Uniwersytet Śląski
Wydział Humanistyczny
Instytut Językoznawstwa

Błażej Wieczorek

IMITACJA FONETYCZNA ANGIELSKICH SAMOGŁOSEK PRZEZ POLSKICH UCZNIÓW JĘZYKA ANGIELSKIEGO

Rozprawa doktorska napisana pod kierunkiem
prof. dr. hab. Arkadiusza Rojczyka

Sosnowiec 2024

University of Silesia
Faculty of Humanities
Institute of Linguistics

Błażej Wieczorek

PHONETIC IMITATION OF ENGLISH VOWELS BY NATIVE POLISH LEARNERS OF ENGLISH

Doctoral dissertation written
under the supervision of
prof. dr hab. Arkadiusz Rojczyk

Sosnowiec 2024

Table of contents

Introduction

The example of Joseph Conrad, a well-known native Polish writer, who reached unparalleled heights in English proficiency, but who was reportedly difficult to understand when he spoke English, is often quoted to illustrate how challenging pronunciation can be to second language learners. While considered as promoting cultural diversity by some, and as an undesirable modification of language norms or even a threat to their cultural identity by others, what remains rather uncontested is that foreign accent can impede the efficiency of communication. There are many factors that contribute to the impression of foreign accent, but what particularly stands out in this respect is the variable nature of vowels – sounds whose inadequate realisation may greatly contribute to the perception of foreign accent. Because of their elusive character, resulting from a lack of precise articulatory identity, there is a limit with regard to the instructions that second language learners may be given to help them produce these sounds adequately. The learner often resorts to their intuition-based judgement as to the quality of foreign vowels, which leads to the underappreciation of qualitative nuances, especially in those cases where their native vowels seemingly resemble those of the target language.

Such challenges are faced by native Polish learners of English, whose native language is characterised by relative simplicity as regards their vowel system. Polish has only six oral vowels, while in English there may be twice as many, depending on the variety or dialect considered. What compounds the issue is that not only are these vowels different in terms of their timbre (or quality), but also the qualitative contrasts are accompanied by durational contrasts, which matter to a much a greater extent in English than in Polish. The speech of those second language learners who are unaware of both qualitative and quantitative vocalic differences between their first and their second language is bound to exhibit profound interference from the former language, especially when the target language has a denser vowel space. The ignorance of qualitative nuances by Polish learners may manifest itself in using the same Polish vowel as a replacement for as many as four different English vowels, such as in the words *bet*, *bat*, *bird*, and *about*. The issue is further compounded by a lack of appreciation for inherent vowel duration contrasts, such as those between *bit* and *beat*, but also for extrinsic contrasts, as in *bet* versus *bed*, which extends the range of potential quasi-homophones in the speech of Polish learners of English. The learners are also likely to display superficial understanding of vowels by considering a vowel monophthong as having a uniform quality, resulting in a disregard for qualitative variation throughout the vowel's duration, which may be particularly relevant in the target language.

The purpose of the current investigation is to objectively measure the extent to which native Polish learners of English can imitate acoustic properties of English vowels. More specifically, what is of interest is whether, having heard a native English speaker utter a word, the learners can reproduce the qualitative and durational parameters of the vowel contained in the word. In view of a somewhat limited role of instructional intervention with regard to proper realisation of vocalic sounds, especially in schools, imitation seems the most optimal way of improving this aspect of pronunciation. In a classroom setting, the learners have been made accustomed to repeating words after a recorded voice of a model speaker, but this has necessarily suffered from the lack of opportunity to obtain relevant feedback on their performance. Even if the teacher can direct their evaluative attention to individual students, their judgment is bound to be subjective, and can therefore be inaccurate. In order to objectively assess whether imitation has taken place, and to determine its degree, it is required that the learners' productions be recorded and subjected to acoustic analysis.

A great deal of research has been conducted in the area of phonetic imitation, including studies focusing on various speech features, and those concentrating on different factors affecting it. The current investigation is intended to enhance our understanding of the process, specifically in the context of second language acquisition, by accounting for factors that have either not been conclusively confirmed to affect imitation, or have been completely overlooked. First, previous studies on vowel imitation seem to have ignored the dynamic aspect of vowel quality, by concentrating solely on steady-state portions of vowels, while it is known that even nominal monophthongs carry relevant information for the listeners at different vowel portions. Second, for the sake of at least partially accounting for considerable variability among second language learners, both primary school students and university students are tested on their imitative performance – two groups which differ not only in terms of age, but also in terms of second language proficiency. Third, the current investigation is arguably the first one to consider the effect of orthographic input in phonetic imitation. Outside the domain of phonetic imitation, previous studies on the influence of orthography on second language processing have shown that it affects second language learners' pronunciation in intricate and penetrating ways, leading to the orthography-induced pronunciation patterns.

This dissertation is divided into five chapters. In Chapter 1, what is offered is a general discussion of the process of imitation, which then focuses specifically on phonetic imitation, with various factors affecting it, which is culminated in a review of studies within the context of second language speech imitation. Chapter 2 explores the differences related to vocalic systems in Polish and English, with special focus on vowel dynamics. Chapter 3 constitutes a brief overview of studies dealing with the impact of orthography on the speech of second language learners. The design of the current study is detailed in Chapter 4, and it is followed by a discussion of the results for each of the three groups of tested parameters, related to vowel durational and qualitative contrasts. Finally, Chapter 5 captures the overall conclusions drawn from the findings, with regard to the research questions asked.

Research partly supported by the National Science Centre Poland grant

Phonetic imitation in a native and non-native language

(UMO-2019/35/B/HS2/02767, PI: Arkadiusz Rojczyk)

Chapter 1

Phonetic imitation in native and non-native speech

The purpose of this chapter is to is to present a brief overview of the research done in the domain of phonetic imitation, also known as phonetic convergence. To offer some perspective, the idea of imitation is first discussed quite generally, and it must be admitted, necessarily superficially, as the scope of research encompassing this phenomenon seems boundless. Following that, the specific kind of imitative behaviour, i.e. the imitation of speech features, is reviewed, factoring in several key aspects of the process concerned. An attempt is made to summarise most commonly reported factors guiding convergence, with particular focus on age, as well as to summarise the issue of what exactly is imitated. Finally, what is considered is speech imitation, specifically in the context of second language acquisition, which is particularly relevant to the current investigation.

