-down linguistic knowledge, among other sources of information. For listeners adapting to L2 speech, their L1 is assumed to bias the used cue-to-category mappings (Pajak et al., 2016).

The extent of adaptation is also modulated by how confident the listener is that the prior cue distribution models the talker's speech well. This is also applicable to L2 learning (Pajak et al., 2016). The main differences between L2 learning and L1 adaptation lie in the presence of L1 influence on priors used in the L2, and the resulting difference between L1-biased prior beliefs and "ideal" beliefs used to process the L2 speech (i.e. those that result in successful adaptation). In the cue-to-category mapping updating process, the confidence that the listener has in the prior distribution plays an important role: it determines the amount of input needed to change the category. In other words, it modulates the flexibility of the cue distribution for a given category. High confidence in a prior results in a greater amount of evidence needed to change the category to a given extent, compared to low confidence, which allows less evidence to change the category to the same extent. Low confidence in a prior produces flexible categories, high confidence more rigid categories.

When first encountering an unknown talker, input from them is processed using prior beliefs specific for the talker's gender, variety, etc., depending on previous experience. As more input is observed, it is integrated with the prior beliefs, constructing a talker-specific cue value distribution. In simple terms, the beliefs about a category at any point in time can be imagined as a normal distribution, defined by its mean and variance, inferred from some tokens of the category perceived in the past together with tokens provided by the current talker. In its mean and variance, the resulting posterior distribution reflects the prior as well as the observed samples. See Figure 1.2 for an illustration of how posterior distributions can differ based on the strength of the prior. Note that the gradient adaptation outcomes illustrated in Figure 1.2 can be similarly achieved using smaller and larger amounts of evidence, as well as a strong or weak prior. The important thing is the proportion of dose of evidence to strength of prior. IAF predicts that once the situation changes, for example, a new talker starts speaking, the situation resumes: a prior deemed appropriate is used, and new evidence is integrated with it to arrive at a posterior distribution that fits the tokens produced by the present talker better than the previous distribution (before updating).

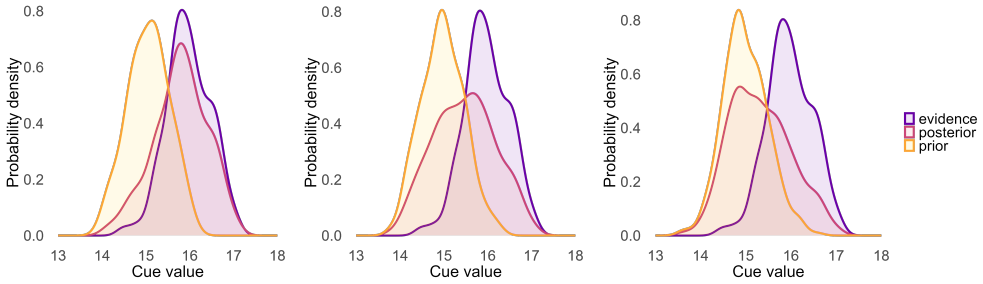

Figure 1.2 Each plot illustrates a prior distribution (generated using a pseudo-count of observations), simulated cues provided in observed evidence, and a posterior that results from combining the respective prior with the evidence. **Left:** a weak prior, characterized by four times fewer observations than present in the evidence. **Middle:** a medium strength prior that uses the same number of observations from experience and as evidence. **Right:** a strong prior, characterized using a pseudo-count of twice as many observations than provided in the evidence.

IAF predicts gradient adaptation effects. In congruence with this, Cummings and Theodore (2023) conceptualize learning as "a measurable difference in perception between listeners groups who received differential biasing exposure." Learning outcomes may be variable be due to different methods, including the use of different exposure dose, i.e. how much evidence listeners are exposed to. The dose of evidence may interact with the specific task used, with the age of the listener, and the distance between the talker's and listener's accents. IAF sees learning outcomes as modulated by the dose of evidence. Cummings and Theodore (2023) used a lexically guided perceptual learning (LGPL) paradigm, where listeners are exposed to speech in a lexical decision task with manipulated speech sounds, followed by a phonetic identification test. Extent of adaptation was found to be graded to reflect quantity of observed evidence, and was not lessened by lack of consistency in the input.

In the view of IAF, both targeted shifts and general criterion relaxing are possible adaptation strategies, but only if evidence points to the relevant change. Adaptation is seen by the IAF as a process in which listeners iteratively update their knowledge about the cue distribution, aiming to approximate the talker's production more closely on each iteration. If cues signal higher variance, the ideal adapter should reflect that in the inferred generative model, which would translate to using more variable categories. If cues signal a different category mean, the category used by the ideal adapter should shift its mean. Note, however, that for both changes in variability and mean, the extent of change will be modulated by the listener's confidence in prior beliefs. Further, it is important that IAF recognizes that listeners do not categorize speech sounds solely based on auditory percepts. Context-based knowledge and visual cues, among others, also modulate adaptation. Listeners can achieve adaptation by either targeted shifts or general criterion relaxing, but these changes are contingent on observing evidence (including contextual evidence, like visual cues) that points to the relevant change.

Where does this approach fit with respect to abstractionist and exemplar theories of speech sound representation? Abstractionist theories (e.g. Hyman, 1970) typically assume talker normalization, by which cue values are transformed so that normalized input from any talker agrees with the normalized values of

other talkers. Exemplar theory accounts (e.g. Pierrehumbert, 2003) do not need to tackle normalization, because exemplars of linguistic units are stored for each talker in at least some acoustic detail, and new talkers are recognized based on acoustic similarity of stored tokens. Talker-specific cue-to-category mappings are employed in exemplar theory-based models. Generalization proceeds on the basis of similarity to exemplars provided by other talkers. The IAF agrees most with a hybrid account: in situations where perception is guided mostly by prior beliefs, predominantly use of abstract categories is used. In situations where prior beliefs are weak and adaptation needs to be guided by evidence, individual exemplars affect the cue-to-category mapping most (Kleinschmidt & Jaeger, 2015).

1.3.2 Generalization over talkers

Among other factors, cue distributions vary based on the talker. IAF (Kleinschmidt & Jaeger, 2015) employs Bayesian thinking in formalizing the inference of which previously encountered generative models (talkers) are most relevant in perceiving current input. Based on information immediately available in the current situation, the listener infers which experience is similar to the current situation and uses it as a starting point for adaptation to the current talker (if adaptation is necessary) to minimize the amount of learning needed upon encountering each new talker. Beliefs about generative models are thought of as distributions over generative models, which attribute each possible model with probability that expresses how well the current observation predicts each of the generative models.

Relevant experience is selected based on the prior distribution over generative models, which allows the listener to focus on generative models that can serve as a good basis for adapting to the current talker. Pajak et al. (2016) describes the process of L2 learning as involving probabilistic inferences which allow learners to hold beliefs about a range of generative models, associating each with some degree of uncertainty in light of the current situation. If chosen appropriately, the prior facilitates faster adaptation and higher comprehension.

The larger the number of models excluded (based on low probability of being applicable in the current situation) before exposure to a new talker, the greater the probability over the remaining models. If the currently perceived model is actually similar to some of the ones with greatest attributed probability, adaptation is predicted to be faster, i.e. require fewer observations. However, more specific prior beliefs are less flexible, they need more evidence to lead to change if the current talker is not sufficiently similar to those initially anticipated.

Importantly, in order to generalize, listeners must be able to recognize familiar talkers or situations, and to make use of structure over talkers (Pajak et al., 2016). In line with both Pajak et al. (2016) and Kleinschmidt and Jaeger (2015), having high confidence in prior beliefs about the types of talkers that exist leads the listener to require less evidence from an unfamiliar talker in order to settle on a generative model that is most fitting to the current situation. If the listeners' experience is wide, the chosen model should already approximate the cue-to-category mappings used by the current talker well. Hence, the listener should closely approximate the talker's cue-to-category mappings faster than a listener with low-confidence experience, who will need more time to adapt (Pajak et al.,

2016). This perspective is able to explain why, for example, multi-accent toddlers in van der Feest et al. (2022) were able to recognize words more efficiently in both their varieties than their mono-accent peers. The multi-accent children had wider experience, which they could exploit for faster adaptation.

Kleinschmidt and Jaeger (2015) mostly discuss talker-specific adaptation, however, the IAF is not limited to talker-specific cue-to-category mappings. Listeners may establish beliefs about any re-encountered context if it provides speech statistics that are systematically different from other contexts. The information that is taken into account for narrowing down the possible range of generative models is assumed to include the situational context as well as acoustic-phonetic information from the speech of the novel talker.

1.3.3 Accent similarity

Although children demonstrate an ability to adapt early on, their adaptability lags behind that of adults (A. Cooper et al., 2023). Differences in adaptation success often arise in non-ideal conditions, e.g. with a distant accent (A. Cooper et al., 2023) or speech shaped-noise over the stimuli (Nathan et al., 1998). The reason behind this is often said to be children's less developed lexicon, especially as greater vocabulary size is frequently found to correlate with more successful adaptation (Harte et al., 2016).

Adaptation to novel L1 accents has been shown to be contingent on the acoustic-phonetic similarity of the novel accent to the listener's accent (A. Cooper et al., 2023) or to previously encountered accents (Xie & Myers, 2017). Xie and Myers (2017) aimed to isolate conditions that facilitate generalization across talkers during accent adaptation. They exposed adult native listeners to Mandarin-accented English from one or multiple talkers. Listeners then judged novel words from an unfamiliar talker with the same accent. Regardless of the number of talkers heard at training, they generalized most successfully when the new talker was similar to talker(s) heard during training: exposure to a talker who showed similar cue distributions boosted word recognition. Talker similarity judgements were not found to affect whether or not generalization occurred.

Xie and Myers (2017) suggest that listeners do not adapt to an accent, but to the acoustic-phonetic properties in the speech of encountered talkers. This is in line with Kleinschmidt and Jaeger (2015), who also do not see generalization as necessarily accent-specific, but rather guided by similar cue distributions, which may not be grouped strictly based on accent. Xie and Myers (2017) conclude that if two talkers show similar cue distributions for a target segment, experience with one of the talkers can be used when processing speech from the other talker.

The relative distance of listener's and talker's accent may have a significant impact on comprehension of accented speech by children. A. Cooper et al. (2023) examined how the distance between a talker's and listener's L1 phonology influences adaptation, testing adults and 27-month-olds using eye-tracking. They employed three artificial accents, which differed with respect to degree of similarity to listeners' L1 Canadian English phonology. Adults adapted to the artificial accent irrespective of its distance from L1. Toddlers, on the other hand, only demonstrated adaptation to the closest of the three accents. Upon being exposed to the two more distant accents, their word recognition was impaired not only

for these accents, but also to their L1 accent. In congruence with the hybrid flexibility hypothesis (E. K. Johnson et al., 2022), children exposed to a distant accent demonstrated slower word recognition in their own accent than their peers without exposure. This indicates that children used the general criterion relaxing strategy to cope with the distant accent upon first encounter.

1.4 Multi-accent experience

Even though many studies portray toddlers as capable of accent adaptation, they often note that listeners do not achieve the same adaptation abilities as adults until adolescence. E. K. Johnson et al. (2022) contrasts this with production studies, where, surprisingly, children usually adopt novel accents faster and with greater success than adults (e.g. Tagliamonte & Molfenter, 2007). Next to a higher impact of accent distance on toddlers' compared to adults' adaptation, an additional factor that modulates accent adaptation is the number of accents a listener has long-term experience with (van der Feest et al., 2022).

Before two years of age, multi-accent children seem to have a disadvantage in accent adaptation compared to their mono-accent peers. In their study, van Heugten and Johnson (2017) report that 12.5 month old children exposed to a single accent showed word recognition in their accent, while their multi-accent peers did not recognize words in one of the accents they were routinely exposed to. The multi-accent toddlers only demonstrated word recognition in that accent at 18 months (which corresponds to the age when toddlers start to recognize words in unfamiliar accents). Vocabulary acquisition reported for the two groups of children was not different. E. K. Johnson et al. (2022) note that it seems that multi-accent input in van Heugten and Johnson (2017) did not affect lexical acquisition, rather the performance that was observed was the result of the way multi-accent children mapped cues onto categories for the input they were exposed to in the lab.

Buckler et al. (2017) reveal that the extent of accent variation children are exposed to on a regular basis impacts their word recognition in a looking-while-listening task. Mono-accent 24-month-olds demonstrated more looking time to the target image than children exposed to the socially dominant accent and a non-native accent. Mono-accent toddlers also shifted their gaze to the targets faster than multi-accent children. However, when 34-month-olds completed the same task, the difference between the two groups was no longer present. Both mono- and multi-accent (exposed to native and non-native talkers) recognized words with similar success rates and shifted to targets with similar speed. Buckler et al. (2017) speculate that mono-accent children use a different word recognition strategy than their multi-accent peers. They posit that the difference originates from mono-accent toddlers receiving less variable input, and therefore using narrower predictions for the acoustic-phonetic forms in the incoming speech, allowing for faster word recognition based on evidence that conforms to their beliefs. This perspective is similar to a scenario described by Kleinschmidt and Jaeger (2015), where a listener holds strong narrow beliefs about cue-to-category mappings in a given situation, and thus adapts fast to predictable input. This effect may dissipate by 34 months due to a large quantity of additional input, allowing multi-accent children to use priors with greater confidence, and thus adapt faster

than at 24 months.

On the other hand, multi-accent children have been reported to, in comparison with to mono-accent peers, adapt with less evidence. At 24 months old, Dutch infants with "mixed input" (nonlocal parents) noticed mispronunciations in speech with the dialect of their parents, while infants with uniform input (local parents) did not detect these, not being familiar with the dialect (van der Feest & Johnson, 2016).

Kartushina et al. (2022) offer insight into how word learning is modulated by toddlers' accent exposure: they test two-and-a-half-year-old toddlers from mono- and multi-accent backgrounds on their word recognition after exposure to accented speech. Toddlers received mono-accent or multi-accent input from a tablet storybook in kindergarten, and were tested on remembering four pseudowords from the story using a 4AFC picture selection task. There was no difference between the groups in word learning. In their exploratory analysis, Kartushina et al. (2022) indicate that out of the nine storybook sessions, both groups of toddlers benefited to a similar extent from the first five sessions, but toddlers who were exposed to multiple accents at home further improved in their word identification after the four following sessions, while those from mono-accent households did not improve further.

Potter and Saffran (2017) addressed the question whether short-term multi-accent experience helps infants comprehend speech in an unfamiliar accent. They exposed 15- and 18-month-olds to stretches of infant-directed speech from multiple talkers. After exposure, 15-month olds discriminated words from non-words in a familiar, but not in an unfamiliar accent. The same was true for 18-month-olds exposed to a single unfamiliar or familiar accent. 18-month-olds, however, when they were exposed to multiple accents, did discriminate words from non-words in an accent they had not heard before. Potter and Saffran (2017) illustrate that the degree to which between-accent variability can be used in learning is modulated by age. At 18 months, infants do seem to benefit from multi-accent exposure when faced with a novel unfamiliar accent, in contrast to 15-month-olds. These findings are interpreted as exposure to accent variability facilitating a wider range of word tokens to map onto lexical representations, in line with general criterion relaxing.

Durrant et al. (2015) reported that multi-accent 18-month-olds recognized familiar words that were canonically produced, but also mispronounced. They used an eye-tracking paradigm to assess the looking time at the target and distractor images before and after auditory form presentation. While multi-accent children looked significantly more to the target image after hearing either a mispronunciation or a canonical pronunciation, their mono-accent peers looked less to the target image after hearing a mispronunciation of the familiar word. Durrant et al. (2015) interpret this as evidence of multi-accent children either using general criterion relaxing to adapt to lab input, or having less precisely specified lexical representations.

The studies above show that both multi- and mono-accent children adapt to L1 accents, with age playing and important role in the manner and efficiency of adaptation. However, lab findings may not be generalizable to exposure outside the lab. Input that is used in the lab often includes a single simple shift (e.g. a front vowels shifted down in Creel, 2012; Maye et al., 2008; White & Aslin, 2011)

as a novel accent. These studies themselves recognize that the extent of exposure needed for adaptation to "natural" accents is likely not extrapolatable from their findings (White & Aslin, 2011, p. 13). Some studies use more ecologically valid designs: perceptual learning has been observed for items heard in a story for adults (Eisner & McQueen, 2006), and also in 8 year old L2 listeners (Hu, 2021). However, there may be other constraints on lab performance. For example, E. K. Johnson et al. (2022) point to the possibility of participants entering the lab with different assumptions about which is the ambient accent in this new environment.

1.5 Accent adaptation by L2 learners

Kleinschmidt and Jaeger (2015) successful at helping to explain L1 accent adaptation, and has been extended to also account for L2 category acquisition by Pajak et al. (2016). L1 listeners, even toddlers (Schmale et al., 2012), adapt fast, and adaptation-triggered changes can last for days after the listener is no longer exposed to the talker (Kraljic & Samuel, 2005). While adaptation in the L2 may use context and distributional learning in a way that is similar to L1 adaptation, other variables, like phonolexical misrepresentation (capturable in prior beliefs), may hinder adaptation in unique ways not consistently observable on L1 listeners. Perceptual adaptation by L2 learners has been examined to a lesser degree than by L1 listeners.

There are some similarities between L1, L2 learning and adaptation. In contrast with the adult L1 listener, and similarly to the L2 learner, an infant does not know the categories of the target language. Both the infant and L2 learner are posited to be able to infer the categories of the language using distributional learning from input, often from unlabeled tokens (Maye et al., 2002; Pajak et al., 2016). However, infants and L2 learners start this process with vastly different prior knowledge.

Viewing L2 adaptation through the lens of Kleinschmidt and Jaeger (2015), L2 learners may be compared to toddler L1 learners with respect to their uncertainty about the incoming input: like the L2 learners in A. Cooper and Bradlow (2018), toddlers may need more evidence to adapt. This is attributed to children holding high confidence prior beliefs, hence changing the cue distribution for the given category may take a larger quantity of instances from input, compared to adult listeners. Following this line of thinking, A. Cooper et al. (2023) note that a strong confidence in the priors may be accompanied by targeted adjustments being made, as opposed to general criterion relaxing. This is presumably because the latter is a more dramatic change, as it results in a greater number of lexical candidates for a given word instance.

When a novel talker is encountered, the ideal adapter infers the cue distribution that corresponds to his speech sound categories to comprehend their speech. The adult ideal adapter has experience to use as a starting point for adaptation, e.g. which categories there are in the varieties he has encountered. L2 learners, however, may not have this knowledge. Or, they may depend on knowledge transferred from the L1, comprehending speech using L1-based prior beliefs, and hence likely different categories. Kleinschmidt and Jaeger (2015) describe the ideal adaptor as an "ideal phonetic categorizer", because it is presumed that the listener is able to label every token from input with the intended category, op-

timally approximating intended categories using observed cue values. This is a reasonable approximation for L1 listeners. However, if used for L2 learners, it can distort our understanding of how they adapt. Pajak et al. (2016) posits that a primary constraint on L2 acquisition is the learners' language experience, which may be mostly experience with the L1. Rooted in the notion that L2 listeners' categorization boundaries may not reflect cue distributions signalled by the observed cue values, this perspective has the potential to provide a better fit to L2 adaptation.

Some frameworks view the initial stage of L2 learning as using a duplicate system of L1 categories (Escudero, 2005; Van Leussen & Escudero, 2015). Pajak et al. (2016) sees as the starting point of L2 acquisition a set of hypotheses about its grammar, rather than a copy of the L1 system. The hierarchical hypotheses, updated in light of new evidence, determine how learners represent the L2, including the cue-to-category mappings they use. Similarities that L2 learners perceive between L1 and L2 are distorted, and this distortion is also affected by the extent of uncertainty about L2 distributions, which is modulated by proficiency, regularity of use and recency. These are expected to have the largest impact during the early stages of learning (due to limited evidence from L2 input). Learning an L2 involves continuous changes of the learner's beliefs about the L2. Thanks to L2 exposure, learners transition from relying to a large degree on beliefs biased by the L1, to beliefs determined largely by L2 experience.

Pajak et al. (2016) takes the perspective that the L1 bias that modulates L2 speech sound adaptation is the result of a trade-off for how well listeners have mastered their L1. From this perspective, it is possible to speculate that children, whose L1 speech sound representations are not yet as robust as adults', may benefit from this trade-off during L2 adaptation. However, L2 learners may be biased by their L1 in which statistical regularities in L2 input they attend to. This may constrain L2 learning.

Some studies have suggested that adaptation is also modulated by whether the talker's speech is native or non-native. There is some evidence that when L1 listeners perceive non-native speech, the strength of the accent modulates comprehension (Wambacq et al., 2023), word recognition speed and accuracy (Porretta et al., 2016). On the segmental level, the cause may be larger category variability or phonological misrepresentation. Pajak et al. (2016) indicates that in the case of L2 listeners and talkers sharing their L1, L2 statistical cues may conflict with L1 characteristics less often, which can facilitate comprehension, as observed evidence should align well with prior expectations. The more similar the target L2 accent to the listener's L2 accent, the larger the possible proportion of L2 characteristics that overlap well with prior beliefs, facilitating adaptation. Non-native talkers employed in the current study share their L1 with the listeners, which may aid adaptation. L2 listeners have been reported as finding L2 speech more intelligible than native listeners when the speech they were exposed to was produced by talkers who shared their L1 (Hayes-Harb et al., 2008).

1.5.1 L2 category learning

Some studies have indicated that exposure to between-talker variability, as part of high-variability phonetic training, facilitates more efficient L2 category estab-

lishment than exposure to a single talker (Hazan et al., 2005). This is presumably due to listeners being better able to separate non-distinctive aspects of L2 speech from relevant cues to the identity of the linguistic. However, many find no benefit of HVPT for adults (Xie & Myers, 2017) or children (Brekelmans, 2020).

Schmale et al. (2015) present evidence that exposing 24-month-olds to indexical (variable age and sex) or social (silent gesturing) stimuli aided word recognition of previously trained words produced by a new accented talker, even when the children were not previously exposed to that accent. They claim that exposure to diversity (indexical or social) increases acceptance rate for word productions that deviate from lexical representations.

Hu (2021) found that multiple talker exposure provided no benefit for child L2 learners' accent adaptation using a referent selection task with non-words. Participants did not show generalization, their word recognition boost was limited to words heard in a story they were exposed to. Hu (2021) interprets this using an exemplar account of representation: word representations that recently received shifted exemplars guided perception, but representations that did not receive shifted exemplars were not matched with their shifted tokens.

Building on the distributional learning hypothesis (Maye et al., 2002), Rost and McMurray (2009) claim that variability helps infants adjust or maintain categories in a way that facilitates use of a robust lexical contrast. This allows infants to learn minimal pairs, which is something that 14-month-olds have been shown to not be able to do in previous research with low-variability familiarization stimuli (e.g. Fennell & Werker, 2003). Rost and McMurray (2009) present evidence that in a switch-task design, 14-month-olds' learning of minimal pairs is aided by acoustic-phonetic variability in the speech of multiple talkers. Their multiple-exemplar task used 54 instances of a given word spoken by 18 people in single word presentation. Infants demonstrated learning of minimal pairs in both trials where they indicated that the minimal pair member was incorrectly used with a mismatching familiar image, and with an unfamiliar image.

When it comes to production, Kartushina and Martin (2019) provide evidence that HVPT with multiple talkers, rather than a single talker that shows the same acoustic-phonetic variability, promotes establishment of L2 speech sound categories in naive adult learners. Participants who were trained on speech from multiple talkers showed more stable realizations of vowel sounds that form a confusable contrast, indicated by greater accurracy and compactness of produced vowel tokens. Only participants trained on multiple talkers generalized their more stable and accurate vowel imitation to a new, unfamiliar talker. Training with a single talker did not lead to such generalization.

In sum, there is some evidence that infants and older children benefit from multi-talker input when it comes to the closely tied processes of phonological category establishment (Hazan et al., 2005), word recognition (Schmale et al., 2015), and minimal pair word learning (Rost & McMurray, 2009). This benefit is in line with both exemplar theories of speech sound representation (Pierrehumbert, 2003) and the distributional learning hypothesis (Maye et al., 2002). Multiple tokens that show acoustic-phonetic variability (as opposed to minimal or no acoustic-phonetic variability) facilitate the use of more robust phonological categories, which in turn lead to robust lexical item encoding.

1.5.2 Type of contrast and phonolexical misrepresentation

When it comes to L2 accents, adaptation may be hindered by processing problems on several levels. On the segmental level, generalization beyond trained talkers may be contingent on the type of L2 contrast (A. Cooper & Bradlow, 2018). Children may be affected differently than adults due to their more broad L1 phonolexical representations (Buckler et al., 2017; Schmale et al., 2011).

To approximate optimal adaptation to L2 speech sound categories, learners may need to utilize a different number of categories than they usually employ in L2 processing. Hence, L2 adaptation can also depend on which speech sound categories the listener has established. In the case that relevant L2 categories have not been established, short-term exposure is not likely to bring about their formation (Flege & Bohn, 2021). In this study, adaptation to a difficult contrast is examined: Czech learners typically struggle with English words that include /ɛ, æ/ (Šimáčková, 2003; Šturm & Skarnitzl, 2011), as both map onto L1 ɛ. Acquisition of the contrast is a prerequisite to successful adaptation. Learners similar to the presently studied population have been shown to differentiate this difficult contrast in production (Kučerová & Šimáčková, 2025; Simon et al., 2016).

Various models try to account for L2 category formation (e.g. Best & Tyler, 2008; Escudero, 2009; Flege, 1995; Flege & Bohn, 2021). The Second Language Linguistic Perception (L2LP) model (Escudero, 2005) and its revised form (Van Leussen & Escudero, 2015) build on Best and Tyler (2008) and Flege (1995) to conceptualize phonological category formation in adults as driven by lexical item learning. The aim is to model the development of L2 speech perception from naive to native-like performance. One of the tenets is the optimal perception hypothesis (Escudero, 2009), which states that at the beginning of acquisition, learners perceive the sounds of L2 how they would perceive them in L1 speech, i.e. they transfer categories from the L1. The degree of acoustic difference between tokens of two categories determines the development of the contrast, resembling to the notion of similar phones in Flege (1995). Relevant to this study, L2LP outlines the development of two contrasting L2 categories whose instances are acoustically close to typical instances of a single L1 category, calling it a "new scenario" and equating it with PAM-L2's "single-category L2 contrast assimilation" (Best & Tyler, 2008). Van Leussen and Escudero (2015), Flege (1995), and Best and Tyler (2008) agree that such a contrast is difficult to acquire.

In congruence with Kleinschmidt and Jaeger (2015), Van Leussen and Escudero (2015) see L2 speech sound learning as supported by utterance meaning, with no direct access to the generative model (or characteristics of the phonological category employed by a talker), which is only inferred from the input. For speech comprehension, it is crucial how well the listener approximates the talker's intended meaning when interpreting an utterance. Van Leussen and Escudero (2015) view perceptual learning as a result of lexical representation updating, which is motivated by the need to adapt (to increase comprehension). Van Leussen and Escudero (2015) recognize four levels of linguistic units: acoustic, phonetic, phonemic and lexical. Neighbouring levels are connected. Word recognition follows from recognition of acoustic forms to recognition of lexical forms. At any level, the path from acoustic to lexical form can branch, possibly returning a new lexical candidate at the end. For each lexical candidate, level links in the path from the phonetic to lexical level are assigned weights. The

resulting word form is chosen based on the strength of the links on its path. In this way, knowing a language is encoded in strengths of the path links for any acoustic word form.

In the revised version, perception and recognition are allowed to interact by making the order of the two levels in the chain flexible, allowing a partly top-down perspective on word-recognition. Learning is thought of as updating the strengths of connections. As long as comprehension is sub-optimal, updating proceeds. This bears resemblance to Kleinschmidt and Jaeger (2015): both approaches view learning as iterative, and both allow top-down influence on speech sound categorization. In other ways, the two accounts are complementary: Kleinschmidt and Jaeger (2015) makes no claims about how speech sounds are represented in the mind, they simply offer a framework that allows for formalization of the starting point and gradual outcomes of adaptation (Kleinschmidt & Jaeger, 2015) and L2 sound learning (Pajak et al., 2016). Adding to Van Leussen and Escudero (2015), Kleinschmidt and Jaeger (2015) describe how learners' utilization of previous experience on the level of generative models may be formalized to describe similarity-based talker generalization.

A. Cooper and Bradlow (2018) use IAF (Kleinschmidt & Jaeger, 2015) to offer two perspectives on how L2 learners may adapt. They may start out with high confidence in L1 prior beliefs, therefore requiring more evidence to adapt, or with low confidence in prior beliefs due to having experience with variable L2 input, therefore adapting quickly. L2 learners may use more variable categories because their lower exposure to L2 leads to higher uncertainty about the identity of linguistic units (A. Cooper & Bradlow, 2018). This higher uncertainty is seen as greater confidence in (L1-based) priors, which means that L2 learners require more L2 input to sufficiently change their cue-to-category mappings to increase comprehension. In the case that naive L2 listeners perceive words involving contrasting units as homophones, L2 learners should show uncertainty about cue distributions signalling those difficult categories. Hence, L2 learners should adapt less efficiently to words with difficult contrasts than to words that contain L2 contrasts which also function in the L1. There is also an alternative account, which states that because L2 learners receive variable input, they may be more flexible in adjusting their cue-to-representation mappings (Weber et al., 2014). This would translate to generally higher acceptance of atypical words, meaning that learners employ a lower confidence parameter, hence adapt faster.

In A. Cooper and Bradlow (2018), Dutch listeners were exposed to an artificial novel English accent whose vowels were shifted to either map onto a different L1 vowel than before manipulation, or to not cross a category boundary. Productions were disambiguated by visual presentation of words. At test, listeners completed a word identification and lexical decision task. With respect to the /ɛ-æ/ contrast, which forms a new scenario for Dutch listeners, A. Cooper and Bradlow (2018) found overall high word endorsement rates from L2 listeners, perhaps due to the categories not being well discriminated. L2 listeners did not improve their lexical endorsement of any English-only contrast after training, but their endorsement rates prior to test were already high. These findings support the hypothesis that L2 learners utilize high-confidence L1-based priors that constrain the degree of adaptation. Further, they highlight problems with difficult contrasts: L2 listeners endorsed variable renderings of words already at pre-test.

Even if learners do establish L2 categories for a difficult contrast, adaptation may not be optimal. The problem may not be inaccurate perception, rather, the issue may be phonolexical, i.e. on the lexical representation level, where a word may be encoded with an unfitting vowel (Llompart & Reinisch, 2019; Šimáčková & Podlipský, 2018). This ties to threvised L2LP model's differentiation of pre-lexical perception and word recognition: the aim of including linked (but separate) representations for lexical and speech sound contrasts is to model how the interaction of these two levels shapes L2 acquisition.

Studies on child L2 accent adaptation are rare. In their L1, children do not seem to adapt as effectively as adults (e.g. Bent, 2014). Child L2 learners may show different adaptation strategies at different stages of L2 learning, when compared to adult L2 learners. However, even in child L2 learners, exposure as brief and limited as presentation of unlabelled isolated words may facilitate adaptation (Hu, 2021). Hu (2021) exposed L1 Mandarin eight-year-old learners of English to a story narrated in Indian English (IE). The study focused on a difficult contrast involving voicing in stops. Before and after exposure to the story, participants completed a referent selection task. After exposure, participants showed better recognition of IE words. However, along with higher word recognition, they showed higher endorsement of non-words, indicating adaptation by increasing category variability, rather than by a targeted shift guided by the input. It could not be established whether the children recalibrated their representations, acquiring the contrast, as referent selection tasks may not asses phonological contrast acquisition, it is possible that participants used a closest-match strategy and at the same time could perceive the difference between the two sounds. Importantly, adaptation to posttest stimuli was found even in the control group, which was exposed to a novel accent only during the pretest, indicating that the unlabelled isolated words were enough to induce adaptation by increasing category variability.

1.6 Cue-to-category mapping predictions

As mentioned above, when learning an L2 with a larger vowel system, 2-to-1 category mappings are typically predicted by models focusing on L2 segment or contrast acquisition (Best & Tyler, 2008; Flege & Bohn, 2021). The vowels studied here belong to such a new scenario contrast (Escudero, 2005). This is predicted to result in category merging, manifesting in low identification and discrimination abilities of the two difficult L2 vowels. In L2 learners, adaptation has been shown to be modulated by whether the shifted vowel traverses a category boundary, mapping onto a different L1 vowel than previously. L2 listeners may more readily adapt to changes that stay within a single L1 category, and deal with sounds that result from traversing an L1 category boundary by general criterion relaxing (A. Cooper & Bradlow, 2018; Hu, 2021).

Below, I offer a brief description of the varieties used in this study during training and test, focusing on vowels corresponding to GBE /ɛ, æ/, and how they may be perceived by the child learners in this study. The English narration used in this study was provided by two Czech talkers (CE), two General British English (GBE), one White South African English (WSAE) talker, and one Singapore English (SE) talker. Apart from the CE talkers, all talkers' native language

was English. I focus on the primary cue to the contrast: vowel quality. Other cues are not taken into account here. Of those that could be exploited by the participants, relative durations of the vowels were measured in recordings of word list reading (words used in this experiment, in variable phonetic context). None of the talkers used in this study exhibited contrastive production of the relevant vowels in length (see 2.2 in the Method section). F3 and higher formants, formant trajectories, and other possible cues are also not considered here.

Descriptions of the target vowels in relevant accents are accompanied by vowel plots (made with McCloy, 2012). Plots demonstrate the production of the target vowels by the six talkers used in this study. Vowels are plotted in the F1-F2 plane, and the most important information contained in these plot is the relative position and overlap of the vowels spoken by each talker. The relative position signals how similar the vowels are in quality.

1.6.1 General British English

The label "General British English" is used here in accordance with Cruttenden (2014) to describe an accent that is used across most of Britain as a "neutral" variety. As described by Bjelaković (2017), GBE has 11 monophthongs differentiated primarily by quality. The two vowels of interest in this study are the the mid front /ɛ/ and the low front /æ/. I adhere to the transcription of the low front vowel as /æ/ (following Ladefoged & Johnson, 2011), even though the vowel is increasingly more fittingly transcribed as /a/ to highlight its shifting to a lower and more central quality relative to the mid-low quality characteristic of RP in the past (Cruttenden, 2014). The traditional transcription is kept to highlight that /æ/ is different from the Czech low central /a/, and, importantly, has not been indicated to assimilate to Czech /a/ in Czech learners even in recent years, despite its lower quality (Šturm & Skarnitzl, 2011).

The contrast between /ɛ, æ/ is difficult for Czech learners to acquire, both vowels have been shown to assimilate to L1 /ɛ/ (Šimáčková & Podlipský, 2018; Skarnitzl & Rumlová, 2019; Šturm & Skarnitzl, 2011). The vowels /ɛ/ and /æ/ are distinguished primarily by quality in many English varieties including GBE, and since there is no qualitative counterpart to /æ/ in Czech, it ends up being assimilated in quality to the closest vowel, /ɛ/. However, this does not mean that Czech learners do not contrast the vowels. They merge the contrast in quality, but they may reuse phonological length from L1 to distinguish the two (Šimáčková, 2003), as Czech uses /ɛ, ɛː/ as distinct categories. See Figure 1.3 for formant measurements from wordlist reading by the GBE talkers.

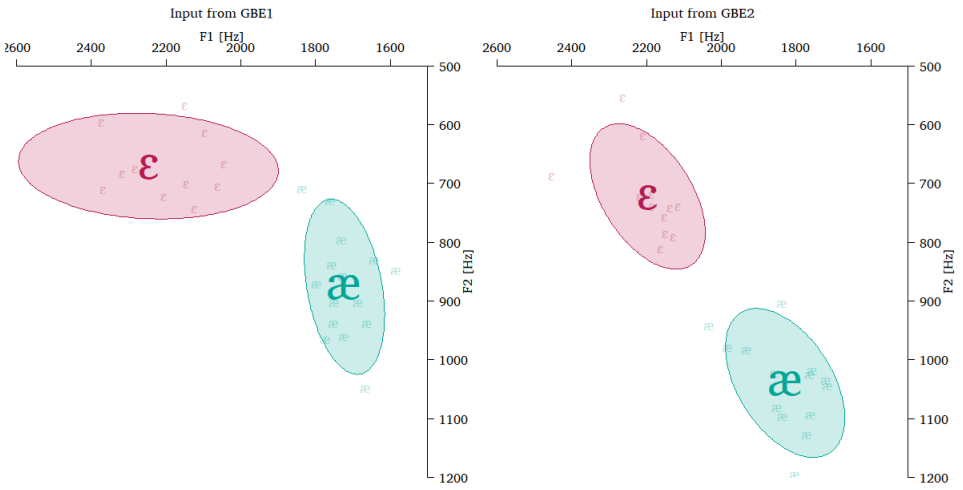

Figure 1.3 The target vowels produced by the two GBE talkers at training. Ellipses include 68% of tokens. The plots are based on 12-15 word tokens produced during the wordlist reading.

1.6.2 Czech English

Czech employs a smaller number of vowel qualities than GBE. The vowel system of the Bohemian variety of Czech distinguishes six vowel qualities. With the exception of the /iː, ɪ/ pair, each vowel quality comes in two degrees of phonological length. In contrast to GBE, vowel length is distinctive in Czech, however, none of the talkers used for stimuli recording in this study showed transfer of this feature into their English speech (see Figure 2.2). Czech talkers have been reported to produce both /ɛ, æ/ as /ɛ/ (Skarnitzl & Rumlová, 2019), even talkers with very high proficiency (Šimáčková & Podlipský, 2018). Acoustic analyses of the recorded wordlist stimuli confirmed that the two Czech talkers in this study also used this merger in their English speech. See Figure 1.4 for formant measurements from wordlist reading by the CE talker.

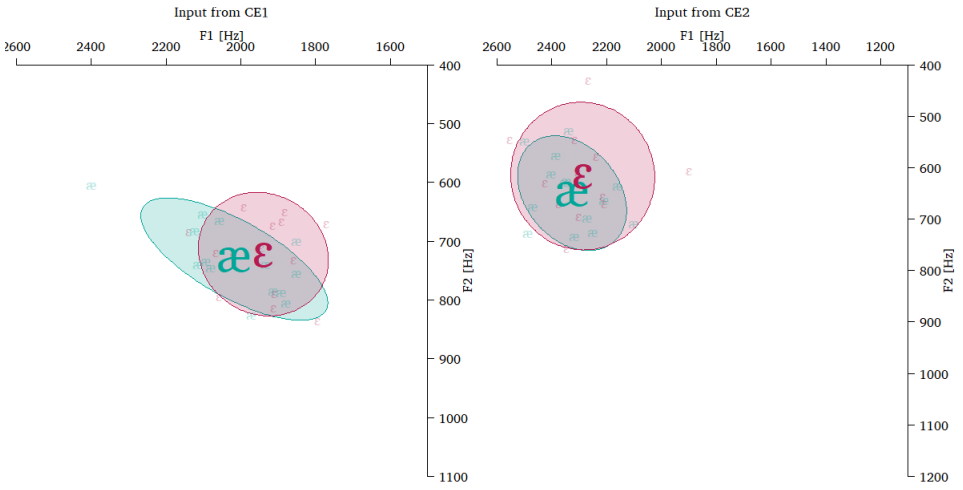

Figure 1.4 The target vowels produced by the two CE talkers at training. Ellipses include 68% of tokens. The plots are based on 12-14 word tokens produced during the wordlist reading.

1.6.3 White South African English

The label White South African English (WSAE) is used in congruence with Bekker and Eley (2007) to describe the accent of one of the talkers recruited for test video narration, who was a female talker from Johannesburg. Bekker and Eley (2007, p. 108) describe WSAE as "a distinct dialect of 'Southern' English," as it is similar in many aspects to GBE. WSAE is described as using 11 monopthongs, which are often characterized using the same vowel symbols used for GBE. Nevertheless, WSAE vowels differ from GBE, predominantly on the phonetic level: relative to GBE, WSAE has slightly shifted some vowels (see Bekker and Eley (2007) for a detailed description). WSAE uses two vowels relevant for this study: it has a counterpart to GBE /æ/, which is slightly raised relative to GBE. Next, WSAE's counterpart to GBE /ɛ/ is also slightly raised, the transcription used by Bekker and Eley (2007) is [e]. Even this slightly higher quality should still map relatively well onto the Czech /ɛ/. See Figure 1.5 for formant measurements from wordlist reading by the WSAE talker.

1.6.4 Singapore English

Leimgruber (2011) describes SE as a "nativized variety of English,". In terms of quality, its vowel system is simpler than that of GBE. Overall, when it comes to quality, the SE vowel system is somewhat similar to the Czech one: three degrees of height are differentiated in front and back vowels, one mid-low central vowel appears, and one mid central. SE does not contrast vowels in length. When compared to GBE, the SE system can be described in terms of mergers. Relevant for this study, SE is described as merging /ɛ, æ/ into /ɛ/. See Figure 1.5 for formant measurements from wordlist reading by the SE talker.

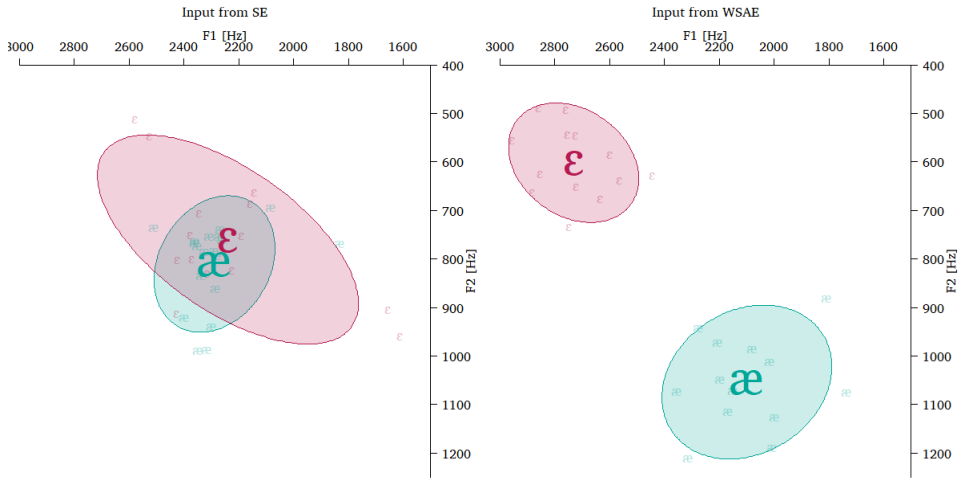

Figure 1.5 The target vowels produced by the WSAE and SE talkers (at test). Ellipses include 68% of tokens. The plots are based on 13-19 word tokens produced during the wordlist reading.

2 This study

2.1 Introduction

In this study, five-to-six-year-old listeners are exposed to L2 English in four animated videos. The children encounter single-accent input from two Czech talkers, or two GBE talkers, or multi-accent input from one Czech talker and one GBE talker. After exposure to the four training videos, they watch a test video, which features a WSAE talker and a SE talker. The target vowels this study focuses on are /ɛ/ and /æ/. The test talkers were chosen on the basis that their target vowel F1 and F2 showed similar distributions to those of training talkers. CE talkers showed unimodal height and retraction distributions, only the /ɛ/ category was present for them, the same pattern occurred with the SE talker. GBE talkers showed bimodal distributions with peaks for /æ/ and /ɛ/ words, like the WSAE talker. This study addresses how short term multi-accent exposure impacts adaptation to a difficult vowel pair in a novel accent. I test one multi-accent group and two single-accent groups using a word identification scores on pairs of dissimilar words, minimal pair items produced contrastively, and GBE minimal pair items produced as homophones (both with /ɛ/).