1.1. The concept of imitation

It must be acknowledged that the term *imitation* is highly context-dependent and its exact definition varies as a result of different authors investigating the phenomenon from different points of view across different disciplines over the years, with many of them offering their unique perspectives on the issue. For example, for some biologists the interest in imitation lies in its adaptive value, crucial to the survival of living organisms, while for some psychologists the focus is placed on the mechanisms by which these organisms learn from others (Zentall 2006: 336). Even within particular fields, such as psychology, imitation has been approached in a multitude of ways, depending on given subfields, resulting in a variety of conceptual phenomena behind the it with little more in common but the name itself (Snow 1989: 73). Despite different venues of investigation that explore the concept concerned, there appear to emerge recurring themes when considering its various definitions proposed by authors over the years, who have defined imitation as:

> … learning to do an act from seeing it done. (Thorndike 1898, cited in Laland and Bateson 2001: 195)

> … purposeful, goal-directed copying of the behavior of one animal by another. (Galef 1988: 21)

> … a process by which individuals learn to perform a behavior pattern as a result of observing another animal performing a similar action. (Laland and Bateson 2001: 195)

> … [a process] involving the recognition of a relationship between a stimulus and the behaviour required to generate it … [and] the subsequent execution (or attempted execution) of the required behaviour. (Markham 1997: 39)

> … consisting of three distinct processes: perception of structural properties in the stimuli being imitated, coding and storage in memory, and regeneration in the form of a motoric code suitable for skilled movement. (Flege and Eefting 1987: 730)

It is acknowledged that there are many neighbouring terms with varying degrees of overlap, which again stems from the multitude of researchers of various disciplines studying it. For example, some authors use the term *mimicry* (Dimberg et al. 2000; Bourgeois and Hess 2007), specifically in the context of unconscious replication of facial expressions, but may it also involve vocal or postural imitation of people with whom we are interacting (Bourgeois and Hess 2007: 1). Some authors have distinguished between *conscious imitation* and *nonconscious mimicry* (Chartrand and Bareen 1999), highlighting the difference in the degree of awareness of the process concerned. Heyes (2011: 478) mentions what she calls *automatic imitation* as something that is "minimally dependent on the actor's intentions," and, contrary to the term *imitation*, it seldom "manifests overt behavioral executions of actions similar to those observed" (Heyes 2011: 467). Also, as regards *mimicry* in the context of second language acquisition, some report the term to be a "disparaging synonym for imitation in the acquisition literature" (Markham 1997: 41).

Another concept, *mimesis*, has been in use since ancient times, alongside its Latin equivalent *imitatio*, in the domains of rhetoric and literature. It refers to "how written and visual arts mimicked or imitated the world" (Fronda 2012: 1), and it involves "an author's conscious use of features and characteristics of earlier works to acknowledge indebtedness to past writers" (ibid). Others have considered mimesis to be a more complex cognitive ability than imitation, defining it as "as imitation, accompanied by intention to communicate, plus creation and/or modification of motor representations"

(Delvaux and Soquet 2007: 4). Chartrand and Bargh (1999: 893) use the term *chameleon effect* to refer to "nonconscious mimicry of the postures, mannerisms, facial expressions, and other behaviors of one's interaction partners, such that one's behavior passively and unintentionally changes to match that of others in one's current social environment." Some other terms related to the phenomenon of imitation include *mirroring*, particularly in the context of mirror mechanism (Rizzolatti and Sinigaglia 2016), posited to serve various cognitive functions, such as understanding actions and emotions of others, and *emulation*, defined as "ongoing internal representations of potential actions and the futures those actions are expected to produce" (Colder 2011: 1).

In light of the above, the concept of imitation has had many names, which themselves have been subject to varying uses, depending on either the researchers' decision to consider or to not consider some of them synonymous, or on the conventions imposed by a given field. At this stage of the dissertation, i.e. before defining the specific type of imitation that is of actual interest, the term imitation is used rather generally, without the slightest intention of disentangling it from neighbouring terms mentioned above. Following the definitions mentioned earlier, imitation here is loosely understood as a process that involves conscious or automatic copying of human behaviour by another human being, not necessarily entirely faithfully, with a potential result of learning that behaviour. While imitation in not restricted to the human species, it is with human beings that imitation reaches its most advanced form (Iacoboni 2009: 654), and is clearly of interest in the current investigation.

The significance of imitation in human development seems uncontested. Meltzoff et al. (2009: 285) write that "learning by observing and imitating experts in the culture is a powerful social learning mechanism." The mechanism is said to accelerate learning by relieving learners of having to engage in the process of individual discovery, and it alleviates the risk involved in learning solely through trial and error. Studies have demonstrated that imitating an action leads to a better understanding of that action and helps in the anticipation of its future instances (Adank et al. 2010). Among a wide range of actions, gestures, and behaviours that have been confirmed to undergo imitation, arguably the most well-known cases involved the imitation of facial gestures of newborns (Meltzoff and Moore 1999). Children were found to follow a range of their parents' mannerisms and speech patterns (Meltzoff et al. 2009), the latter point being evidenced by infants of 12 weeks of age having been found capable of vocal imitation (Kuhl and Meltzoff 1996). The ability to imitate remains available for humans throughout their lives

(Delvaux and Soquet 2007: 3), allowing them to enhance liking, rapport, and affiliation when interacting with others (Lakin et al. 2003). Similarly to children, adults have been found to unwittingly follow others' facial expressions, gestures and mannerisms, including face-touching or foot-tapping (Chartrand and Bargh 1999). Increased muscle tension at the sight of arm wrestling (Berger and Hadley 1975) or wincing at the sight of people in pain (Vaughan and Lanzetta 1980) are but a few examples of vicarious experiences or at least partially copied behaviours that human beings have manifested. What seems sparsely contested is the importance of imitation for the growth of cognitive and social behaviours in humans and, consequently, in language acquisition (Nielsen 2014: 2065), with some authors considering language acquisition as a strongly imitative phenomenon (Markham 1997). Imitation has been recognised with respect to the acquisition of such areas as lexis (Leonard and Kaplan 1976), syntax and grammar (Snow 1979; Branigan et al. 2000), pragmatics and discourse (Pickering and Garrod 2004). More pertinently to the current investigation, imitation has been considered from the point of view of phonetics and phonology, both in the context of language considered generally and in the context of second language acquisition. Imitation is a universally recognised factor in acquiring second-language speech (Rojczyk et al. 2013:5), and it has played a crucial role in pronunciation teaching (Dziubalska-Kołaczyk et al. 2014: 238), by prompting second language learners to assume native-like speech characteristics.

1.2. Phonetic imitation (convergence)

Among the multitude of behaviours that have been found subject to imitation as a result of exposure to the behaviours of other human beings, the one that is concerned in the current investigation is the imitation of speech features. As it was the case with the term imitation, its specific type, *phonetic imitation*, is itself largely multifaceted. As Schertz et al. (2023: 2) remark the term "encompasses a wide range of phenomena, and it is used in studies differing in the target of imitation (e.g., words in isolation vs. natural stretches of speech by an interlocutor), instructions (e.g., with vs. without an explicit directive to imitate), and task (e.g., speech shadowing vs. natural conversation)." Still, on the fundamental level, what these phenomena have in common is that they involve more or less temporary changes in one's speech patterns that are due to some form of exposure to the speech of other human beings. Across the literature, phonetic imitation has been defined, rather consistently, as:

> … the auditory-motor reproduction of an acoustic speech stimulus. (Kent 1979: 43)

> … the process in which a talker takes on acoustic characteristics of the individual that he or she is interacting with. (Babel 2012: 178)

> … the unintentional, spontaneous acquisition of speech characteristics of another talker. (Babel et al. 2014: 123)

> … process by which the production patterns of an individual become more similar on some phonetic or acoustic dimension to those of her interlocutor. (Yu et al. 2013: 1)

> … [a process that] occurs when talkers alter their production toward speech they hear and can occur in lab settings without explicit instruction to imitate. (Hauser et al. 2023: 1)

The phenomenon in question has been referred to, often but not always interchangeably, in various ways, including such terms as *phonetic convergence* (Kim et al. 2011; Pardo 2006), *vocal imitation* (Studdert-Kennedy and Terrace 2017), *replication* (Tahta et al. 1981), *spontaneous imitation* (Kwon 2019), *speech repetition* (Brady et al. 1983), *speech entrainment* (Beňuš 2014), *speech alignment* (Miller et al. 2010), and *accommodation* (Shepard et al. 2001). While it is not the intention here to consider possible nuances between these terms, or contexts in which they originated, the last term, employed within the framework of Communication Accommodation Theory (Giles et al. 1991), discussed later, must be acknowledged to be arguably the most general one, as it also considers scenarios in which interlocutors become unlike each other in terms of speech patterns. In the current dissertation, both phonetic imitation and phonetic convergence are used interchangeably, although some authors associate controlled intent with the former term (Lewandowski 2012). It should also be noted that, unlike what some of the above definitions may suggest, actual interaction is not a prerequisite for imitation to take place. Indeed, such definitions have been driven by the experiments in which participants' speech was judged to have grown increasingly similar over the course of some form of a conversation (e.g. Pardo 2006). However, more pertinently to the current investigation, the process is also known to take place in non-interactive settings, as a result of one-sided exposure to a model speaker rather than through interaction where both parties tend to respond to each other in real-time. A number of aspects have been accounted for in the investigation of speech imitation. These have included analyses at a global or suprasegmental level, where convergence has been found for accent (Paquette-Smith et al. 2022), accent variation (Mitterer and Müsseler 2013), speech rhythm (Bilous and Krauss 1988), speech rate (Street 1983), phonetic style (Kappes et al. 2009), vocal intensity (Natale 1975), pitch (Kappes et al. 2009). At a more segmental level, many studies have focused on the imitation of features pertaining to consonants, such as lengthened VOT (Shockley et al. 2004; Nielsen 2011), pre-voicing (Mitterer and Ernestus 2008), variants of (Dutch) /r/ (ibid), variants of American English /l/ (Honorof et al. 2011), and the lack of release of final plosives (Rojczyk et al. 2013). As for vowels, convergence has been reported for vowel duration (Lehiste and Shockey 1980; Zając 2013; Podlipský and Šimáčková 2015), spectral characteristics of vowels (Repp and Williams 1985, 1987; Evans and Iverson 2007; Alivuotila et al. 2007; Tilsen 2009; Babel 2012), and coarticulatory vowel nasality (Zellou et al. 2016; Zellou et al. 2017).

1.3. Factors driving phonetic imitation

There are two broad theories of imitation, which, not being altogether mutually exclusive, focus on different aspects of the phenomenon (Black 2012: 16). The first category maintains the position that imitation is an automatic process, which is simply a result of the way language is processed in the brain (Trudgill 2008; Goldinger 1998). Here, it is posited that perception and production are directly linked, with our brains processing linguistic input, which directly influences our linguistic output without any intervening processes, such as conscious thought, decision or effort. This position tends to downplay the role of social factors as intervening, modulating factors in the process of imitation, and it has found support in a number of studies, in which the participants have been found to imitate model speakers after short one-sided auditory exposure in controlled laboratory conditions (e.g. Goldinger 1998; Nielsen 2011; Shockley et al. 2004), proving the phenomenon takes place in non-interactive settings and without being explicitly instructed to imitate. A typically employed research paradigm, and one whose form is employed in the current investigation, involves the participants being presented with a list of words, which they are asked to read according to their natural speech habits, upon which they are supposed to repeat the same words after a model speaker. In order to ascertain convergence or lack thereof, the participants' productions, particularly their acoustic parameters, whether spectral or temporal, are measured and compared across the two tasks. Alternatively, the comparison is made more globally within the AXB format, by having listeners judge whether the baseline token or the imitated token, without them knowing which is which, is more similar to the original, model token.

According to another perspective, imitation is a socially-mediated process and therefore serves social functions. Here also, as it is assumed to be the case with automatic imitation mentioned above, the process can occur at the unintentional or subconscious level (Pickering and Garrod 2004), but it is the intervening social factors that are posited to be the key ones that are the driving force behind accommodation. This perspective has most commonly been investigated withing the framework of Communicative Accommodation Theory or CAT (Giles et al. 1991), according to which, speakers often subconsciously regulate the social distance between themselves by adjusting their communicative behaviour. One of the major tenets of CAT is that communication is not only about exchanging facts, ideas, and emotions, but also about "salient social category memberships [which] are often negotiated during an interaction through the process of accommodation" (Giles and Ogay 2007: 294). When speakers adopt their interlocutors' communicative behaviours, whether linguistic, paralinguistic, or nonverbal, they are said to *converge* to them, due to various reasons, including the desire to signal social affiliation, empathy, attraction or approval of the interlocutor. Contrarily, they may show *divergence* if their communicative behaviours shift away from those of their interlocutors, for example, for the purpose of manifesting the opposite of the above, i.e. to highlight distinct group membership. Alternatively, an intermediate or neutral position is possible, when the speakers make no discernible shifts in either direction, in which case they are posited to exhibit *maintenance*.