Given the reported duration and intensity of exposure to English, it is possible that children with this amount of non-immersion experience have acquired the /ɛ, æ/ contrast including the ability to encode it lexically. If this is the case, the participants in the GBE group are expected to perform above chance on the contrastively produced MPs at training. The Multi group, if their adaptation is not constrained by the increased processing cost of multi-accent exposure, should also perform above chance on contrastively produced MPs at training. The CE group were not exposed to these items at training. No difference in filler identification accuracy is predicted for either group at training or test.

Adaptation to test stimuli is assessed by analyzing answers to MPs produced contrastively: above chance performance indicates targeted shifts adaptation. As the test stimuli are provided in two accents, it is possible that due to their more diverse short-term experience, the Multi children will adapt more successfully (i.e. show higher word identification accuracy) at test than GBE-exposed children, who may show lower identification accuracy due to having to process two different accents during the task. If their short-term exposure influences their adaptation to a noticeable extent, the CE-exposed children should be less accurate at identifying the contrastively produced MPs than the GBE children, who have been exposed to contrastive MP production at training, unlike CE-exposed children. The same may be true for the comparison between CE and Multi children, provided that the multi-accent experience does not hinder the multi-children's processing of speech.

In general, toddlers' processing of unfamiliar accent seems to make use of mechanisms that are also used by adults. However, the research shows that without previous experience with an accent, it is more difficult for infants to process an unfamiliar accent than it is for adults (Creel et al., 2016). The theoretical account of adaptation from Kleinschmidt and Jaeger (2015) can be used to understand adaptation by the currently studied population. Due to the lack of

research on adaptation by L2 child learners, concrete predictions concerning degree of category updating, and hence performance on the word identification task, cannot be formulated. However, based on previous research on adult L2 learners, a description of two extreme outcomes is possible.

Using an IAF perspective, the fact that the learners in this study are learners of English should be captured by their prior beliefs. The confidence parameter is key in determining the degree of category flexibility, hence also the extent of adaptation. A conservative estimate of the degree of adaptation would be characterized by strong prior beliefs in height and retraction values that are typical of Czech /ɛ/. These prior beliefs would be used for both /ɛ, æ/ in GBE and WSAE input. This prior would express that the child has a single category in the region of perceptual space where talkers of GBE and WSAE have two categories.

The extent of learning between training and test can be variable based on the children's confidence in the initial prior, i.e. how relevant the they believe their previous experience to be for perceiving the current speech. For an English animated video, the participants may consider previous English videos they have seen to be informative, while real life interactions in Czech may be considered less relevant. If this is the case, the multi-accent children should be equipped with experience most relevant for both test talkers. The degree of acoustic-phonetic similarity between the training and test talker should predict generalization of phonetic adaptation (Xie & Myers, 2017).

At one hypothetical extreme, participants' confidence in the prior distribution would be sufficiently low to allow for greater influence of training on the final posterior distribution. If this were the case, the groups would differ in whether they use bimodal or unimodal distributions for the cues tied to target vowels after training. The test input would also determine their post-training distributions, but to a smaller degree than training. Hence, in the case of lower confidence in the initial prior, word identification scores for the minimal pairs produced contrastively should differ between groups: CE children should perform at chance, GBE children and possibly also Multi children above chance.

However, at the other extreme, the participants' confidence in their Czech prior distribution would be high. Hence, the combined effect of training and test input would be small compared to prior experience. In this case, no difference should be observed in word identification scores for the contrastively produced minimal pairs between groups, and no group should show a change in score between training and test.

If listeners use talker-specific cues to speech sound identity, they should be able to utilize experience with the speech of a previously encountered talker when faced with a new similar talker. Having been exposed to a talker with a similar accent can be used to headstart adaptation to a current talker. Listeners should use their previous exposure when trying to adapt to test talkers, though they need not, if for some reason, the previous exposure is not found to be informative when it comes to the test talkers. Listeners with a wider experience (having encountered more variable speakers) are predicted to show more adapt faster than those with narrower experience (Kleinschmidt & Jaeger, 2015).

Previous exposure to English may boost adaptation, however, a small L2 vocabulary tied to low L2 proficiency may hinder it. The learners in this study are in the process of establishing an L2 lexicon. They likely do not know many

words provided as stimuli, these have to be learned from the videos. Dissimilar filler words were included in word identification tasks to assess whether children are able to retain words from the videos in memory.

The theoretical framework put forward by Kleinschmidt and Jaeger (2015) is valuable in that it can also, to some extent, be applied to child L2 adaptors. However, it cannot be relied upon to provide the basis for establishing clear predictions. For example, IAF uses phone categorization functions to formalize prior beliefs of listeners. Children generally show shallow identification function slopes, hence more variable categories (Walley & Flege, 1999). Further, even the design of the identification task may be affect responses: some are finding that children's performance on similar tasks is modulated by whether the forced choice involves two images or two sounds (Wagner et al., 2024). This study utilizes an identification task with a continuum between [bæɹi] and [bɛɹi]. At this point in their learning, a task focused on such fine-grained differences may not be productive. Further, it may overtax children's attention and decrease motivation to participate. It was included as an exploratory feature.

The category boundary task was included as some participants had been exposed to English since infancy. They were expected to show reliable identification of continuum stimuli. The CE children, who were not exposed to contrastive minimal productions of the target vowels, also completed the task, with stimuli that did not agree with the talker's production. It is possible that even those children have acquired the target contrast, and may accept contrastive production even from a talker who did not demonstrate it at training. However, it was also possible that participants in no group, irrespective of their length and intensity of exposure to English, distinguish the contrast. This is a difficult contrast that poses challenges even to experienced adult learners (Šimáčková & Podlipský, 2018). It is also possible that the task will be too demanding for the children. Hence, the main interest with regard to this task was whether participants in different groups show different performance.

In many phonetic training and accent adaptation studies, participants are exposed to isolated words. In this study, learners are also provided with isolated word tokens as part of the procedure, but the main component of the training is story narration accompanied by video, where words are anchored contextually. Target words are presented also outside the narration stream, as isolated tokens in an interactive exercise (image tapping), to connect the target word sound with a visual cue to the meaning of the word in a non-ambiguous way.

Acoustic-phonetic and indexical variability has been demonstrated to aid minimal pair learning by supporting robust encoding of the segment that differs in a given minimal pair. However, as (Rost & McMurray, 2009) point out, it is not that variability, in general, is beneficial for category establishment and maintenance. The paramount requirement is that input provide statistical structure that is appropriate for a given learning mechanism. Phonetic variability is commonly a part of this statistical structure. The variability presented in this study, however, is not solely phonetic. By including multiple varieties, phonological variability is introduced: different talkers signal either a merger or a split of the same two categories.

In agreement with Kleinschmidt and Jaeger (2015), Rost and McMurray (2009) stress that children's performance in tasks involving minimal pair (MP)

learning is determined by the input statistics they have previously experienced, the evidence presented in the task, and top-down influences like the structure of the lexicon and task demands. Phonological variability in the input means that the multi-accent children are faced with increased task demands while also getting a smaller quantity of input signalling the target vowel status in each variety.

In sum, it is possible that multi-accent input does not provide a short-term benefit for beginner learners' adaptation to a difficult vowel. From this point of view, the mono-accent-exposed children should face smaller task demands and may benefit from larger quantity of input from a single variety. The GBE is group is considered most likely to identify minimal pair words with above chance accuracy, especially during training. Similarly, if the Multi children have acquired the target contrast and are not overtaxed by task demands, they should identify minimal pair words with above chance accuracy at training. No group is expected to perform above chance on minimal pairs produced as homophones. When it comes to test performance, formulating specific predictions is difficult. The main objective this study addresses is whether multi-accent children differ from mono-accent children in identification of minimal pair and dissimilar words presented in two unfamiliar dialects. However, and this holds for any of the three groups, if the difference between a given MP is not yet stored in the lexicon, it is possible that hearing either item from the MP would activate the other item. It is not precluded that even these experienced learners have not established lexical representations with sufficiently robust encoding of the difficult /æ/ vowel.

2.2 Method

2.2.1 Participants

Data from 51 children aged 4;11 to 6;11 were included in the analyses. The total number of participants from which data were collected was 63. Of the excluded participants, three had been diagnosed with developmental language disorder, one with attention deficit disorder, one with autism spectrum disorder, and seven were reported to have no experience with English. I further excluded data from the last training video from one participant from the GBE group, because he accidentally turned of the sound during the task and finished it with no sound. One participant from the Multi group completed only the last training video, her data are included.

Each participant was randomly assigned to one of three conditions, see Table 2.1 for characteristics of the three groups. All children had previous experience with English, ranging from only media exposure to exposure during weekly English lessons. All were acquiring Czech as their native language, lived in Prague or central Bohemia, and were predominantly exposed to Bohemian Czech. See Table 2.1 for further information about the participant groups. Detailed information from the questionnaire for each participant is available at https://osf.io/6ptc8/?view_only=e02bc34e428e43f29bfb5bb8e71d7774.

Condition	Mean age	Females	Males	Onset	Exposure
CE	68.4 (7.74)	9	7	33.9 (17.5)	3.95 (3.36)
GBE	66.9 (7.48)	7	9	37.2 (20.5)	7.63 (6.81)
Multi	67.8 (6.93)	12	7	23.2 (20.3)	5.91 (6.96)

Table 2.1 Group characteristics: condition, mean age in months with standard deviation (SD) in months in parentheses, number of females in the group, number of males in the group, mean onset of exposure to English in months with SD in parentheses, mean reported percentage of current exposure to English with SD in parentheses.

At training, participants assigned to the CE condition heard the English narrative from two Czech talkers, those in group GBE from two talkers of GBE, and those in group Multi from one Czech talker and one GBE talker. At test, all participants heard one talker of SE and one of WSAE.

2.2.2 Stimuli

Participants watched five animated videos in total. After each video, they completed a word identification task. Data from task completed after the last training and test video were used to asses whether participants were able to use the targeted shifts mechanism to adapt to the MPs in novel input. Within this task, filler items were used to test whether children had learned dissimilar words from the story. In the lab, some participants also completed a category boundary task, and all completed the Dimensional Change Card Sort task (DCCS, Zelazo, 2006). The at-home sessions used an experiment built using jspsych (De Leeuw, 2015), which ran from files stored on GitHub in private mode (accessible only via a Personal Page link after entering a password). The final training, test task, and

DCCS task were also built in jspsych (De Leeuw, 2015) and ran from local files in the lab.

2.2.3 Videos

Video observation was the initial part of both the training and test sessions. Each participant watched four animated training videos accompanied by English speech from two talkers, and one test video accompanied by speech of two novel talkers. All videos were under six minutes in duration. Videos were interrupted by an interactive task featuring isolated word presentation accompanied by a backgroundless image taken from the video.

Six female talkers were recruited to provide the narration and isolated words: four native English talkers and two Bohemian Czech talkers highly proficient in English. Two native talkers spoke General British English (GBE1, GBE2), one spoke Singapore English (SE) and one spoke White South African English (WSAE). CE talkers spoke the Bohemian variety of Czech. They were university students with at least a B2 proficiency level according to CEFR (Council of Europe), 2001). They were chosen based on their Czech-accented production of both /ɛ, æ/ as /ɛ/. Both GBE talkers were from England. The WSAE talker was from Johannesburg. Target vowel duration was measured in all talkers' productions of isolated MP words. No talker seemed to distinguish the target vowels using duration (see Figure 2.2).

For video observation stimuli, the talkers were recorded reading the full stories and a wordlist of isolated target and filler words. They were instructed to read as if they were reading to a child. Hesitations, false starts and repetitions were removed from the recordings. Errors in and substitutions of synsemantic words were retained in the final story stimuli, including errors which could be considered agrammatical in some varieties (i.e. elision of a final [s] in a regular plural noun). Target words studied here were produced consistently by talkers of the same L1 variety.

Talkers were recorded using a head mounted AKG C 520 microphone and a Roland Rubix22 audio interface connected to a HP ProBook 450 G5 laptop running Audacity software (version 3.0.2, Audacity Team, 2019). Recordings were saved in the wav format with a 44.1 kHz sample rate and 16 bit quantization.

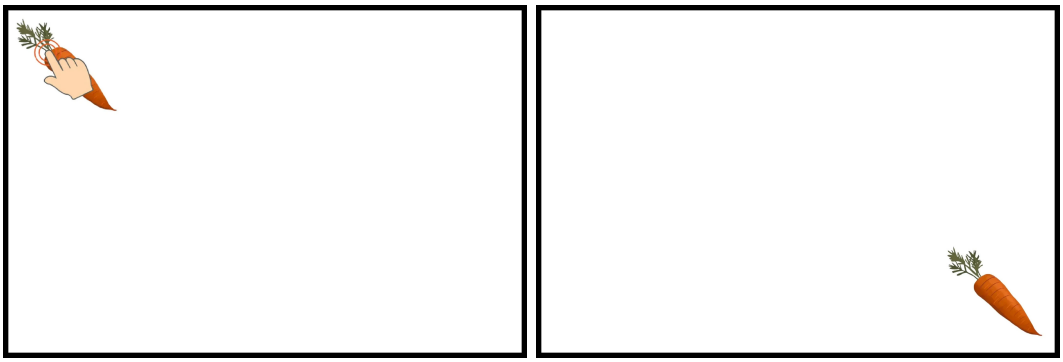

Figure 2.1 **Left:** The first appearance of an image in an interrupting image trial. **Right:** The second appearance of an image. The order of the appearances in each corner was random, without repetition, on each trial. Black border shows the bounds of the screen.

Table 2.2 shows the number of instances that each target vowel was heard by a participant during training and test in target words. The participants did observe more tokens of the target vowels than disclosed here, but they appeared in non-target words (e.g. adjectives and verbs). As opposed to the targets, these were not explicitly visually labelled with a static image, which is why they are omitted from the summary.

GBE vowel	Video(s)	Narrator count	Characters count	Narrator	Characters
/æ/	hen, ice	20	16	talker 1	talker 2
/æ/	lion, oak	22	24	talker 2	talker 1
/æ/	mouse	21	20	talker 3	talker 4
/ɛ/	hen, ice	21	20	talker 1	talker 2
/ɛ/	lion, oak	15	15	talker 2	talker 1
/ɛ/	mouse	5	5	talker 3	talker 4

Table 2.2 Vowel token counts for training and test talkers from target words. Talker 1 and talker 2 stand for any training talker pair without repetition, talker 3 and talker 4 stand for either of the two possible pairs of the test talkers, i.e. WSAE and SE or SE and WSAE.

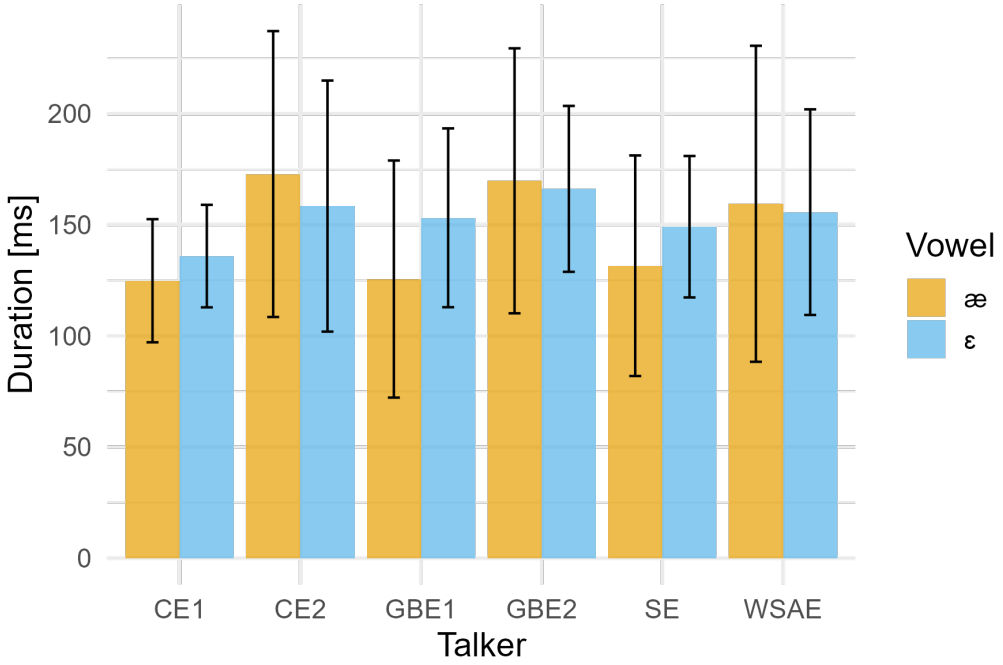

Figure 2.2 Duration of [ɛ]s and [æ]s produced by the talkers. Error bars extend one standard deviation away from the mean.

Throughout the four training videos, each participant heard one talker twice as the narrator of the story and twice as the characters, in alternating order. The sequences in which participants watched videos were pseudo-randomized, to ensure that the number of words spoken by the talkers was balanced. To satisfy this criterion, in all video orderings, the narrator for the videos "Lion" and "Oak" is the same talker, ensuring that one talker provided 47.5% of all words

Video	Narrator	Characters	Video total
Lion	164	119	283
Hen	346	178	524
Oak	185	76	261
Ice	142	90	232
Role total	837	463	1300

Table 2.3 Total words provided by the narrator and the characters for each video. Here, a word is a string of characters surrounded by spaces or punctuation.

(617 words), and the other 52.5% (683 words). See Table 2.3 for word counts for each role, separately for each video. There were 16 permutations specifying video orders that fulfilled this criterion (see https://osf.io/7w2xk for the video orderings).

Processing of story recordings

The raw recording files were segmented into intervals of narrator's and characters' speech. The interval boundaries of the UTF-8 encoded TextGrid were moved to zero crossings using a script (Atria, 2015) in Praat (Boersma & Weenink, 2024). The recordings were cut based on the segmented intervals, and peak normalized using a Praat script (DiCanio, 2014). Then, each talker's turns were combined with another talker's turns using a script (available at https://osf.io/7w2xk) in R (R Core Team, 2024) with the libraries tuneR (Ligges et al., 2023) and seewave (Sueur et al., 2008). Zero-padding filled the time between talkers' turns. Using FFmpeg (Tomar, 2006), audio tracks were converted to MP3 (to satisfy GitHub upload criteria for file size).

Videos were obtained from YouTube (TheFableCottage.com, 2021, 2022, 2023a, 2023b, 2023c). Their original audio tracks were removed. The narration used in this experiment is an adjusted version of the original narration on YouTube, adjustments included renaming characters to serve as target words, and reducing sentence complexity. The videos were cut to fit the new narration in DaVinci Resolve (Milovanovic, 2024). They were then rendered at a 1920x1080 resolution, 25 fps frame rate, in the mkv format. The silent videos were combined with the MP3 stereo tracks into an MP4 file using FFmpeg (Tomar, 2006).

2.2.4 Training and test tasks

A word identification task followed each of the five videos, but only data collected from the last training and test session, which were conducted in the lab, were analyzed. Word identification as part of the at-home training sessions served primarily to familiarize the children with the task. For the word identification task stimuli, I used the same recordings from the wordlist read by the six talkers as were used for interrupting image trials during video observation. Additionally, the talkers were recorded reading a list of carrier phrases, and a list of feedback phrases (see https://osf.io/7w2xk for carrier phrases, feedback phrases, and the wordlist). The task involved presentation of isolated words (from the read wordlist), or isolated words preceded by a carrier phrase.

Each word was presented along with its single corresponding image, which was a png file taken from the YouTube video, with the background removed (the same image as was used in the interrupting trials during video observation). Additionally, two megaphone buttons for playing sound, one grey and one white, were displayed under the image. A green button in the lower right corner of the screen was used for advancing through trials. Feedback trials displayed the target image again, along with a single green megaphone image for playing the correct sound, and the green arrow button for advancing through trials. Concurrently presented feedback phrases were accompanied by an image of a green tick or red cross displayed in the upper right corner of the screen.

The first pair of sounds in each run of the task included dissimilar filler words that did not contain target vowels. After that, word pairs featuring two fillers, one filler and one member of a minimal pair, and minimal pairs followed in a pseudo-random order. If the number of the pairs from each type allowed it, minimal pair (MP) trials were followed by trials with at least one filler. At test and during training sessions of the Multi group, MP trials were divided into two types: MPs produced contrastively and MP produced as homophones. MPs produced contrastively featured one of the talkers who used /æ/ producing an item with [æ], and the other talker (with the merger) producing the /ɛ/ item. MPs produced as homophones featured a talker with a merger producing the item which in GBE and WSAE contains /æ/ with [ɛ] and a talker with a split producing the /ɛ/ item, leading to a homophone-like pair. Table 2.4 illustrates the types of minimal word pairs that appeared in the word identification task. Three minimal pairs were used: *berry-Barry*, *bread-Brad*, *Jen-Jan*.

MP type	**Talker 1**	**Word 1**	**Vowel 1**	**Talker 2**	**Word 2**	**Vowel 2**
Different	GBE/ WSAE	Barry	[æ]	CE/SE	berry	[ɛ]
Same	CE/SE	Barry	[ɛ]	GBE/ WSAE	berry	[ɛ]

Table 2.4 Examples of sound combinations used for minimal pairs. Minimal pair types are *different* (contrastive production), and *same* (homophone-like). Talker 1 produces word 1 with vowel 1, talker 2 produces the word 2 with vowel 2.

Category boundary task

A two-alternative forced-choice identification task was used. The auditory stimuli for the category boundary task were created using one natural token of the word *berry* from talkers GBE1, GBE2, CE2, and WSAE. Manipulations based on a production from GBE1 or GBE2 were used with the GBE and Multi groups after training, and a production from CE2 was used with the CE group after training. At test, all groups were exposed to manipulations based on a production of *berry* by the WSAE talker. The natural token was manipulated using Source-filter synthesis of an existing *berry* sound using a Praat script (available at https://osf.io/7w2xk) based on Podlipský (2024).

To create the continuum stimuli, the script used manually measured F1 and F2 from an instance of *berry* and *Barry* by the relevant talker. For CE2, F1

and F2 for [bæɹi] were estimated based on [æ] productions of a GBE talker from Bjelaković (2017) who had similar F1 and F2 values to CE2 for [ɛ]. F1 and F2 were manipulated inside a manually segmented vowel interval. Seven sounds were created for each talker, with equidistant steps in ERB-transformed F1 and F2 between each adjacent pair of continuum sounds. The first stimulus corresponded to the word *berry*, the seventh to the word *Barry*.