There are a variety of socially driven motives behind convergence. Studies have shown that positive bias towards the representative of a particular group, for example, towards a member of a particular nationality, can incur more likely shifts towards the interlocutor (Babel 2010). Along these lines, Yu et al. (2013) revealed that the participants' tendency to converge to the model narrator correlated with the (positive) impression that was made on the participants. What is more, individuals with higher social desirability, indicating their higher concern with being socially approved, adjusted their vocal intensity to the interviewer more than those that showed lower scores on the social desirability scale (Natale 1975). Convergence was also found to positively correlate with how attractive the model talkers were perceived to be by the imitators (Babel 2012). The results of Gregory and Webster (1996) showed convergence is more likely when the interlocutor is a representative of a higher social class, while the opposite process, divergence, may happen when they are of a lower social status. What may be connected with this is a greater tendency to converge to those with perceived authority, as the findings of Giles (1973) suggest, where a traveller was judged to accommodate more towards the ticket inspector than it was the case the other way around. Convergent behaviour may also stem from the desire to improve the effectiveness communication, by speakers adjusting their linguistic style towards that of the audience (Bell 1984).

The presence versus absence of phonetic convergence, along with its degree, has also been known to be affected by factors pertaining to linguistic, acoustic or cognitive considerations. Goldinger (1998) found greater imitative effect in the case of low-frequency words as stimuli (but see the case of vowels in Black 2012). Goldinger (1998) also found that imitation increases together with the number of repetitions of words (see Babel 2012; Shockley et al. 2004 for exceptions), and it decreases in delayed shadowing conditions. The degree of convergence can be selective in terms of specific acoustic parameters being measured, as was found by Babel (2012), where low vowels tended to show greater imitative effect than high vowels. Another factor mediating the process of accommodation includes perceptual salience related to dialectal differences, which was determined to positively affect the magnitude of imitative shifts (MacLeod 2014). This may be connected with the findings of Nycz and Mooney (2017), who ascertained greater phonetic convergence with increased initial phonetic distances between speakers. The results of Gregory et al. (1997) proved the importance of low-frequency signal in phonetic convergence, the absence thereof (as a result of signal filtration) may downgrade positive qualitative evaluations of communication. More cognitively oriented studies found that taxing the participants' working memory through distraction, and consequently limiting their ability to attend to aural stimuli, negatively impacts convergence (Heath 2017). Moreover, phonetic talent, said to be composed of a bundle of abilities, including those located at the input, central, and output processing stages (Lewandowski 2012: 65), may influence the core of mechanisms behind phonetic convergence. Dias et al. (2013) confirmed the importance of the role of additional visual input in the form of the model's mouth, which enhanced speech alignment. Finally, what has an impact on phonetic convergence may lie in whether or not the participants are explicitly instructed to imitate the models, as opposed to just repeat after them (Adank et al. 2010; but see Zając and Rojczyk 2014).

1.4. Abstractionist and exemplar accounts of phonetic imitation

Another question that has been frequently raised in the literature is whether or not phonological encoding mediates perception and subsequent production in the process of imitation. According to the first position, the incoming acoustic signal is warped by our perceptual system, which results in all the seemingly redundant, non-linguistic, sub-phonemic, or speaker-specific information contained in the signal being filtered out. In consequence, the output signal should be deprived of all this content and should undergo reorganization according to our relatively constant phonological categories. In support of this abstractionist view, Flege and Eefting (1987), who investigated the imitation of VOT /da/-/ta/ continuum among native English and Spanish speakers, observed abrupt shifts in the stimuli reproduction of their participants, whose locations corresponded with their phonemic boundaries, established earlier for the two languages. The authors ascribed this to the view that stimuli are categorised prior to imitation. Similarly, imitation was concluded to be driven by abstract categories in Kwon (2019), who found their Seoul Korean imitators to either supplement or substitute long VOT with high fundamental frequency, the two acoustic properties determining the voicing status of initial stops in that language. Another piece of evidence in favour of the abstractionist position seems to come from Brouwer et al. (2010), where listeners, tested for whether or not they could align to reduced speech, were found to reconstruct canonical form of the reduced forms.

The abstractionist view seems in line with the contrast-preservation hypothesis (Podlipský and Šimáčková 2015), according to which the speaker should be unlikely to imitate phonetic features if it leads to a change of the imitated phoneme. Mitterer and Ernestus (2008) found their Dutch participants to imitate the phonologically relevant presence versus absence of prevoicing but not the phonologically irrelevant amount of prevoicing, supporting the phonological account, rather than the gestural account of speech perception (Fowler 1996). Partly contrary to this, Nielsen (2011) observed native-English speakers successfully imitating artificially extended VOT for /p/, even though this manipulation did not change the phonological status of /p/. However, shortened VOT were not imitated, which was possibly motivated by the desire to not encroach on the phonemic boundary between long-lag /p/ and short-lag /b/, showing that phonological representations had an effect. Somewhat in agreement with Nielsen (2011) but in disagreement with Mitterer and Ernestus (2008), Podlipský and Šimáčková (2015) observed imitation of extended but not of reduced prevoicing among their Czech participants, with the lack of the latter effect being in line with the contrast-preservation hypothesis. To complicate matters, English participants in Schertz et al. (2023) imitated both lengthened and shortened VOT, but the latter finding was at least partly due to the fact that some imitated tokens exhibited prevoicing, which lead to the imitation of the opposite phonological category (i.e. /b/ instead of /p/). Imitation of shortened VOT for the voiceless category was also observed in the case of native English speakers imitating Spanish-accented English in Flege and Hammond (1982), where they showed frequent underaspiration of voiceless /t/, which is typical for Spanish.

Some of the aforementioned evidence suggest an alternative position to the abstractionist view, namely one which holds that seemingly linguistically irrelevant details in the speech signal are not discarded in speech perception. Considerable variability in the speech signal, stemming from a given speaker's specific information, such as age, gender, and emotional state, are posited to be preserved in memory (Goldinger 1998). This is based on episodic theory MINERVA 2 (Hintzman 1986), according to which, whenever a new word is auditorily presented, an echo, i.e. an aggregate of all previously acquired traces or instances of that word that bear resemblance to the current stimulus, is activated and sent from long-term memory to working memory (Goldinger 2013: 2). In consequence, these prior activated traces contribute to the word's perception, and may therefore also emerge in its subsequent production. Goldinger (1998: 264) goes as far as to say that "an episodic lexicon should support direct matching of words to traces, without normalization," the process whereby speakers supposedly filter out variable, phonetically irrelevant information (Joos 1948). It should be noted that the episodic model of convergence does not necessarily contradict the abstractionist account entirely, in that the features of the incoming auditory input are still, in a sense, categorised, which may become evident in the potential imitators' productions. In the context of an episodic model of convergence, Heath (2017: 8) compares phonetic features to stereotypes, saying that "people form expectations of what they will hear, and if they hear speech that meets those expectations, their stereotypes are reinforced, and they rely more heavily on those stereotyped features, resulting in convergence."

The position that sub-phonemic details are imitable has been confirmed by many studies, some of which have already been mentioned at the end of Section 1.2 and earlier in this section. To quote a few more, Tilsen (2009) observed significant effects of sub-phonemic priming on vowel formants in a shadowing experiment, attributing these effects to the role of episodic memory in speech perception. Similarly to Nielsen (2011), Shockley et al. (2004) demonstrated that English speakers can imitate (phonologically irrelevant) artificially lengthened VOT of voiceless plosives. The authors, however, concluded their findings to be more compatible with one of the gestural accounts of speech perception, rather than an episodic one. More support in favour of fine details of speech being preserved for further manifestation comes from Babel's (2012), whose investigation into the accommodation of vowels was concluded to be at odds with the results of Mitterer and Ernestus (2008), mentioned earlier. Specifically, the author argued that gradient acoustic information contained in the signal were reflected in subsequent speech production. Finally, despite the majority of evidence pointing to the fact that fine phonetic features are prone to being imitated, the other perspective should not be discarded, as Nielsen (2011: 141) herself concluded that both phonetic details and various levels phonological representations, such as word and phoneme, act together in the process of phonetic imitation, with words not being "purely made up of discrete abstract units, but [they] can be episodic and abstract at the same time."

1.5. The factor of age in phonetic imitation

One factor that merits special attention, and one which may be related to the abstractionist-exemplar debate, is the factor of age. Studies on phonetic imitation have mostly involved adult participants, neglecting younger age groups, with even fewer studies actually investigating the effect of age on the degree of phonetic imitation within single studies, leaving this factor sparsely explored. The importance of imitation in early language acquisition seems uncontested. Kuhl and Meltzoff (1996) demonstrated that even infants as young as 12-20 weeks old exhibit imitation of vowels, by making use of both auditory and visual stimuli. The authors remark that infants' early attempts at articulations are not random events, but rather they are a part of the process of relating auditory information to their own evolving vocal configurations. Infants' speech is clearly affected by the speech of their caretakers, but it was also shown to be influenced by the language used by the local community by as early as two years of age (Floccia et al. 2012). Studies have found that, by the age of five, children seem to be capable of recognising different accents in their native language, but they are more attuned to the features of foreign-accented speech compared to regional accents (Girard et al. 2008). As far as adults are concerned, research shows that while they may adopt certain aspects of a new dialect, they seldom fully master it (Babel et al. 2014: 124). Some authors (e.g.

Nielsen 2014) link the decline of adaptive ability, specifically at around the age of the end of puberty, to the critical period hypothesis (Johnson and Newport 1989), although it is generally agreed that the capacity to tune one's perception (and consequent production) to foreign speech, despite being influenced by the already established phonological systems, should remain available throughout lifetime (Flege et al. 1997).

As mentioned earlier, only a handful of studies seem to have been conducted that compared age groups in the context of phonetic imitation, some of which belong to the realm of non-native speech imitation, which is further expanded upon in the Section 1.6. One of the earlier attempts comes from Cochrane and Sachs (1979), whose study compared the performance of non-native Spanish child and adult participants, who had no previous knowledge of Spanish, in a task involving imitation of Spanish words. On the whole, the children's performance was judged to excel that of adult participants, according to native Spanish judges. While there were no significant differences between adults and children in how well they applied the Spanish stress rules, the younger group exhibited less interference from English patterns when reading Spanish words. The authors noted the difference between the age groups may have been due to the adults' partial loss of articulatory flexibility relative to the younger group, or their having undergone neural specialisation involved in the production and perception of language The study of Tahta et al. (1981) focused on more minute age differences, when they examined the ability of monolingual English-speaking schoolchildren, aged 5 to 15, to replicate French and Armenian sounds and intonation patterns. The ability to replicate word pronunciation showed a fairly consistent decline across the entire age range studied, while in the case of intonation, there was a rapid drop in imitative performance roughly in the middle of the age spectrum. No noticeable differences were ascertained in the participants' replication ability with respect to the foreign language considered (familiar French vs unfamiliar Armenian). The study by Nielsen (2014), involving three native American English age groups (preschoolers, third graders, and college students), revealed the younger groups to be better able to imitate artificially extended VOTs when compared to adults, although the differences where not found within the two young age groups. The author interpreted the results in terms of exemplar-based theories, which predict greater imitation for children than adults as a result of having accumulated fewer exemplars in memory. Nielsen speculates that the supposed alteration in the plasticity of phonological categories around adolescence leads to fewer phonetic details being perceived or retained with increased age, resulting in attenuated imitative ability.

Despite this evidence that is in favour of younger imitators being superior to adults, other studies seem to support the opposite tendency. Kent (1979), who tested the ability of both native English children and adults to imitate 15 synthesised vowels, including English and non-English vowels. They found that although both groups reproduced their native vowels more reliably, the performance of the younger group was more dependent on whether the vowels were native or not, showing their greater reliance on familiarity with the sounds to be imitated. Snow and Hoefnagel-Höhle (1977) investigated imitative tendencies in the case of native English speakers imitating Dutch words. In the laboratory study, the 5- to 31-year-old participants, with no prior Dutch knowledge, were rated as better imitators as their age progressed. The better performance of older participants was ascribed to their already better baseline performance or, alternatively, to quick adaptation to the auditory stimuli. The authors also conducted a naturalistic study, in which Dutch was the second language of participants, who were 3 to 60 years old. No effect of age on imitative performance in the second experiment was noted, but the participants obtained better, age-correlated ratings for a picture naming task. Alivuotila et al. (2007) compared how Finnish preschool children, naïve adults, and expert phoneticians imitate synthetic vowels presented along the [æ] to [ɑ] continuum. Children turned out to be most affected by their native vowel system, as evidenced by their visibly categorical productions in the imitation task, as was the case with naïve adults, although to a lesser extent. The trained phoneticians, on the other hand, unsurprisingly showed the least amount of categorisation by imitating ambiguous vowels along the continuum. The authors concluded that imitation sharpens with age and phonetic experience. More recently, Schertz and Johnson (2022) compared the degree of VOT imitation (both lengthened and shortened), extracted from sentence material, between Canadian teenagers and adults. Their results revealed that both shortened and lengthened VOT were imitated by both groups, but adults showed significantly more imitative effect in the lengthened VOT condition.