An earlier version of the category boundary task based on Hazan and Barrett (2000) was used for the first six children who participated in the experiment. The same Praat script (available at https://osf.io/7w2xk) was used to generate 51 continuum steps from the *berry* to *Barry* endpoints. Other than that, the stimuli preparation was the same as for the new version described above. The procedure differed in that steps were presented in a non-random order as a single adaptive track starting from a random endpoint. The response on a trial determined the sound presented at the next trial: a sound closer to the endpoint corresponding to the response image was given on a following trial. The difference between a preceding sound and a current trial sound started at 10 (i.e. there was a distance of 10 steps between the preceding and current sound), and decreased on every reversal of the adaptive track. Difference sizes were 10, 5, 3, 2 and 1. The task ended after seven reversals of the adaptive track, or after 20 trials. Every fifth trial was an interleaved random endpoint.

Participants were struggling with the duration of this task, so it was replaced by the newer version described above. To analyze the responses of the six participants who completed the older version, the 51 responses were rescaled to correspond to a seven-point response scale.

The Dimensional Change Card Sort task

An adjusted version of the Dimensional Change Card Sort (DCCS) task (Zelazo, 2006) was adapted for administration on a tablet using jspsych (De Leeuw, 2015). The task instructions were translated to Czech. A young female talker from the Bohemia region was recorded reading the instructions. The visual aspect of the task was adjusted: the image of a boat originally used in Zelazo (2006) was changed to a picture of a tractor, so that both images were matched in grammatical gender in Czech (rabbit and tractor). Each trial showed one stimuli image, either a red rabbit or a blue tractor. Below it, two image buttons were displayed, one of a blue rabbit, the other of a red tractor. See Zelazo (2006) for details about the task.

2.2.5 Procedure

The procedure was originally intended to involve in-person administration in preschools, but was later adapted for hybrid online and laboratory administration. The experiment was conducted on three consecutive days. At home, the child watched a single training video on the first day and two videos on the second day, all three were followed by a word identification task. In the lab, on the third day, they watched the fourth training video, the test video, and completed DCCS task (Zelazo, 2006). Some participants failed to watch the first video on the first day, and so watched all three videos a day before visiting the lab. Before the

in-lab session, the parent or caretaker signed an informed consent, and filled in a language background questionnaire (both available at https://osf.io/7w2xk).

The parents were instructed to complete the at-home sessions with the children in a quiet room on a device with a sufficiently large screen (at least a tablet or laptop) using the Chrome web browser. They were sent a GitHub Personal Page link that lead them to the online version of the experiment. The parent was asked to communicate as little as possible with the child during the experiment, and to not practice any words from the tasks with the child. Parents received information on how to interact with the experiment interface via e-mail. They were told that after they enter the password and ID, they should let the child start the first task by clicking on (or tapping) a green arrow button, and let her operate the device for the remainder of the experiment. If the child did not want to or could not interact with the device, the parent was instructed to do so, as the central aim of the training sessions was simply to expose the child to the stimuli and familiarize her with the tasks. Each session began with the video observation task and was followed by the word identification task.

In the lab, the child completed the tasks in a quiet room with the experiment administrator. For most participants, a parent was also present in the room. The experiment was administered in the Chrome web browser on a Microsoft Surface Go tablet, version 2 or 4, which have the same physical dimensions (10.5 inch display). Most participants wore Bose QC35 II headphones for all tasks with noise cancellation set to high. Some refused the headphones, data from them are also included in the analysis. In both cases, the volume was set to 60 on the tablet. The administrator entered the session number and child's randomly assigned ID into the experiment's interface, and gave the tablet to the child, instructing them to press the green arrow button to start the experiment.

The last training session and test session each consisted of a video observation task and a word identification task. Some children completed a category boundary task after the last training session and test session; this task was later omitted completely, because it was too long and demanding. The child was given at least a 20-minute break between the last training and test session. After the test session, the DCCS task in Czech followed. The children were offered a break between the test session and DCCS task, but many chose to complete the DCCS task almost immediately after the test session. Including the breaks, the child spent around 50 minutes in the lab. Parents received a drugstore voucher for 500 CZK as a reward for their child's participation.

Video observation

To start playing a video, the child pressed a green arrow on the tablet screen. At a number of pre-defined moments, the video was interrupted, and a white screen with a single target or filler image in one of the corners of the screen appeared, with the sound corresponding to the image playing at the same time (see Figure 2.1 for screenshots from the trial). Concurrently, an image of a tapping hand appeared over the image for 800 ms, prompting the child to tap the image. After the word sound stopped playing and an image was tapped, the same image appeared again in a different corner of the screen, and the sound played simultaneously. Each image appeared four times before the video resumed. If the child did not tap the image, after 15 seconds, it changed position and its

sound was played automatically. The number of instances where the video was stopped corresponded to the total number of targets and fillers for that video; the minimum was six interrupting trials in the "Hen" video, maximum nine trials in the "Mouse" test video. The target and filler images appeared in a given order at fixed times which were the same for all participants (see https://osf.io/7w2xk for word orders and timestamps within each video).

Word identification

On each trial of the word identification task, the participant saw a picture corresponding to a target or filler word, and two megaphone buttons below it, one white and one gray. When the participant clicked one of the megaphone images, a word from a pre-defined word pair was played (see https://osf.io/7w2xk for a list of all word pairs). The other megaphone button played the other word from the pair. The goal was to select the sound that corresponds to the displayed image, the parents were asked to not explicitly instruct children to do this, but let them infer it from the task during the at-home sessions. Both talkers known from the story appeared on each trial, i.e. one of the two talkers always produced one word from the pair, and the other produced the other word. When clicked, a megaphone button played the word in isolation, or preceded by a carrier phrase; these alternatives occurred with the same probability. If clicked once again, a button could play the word in a different form: a different instance of the isolated word, or the word in different carrier phrases (see https://osf.io/7w2xk for a list of used carrier phrases). There was no limit on how many times a button could be clicked. Once clicked, a button became highlighted in yellow, indicating that it was chosen as the one that plays the sound corresponding to the image.

After each of the sound buttons was pressed at least once, a green button appeared in the lower right corner of the screen. The participant confirmed her sound selection by clicking the green arrow button. See Figure 2.3 for screenshots from the task. A 500 ms pause followed each green arrow press, after which a feedback trial started, where the image appeared again. At the same time, a feedback phrase was played, depending on whether the participant chose the sound button correctly (the list of used feedback phrases is available at https://osf.io/7w2xk). The feedback was accompanied by an image of a green tick or a red cross in the upper right corner of the screen. The displayed target or filler image had a single green megaphone button below it. Once pressed, the button played the isolated word corresponding to the image. This button could be clicked repeatedly, and always provided the same sound. If pressed at least once, a green button again appeared in the lower right corner, by which the participant moved to the next trial with two sound buttons. See Figure 2.4 for screenshots from the feedback portion of the task.

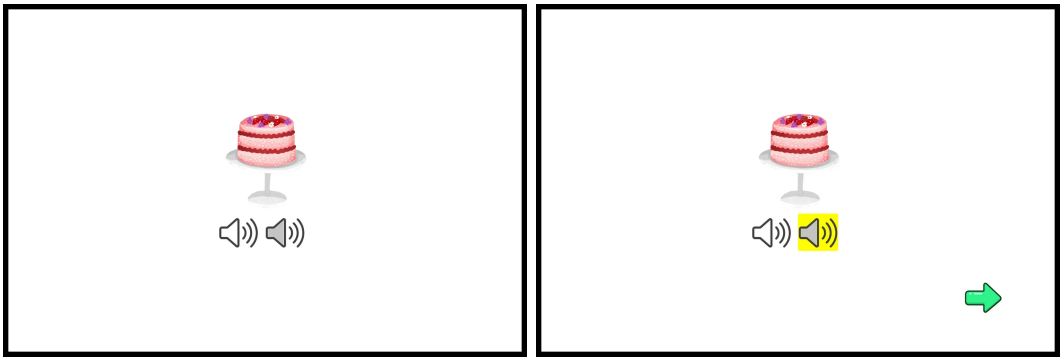

Figure 2.3 **Left:** A screenshot from the onset of a word identification task, before any buttons had been pressed. **Right:** Word identification task with a selected megaphone button (gray), indicating the sound choice. The green arrow button in the lower right corner only appeared after each sound button had been clicked at least once. Black border shows the bounds of the screen.

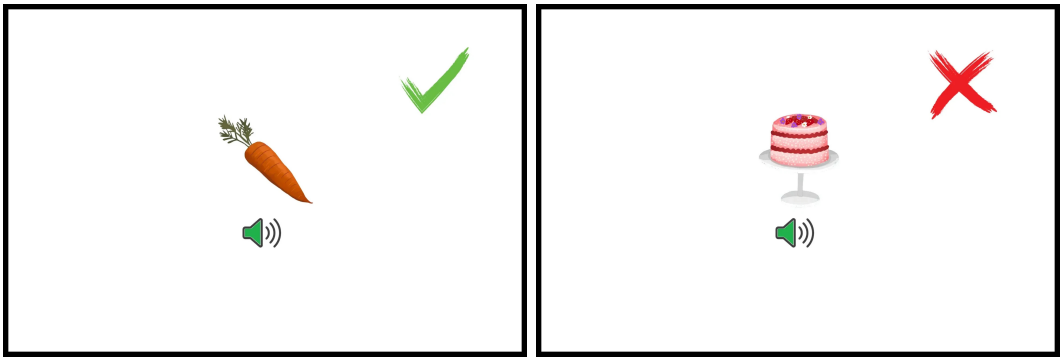

Figure 2.4 **Left:** Feedback screen for a word identification trial where the participant selected the sound matching the image. **Right:** Feedback for a trial where there was a mismatch between displayed image and selected sound. Irrespective of response correctness, the green arrow button in the lower right corner only appeared after the green megaphone button had been clicked at least once. Black border shows the bounds of the screen.

Category boundary task

The category boundary task was administered on the same tablet after the last training session, and after the test session. It featured a single megaphone button and two images on the sides of the screen below it. The relevant continuum sound played automatically upon trial start. The child could choose to listen to it again by pressing the megaphone button. The number of replays was not limited. After pressing one of the picture buttons, it highlighted in yellow, indicating that this image was chosen as matching the sound. The participant confirmed their choice by clicking a green arrow in the right bottom corner of the screen. The arrow button was only displayed once the participant clicked an image button. The task used an ISI of 500 ms. On each new trial, the background color changed, indicating that a new stimulus was playing. Apart from the first six participants who took the earlier version of the task, each child, if they did not choose to end the task earlier, responded to two sets of seven trials, i.e. participants responded

to each continuum step twice, generating a total of 14 trials. Sounds appeared in random order within the two sets.

The 14 trials were preceded by four task familiarization trials in the same design, but featuring only the two continuum endpoints. Each endpoint was played twice, and the order of sounds was randomized. Feedback to choices was provided at familiarization trials. Once the child confirmed their choice of image, a blank screen with a green tick or a red cross appeared along with a feedback phrase. The feedback images were the same as those used in the word identification task, and were displayed in the center of the screen. The content of the feedback phrase depended on whether the participant chose the correct button (see https://osf.io/7w2xk for a list of feedback phrases). Three positive and three negative feedback phrases were used. After the feedback sound finished playing, a new trial started. The background color was the same for a trial and its corresponding feedback portion, but changed between trials.

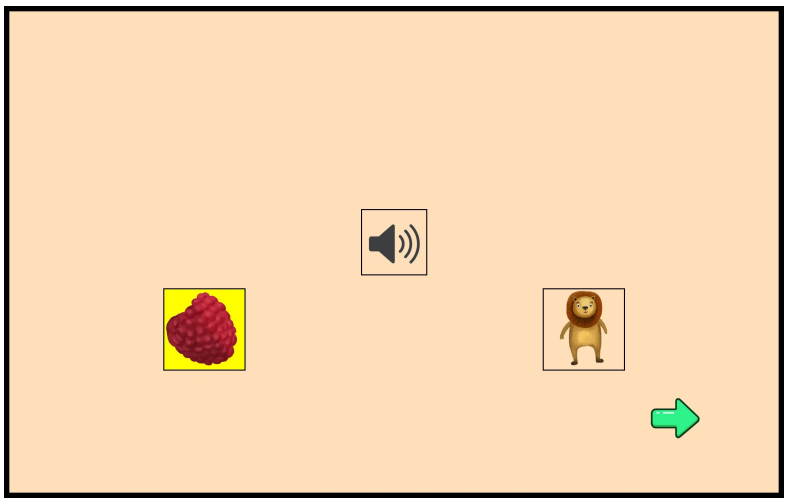

Figure 2.5 A screenshot from the category boundary task. Here, the *berry* button is selected (highlighted in yellow) as matching the sound. Black border shows the bounds of the screen.

The DCCS task

The participant's task was to match the stimuli image with a response image according to the instructed dimension. As in Zelazo (2006), the children first completed six trials focused on color matching; the following six trials focused on shape matching. In the third and final subsection, participants completed 12 trials where images were to be sorted by color or shape depending on whether they appeared with a black border. See Figure 2.6 for screenshots from the video with instructions and from the final portion of the task. Before each of the three parts of the experiment, a short video gave instructions on how to sort the following images. Image button placement was counterbalanced between participants. To eliminate the possibility of participants forgetting instructions, the same instructions were played on every trial start (following Zelazo, 2006). If the participant chose to respond to the trial before the instructions ended, the audio track was interrupted, and a new trial began. No feedback was provided on the DCCS task. See Zelazo (2006) for a detailed description of the procedure.

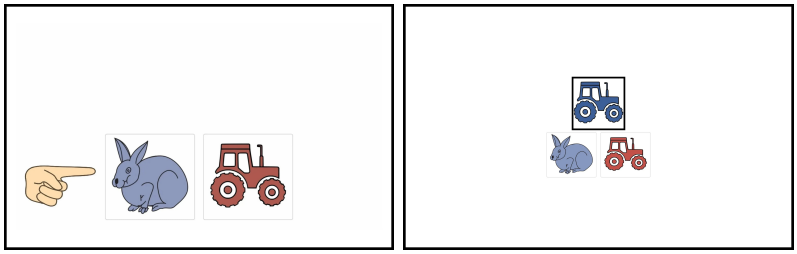

Figure 2.6 **Left:** A screenshot from one of the three instructional videos in the DCCS task. **Right:** A trial from the final subsection from the task. Black border shows the bounds of the screen.

2.2.6 Statistical analysis

To analyze word identification accuracy, a generalized linear mixed-effects model (GLMM) was used with a binomial distribution and logit link function. The model was fit using the Laplace Approximation method. Next to accuracy, I also used a GLMM to analyze stimulus replays at word identification trials. Responses to continuum sounds from the category boundary task were plotted for visual analysis to assess whether participants showed reliable identification of the minimal pair words. Performance on the "border version" (last of the three sections) of the DCCS task (Zelazo, 2006) was modelled using a linear model (LM) to analyze executive function differences between the participants.

To fit the models and compute p-values, the packages lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) were used in R (R Core Team, 2024). The Bobyqa optimization (Powell, 2009) was used to fit the GLMMs. Alpha, the threshold of significance, was set to 0.05. Visualizations were made using the packages ggplot2 (Wickham, 2016), ggeffects (Lüdecke, 2018), emmeans (Lenth, 2024), and ggsignif (Constantin & Patil, 2021).

2.3 Results

2.3.1 Word identification

This study investigates the impact of short-term multi-accent input on the degree of vowel adaptation indicated by word identification accuracy. The more successful the adaptation, the higher the proportion of correctly identified MP words at test should be. Each word identification trial featured a pair of words. Each word in a pair was produced by a different talker. The pairs of words included MPs produced contrastively, or produced as homophones. Pairs of dissimilar words, which differed in at least one consonant, were included as fillers. Dissimilar word pairs assess whether participants remember words from the immediately preceding videos. Contrastively produced MP items serve as indicators of the ability to encode words with the target vowels /ɛ, æ/ as contrastive segments. For MPs produced as homophones, all groups are expected to perform at chance during both the last of four training sessions and the test session. The question addressed here is whether identification of MP words or dissimilar words differs between children exposed to multi-accent input and children exposed to mono-accent input.

A GLMM was built to compare word identification accuracy of the multi-accent group to the two mono-accent groups. The levels of the outcome variable, correctness of response, were FALSE (0) for incorrect answers and TRUE (1) for correct answers. Condition was a between-subjects fixed effect predictor with levels *CE* for training input from two Czech talkers, *GBE* for training input from two GBE talkers, and *Multi* for training input from one Czech and one GBE talker. Pair type was a within-subjects fixed effect predictor with levels *fill* for dissimilar words (a pair of words that differ in consonants), *same* for GBE minimal pairs produced as homophones, and *diff* for GBE minimal pairs produced contrastively. The model also contained an interaction parameter of condition by pair type. Progress was a within-subjects fixed effect predictor with levels *train*, corresponding to responses from the last training session, and *test* for responses from the test session. Varying intercepts were fitted for participants and word pairs. The model formula was *correct ~pair_type ∗ condition + progress + (1 + participant) + (1 + pair)*.

At test, each participant responded to contrastive productions of MPs (diff), homophone-like productions of MPs (same), and filler items (fill). At training, participants from the Multi group also responded to all three word pair types, those in the CE group responded only to types *same* and *fill* (due to CE talkers not producing minimal pairs contrastively); and those in the GBE group responded only to *diff* and *fill* pair types (due to GBE talkers not producing minimal pairs as homophones). These missing levels were also estimated by the model, but are not reported. The total number of observations provided to the model was 662 from 51 participants. Figure 2.8 shows the raw data for a rough idea of the success rate in the word identification task separately for each of the three groups and word pair types.

Even though general criterion relaxing adaptation cannot be inferred from the present data, it may be valuable to look separately at performance on trials with a member of an MP and a filler word (a subset of the filler trials). As is apparent from Figure 2.7, the participants performed well on these trials, meaning that they likely could differentiate MP members from dissimilar words.

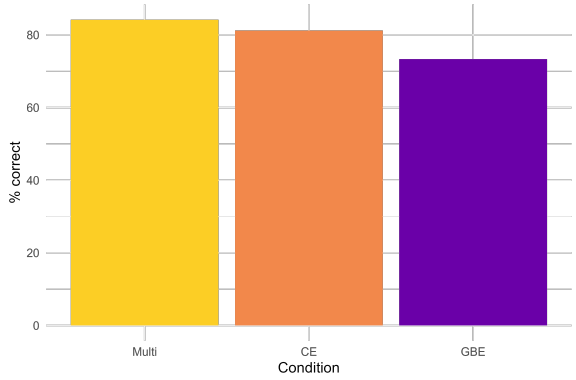

Figure 2.7 Plot of raw data subset showing the proportion of correct answers on trials with one member of an MP and a dissimilar word. Separated by condition.

Table 2.5 shows the coefficient table for the model estimating word identification accuracy. Note that the model fit is singular. Singularity may be caused by collinearity in fixed effects, in which case, removing one or multiple fixed ef-

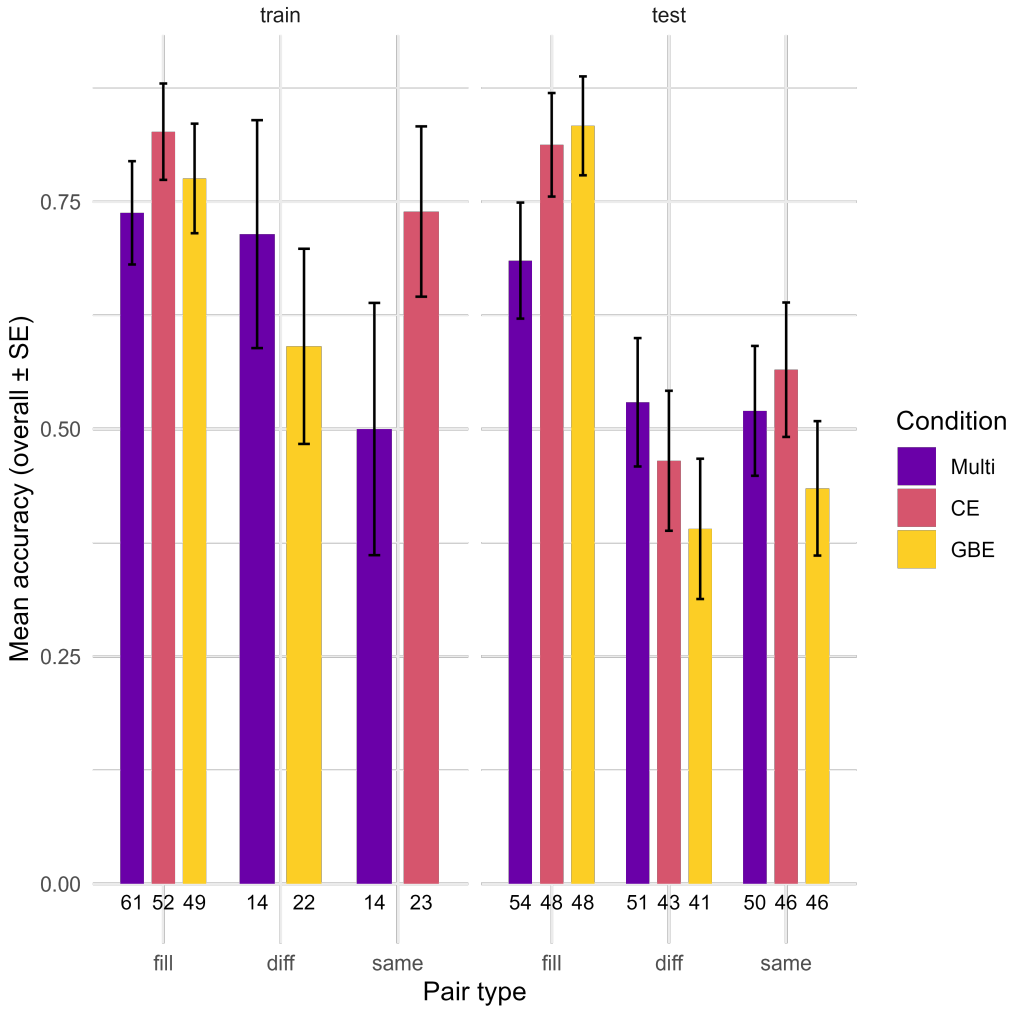

Figure 2.8 Mean word identification accuracy across participants in the three exposure groups, separately for the three word pair types. Plotted from raw data. The left panel shows performance on the last training session, the right panel on the test session. Color indicates condition. Error bars stretch 1 standard error away from the mean. Numbers represent the count of observations on which the mean is based.

fects would resolve the singularity. However, in this case, even the simplest model with a single fixed effect and the simplest random effects structure (by-participant varying intercepts) results in singularity. This is because singularity is triggered by the variance of the by-participant varying intercepts being estimated as zero. This random effect is a an indispensable term when it comes to modeling non-independent data, i.e. when a single participant provides several responses. This is the reason why the random effect was retained at least formally, even though the model was not able to find any by-participant variance.

Inspection of the raw data from the 51 participants shows that participants overwhelmingly identify fillers correctly. Indeed, including a single participant from the pool of excluded participants for completely lacking any prior English exposure, who provided mostly incorrect answers (unlike the participants who had experience with English), resulted in a non-singular model fit. Further, comparing the singular model with 51 participants to the regular model with 52 participants

showed that the coefficient estimates differed only slightly between the models, and they agreed in which effects were found to be significant and non-significant. The estimates from the singular model with the 51 participants are reported here. Further discussion of the singularity issue with respect to this data can be found at https://osf.io/7w2xk.

In Table 2.5, the estimates for the levels of the fixed effects *pair type*, *condition*, and *progress* represent the predictors' influence on the log-odds of the dependent variable (correct). All factors were treatment-coded, the reference levels were *Multi*, *test* and *fill*. Figure 2.9 visualizes of the estimated accuracy in the word identification task.

term	**estimate**	**std.error**	**statistic**	**p.value**
(Intercept)	0.763	0.248	3.074	**0.002**
pair_typediff	-0.514	0.355	-1.446	0.148
pair_typesame	-0.734	0.354	-2.072	**0.038**
condCE	0.615	0.333	1.844	0.065
condGBE	0.513	0.330	1.555	0.120
progresstrain	0.275	0.222	1.241	0.215
pair_typediff:condCE	-0.949	0.521	-1.824	0.068
pair_typesame:condCE	-0.214	0.488	-0.440	0.660
pair_typediff:condGBE	-0.998	0.489	-2.041	**0.041**
pair_typesame:condGBE	-0.763	0.513	-1.486	0.137

Table 2.5 Coefficient table for the model estimating word identification accuracy. The reference level for pair type is *fill*, for condition it is *Multi*, and for progress it is *test*. Hence, the intercept represents the estimated accuracy by the Multi group on dissimilar words at test in log odds. This model fit was singular due to the variance attributed to the by-participant varying intercepts being estimated as 0.

The intercept in Table 2.5 represents the accuracy of the Multi group for dissimilar words (*fill*). The p-value for the intercept (estimate = 0.763, p = 0.002) indicates that the Multi group identified dissimilar words reliably above chance (chance corresponds to a probability of 0.5 here, or log odds of 0). The data do not serve as sufficient evidence that responses to trials with contrastively produced MPs were reliably different from responses to dissimilar words. However, this does not mean that performance on contrastively produced MPs by the Multi group was above chance. In fact, inspection of Figure 2.9 reveals that the 95% confidence intervals (CIs) for contrastive MPs at test include the chance level, hence, it cannot be concluded that performance was above chance for contrastively produced MPs. The difference between identification accuracy for dissimilar words and MPs produced as homophones, did, however, reach significance with a negative estimate (-0.734, p = 0.038), meaning that the Multi group performed reliably worse on homophone-like MPs than dissimilar words. Inspection of Figure 2.9 reveals that word identification on trials with MPs produced as homophones did not differ reliably from chance.

When it comes to assessing whether performance differs from chance, Figure 2.9 shows that all groups identified dissimilar words reliably above chance, at both the last training session and the test session. Other than that, no group seems to have performed above chance on either of the MP types, with one no-

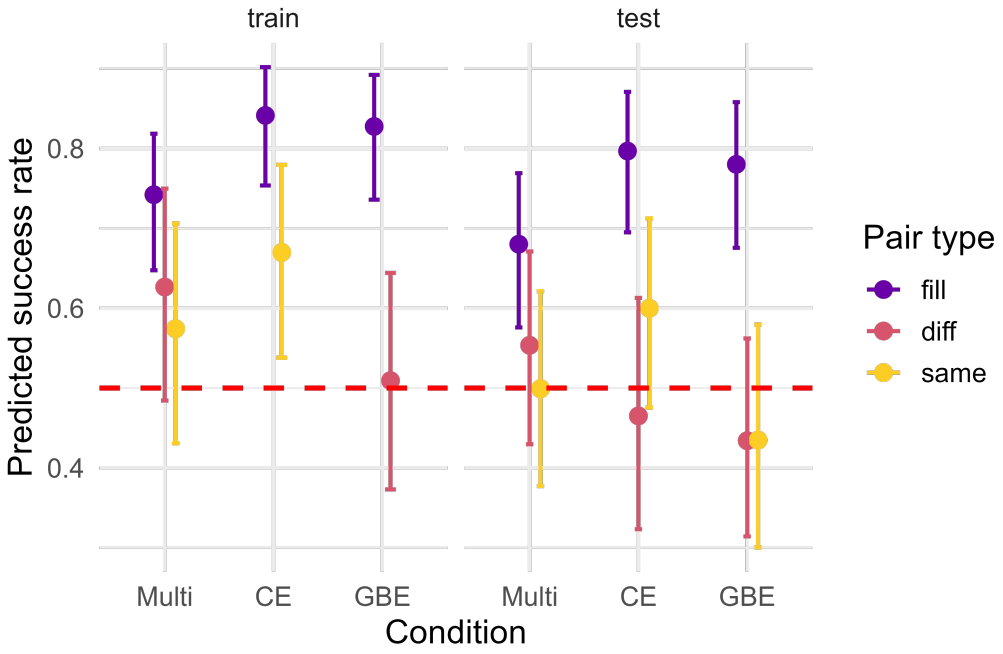

Figure 2.9 Estimated word identification accuracy for each group of participants at training (left panel) and test (right panel). Points in error bars show mean success rates for fillers, minimal pairs produced contrastively (diff) and as homophones (same). Error bars show 95% CIs. The level of chance is highlighted by a dashed line.

table exception: the CE group at training. The homophone-like training pairs for the CE group are pairs consisting of the words *Barry* and *berry*, each spoken by one of the talkers CE1 and CE2. The productions of these words that the participants responded correctly to during the last training session were analyzed acoustically to determine what cues may have been used to identify the produced word with the correct image: the analyzed dimensions were vowel midpoint F1, F2, f0, as well as their tracks, and duration. None of these cues was identified as being able to reliably separate the *Barry* from *berry* items.

Comparing the CE group to the Multi group, Table 2.5 shows a positive estimate for CE group's performance on dissimilar words compared to Multi group's performance on dissimilar words. However, this comparison did not reach significance, hence, it cannot be concluded that their performance was reliably different (estimate = 0.615, p = 0.065). The same holds for this comparison with the Multi and GBE group (estimate = 0.513, p = 0.120). Hence, there is no sufficient evidence for either of the mono-accent groups performing differently from the Multi group on dissimilar words at test.

Comparing the CE group's identification accuracy for dissimilar words and contrastively produced MPs also did not reach significance (estimate = -0.949, p = 0.068), though the negative estimate indicates that this group may have performed worse on contrastively produced MPs than dissimilar words. No reliable difference was found in the CE group's identification accuracy for dissimilar words compared to homophone-like MPs (estimate = -0.214, p = 0.660).

Unlike the CE and Multi group, the GBE group showed significantly lower identification accuracy for contrastively produced MPs when compared to dissim-

ilar words (estimate = -0.998, p = 0.041). Like the other two groups, the GBE group was not found to show reliably different performance on homophone-like MPs compared to dissimilar words (-0.763, p = 0.137).

Figure 2.9 seems to suggest that the Multi children may be performing worse on dissimilar words than the other groups, and better on contrastively produced MPs than other groups. This is not, however, confirmed by pairwise comparisons (see Table 2.6). Hence, there is no sufficient evidence to support different performance of the multi-accent compared to the mono-accent groups when it came to MPs produced contrastively or dissimilar words.

contrast	**estimate**	**SE**	**z.ratio**	**p.value**
Multi fill - CE fill	-0.606	0.332	-1.826	0.449
Multi fill - GBE fill	-0.502	0.329	-1.528	0.646
Multi diff - CE diff	0.418	0.395	1.059	0.898
Multi diff - GBE diff	0.438	0.356	1.230	0.822

Table 2.6 Emmeans table comparing the multi-accent group to the mono-accent groups with respect to performance on dissimilar words and contrastively produced MPs. The non-significant effect of progress is not taken into account here. P-values are adjusted using the tukey method for comparing a family of 6 estimates.

The effect of progress did not reach significance (estimate = 0.275, p = 0.215), signalling that the data do not serve as sufficient evidence for different performance on the last training session compared to the test session. Hence, the comparisons described above hold for word identification performance at both test and training. Figure 2.10 illustrates performance regardless of progress.

In sum, participants in all groups performed above chance for dissimilar words, at both training and test. No group performed reliably above chance for either type of MP at either test or training, except for CE participants at training, who seem to have identified MPs produced as homophones with above chance success. For the Multi group, sufficiently better accuracy for dissimilar words compared to contrastively produced MPs could not be confirmed, and the same was true for the CE group. For the GBE group, however, the accuracy on dissimilar words was significantly higher than on contrastively produced MPs.

2.3.2 Sound replays

Next to word identification performance, the number of sound replays during the word identification task was analyzed to see if there were any differences in how many repetitions of words participants listened to before they confirmed their answer. No specific predictions were formulated for the number of replays.

Figure 2.11 illustrates the number of replays separately for condition, pair type and progress. The minimum number of plays that allowed a participant to move to the next trial was 2. Hence, 2 was subtracted from each trial's number of sound plays. Further, a number of plays equal to 3 is likely to simply indicate that the child wanted to mark the sound she had listened to first as her final choice. Hence, this "replay" probably does not indicate that the child wanted to hear the sound again. Figure 2.12 shows the proportion of trials where the number of repetitions was 0 or 1 (i.e. the minimum number or the minimum

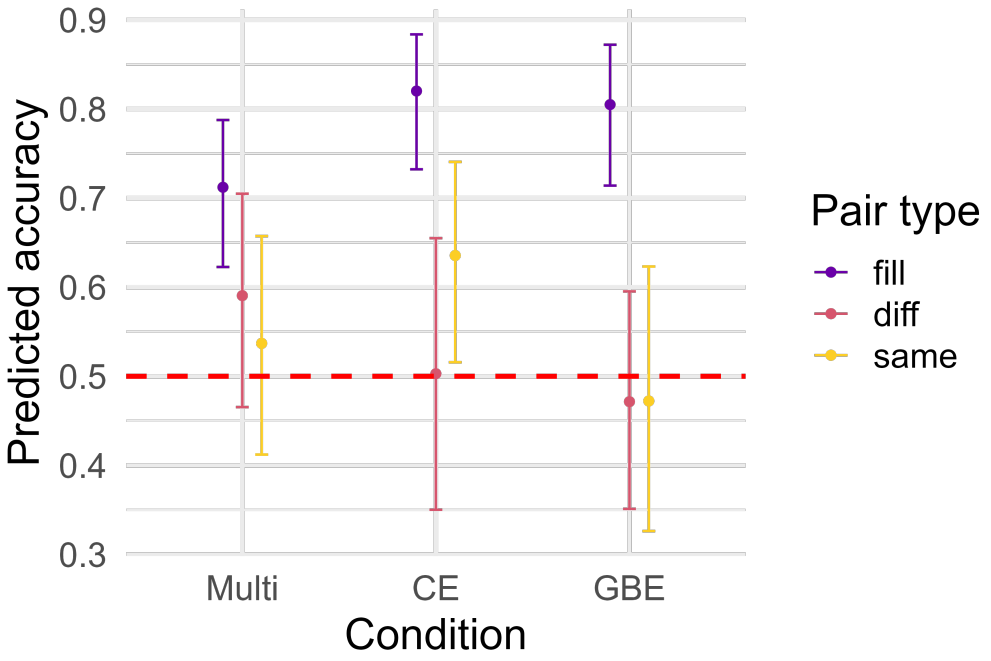

Figure 2.10 Estimated word identification accuracy for each group of participants regardless of progress. Points in error bars show mean success rates for fillers, minimal pairs produced contrastively (diff) and as homophones (same). Error bars show 95% CIs. The level of chance is highlighted by a dashed line.

number with an additional sound button tap to indicate final choice) to the trials where a participant chose to listen to a sound repeatedly (i.e. more than one replay). The highest proportion of trials where a child chose to repeatedly listen to a sound was 19.6% by CE children at test for same items; the smallest was 0% by the CE children at test for dissimilar word trials.

A GLMM was built to see if the probability of sound replays was different between the groups for MPs and dissimilar words. The two MP types were modelled together. The dependent variable, replay count, was binary and had levels *0* for no replays (i.e. the obligatory 2, or 3 button presses) and *1* for some replays on a given trial (i.e. more than three button presses). Like in the word identification model, condition was a factor variable with levels *CE*, *GBE*, *Multi*. Progress was a factor variable with levels *train* and *test*. *Stim* was a factor variable with levels *fill* for dissimilar words and *mp* for minimal pairs produced contrastively or as homophones. All factors were treatment-coded, the reference levels were *Multi*, *mp*, and *train*. The random factors remained the same as in the word identification model. The model formula was *replay_count ~cond * stim + progress + (1 | participant_id) + (1 | pair)*. The coefficient table is shown in Table 2.7.

The estimate of the intercept is negative and associated with a significant p-value (estimate = -3.204, $p < 0.001$), indicating that Multi children at test were reliably more likely to not replay the sounds on MP trials than to replay them. This is true for both dissimilar words and MPs across all groups at both training and test: Figure 2.13 shows that upper bounds of CIs never even reach 0.3. The comparison of the Multi group with either of the mono-accent groups did not

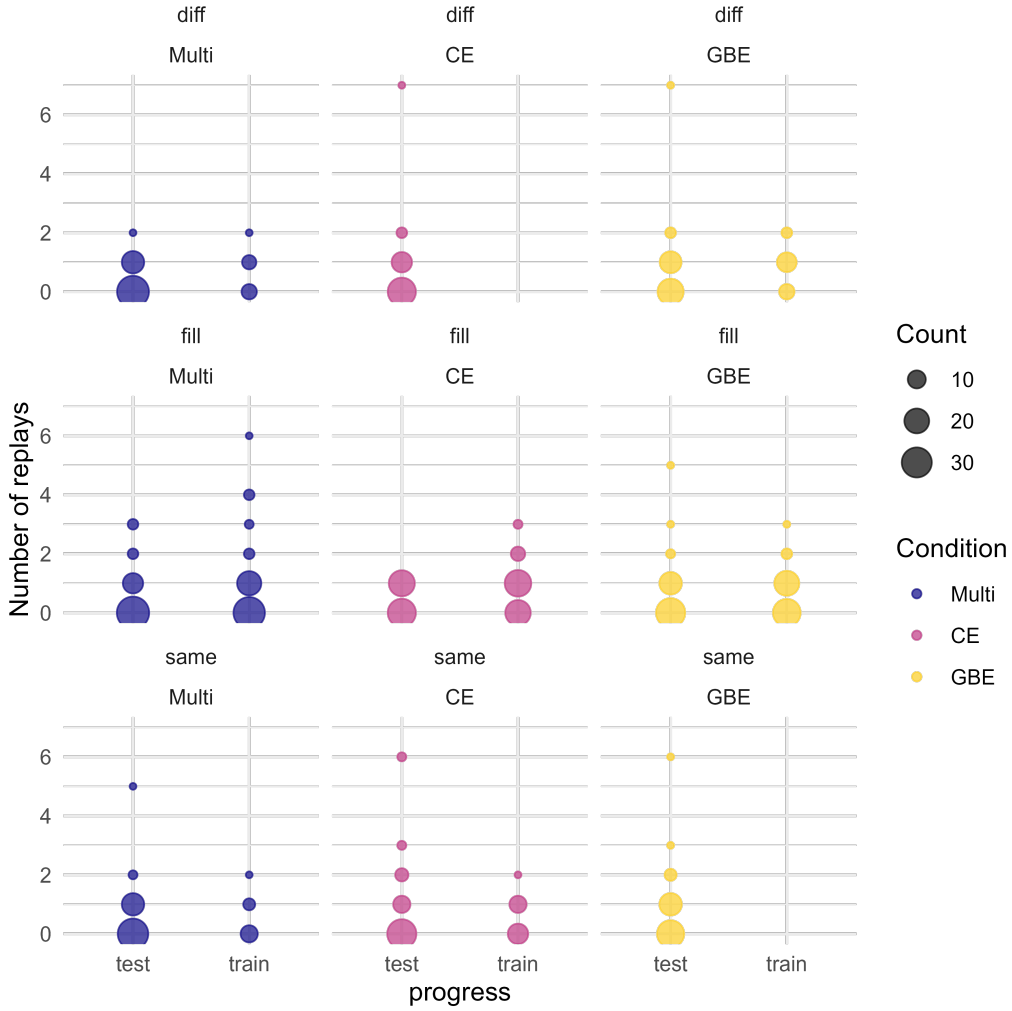

Figure 2.11 Plot of raw data summarizing the number of replays per condition, pair type, and progress.

term	estimate	std.error	statistic	p.value
(Intercept)	-3.204	0.603	-5.315	**0.000**
condCE	1.075	0.720	1.492	0.136
condGBE	1.186	0.721	1.645	0.100
stimfill	1.050	0.537	1.957	0.050
progresstest	-0.523	0.315	-1.661	0.097
condCE:stimfill	-1.826	0.733	-2.490	**0.013**
condGBE:stimfill	-1.587	0.734	-2.163	**0.031**

Table 2.7 Coefficient table for the model estimating whether children replayed or did not replay sounds during word identification. The reference of the outcome variable is 0 (no replays), the other level is 1 (the child replayed sounds). All fixed effects are treatment-coded, the reference level for stimulus is *mp*, for condition it is Multi, and for progress it is *train*. Hence, the intercept represents the estimated log odds of replaying MPs after training by the Multi group.

reach significance (CE: estimate = 1.075, p = 0.136; GBE: estimate = 1.186, p = 0.100), indicating that the Multi group did not reliably differ from either of the

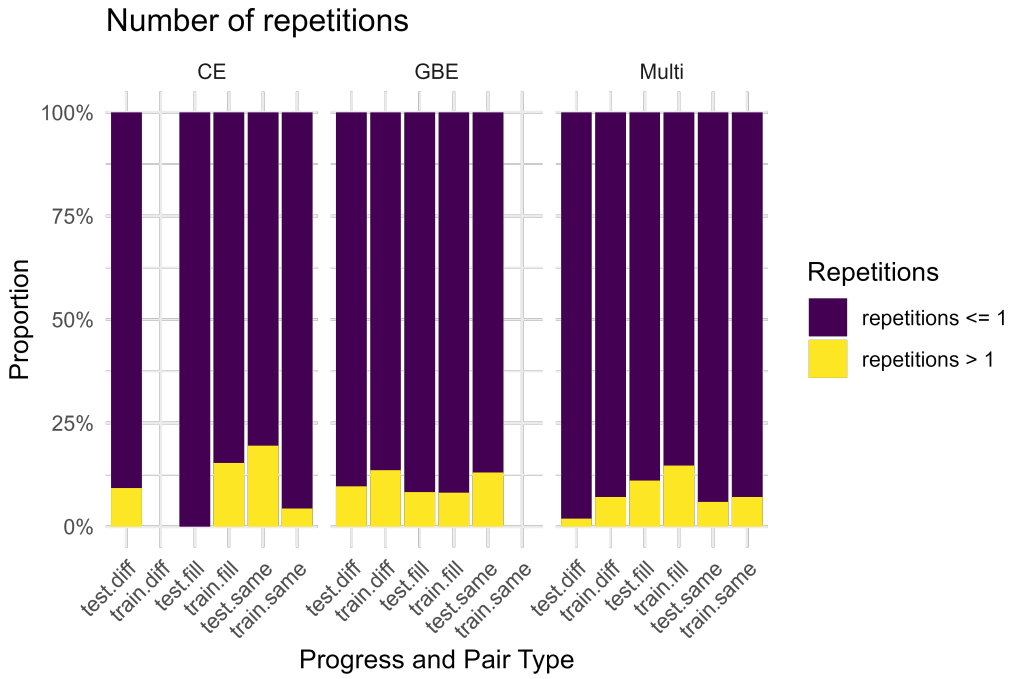

Figure 2.12 Plot of raw data summarizing number of replays larger than one in proportion to trials with one or no repetitions. Plotted separately for each group after training and at test.