Jia et al. (2006) examined age-related performance in American vowel perception and production (in an immediate imitation task) by native Chinese speakers, who differed in terms of the amount of exposure to English. For the participants residing in China, who had not had prior L2 immersion experience, age correlated positively with both perception and production accuracy of vowels. The effect of age was not ascertained for the recent arrivals in the US (2 years or less), while among past arrivals (3–5 years), there was a younger-learner advantage. The authors concluded that both age and the amount of immersion jointly influence second language learning, as shown by the age-related differences being dependent on varying exposure to the target language. Finally, Paquette-Smith et al. (2022) exposed both Canadian-English children and adults to a delayed-imitation task and to a shadowing task, where the model stimuli included elongated (VOT) of speakers with different accents. While neither of the groups showed signs of imitative effect in the delayed-imitation task, contrary to the results of Nielsen (2014), where a similar paradigm was employed, both groups exhibited robust imitation in the immediate shadowing task with no difference in the degree of imitation found across the two age groups.

The review of the studies that have included age as a potential factor that may influence the degree of phonetic imitation does not warrant definitive conclusions, with young imitators appearing advantageous in some contexts, but not in others. Some evidence points to the importance of considering lesser categorisation in the case of children, as a result of having acquired less exemplars throughout their lives, suggesting they should be more flexible in the reproduction of new or foreign phonetic input, but their tendency to imitate continua rather categorically does not seem to support it. In line with the earlier observation that children are generally good at recognising different accents and they are attuned to the features of foreign-accented speech, some of the cited studies indicate they may have the imitative advantage in a naturalistic setting, especially having been exposed to it for at least few years.

1.6. Phonetic imitation in L2 speech

So far, phonetic imitation has been considered in the current dissertation rather generally, as it is within the wide-ranging domain where the bulk of research has been conducted over the years. What many of these studies seem to have in common is the prerequisite for phonetic imitation to take place, in the form of some type of contrast between the baseline speech of a potential imitator and the speech of the model talker. This has either meant that, for example, some form of model stimuli manipulation, such as VOT lengthening. In some other cases, it has meant exposing the participants to a model speaker from a different dialectal or linguistic background. Some of the latter type of studies that have focused on the imitation of dialectal characteristics have considered different regiolects of Belgium (Delvaux and Soquet 2007), France (Dufour and Nguyen 2013), Italy (D'Imperio et al. 2014), England (Evans and Iverson 2007), the USA (Phillips and Clopper 2011). These studies can be considered as part of the broader domain of second dialect acquisition (SDA), a process by which individuals who relocate to different geographic areas adjust to the dialect variations within their native language (Riverin-Coutlée et al. 2023: 291).

It is necessary now to limit the scope of consideration to the context of second language acquisition (SLA), and more specifically, to the context of second language (L2) speech imitation. While related to dialectal imitation, the two contexts should be distinguished, one reason being that, in the case of L2 speech, the potential imitators and model native speakers are naturally likely to differ in more respects and to greater extents, especially if L2 learners have limited proficiency in the target (L2) language. The L2 learners' productions will inevitably exhibit cross-linguistic interference (CLI) at multiple levels of language competence, including pronunciation, stemming from their already established L1 systems interacting with the developing L2 systems. It should be recognised that L2 imitation constitutes a part of the broad domain of L2 speech acquisition, which itself belongs to the extensive area of SLA, along with the long tradition and considerable amount of research done in those areas. Therefore, the discussion of phonetic imitation of L2 speech should not proceed without at least a passing mention of some of the most recognised models of L2 sound learning that have been proposed over the years. These range from older models, such as the Contrastive Analysis Hypothesis (CAH, Lado 1957) and the Phonological Interference Model (PIM, Brown 1998), to more recent and influential ones, including the Perceptual Assimilation Model (PAM, Best 1995; PAM-L2, Best and Tyler 2007), the Speech Learning Model (SLM, Flege 1995; SLM-r, Flege and Bohn 2021), the Second Language Linguistic Perception model (L2LP, Escudero and Boersma 2004). These models have inspired an abundance of L2 perception studies, as well as numerous production studies, across a variety of L1–L2 combinations. What these models share is their focus on predicting the degree of success in the target language based on the similarities and differences between a learner's native and target language categories.

In short, the older models predict that difficulty in the acquisition of foreign sounds increases when target phonemes (in CAH) or particular distinctive features (in PIM) are absent in the learner's native language. Among more recent models, PAM (PAM-L2) and L2LP have emphasised the role of contrasts, predicting that if a particular contrast between two phonemes is realised similarly in both L1 and L2, it will be easier to discriminate them in the target language, with the latter model additionally accounting for the variability among learners with respect to the particular L1 variety they represent and the L2 variety they are exposed to. Last but most certainly not least, in the case of SLM model (later revised as SLM-r), it is posited that sufficient acoustic distance between a given L2 sound and its closest counterpart in the L1 repertoire facilitates the formation of a new category, and consequently the acquisition of the target sound. Whether the new category is formed also depends on the quantity and quality of input received by the learner, as well as how well the closest L1 category is specified in the learner's phonetic repertoire at the onset of acquisition. According to SLM, the accuracy of L2 sounds' production is contingent on the accuracy of their perceptual representations, although the revised version of the model proposes that perception and production may develop concurrently. In light of the above, in the context of acquisition of L2 speech, phonetic imitation can be viewed as a process whereby L2 learners have the opportunity to bridge the gap between their L1 (or interlanguage) categories and the native-like L2 categories, if only temporarily. In the context of SLA, the imitative effect is expected not only because the learner was simply told to repeat or imitate words, but they may also be motivated by the desire to sound native-like, and therefore they may be expected to approximate the model as best they can.[1] However, in order for this to take place, two main subprocess involved in imitation must be successfully executed, i.e. the perception and subsequent production of the L2 category, both of which have their own challenges likely stemming from either too subtle differences between the L1 and L2 category or some form of articulatory constraint involved in the production of L2 category.

Some of the evidence supporting imitative effect in the context of L2 speech, or more generally, of non-native speech[2] has already been reported in Section 1.5 (Cochrane and Sachs 1979; Tahta et al. 1981; Snow and Hoefnagel-Höhle 1977; Jia et al. 2006). With regard to studies testing the imitation of sound continua[3], Schouten (1977) studied imitation of synthetic vowel stimuli among four Dutch-English bilinguals, and found the resultant imitation to be quite categorical, with the participants' responses forming clusters that corresponded to vowel phonemes from both languages, as indicated by the

---

[1] Although it must be recognised that there is expected variation in terms of the degree to which particular L2 learners may be motivated to acquire the target language or native-like pronunciation specifically. Not to mention that there are many other sources of inter-learner variability accounted for in the literature on SLA.