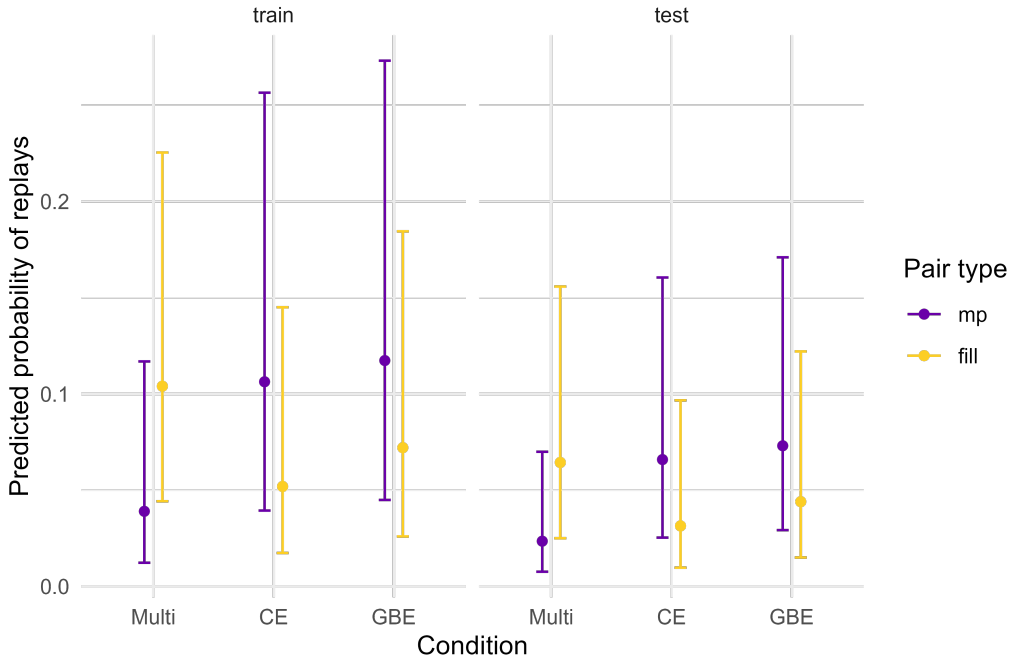

Figure 2.13 Plotted estimates for trials with more than one replay and trials with a smaller number of replays. Color represents pair type. Panels show progress, the horizontal axis shows condition, the vertical axis shows probability of replaying sounds on a trial.

groups in whether or not MP sounds were replayed on word identification trials. The comparison between the Multi group's replays of MPs and dissimilar words

does not reach significance, but its positive effect does not seem to be negligible (1.050, p = 0.050), the Multi children may show a tendency of replaying MPs more often than dissimilar words.

Progress is not associated with a significant p-value, indicating that the data did not serve as sufficient evidence for different performance on test compared to training. The interaction between the CE condition and dissimilar words reached significance with a negative estimate (estimate = -1.826, p = 0.013). The same was observed for the GBE condition (estimate = -1.587, p = 0.031). This indicates that the CE and GBE groups replayed MPs reliably more often than they did dissimilar words. The model estimates are visualized in Figure 2.13. It illustrates the pattern that GBE and CE children at test replay MPs more often than dissimilar words, while Multi children may show the opposite tendency: replaying words on dissimilar pair trials more often than those on MP trials.

2.3.3 Category boundary task

The category boundary task was administered to 31 children, ten of which were excluded: five due to atypical language development, five due to skipping the task before providing more than three responses. This task was later completely omitted from the procedure, as mentioned in the Method subsection. The main reasons for omitting the task were that it did not seem appropriate for the participants: many complained that the task is giving them the same sound repeatedly, they found the task boring and repetitive.

There were 21 participants remaining for analysis, eight from the GBE group, seven from the CE group, and six from the Multi group. Responses from the longer version of the task, completed by six of the 21 participants, were rescaled to fit the seven-point scale (see https://osf.io/7w2xk for details).[1] The responses on the seven point scale are plotted in Figure 2.15 for participants in the GBE group, Figure 2.16 for the CE group, Figure 2.17 for the Multi group. Continuum step number 1 corresponds to a production of *Barry*, step number 7 a production of *berry*.

I first check whether there was a preference for either of the endpoints, especially as the words differ in animacy. Overall, the participants seemed to be using both buttons to respond to the continuum stimuli, without a clear preference (see Figure 2.14).

The plotted data in Figure 2.15 suggest that none of the GBE-exposed participants categorized the continuum stimuli as percepts ranging from a prototypical exemplar of one category to a prototypical exemplar of another consistently. Performance like that of participant 8 and participant 11 at both training and test suggests a merger, i.e. the two words were likely being perceived as homophones, and the task did not offer contextual information that could help disambiguate their meanings. Performance like that of participant 2, participant 17 at training, and participant 23 at test is also a possible signal of a merger, but may additionally point to a difference in word familiarity, or may signal that some aspect

[1]The formula used for rescaling was $rescaled\_sound = round((original\_sound - 1)/(51 - 1) * (7 - 1) + 1)$. This assigned values from 1 to 51 to seven bins, overwriting the original values according to the bin they were assigned to. The first and last bins were assigned five values, bins 2, 3, 5, and 6 were assigned eight values, and bin 4 nine values.

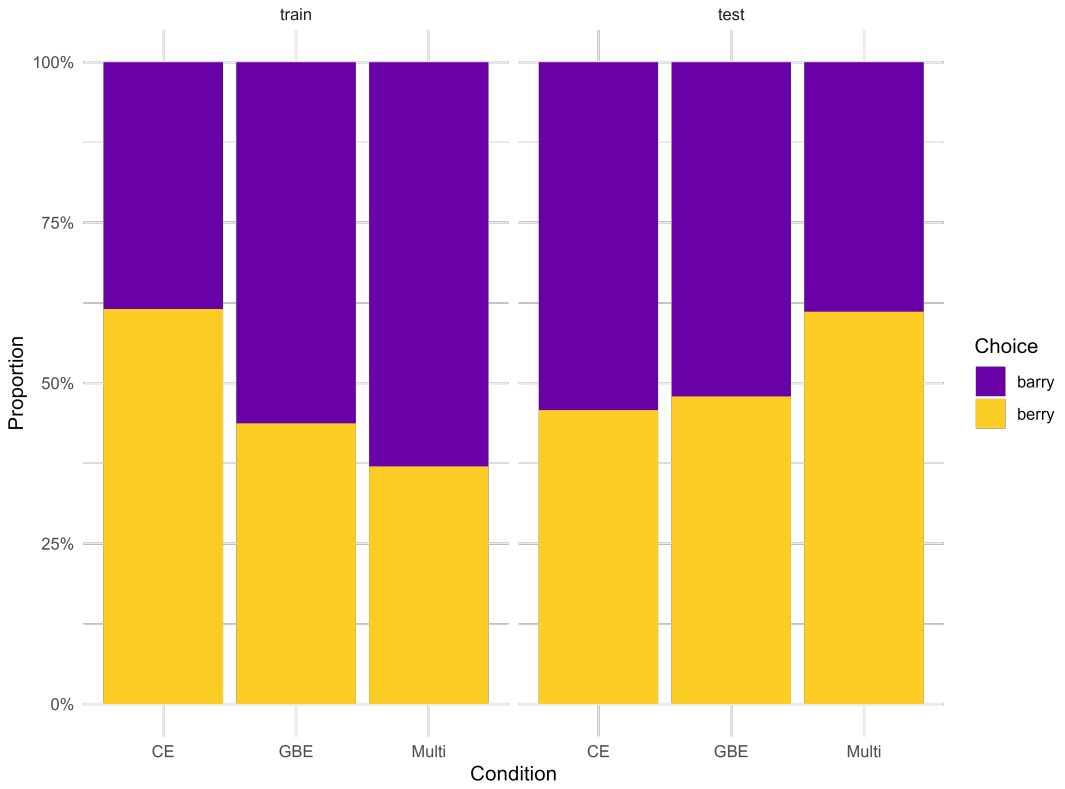

Figure 2.14 Raw classification data for the three groups. Proportion of answers irrespective of the sound.

of the context, lead the participant to overwhelmingly prefer one word over the other. The performance of participant 14 at training may indicate categorical perception with a boundary between steps 5 and 6. At test, only four datapoints are available for this participant, and these seem to go against perception of two categories in the continuum.

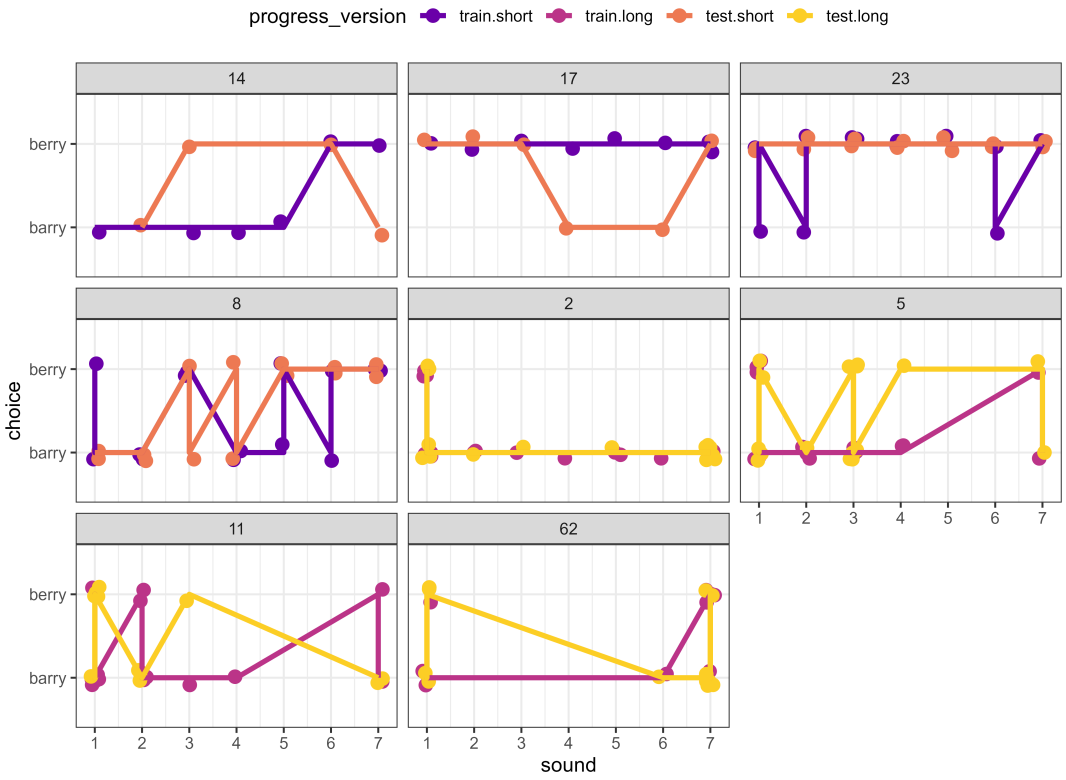

Figure 2.15 Raw classification data for the GBE group. The last four facets, which use the pink-yellow color pair, mark the participants who completed the longer task version with 51 stimuli (rescaled in this plot).

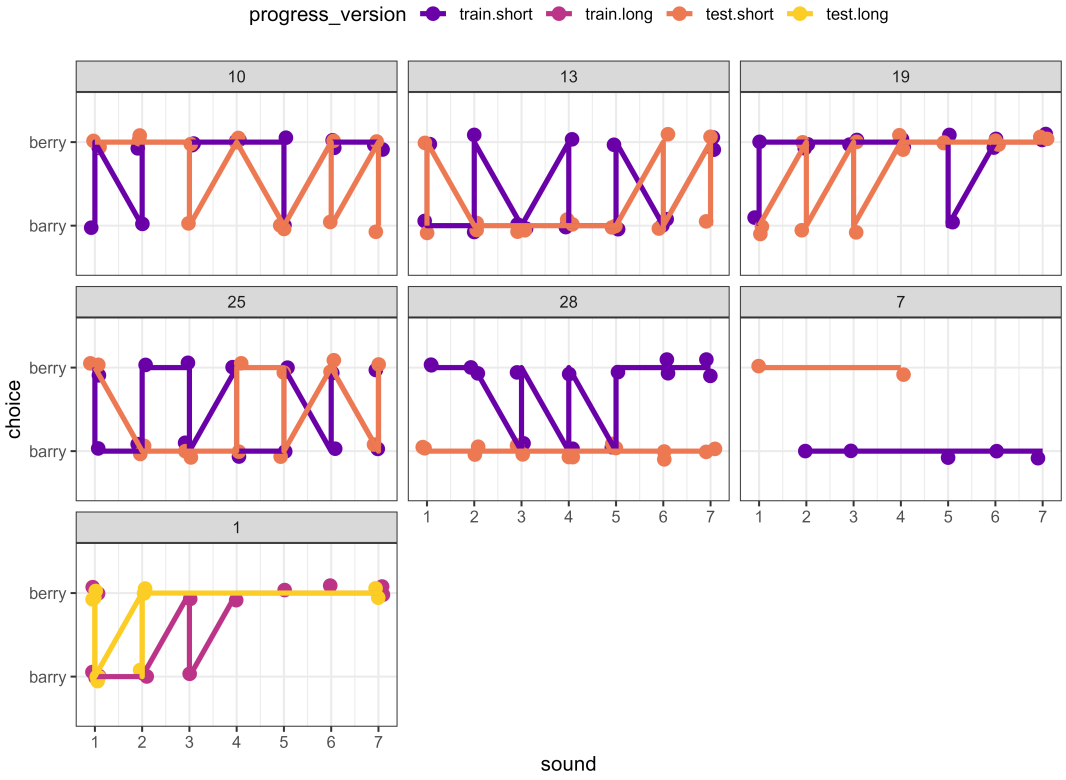

Figure 2.16 Raw classification data for the CE group. The last facet, which uses the pink-yellow color pair, marks the participants who completed the longer task version with 51 stimuli (rescaled in this plot).

Figure 2.16 shows patterns similar to Figure 2.15. No CE-exposed participant's performance indicates perception of two categories distinct in quality. Perception of homophones is indicated by all participants except perhaps participant 7. This participant provides little data, and may pattern with the former group. Notably, categorization of participant 7 is in conflict with the intended meaning of the words: continuum step 1 corresponds to a production of *Barry*, step 7 to *berry*.

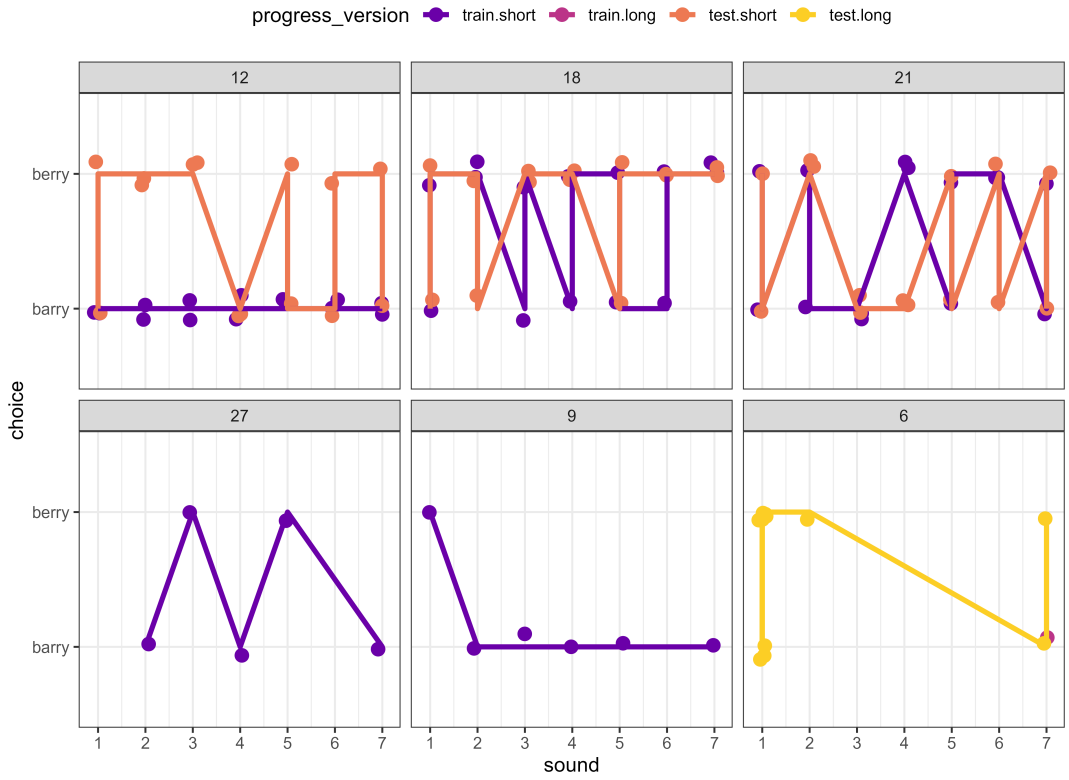

Figure 2.17 Raw classification data for the Multi group. The last facet, which uses the pink-yellow color pair, marks the participants who completed the longer task version with 51 stimuli (rescaled in this plot).

From visual inspection of Figure 2.17, similar patterns as those found in Figure 2.15 can be seen: no Multi-exposed participant except maybe participant 9 seems to indicate perception of two categories. However, participant 9 does not provide many datapoints, and classifies continuum step 1 a *berry*, and further steps as *Barry*, even though step 1 corresponds to a production of *Barry*. Hence, this participant's performance may be more in line with that of e.g. participant 12 at training. Overall, performance of participants in the Multi group does not seem to indicate perception of two distinct categories with continuum endpoints serving as prototypical exemplars of the two categories.

Even though most of the classification functions do not indicate perception of two distinct categories, I decided to also inspect the overall proportions of *Barry* vs. *berry* responses to each of the seven continuum steps, to check for a possible presence of a trend towards classifying the continuum endpoints as the intended words. Figure 2.18 shows the proportions with which the two responses were chosen for each of the seven continuum steps, separately for each condition

after the last training video and after the test video. An ideal categorizer would respond with *Barry* to step 1, and with *berry* to step 7. For listeners who have acquired the contrast, continuum sounds closer to an endpoint should make more satisfactory exemplars of a category, which would be reflected in high proportion of answers indicating this category. This is not a pattern suggested by the present data.

Generally, an increase in *berry* answers with increasing continuum sound index may be present for GBE and CE listeners at training. Note, however, that this data is simply averaged over participants without any measure of between-participant variability. An increasing trend of *berry* answers does not seem to be indicated by any group at test, and also by the Multi group at training. At training, the GBE and CE group seem to give more *berry* answers to the appropriate endpoint than *barry* answers, and a corresponding pattern is present for the *Barry* endpoint. A similar trend is not present for the Multi group, who seem to be just as likely to answer *berry* to either endpoint. This also may be true of the CE and Multi group at test. The GBE group at test seems to show the inverse pattern: more frequent *berry* responses to the *Barry* endpoint, and vice versa.

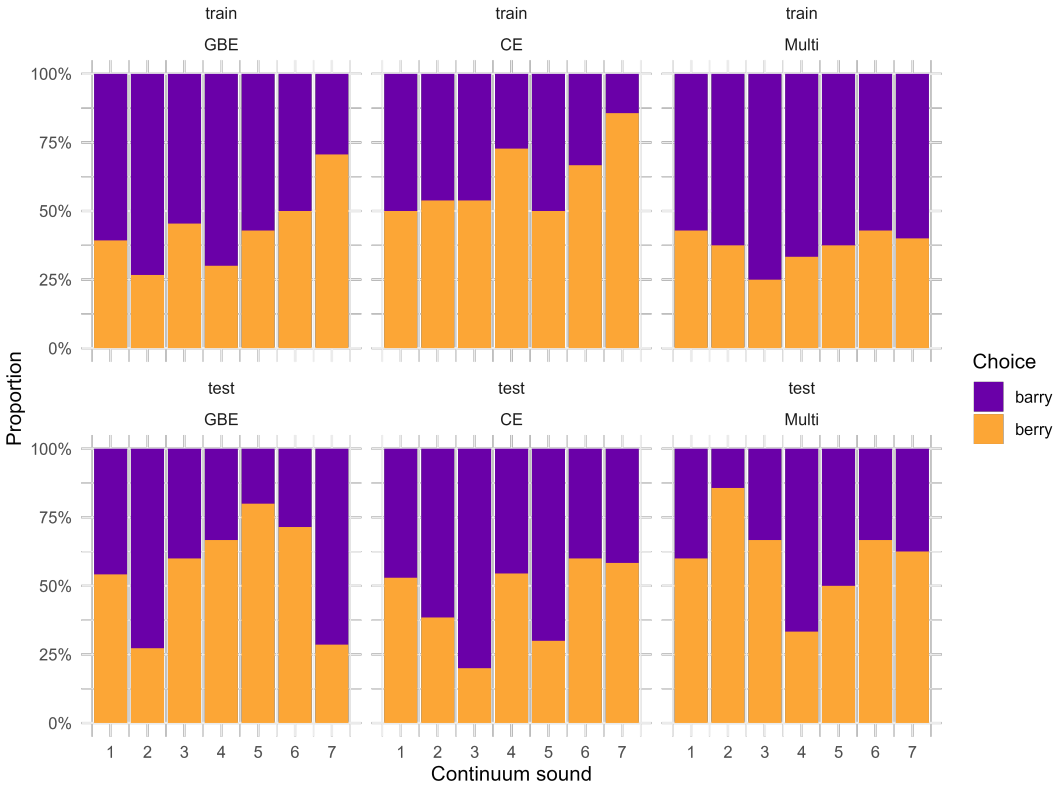

Figure 2.18 Proportions of *Barry* and *berry* responses to each continuum step from raw data. Plotted for each condition at training and test.

Figure 2.18 shows proportions of the two responses calculated from raw data. Generally, at training, it looks like *berry* responses may be increasing for the GBE and CE groups with continuum sounds that approach the *berry* endpoint. This trend does not seem to be present for the CE and Multi group at test. GBE children responded more with *Barry* on continuum step 7 (corresponding

to *berry*), which also indicates a lack of preference of *berry* as the continuum nears the *berry* production. Overall, the categorization data seem to suggest that children in any group perceived the members of the minimal pair as homophones.

2.3.4 The Dimensional Change Card Sort task

The DCCS task was used as a independent indicator of participants' cognitive flexibility. Performance was not expected to differ between groups, given that they were balanced in age. An LM was built to assess whether scores on the border version of the task differed between groups. The outcome variable was *border score*, *condition* was a three-level factor predictor (levels CE, GBE and Multi, as in previous models), age was a numerical predictor (expressed in months). Condition was treatment coded, the Multi group served as the reference level. The model formula was specified as *border_score ~condition + age*. Age was included because it modulates performance on the DCCS task (Zelazo, 2006), although the relatively narrow age range of participants who took part in this study may not produce sufficiently large differences in border scores. The score was formalized as the number of correct responses given by the participant divided by the total number of responses given by the participant. The task included 12 trials, however, one participant only completed 11 trials. Data from 50 participants (out of the 51) were included (as mentioned above, one participant only provided data from training).

As the coefficients in Table 2.8 indicate, the data did not serve as evidence of a difference in border scores based on condition. The effect of age also did not reach significance, likely because all participants were all almost five years old or older: at the studied age interval, differences may not arise due to similarly developed cognitive function. Figure 2.19 shows border scores as a function of age, with regression lines and their 95% CIs for each group. In sum, a difference between the groups on children in cognitive function cannot be concluded based on the present data.

term	**estimate**	**std.error**	**statistic**	**p.value**
(Intercept)	0.305	0.247	1.234	0.223
condCE	0.069	0.063	1.104	0.275
condGBE	-0.041	0.063	-0.645	0.522
age	0.007	0.004	1.831	0.074

Table 2.8 Coefficient table for the model estimating DCCS score as a function of condition and age. The reference level for condition is *Multi*. DCCS score is the number of correct responses divided by the number of total responses given.

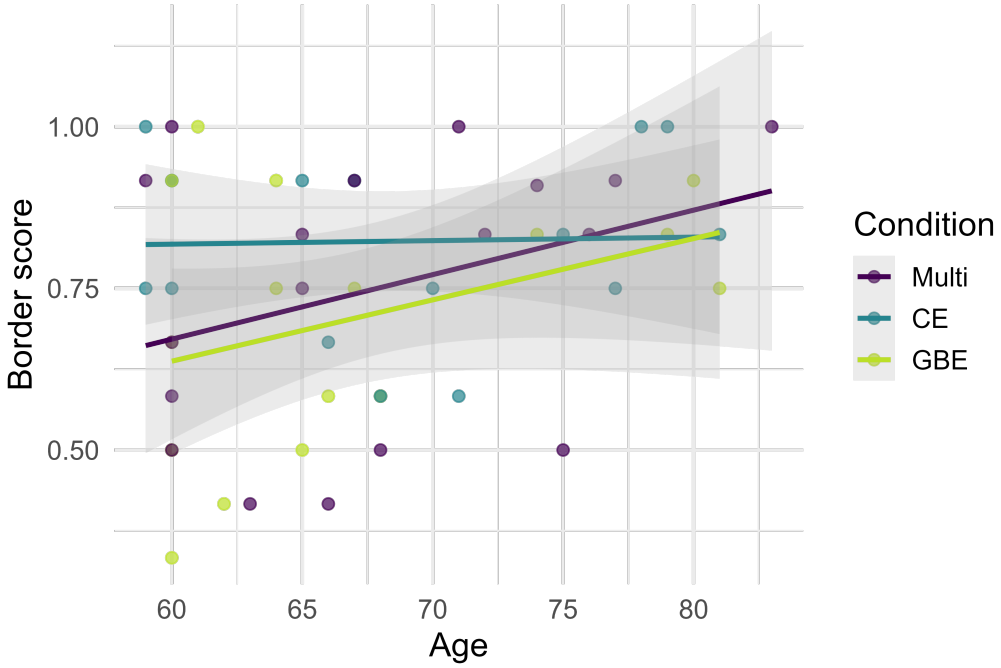

Figure 2.19 Plotted DCCS border version scores (as proportion of correct responses to all responses). Regression lines for the different conditions are plotted with shading representing 95% CIs.

2.4 Discussion

Long-term exposure to multiple L1 accents has been reported to facilitate processing of these accents in two year olds (van der Feest et al., 2022). Adults have been shown to adapt in the short-term using acoustic-phonetic similarity rather than accent-based groupings (Xie & Myers, 2017). Child L2 learners adapt using general criterion relaxing following brief exposure to an unfamiliar accent (Hu, 2021). The question addressed in the present study is whether training with short-term multi-accent exposure helps child L2 learners adapt to vowels in unknown L2 accents using the targeted shifts mechanism. This question is tested on word identification of minimal pair (MP) items that differ in a difficult vowel (/ɛ/-/æ/). Talkers of unfamiliar accents were chosen based on acoustic-phonetic similarity regarding the production of the two difficult vowels: a General British English talker produced vowels [ɛ], [æ] at training, similarly to a White Southern African English talker at test; a Czech English talker produced words with a merged category [ɛ] at training, similarly to a Singapore English talker at test.

2.4.1 Word identification

The word identification task administered after the last training video assessed whether children had learned the /æ, ɛ/ contrast. For CE-exposed children, the aim was to confirm that they perceived GBE MPs produced by CE talkers as homophones. Having acquired the relevant contrast would be signalled by above-chance performance on trials with minimal pair words produced contrastively (i.e. one word containing [æ] and the other [ɛ]). The GBE-exposed children were ex-

pected to perform above chance on these items at training. CE-exposed children only encountered homophone productions at training, where they were expected to perform at chance. The multi-accent-exposed children were expected to perform above chance on contrastively produced MPs, unless the exposure to diverse input hindered their word processing or learning. At test, trials with MPs produced as homophones were included for comparison with contrastively produced MPs. For these items, performance was expected to be at chance in all groups. Filler trials were included at test to assess whether children remembered dissimilar words from the novel-accent input. The expectation was that all groups would identify dissimilar words with above-chance success, both for training accents and test accents. The question this study addresses is whether short-term multi-accent exposure modulates identification of words from difficult minimal pairs in a novel accent.

Dissimilar words

In contrast with Kartushina et al. (2022), who reported multi-accent child listeners to be able to learn words more efficiently than mono-accent peers when they are presented in two accents, this study did not find any difference in the identification of dissimilar words following short-term mono-accent or multi-accent exposure. The children Kartushina et al. (2022) reported on were exposed to L1 accents in the long-term at home, which is a situation that is difficult to compare to L2 learning: a direct comparison between this study and Kartushina et al. (2022) is not possible due to the focus on populations that differ in their long-term exposure to accent variability. In this study, after both training and test, all groups performed above chance for word identification in dissimilar word pairs. This indicates that the participants were able to learn words from the short-term video exposure even from the novel accents. There was no sufficient evidence of the groups differing in their identification accuracy for dissimilar words. All groups demonstrated similarly high identification of dissimilar words.

Homophone-like MPs

Unlike Creel (2012), Maye et al. (2008), and White and Aslin (2011), this study uses natural accents with the aim of increasing the ecological validity of its findings. However, using natural accents has its pitfalls. Among them is not being able to control all cues present in the speech: it was previously stated that the CE talkers produced minimal pairs as homophones, and hence, CE-exposed children were expected to perform at chance for these items. This is, however, not what was found. CE-exposed children at training performed reliably above chance on homophone-like minimal pair items produced by two CE talkers. The 95% confidence interval for the CE group extended up from 53% accuracy. This is surprising, as acoustic measurements revealed no differences in F1, F2, duration or f0. Which cue(s) the children attended to successfully identify these words is unclear. While it is possible that the 95% confidence intervals do not contain the true accuracy of the population, other possible explanations remain. One is exemplar representation of words, and the other is use of cues overlooked by the analysis.

An exemplar representation-based explanation is possible, although it does not seem likely. During exposure, video presentation was repeatedly interrupted by single word-image presentation to help participants associate a given referent with a single target word. Each interleaving word presentation featured four word tokens, two from each talker. The same word tokens were presented in the word identification task in carrier phrases. Assuming at least a partially exemplar-based account of lexical representation, it may be possible that the participants stored the word tokens in sufficient acoustic-phonetic detail as instances of a given word, and were thus able to recognize the items in the word identification task as the intended lexical items, even if acoustic-phonetic cues did not signal a given item reliably.

However, if this were the case, why would the children in the other exposure groups, or the CE group at test, not be able to exploit this mechanism? Multi-accent-exposed participants also heard MPs produced as homophones, not from two CE talkers, but from one CE talker and one GBE talker (both produced [ɛ]), and this was not sufficient to boost their identification of MP items. From the perspective of Pajak et al. (2016), L2 comprehension should be higher if the talker and listener share an L1, because in L1-biased input, statistical cues conflict less often with the listeners' L1-biased beliefs. Sharing the bias may allow listeners to process minimal pair words more easily. An interlanguage benefit for L2 intelligibility has been reported in adults (Hayes-Harb et al., 2008). Processing Czech-accented input may have eased the cognitive load of contending with L2 input, and this ease of processing may have given way to storage of more detailed word exemplars, allowing participants to identify words based on their more detailed representations. Children in the other groups had to contend with at least one GBE talker, which may have increased the processing load of L2 speech.

While it may be possible that exposure to English from talkers who share the L1 with the listeners facilitated easier processing of speech for the CE group, there is an alternative explanation to the one proposed above. The combined effect of exemplar representation and cognitive ease of processing L1-accented speech may not have been what helped them achieve above-chance word identification. Word identification may have been facilitated by a combination of acoustic cues that was not uncovered by the acoustic analysis, especially as it was based on few word tokens used in the word identification task. Participants may have used any combination of the four most likely cues (F1, F2, f0, duration), in the relevant vowel or even in a different segment, to infer the lexical identity of the auditory stimulus.

Though they did not show reliably better accuracy at identifying dissimilar words than MPs, CE children show a similar pattern as GBE children in sound replays: they replay minimal pairs more often than dissimilar words. This is despite CE children identifying MPs produced as homophones reliably at training, while GBE children did not identify contrastively produced MPs reliably above chance; and also despite CE children not showing reliably greater identification accuracy on dissimilar words than minimal pairs at test. Why the CE children pattern with the GBE children in replaying MPs more often than dissimilar words is not clear. If they are relying on detailed exemplar representation, is possible that they need to pay increased attention to the auditory form of the MPs to match them with those detailed exemplars. Replaying sounds would allow participants

to carefully assess the tokens.

The division between general criterion relaxing in adaptation and equivalence classification in perception itself does not seem clear-cut when dealing with L2 adaptation. It is not clear whether successful minimal pair identification using L1-based cues in the L2 qualifies as acquiring the L2-specific contrast, or whether it is more appropriately considered acquisition of an interlanguage-specific contrast. If classified as an interlanguage contrast, would it then be possible to say that CE-exposed listeners adapted to test stimuli using targeted shifts, having acquired a non-native accent-specific contrast, rather than an L2-specific contrast?

Contrastively produced MPs

Results of the present study show that at training, two of the three groups, and at test, all groups failed to demonstrate reliably above-chance performance on minimal pair trials, even when minimal pairs were produced with contrasting vowels. Despite their experience with English, it cannot be concluded that the children could identify the MP words as different based on containing the two contrasting vowels. Neither the multi-accent children nor the GBE-exposed mono-accent children performed reliably above chance for MPs even at training, indicating that they may not have yet acquired the /ɛ-æ/ contrast, which, depending on the theoretical perspective one takes, can be seen as a prerequisite for targeted shifts adaptation in this case.

No group yet having acquired the target contrast is evidenced by performance on the category boundary task. Participants did not demonstrate perception of two categories best exemplified by qualitatively distinct continuum endpoints. The lack of evidence for categorization of [ɛ] and /æ/ as distinct is in conflict with one of the core assumptions of Kleinschmidt and Jaeger (2015). Unlike L1 adaptors, beginner L2 learners cannot be labelled ideal phonetic categorizers. The currently studied population does likely depend on knowledge transferred from L1. When it came to minimal pairs, participants did not show optimal inference of intended categories provided the cue values. Hence, at this stage of L2 learning, L1 experience seems to constrain distributional learning, blocking perception of two categories given relevant bimodally distributed cues. The categorization observed indicates lexical misrepresentation of /æ/ words as containing /ɛ/, often reported in even adult L2 learners (e.g. S. L. Barrios et al., 2023; Llompart & Reinisch, 2019).

Word identification accuracy was complemented by data on whether the participants replayed sounds on word identification trials or listened to the minimum amount of sounds. The results indicate that all children, at both test and training, and for both MPs and dissimilar words, were more likely to not replay sounds than replay them. Further, the GBE group was reliably more likely to replay MPs than dissimilar words. This complements word identification data: GBE children replay dissimilar words, which they identify with reliably greater accuracy, more than contrastively produced MPs, for which they are at chance performance. The data on sound replays seem to suggest that children are unsure about which word they should pair with the displayed image, and they seem to be aware that the auditory form includes some disambiguating information, hence they replay the sound. However, they are either not able to tease apart the auditory forms of the two words, or perhaps not able to associate them with sufficiently specific

lexical representations. Importantly, the differences in replays do not seem to be connected to disparities in executive function, as analysis of the DCCS scores did not indicate between-group differences in executive function.

The *hybrid flexibility hypothesis*, formulated for child L1 adaptation, states that targeted shifts are used to adapt to a novel accent only after sufficient evidence about the shifts has been collected. Here, there is one more barrier blocking targeted shifts adaptation: L2 category establishment.

Unlike White and Aslin (2011) this study did not find adaptation by targeted shifts. However, there are enormous differences between this study and L1 accent adaptation studies on infants like Buckler et al. (2017) and White and Aslin (2011). Apart from White and Aslin (2011) focusing on L1 adaptation in much younger children (18-20 month-olds), the accent White and Aslin (2011) used was a manipulated version of the children's L1 variety with a shift in front vowels.

Since the participants demonstrated successful identification of MP members on trials where they were contrasted with dissimilar words, it is unlikely that the children had not learned any of the MP members. The most likely scenario seems to be the use of a wider L1-biased /ɛ/ category into which tokens of both L2 [æ] and L2 [ɛ] are accepted. In SLM-r (Flege & Bohn, 2021) terms, this is a demonstration of equivalence classification for [æ] and [ɛ]. If the currently studied population does use an /ɛ/-/æ/ merger, comparing toddler L1 adaptation and child L2 adaptation may be misleading, as, strictly speaking, the reason behind using more variable categories may not be general criterion relaxing adaptation, but equivalence classification supporting the use of a merged category in perception. However, the conceptual separation of the general criterion relaxing mechanism and equivalence classification may not be so clear-cut to allow the consideration of one as a prerequisite for the other.

Not finding the use of the targeted shifts mechanism, does not preclude adaptation using general criterion relaxing. I cannot directly address the use of general criterion relaxing due to the design of the word identification task used at test, but it is possible that the participants made use of this mechanism. The children in this study may have well adapted to non-minimal pair stimuli like that from Buckler et al. (2017) and van der Feest et al. (2022), but this is something that cannot be reliably claimed based on the collected data, as no items at test probed adaptation to non-words or pairs with a member of a minimal pair and a dissimilar word. Items like these were included at training, and performance on them does indicate general criterion relaxing adaptation, or learning of minimal pair members with underspecified target vowels as a result of equivalence classification.

If input signals categories that a listener has not acquired, targeted shifts adaptation is precluded. However, adaptation using general criterion relaxing is still possible, depending on the theoretical perspective one takes. There are similarities in accent adaptation and L2 learning when it comes to using variable categories in perception. The general criterion relaxation mechanism observed in child L1 adaptation (e.g. Schmale et al., 2015) has some commonalities with the mechanism of equivalence classification adopted by Flege (1995) to describe L2 category learning based on the L1 category acquisition concept of *equivalence classes* (Kuhl, 1983). General criterion relaxing is a mechanism by which a listener increases the likelihood of word tokens from input mapping onto lexical

representations by making speech sound representations more variable. Equivalence classification in the L2, is viewed as a consequence of having established a speech sound inventory through which novel input is perceived. Equivalence classification is seen as blocking L2 category formation by perceptually linking sounds signalling distinct L2 categories (Flege, 1995). Both general criterion relaxing and equivalence classification pertain to mapping incoming sound tokens indicating different categories onto a single category. They also share implications about high minimal pair confusability.

A. Cooper et al. (2023) claims that children use the general criterion relaxing strategy to cope with a distant accent upon first encounter due to not having enough evidence of specific cue-to-category mappings (congruent with E. K. Johnson et al., 2022). Schmale et al. (2015) claims that children's categories are more yielding to general criterion relaxing than targeted shifts when compared to adult learners due to children's categories being less robust. L2 learners' representations may also be less robust than L1 learners'. This would support the use of wide categories in the currently studied population. It is possible that the results of this study point to either equivalence classification of the two target vowels, or that this equivalence classification has been overcome, but the learners adapted to both training and test stimuli using general criterion relaxing, except for the CE group, who would be assumed to have adapted to minimal pair members produced by CE talkers using targeted shifts. This seems unlikely, as most children reported being familiar with GBE. If they had acquired the GBE /æ-ɛ/ contrast, at least the GBE-exposed group would have shown successful identification of minimal pair members following exposure to four videos featuring GBE talkers. The fact that all three groups, including the GBE-exposed group, did not perform reliably above chance on identification of contrastively produced MP members suggests that this population has not yet lexically encoded the difficult /ɛ-æ/ contrast.

Buckler et al. (2017) expands on the idea that listeners need to reach a minimum amount of exposure to adapt using targeted shifts (E. K. Johnson et al., 2022): before the exposure threshold is reached, children may be using underspecified lexical representations, hence adapting using general criterion relaxing. A question that arises from this is whether this exposure threshold would be lower for mono-accent children (here, GBE-exposed children) than for multi-accent children, as Buckler et al. (2017) suggest. A. Cooper et al. (2023) suggest that the general expansion strategy is used when the context does not suffice to disambiguate words, or when the learner has increased uncertainty about the cue-to-category mapping being used. Both the unsupportive context and the increased uncertainty seem to be at play for participants in this study.

In general, participants were able to identify words heard during the short-term video exposure even in the novel accents, but only if the words were sufficiently dissimilar from one another. Participants in the current study cannot be compared to L1 users in a straightforward way. However, the results of this study do highlight some parallels between early stages of L1 and L2 learning. Children in the current study did not demonstrate the ability to identify minimal pair members that differed in a difficult vowel. The number of varieties they were exposed to in the short term was not confirmed to affect their identification of either dissimilar words or minimal pair members. This agrees with findings in L1

speech processing research by infants: Kemp et al. (2017), who assessed minimal pair word learning in 18-20 month olds using a switch task, found that they could not, as a group, learn members of minimal pairs. When vocabulary size was taken into account, it was found that children with relatively larger vocabulary size did learn the members of the minimal pairs, while those with smaller vocabularies did not. This is a pattern that is apparent in both L1 acquisition and L2 learning (Llompart, 2021). The children in the current study may not have formed sufficiently large vocabularies to be able to encode minimal pair members as distinct lexical items.

Strong prior beliefs

In their L1, both adults and children adapt fast (Bradlow & Bent, 2008; White & Aslin, 2011). The results of this study highlight that L2 learners' experience is typically dramatically different from that of native listeners. Implicit knowledge of cue-to-category mappings aids efficient word recognition only if listeners' implicit assumptions align with the statistical patterns in the current input. In the case of the population studied here, experience mostly comes from the L1, and results in L1-biased perception of L2 words. Listeners recognize phonological categories indicated in input by using and updating their knowledge about these cue distributions only if their prior beliefs approximate the characteristics of the underlying categories from which evidence was generated (Kleinschmidt & Jaeger, 2015). Results of the current study point to a lack of targeted shifts adaptation to novel accents by all listener group irrespective of the number of accents present in the input.

Due to the confidence parameter being able to stand for various influences on adaptation, the IAF (Kleinschmidt & Jaeger, 2015) is able to provide theoretical support for the observed lack of targeted shifts adaptation. The present findings can be interpreted as the learners having high confidence in their prior beliefs (that specify a single /ɛ/ category in the vowel space region of interest). An L2-specific view provided by Pajak et al. (2016) would express this issue as the learners' hypotheses about English input being strongly biased by the L1, e.g. the learners may (uncounsciously) hold the hypothesis that GBE and WSAE talkers use a single vowel quality in the mid- to low- front region of the vowel space.

Not only was the short-term exposure not sufficient to change their beliefs about category characteristics, but it seems that an important assumption of Kleinschmidt and Jaeger (2015) was violated: the listeners did expect the appropriate number of categories in the input. The findings of this study underscore that distributional learning is constrained by previous experience, including the already established categories (Chládková & Šimáčková, 2021; Chládková et al., 2022). L2 learners make non ideal inferences about cue-to-category mappings, which may be caused, among other things, by expecting a different number of categories that the talker uses. The Czech-based unimodal prior belief about vowel height and retraction cues seems to outweight the bimodal cues present in English evidence. The multi-accent learners in this study do have wider experience than their mono-accent peers, but the quantity of that wider evidence of various generative models does not seem to have been sufficient to shift their beliefs.

Phonetic and phonological variability

Like the eight year old L2 learners in Hu (2021), participants in this study did not adapt using targeted shifts. Buckler et al. (2017) and van der Feest et al. (2022) report results of word recognition with no minimal pair stimuli included. As supported by the current study, minimal pairs are difficult to acquire. This is true even for children in their L1. While it has been found that idexical variability in input facilitates speech sound category learning (Rost & McMurray, 2009), it also leads to more effortful processing and identification, because listeners need to adapt to each new talker including the use of unique contextual and indexical information to successfully identify speech sounds (Pisoni, 1993). There is a trade-off here: support of robust speech sound categories is accompanied by higher processing load, even if talkers only differ in phonetic aspects of their speech.

The test stimuli in this study additionally differed in how many phonological categories were signalled by the input. The multi-accent children experienced phonological variability also during training. Simulations by Lev-Ari (2018) suggest that when L2 speech sound categories in input are known by the listener, performance improves with multiple talker input, but at the earliest stage of L2 learning, when categories are not known, variability does not improve performance. "Early stage" of L2 learning is not a clearly defined time window, but judging by the present learners' inability to differentiate minimal pair words, they may still be considered to be at an early stage of learning. In this case, even a benefit of mono-accent multiple talker input may not be present

Equivalence classification and general criterion relaxing

In the adaptation literature, successful word identification is viewed as a result of appropriately changing cue-to-category mapping. In the L2 category learning literature, successful word identification in the case of a difficult contrast is the result of establishing a new category that cues can be mapped onto. Increases in category variability, which in L1 adaptation point to learning, are in L2 acquisition are labelled the result of equivalence classification, hence suboptimal learning. But increasing category variability in reaction to L2 stimuli with a difficult speech sound is also adaptive to the input, it does point to learning. A variable L1-reused category in the L2 is indicative of L2 learning, even if this strategy may lead to minimal pair confusability.