[2] In some cases, the participants can hardly be considered L2 learners, as they had no previous knowledge of the non-native languages in question, nor did they possibly have any intention of acquiring such knowledge after the experiments.

[3] These studies do not actually expose the participants to L2 native models, but to a series of synthetically generated tokens, some of which are likely to coincide with L1-like categories, while others with L2-like categories.

clusters' being directly related to the categories established in the identification task. The participants displayed much more ease with the execution of the imitation task than in the identification task. Somewhat in line with the idea of a merged L1-L2 category in SLM, based on their results, the authors make an observation that due to a potential limit of how many vowel phonemes can be easily accommodated within a specific area, native phonemes may lose stability in the face of exposure to close (but distinct) non-native counterparts. Flege and Eefting (1987) tested how native English and native Spanish speakers of different ages and linguistic experience imitated the VOT continuum ranging from /da/ to /ta/ (from lead to long-lag VOT). The participants displayed categorisation prior to imitation, with their responses showing abrupt VOT shifts at location along the continuum that corresponded to phoneme boundaries for the given languages. Monolingual English children and adults showed similar patterns, with their VOT productions mostly concentrated in the short-lag and long-lag ranges, as is expected in English. Monolingual Spanish speakers generally followed their native habits of lead and short-lag values, but some the adult speakers showed more English-like long VOT values, possibly due to some experience in English despite being labelled as monolingual. Native Spanish participants who actually spoke English produced stops that exhibited all three modal VOT ranges, including their non-native {t^h} category.

In their comparison of adult English speakers' performance in the realisation of Mandarin Chinese tones in three tasks (identification, read-aloud, imitation), Hao and Jong (2016) found the participants to cope better in the imitation than in the other two tasks. Their second experiment, in which adult Koreans were engaged in similar tasks but involved English consonants, did not align with the results above, as the participants imitative performance was worse than that in the reading task (for plosives) and worse than that in the two other tasks (for fricatives). The authors concluded that the exceeding performance in the imitation task is related to the diminished effect of phonological categorisation in that task. The findings from both experiments were interpreted to suggest that L2 imitation may circumvent certain effects of phonological categorisation. Burin (2018) analysed French speakers' imitation of British English vowels duration and quality in the context of sentence productions. Vowel duration seems to have been accounted for globally only, as a metric of speech rate, with the participants showing decreased vowel duration after auditory exposure, following the models' faster rate of speech. However, the author notes that short vowels tended to be imitated more than long vowels. As for quality, the authors' hypothesis that low vowels should be more imitable than high vowels, particularly in the F1 dimension, as previously found by Babel (2012), was partially validated. The vowels KIT, FLEECE, and HAPPY were discriminated less successfully than they were in the baseline task, although KIT was subject to imitation in most cases. The highest degree of imitation as regards vowel quality was observed in the case of GOOSE and NURSE. The proficiency level of the participants was also considered, but it did not seem to have a significant effect on imitative performance, although some differences were observed in favour of the more advanced group

With regard to native Polish learners of English imitating the features of the latter language, studies have considered either VOT, lack of coda release, and vowel duration contrasts. As regards VOT, native Polish advanced learners of English in Rojczyk (2012) were tested for their ability to imitate English-like long-lag VOT values for /p, t, k/, which are characterised by short-lag values in Polish, in both immediate and distracted imitation conditions. Significant increase in VOTs were observed in the immediate imitation condition relative to the baseline condition. The values obtained in the delayed condition fell in between those obtained in the other two conditions. The author concluded that immediate imitation may temporarily bypass the influence of native pronunciation habits. In contrast to these findings, Wieczorek and Rojczyk (2021) did not observe VOT imitation in their primary school participants. Their native Polish tendency to prevoice /b, d, g/-initial words was maintained in the imitation task, despite the models' invariably positive VOTs for these tokens. This was interpreted to suggest that prevoicing constituted too subtle a cue, which was possibly overridden by the presence of orthographic forms of the stimuli in the imitation task. Those forms may have triggered their native Polish (prevoiced) realisations of the words, which the subsequent auditory model stimuli failed to alter. Similarly, no significant effect of imitative exposure was ascertained for long-lag VOT in /p, t, k/-initial words, despite the model values exceeding the participants' baseline values, although considerable variability was detected across the participants. Still, the participants' baseline values were quite considerable, which could be the result of them already having lengthened their VOTs in the baseline task, as a consequence of simply being instructed to produce words in English, precluding further possibility of such lengthening in the imitation task. Rojczyk et al. (2022) engaged Polish and Czech adult students in a long-lag VOT imitation experiment. Surprisingly, both groups' baseline values exceeded those of the models, but convergence was still observed, as most of the participants reduced their aspiration, and as a result also the participant-to-model distances, in the imitation task. Finally, as regards consonants, Rojczyk et al.

(2013) investigated immediate and distracted imitation of unreleased plosives across word boundary in two-word noun phrases (e.g. *that-tap*, *black-pack*) by Polish advanced learners of English. The feature, frequent in English but much less so in Polish, operationalised as frequency of bursts and their duration, became more English-like in the immediate imitation task. Homorganic clusters (*that-tap*) contributed to the imitative effect more than heterorganic clusters (*black-pack*), due to greater tendency to unrelease the former type of clusters in Polish. In the distracted imitation task, impeded imitative performance was ascertained, as a result of an additional time interval, cognitive taxing, and articulatory resetting between exposure and production.

Most pertinently to the current investigation, a few studies have focused on the imitation of temporal and spectral characteristics of vowels. Zając (2013) examined Polish speakers' imitation of English front vowels in both shortening and lengthening /b_t/ and /b_d/ contexts, as produced by a native and a non-native model speaker. Overall, the participants were found to shorten their vowels relative to the baseline productions, following the models' shorter vowels, which was likely the result of discrepant speech rates at which the participants and the models produced the words. As for the durational contrast related to pre-fortis clipping, some positive convergence was observed for the *bit-bid* pair, when imitating the native model speaker. The participants appeared to have diverged from the non-native model speaker, possibly due to their desire to distance themselves from foreign-sounding pronunciation. A follow-up study by Zając and Rojczyk (2014) revealed that Polish learners of English had already established the correct realisation of vowel duration contrast as a cue to the voicing of the following consonant, as they produced significantly longer vowels before voiced stops than before voiceless stops in the word-list task, prior to the imitation task. The durational distinction was preserved even when they imitated the non-native model speaker, who exhibited no vowel duration contrasts. The authors also investigated the effect of instruction given to the participants before the imitation task, hypothesising that a more explicit instruction would prompt more faithful imitation, but such an effect was not found to mediate the participants' productions. The already mentioned study by Rojczyk et al. (2022) also included the analysis of vowel clipping imitation by Polish and Czech adult participants. Contrary to their findings regarding aspiration, the participants' baseline performance in distinguishing the durational contrasts was worse than that of the model native speakers. Both groups, however, showed a clear pattern of convergence towards the models with respect to vowel durational contrasts, but the effect diminished in the post-task.