Unlike L2 category acquisition research, child L1 accent adaptation studies typically do not deal with merged categories, rather, they focus on category shifts. In line with this, the account of Kleinschmidt and Jaeger (2015) relies on listeners knowing what categories there are in the input. Adaptation is mainly discussed as shifts (or variability increases). But adaptation, even in the L1, is closely tied to category formation, as varieties often differ in their phonological inventories. From this perspective, adaptation closely interacts with category formation, or may even include category formation as a prerequisite to targeted shifts adaptation. To be able to understand L1 or L2 accent acquisition in depth, future research would benefit from exploring adaptation to novel accents featuring split or merged categories next to category shifts.

2.4.2 Limitations and future research

Given the listeners' early stage of L2 learning, testing adaptation to a difficult L2 contrast may not have been appropriate. Had learners been tested on a vowel shift that traversed L1 vowel category boundaries, the learners may have been able to adapt using targeted shifts.

Both Van Leussen and Escudero (2015) and Kleinschmidt and Jaeger (2015) predict that with greater experience, multi-accent participants should be able to identify words successfully. With enough input, both the multi-accent-exposed and GBE-exposed children in this study should be able to correctly identify contrastively produced MPs with enough input. Maybe then, they may also be able to adapt to the novel accents after short-term exposure. This study could not satisfactorily assess the quality of input that the children were getting. The sample examined here is possibly too heterogeneous, resulting in high between-participant variability in word identification. Determining the amount of experience or exposure that would allow the children to use two categories to process the minimal pairs is not straightforward, as L2 category formation is modulated by a myriad of factors and is highly variable between individuals. Though some have indicated it possible (Simon et al., 2016), it is not entirely uncontroversial that children may actually not be able to learn efficiently from media input (based on research with naive infant listeners Kuhl, 2004). The non-immersion population of interest seems to require higher input quantity, and possibly also more interactive input, to acquire L2 categories.

Multiple questions arise from the lack of confirmed differences between mono- and multi-accent-exposed L2 learners. Since multi-accent 24 month-olds show processing disadvantages, but these are no longer present and 34 months old, would older (or more experienced) children be able to process multi-accent input as efficiently as mono-accent input when it comes to learning of dissimilar words? I do not know how much it would take for the participants to show targeted shifts for this specific contrast, but acquisition of the difficult contrast is a prerequisite.

Due to having to keep the word identification task to a minimal number of trials to not overtax the children's attention, the test did not include filler items heard during training but not at test. This is why I could not assess general criterion relaxing adaptation using filler items. Further studies would benefit from a design incorporating these items in a way that would not overtax the participants' attention.

The training exposure up until the last training video was not controlled to a large degree, as the participants watched the videos online from home. Hence, it is not certain that all participants watched on a large enough screen, had seen each video only once, or completed the task that followed the video. Similarly, they may have been exposed to the videos in different environments with different noise levels, the degree of intervention of e.g. their parents into exposure may differ, as well as the degree to which the children themselves could interact with the task (e.g. some may not have been able to use a mouse, in which case, the parents were instructed to click images indicated by the child). Assessing how at-home exposure may impact performance would be beneficial.

Xie and Myers (2017) noted that generalization of adaptation to novel talkers does not occur if listeners' experience with a prior talker's acoustic-phonetic realization of a segment does not match that of the current talker. If talkers do

demonstrate similarity, generalization occurs. However, it is not clear to what extent the talkers' acoustic-phonetic realizations of the target segment need to be similar for generalization to occur. Further, it is not clear whether other segments can modulate the extent of adaptation, if they differ dramatically between talkers, even though the target segments show great similarity.

There are variables that affect accent adaptation that are not addressed in this study, for example, vocabulary size (Bent, 2014), phonological awareness, and phonological memory (Hu, 2021), all of which may facilitate adaptation for some accents, and inhibit it for others. They may also interact in unique ways. Including an assessment of these effects in future work would improve our understanding of L2 accent adaptation by beginner learners.

Conclusion

Both mono- and multi-accent-exposed participants demonstrated their ability to learn non-minimal pair words from the short-term exposure to both the familiar GBE and CE talkers, as well as the novel WSAE and SE talkers. They were able to identify these words when presented with two auditory form alternatives that differed in their acoustic-phonetic form substantially (e.g. *cake* vs. *hen*). Despite this ability, children exposed to two GBE talkers and those exposed to one CE and GBE talker were not able to learn minimal pair members, even after the four training videos. This pattern was implied at test for all three groups: children did not demonstrate a reliable ability to differentiate minimal pair members when presented with two auditory forms. Surprisingly, the CE children showed above-chance performance after training for minimal pairs produced as homophones. A possible explanation seems to be that the two CE talkers were the only talkers who cued the contrast in a way that could be exploited by the CE children. Alternatively, CE children may have exploited the shared-L1 benefit, which allowed them to allocate more attention to auditory word forms that they stored in enough phonetic detail to be able to use them successfully for word identification.

Participants did not demonstrate reliably above-chance word identification when it came to minimal pairs. This was true for both the multi-accent and mono-accent groups at test, which signals that the input did not provide enough evidence for the children to use non-homophone-like lexical representations for minimal pairs. Data here suggest that low-experience learners who have to contend with a novel accent do not yet make use of the targeted shifts mechanism for difficult contrasts. Importantly, the use of this mechanism is constrained by how the relevant categories are represented, which is most likely what prevented the learners in this study from utilizing targeted shifts. Here, the vastly different experience of infants learning their L1 and children learning an L2 is apparent: high confidence L1-biased priors determine how these learners adapt. The results indicate that children with months of EFL experience struggle to acquire the /ɛ, æ/ contrast.

This study supports the view that beginner learners start with high confidence in L1 prior beliefs and thus require more evidence to adapt in comparison to learners with lower-confidence beliefs. Experience with variable L2 input may provide a flexibility benefit only at a later stage of acquisition. However, even if participants did not adapt using targeted shifts, it is possible that they used a different adaptive strategy. It is important to recognize that despite causing higher minimal pair confusability, even processes like equivalence classification and general criterion relaxing facilitate effective speech comprehension. Not showing evidence of having acquired a minimal pair does not mean that listeners did not show learning: they may have adapted to input using a different mechanism.

Bibliography

Appelbaum, I. (1996). The lack of invariance problem and the goal of speech perception. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96*, *3*, 1541–1544.

Atria, J. J. (2015). Plugin_jjatools [Commit 9daf396]. https://github.com/jjatria/plugin_jjatools/blob/master/textgrid/move_to_zero_crossings.praat

Audacity Team. (2019). Audacity(r): Free audio editor and recorder. https://audacityteam.org/

Barrios, S., & Hayes-Harb, R. (2021). L2 processing of words containing English /æ/-/ɛ/ and /l/-/r/ contrasts, and the uses and limits of the auditory lexical decision task for understanding the locus of difficulty. *Frontiers in Communication*, *6*, 689470.

Barrios, S. L., Rodriguez, J. M., & Barriuso, T. A. (2023). The acquisition of L2 allophonic variants: The role of phonological distribution and lexical cues. *Second Language Research*, *39*(3), 899–924.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bekker, I., & Eley, G. (2007). An acoustic analysis of White South African English (WSAfE) monophthongs. *Southern African linguistics and applied language studies*, *25*(1), 107–114.

Bent, T. (2014). Children's perception of foreign-accented words. *Journal of child language*, *41*(6), 1334–1355.

Best, C. T., & Tyler, M. D. (2008). Nonnative and second-language speech perception: Commonalities and complementarities. In *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34).

Best, C. T., Tyler, M. D., Gooding, T. N., Orlando, C. B., & Quann, C. A. (2009). Development of phonological constancy: Toddlers' perception of native-and Jamaican-accented words. *Psychological science*, *20*(5), 539–542.

Bjelaković, A. (2017). The vowels of contemporary RP: Vowel formant measurements for BBC newsreaders1. *English Language & Linguistics*, *21*(3), 501–532.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* [Version 6.4.23]. http://www.praat.org/

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, *106*(2), 707–729.

Brekelmans, G. (2020). *Phonetic vowel training for child second language learners: The role of input variability and training task* [Doctoral dissertation, UCL (University College London)].

Broadbent, D. E., Ladefoged, P., & Lawrence, W. (1956). Vowel sounds and perceptual constancy. *Nature*, *178*(4537), 815–816.

Buckler, H., Oczak-Arsic, S., Siddiqui, N., & Johnson, E. K. (2017). Input matters: Speed of word recognition in 2-year-olds exposed to multiple accents. *Journal of Experimental Child Psychology*, *164*, 87–100.

Chládková, K., Boersma, P., & Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingualism: Language and cognition*, *25*(5), 827–840.

Chládková, K., & Šimáčková, Š. (2021). Distributional learning of speech sounds: An exploratory study into the effects of prior language experience. *Language Learning*, *71*(1), 131–161.

Constantin, A.-E., & Patil, I. (2021). ggsignif: R Package for Displaying Significance Brackets for "ggplot2". *PsyArxiv*. https://doi.org/10.31234/osf.io/7awm6

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL quarterly*, *33*(2), 185–209.

Cooper, A., & Bradlow, A. (2018). Training-induced pattern-specific phonetic adjustments by first and second language listeners. *Journal of Phonetics*, *68*, 32–49.

Cooper, A., Paquette-Smith, M., Bordignon, C., & Johnson, E. K. (2023). The influence of accent distance on perceptual adaptation in toddlers and adults. *Language Learning and Development*, *19*(1), 74–94.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*, *24*(6), 597–606.

Council of Europe). (2001). *Common European framework of reference for languages: Learning, teaching, assessment*.

Creel, S. C. (2012). Phonological similarity and mutual exclusivity: On-line recognition of atypical pronunciations in 3–5-year-olds. *Developmental Science*, *15*(5), 697–713.

Creel, S. C., Rojo, D., & Paullada, A. (2016). Effects of contextual support on preschoolers' accented speech comprehension. *Journal of Experimental Child Psychology*, *146*, 156–180.

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Routledge.

Cummings, S. N., & Theodore, R. M. (2023). Hearing is believing: Lexically guided perceptual learning is graded to reflect the quantity of evidence in speech input. *Cognition*, *235*, 105404.

De Leeuw, J. R. (2015). JsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior research methods*, *47*, 1–12.

DiCanio, C. (2014, July). Rescale peak [Praat scripts - Christian DiCanio]. https://www.acsu.buffalo.edu/~cdicanio/scripts/Rescale_peak.praat

Durrant, S., Delle Luche, C., Cattani, A., & Floccia, C. (2015). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, *42*(2), 447–465.

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2024). *Ethnologue: Languages of the World* (27th ed.). SIL International. http://www.ethnologue.com

Eisner, F., & McQueen, J. M. (2006). Perceptual learning in speech: Stability over time. *The Journal of the Acoustical Society of America*, *119*(4), 1950–1953.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* (Publication No. 113) [Doctoral dissertation, Utrecht University]. LOT. https://www.lotpublications.nl/linguistic-perception-and-second-language-acquisition-linguistic-perception-and-second-language-acquisition-explaining-the-attainment-of-optimal-phonological-categorization

Escudero, P. (2009). The linguistic perception of SIMILAR L2 sounds. In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 151–170). De Gruyter Mouton. https://doi.org/10.1515/9783110219234.151

Escudero, P., & Chládková, K. (2010). Spanish listeners' perception of American and Southern British English vowels. *The Journal of the Acoustical Society of America*, *128*(5), EL254–EL260.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, *40*(2), 280–288.

Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and speech*, *46*(2-3), 245–264.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, *92*(1), 233–277.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In O.-S. Bohn & J. E. Flege (Eds.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–38). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Harte, J., Oliveira, A., Frizelle, P., & Gibbon, F. (2016). Children's comprehension of an unfamiliar speaker accent: A review. *International Journal of Language & Communication Disorders*, *51*(3), 221–235.

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English word-final voicing contrasts. *Journal of phonetics*, *36*(4), 664–679.

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of phonetics*, *28*(4), 377–396.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech communication*, *47*(3), 360–378.

Hu, C.-F. (2021). Adaptation to an unfamiliar accent by child L2 listeners. *Language and Speech*, *64*(3), 491–514.

Hyman, L. M. (1970). How concrete is phonology? *Language*, 58–76.

Johnson, E. K., van Heugten, M., & Buckler, H. (2022). Navigating accent variation: A developmental perspective. *Annual Review of Linguistics*, *8*(1), 365–387.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of phonetics*, *27*(4), 359–384.

Kartushina, N., & Martin, C. D. (2019). Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning*, *69*(1), 71–105.

Kartushina, N., Rosslund, A., & Mayor, J. (2022). Toddlers raised in multidialectal families learn words better in accented speech than those raised in monodialectal families. *Journal of Child Language*, *49*(6), 1093–1118.

Kemp, N., Scott, J., Bernhardt, B. M., Johnson, C. E., Siegel, L. S., & Werker, J. F. (2017). Minimal pair word learning and vocabulary size: Links with later language skills. *Applied Psycholinguistics*, *38*(2), 289–314.

Kleinschmidt, D. F. (2020). What constrains distributional learning in adults? [Manuscript available on OSF]. https://osf.io/3wdp2/

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review*, *122*(2), 148.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive psychology*, *51*(2), 141–178.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1–15.

Kučerová, M., & Šimáčková, Š. (2025). Preschoolers' production of L2 vowels is affected by input quality: A longitudinal study. *Journal of the European Second Language Association*, *9*(1), 19–35. https://doi.org/https://doi.org/10.22599/jesla.125

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant behavior and development*, *6*(2-3), 263–285.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature reviews neuroscience*, *5*(11), 831–843.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Ladefoged, P., & Johnson, K. (2011). *A Course in Phonetics* (6th). Wadsworth, Cengage Learning. https://api.semanticscholar.org/CorpusID:150134828

Lee, S. A. S., & Iverson, G. K. (2012). Vowel category formation in Korean–English bilingual children. *Journal of Speech, Language, and Hearing Research*, *55*(5), 1449–1462.

Leimgruber, J. R. (2011). Singapore English. *Language and Linguistics Compass*, *5*(1), 47–62.

Lenth, R. V. (2024). *Emmeans: Estimated Marginal Means, aka Least-Squares Means* [R package version 1.10.6]. https://CRAN.R-project.org/package=emmeans

Lev-Ari, S. (2018). The influence of social network size on speech perception. *Quarterly Journal of Experimental Psychology*, *71*(10), 2249–2260.

Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2023). *TuneR: Analysis of Music and Speech* [R package version 1.0.0]. https://CRAN.R-project.org/package=tuneR

Llompart, M. (2021). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition*, *24*(3), 481–496.

Llompart, M., & Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*, *22*(5), 1085–1100.

Lüdecke, D. (2018). Ggeffects: Tidy Data Frames of Marginal Effects from Regression Models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Maye, J., Aslin, R. N., & Tanenhaus, M. K. (2008). The weckud wetch of the wast: Lexical adaptation to a novel accent. *Cognitive science*, *32*(3), 543–562.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101–B111.

McCloy, D. R. (2012). Vowel normalization and plotting with the phonR package. *Technical Reports of the UW Linguistic Phonetics Laboratory*, *1*, 1–8.

McLeod, S., & Crowe, K. (2018). Children's consonant acquisition in 27 languages: A cross-linguistic review. *American journal of speech-language pathology*, *27*(4), 1546–1571.

Miller, R. L. (1953). Auditory tests with synthetic vowels. *The Journal of the Acoustical society of America*, *25*(1), 114–121.

Milovanovic, F. (2024, September). DaVinci Resolve [DaVinci Resolve 19, Blackmagic Design. Available at: https://www.blackmagicdesign.com/products/davinciresolve].

Ministry of Education, Y., & Sports. (2024). Revision of the Framework Educational Program for Primary Education. https://prohlednout.rvp.cz/

Nathan, L., Wells, B., & Donlan, C. (1998). Children's comprehension of unfamiliar regional accents: A preliminary investigation. *Journal of Child Language*, *25*(2), 343–365.

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning additional languages as hierarchical probabilistic inference: Insights from first language processing. *Language Learning*, *66*(4), 900–944.

Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and speech*, *46*(2-3), 115–154.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech communication*, *13*(1-2), 109–125.

Podlipský, V. J. (2024). *Praat for child speech researchers: Acoustic analysis, stimulus preparation, and running experiments* [XVIth International Congress for the Study of Child Language]. https://iascl2024.com/program/tutorials/

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics*, *58*, 1–21.

Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition*, *166*, 67–72.

Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. https://api.semanticscholar.org/CorpusID:2488733

R Core Team. (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science*, *12*(2), 339–349.

Schlechtweg, M., Peters, J., & Frank, M. (2023). L1 variation and L2 acquisition: L1 German/e:/-/ɛ:/overlap and its effect on the acquisition of L2 English/ɛ/-/æ/. *Frontiers in Psychology*, *14*, 1133859.

Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, *15*(6), 732–738.

Schmale, R., Hollich, G., & Seidl, A. (2011). Contending with foreign accent in early word learning. *Journal of Child Language*, *38*(5), 1096–1108. https://doi.org/10.1017/{S}0305000910000619

Schmale, R., Seidl, A., & Cristia, A. (2015). Mechanisms underlying accent accommodation in early word learning: Evidence for general expansion. *Developmental science*, *18*(4), 664–670.

Šimáčková, Š. (2003). Engela's Fleshes": Cross-linguistic perception and production of English [æ] and [ɛ] by Czech EFL learners trained in phonetics. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS 2003), Barcelona*, 3–9.

Šimáčková, Š., & Podlipský, V. J. (2018). Production accuracy of L2 vowels: Phonological parsimony and phonetic flexibility. *Research in Language*, *16*(2).

Simon, E., Lima Jr, R., & De Cuypere, L. (2016). Acquiring non-native speech through early media exposure: Belgian children's productions of English vowels. *Poznan Studies in Contemporary Linguistics*, *52*(4), 719–743.

Skarnitzl, R., & Rumlová, J. (2019). Phonetic aspects of strongly-accented Czech speakers of English. *AUC PHILOLOGICA*, *2019*(2), 109–128.

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. *KONVENS*, 14–26.

Šturm, P., & Skarnitzl, R. (2011). The open front vowel/æ/in the production and perception of Czech students of English. *Proceedings of Interspeech 2011*, 1161–1164.

Sueur, J., Aubin, T., & Simonis, C. (2008). Seewave: A free modular tool for sound analysis and synthesis. *Bioacoustics*, *18*, 213–226.

Tagliamonte, S. A., & Molfenter, S. (2007). How'd you get that accent?: Acquiring a second dialect of the same language. *Language in Society*, *36*(5), 649–675.

TheFableCottage.com. (2021, May). The Frightened Lion – US English accent (TheFableCottage.com). https://youtu.be/17o2rPGryuk

TheFableCottage.com. (2022, February). The Girl and the Ice Cream Truck - US English accent (TheFableCottage.com). https://www.youtube.com/watch?v=1DeQVnSxcLk%5C&t=12s

TheFableCottage.com. (2023a, May). The City Mouse and the Country Mouse (UK English accent). https://www.youtube.com/watch?v=Atkt-vhxFIc

TheFableCottage.com. (2023b, June). The Oak Tree — US English accent (TheFableCottage.com). https://www.youtube.com/watch?v=_6mlmRJHxNo

TheFableCottage.com. (2023c, November). The Little Red Hen (US English accent) - TheFableCottage.com. https://www.youtube.com/watch?v=VjTJvwgJnGc%5C&t=88s

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: A two-process account of statistical learning. *Psychological bulletin*, *139*(4), 792.

Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal*, *146*(10).

van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology: General*, *143*(1), 340.

van Heugten, M., & Johnson, E. K. (2017). Input matters: Multi-accent language exposure affects word form recognition in infancy. *The Journal of the Acoustical Society of America*, *142*(2), EL196–EL200.

van der Feest, S. V., & Johnson, E. K. (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*, *23*(2), 89–111.

van der Feest, S. V., Rose, M. C., & Johnson, E. K. (2022). Showing strength through flexibility: Multi-accent toddlers recognize words quickly and efficiently. *Brain and Language*, *227*, 105083.

Van Leussen, J., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in psychology*, *6*, 1000.

Vroomen, J., van Linden, S., De Gelder, B., & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory–visual speech perception: Contrasting build-up courses. *Neuropsychologia*, *45*(3), 572–577.

Wagner, L., Jin Song, Y., Speer, S., White, S., Stanhope, R., & Frush Holt, R. (2024). *The Prosody of Compound Nouns: Direct Comparison of Tunes Helps Children (and Adults)* [IASCL 2024: XVIth International Congress for the Study of Child Language]. https://iascl2024.com/2024/07/15/iascl-2024-individual-oral-presentations/

Walley, A. C., & Flege, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, *27*(3), 307–332.

Wambacq, I., Ghanim, I., Greenfield, S., Koehnke, J., Besing, J., Chauvette, C., & Yesis, C. (2023). Neurophysiologic patterns of semantic processing of accented speech. *Journal of Neurolinguistics*, *65*, 101117.

Weber, A., Di Betta, A. M., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of phonetics*, *46*, 34–51.

White, K. S., & Aslin, R. N. (2011). Adaptation to novel accents by toddlers. *Developmental science*, *14*(2), 372–384.

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Xie, X., & Myers, E. B. (2017). Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers. *Journal of Memory and Language*, *97*, 30–46.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature protocols*, *1*(1), 297–301.

List of Figures

List of Tables

List of Abbreviations

CE	Czech English
CI	confidence interval
DCCS	Dimensional Change Card Sort
EFL	English as a foreign language
ELF	English as a lingua franca
f0	fundamental frequency
F1	first formant
F2	second formant
GAE	General American English
GBE	General British English
GLMM	generalized linear mixed-effects model
IAF	ideal adaptor framework
L1	first language
L2	second language
L2LP	Second Language Linguistic Perception
LM	linear model
MP	minimal pair
Multi	multi-accent
PAM-L2	Perceptual Assimilation Model of L2 speech learning
SE	Singapore English
SLM(-r)	(Revised) Speech Learning Model
WSAE	White South African English