Moreover, the authors considered the effect of the type of instruction given to the participants (implicit vs explicit imitation), but similarly to Zając and Rojczyk (2014), it proved to be negligible. Finally, one study considered the imitation of spectral characteristics of vowels. Rojczyk (2013) looked at the degree to which Poles modify their baseline production of the English TRAP vowel a result of exposure to the native English model. Despite their general tendency to yield to the assimilatory impact of their neighbouring native sound categories, namely that of Polish /a/ and /ɛ/, the author observed the participants' ability to overcome this impact, at least temporarily, as they had lower participant-to-model Euclidean distances in the shadowing task. The results were interpreted to support the position that fine-phonetic details are not discarded in perception, as evidenced in their manifestation in subsequent production.

Overall, the reported studies support the position that exposure to L2 native speech in imitation tasks elicits changes in L2 learners' pronunciation in the direction of the target language. These changes are unlikely to be long lasting after such singular experimental sessions as those in most of the studies quoted, as showed by the L2 learners' worse performance in delayed or distracted imitation tasks or in post-exposure tasks. It appears that the role of instructions given to the learners is far from critical, as indicted by comparable degrees of imitation in explicit and implicit conditions. What has also been considered, albeit not frequently, is the level of L2 experience, with some imitative advantage in the case of more adept learners having been reported, although it is not clear how to reconcile this with that such learners should have a lesser initial participant-to-model distance, and therefore they should have less room to converge to the models.[4] With regard to Polish learners of English, they have shown varying tendencies of VOT imitation, but they exhibited more consistency in the imitation of vowel duration contrasts.

[4] Kim (2012: 37-38) writes that "if there is no difference between the model and the pretest [baseline] values, the pretest and the model are already aligned and there is no room for convergence or divergence to occur." However, the question is whether we should consider the imitative effect only in terms of bridging the potential participant-to-model distances. In the case such distances are negligible at both the baseline and imitation level, it could, perhaps, be argued that there still was an effect, in that the participants did not diverge from the model. Still, such a scenario is not easy to disentangle from the potential case of maintenance, where participants persist in the use of their own speech patterns.

1.7. Summary

This chapter could by itself be considered only a cursory and selective summary in the face of an enormous number of venues researchers have explored in their investigation of imitation, one of the fundamental behaviours, whose significance in human development seems uncontested. The mechanism is believed to enhance learning by sparing learners the need to engage in individual discovery and reducing the risks associated with learning purely through trial and error, allowing the imitators to better comprehend the actions they perceive and enact, as well as better predict their potential consequences. The discussion has been selective in the sense that it has primarily focused on phonetic imitation or convergence, a process whereby one's speech undergoes intended or automatic changes, which are more or less temporary, as a result of auditory exposure to the speech of a model speaker.

The process has been found to occur in both socially-rich, interactive settings, suggesting that phonetic imitation is a socially-mediated process, although convergence has also been ascertained in non-interactive settings as well, where the influence of socio-motivational factors is limited. The degree of imitation is also known to vary as a function of factors that are linguistic or cognitive in nature. One of the factors whose influence remains rather unresolved is that of age, with both child and adult age groups surpassing one another from study to study, but with the tentative overall advantage of the younger group, possibly stemming from their attested greater plasticity and quickness to adapt to the features of foreign accents. The debate has also revolved around the issue of whether or not linguistically irrelevant details in the speech signal are discarded in perception and in subsequent production. While some evidence points to the imitation of phonologically relevant features, considerable amount of research has supported the position that fine sub-phonemic speech details are subject to imitation.

Finally, what has been considered is phonetic imitation in the context most pertinent to the current investigation, i.e. the context of L2 speech acquisition. While L2 speech imitation can be regarded along similar lines to those found in L1 speech imitation, it should not be ignored that the former constitutes a part of a broader research area related to second language acquisition. Here, in view of the most influential L2 sound acquisition models, phonetic imitation may be seen as one of the ways of bridging the gap between the L2 learner's default representation of a given target L2 category, likely influenced by their closest L1 sound category or categories, and the true native-like L2 category.

Chapter 2

Differences between Polish and English vowel systems

The purpose of this section is to provide a background on the differences between Polish and English vowel systems, which will help understand the challenges that learners coming from such linguistic backgrounds as Polish may face in their endeavour to enhance their pronunciation in more vocalically complex languages, such as English. While, unlike consonants, vowels do not have precise articulatory identities, which restricts the description of these sounds in terms of similarities and differences between the native and the target language, research has shed light on how the vocalic cross-linguistic influence might manifest itself, and how it may affect the pronunciation in the target language.

The chapter considers two aspects of Polish-English vowel contrasts that are known to be problematic for Poles, including vowel durational and qualitative contrasts. The former aspect is discussed with regard to both inherent and relative or context-dependent vowel temporal characteristics. The latter takes into account two perspectives, with one regarding the quality of vowel monophthongs rather holistically, as if they were constant throughout their duration, the position which traditional description seem to have favoured. Another perspective recognises the fact that vowels are not uniform entities, but their quality changes over their course, much like in diphthongs, although to a lesser extent.

2.1. Vocalic systems in English and Polish

The most notable disparity concerns the numbers of vowels that the two languages feature, with Polish boasting six oral monophthongs, while English, depending on the variety, between nine and eleven (Weckwerth 2010: 542), or up to twelve with short schwa included (Rojczyk and Porzuczek 2012: 4). Such a number of vowels as is present in English places it in the middle in terms of the complexity of vocalic patterns across languages (Gonet 2017: 103). Therefore, with roughly comparable vowel space dimensions for the two languages and a smaller number of vowels for Polish, more vowel dispersion is inevitable in Polish, the consequence of which are greater distances between the vowels. On the other hand, since the English vowels are less separated, they require greater precision in their articulation than it is the case in Polish (Gonet 2017: 111).

Moreover, English vowels are typically found between Polish ones in the vowel space (Sobkowiak 2008: 54), and there appears not to be a single vowel in Polish, whose spectral properties would match those of any vowels found in the English repertoire (Rojczyk and Porzuczek 2012: 4). The close vicinity of English vowels to the Polish ones, and their deceptive similarity often results in Poles' resorting to their native vowel inventory when speaking English. In line with the predictions of the SLM model, the misleadingly equivalent English vowels are likely to be more difficult for the Polish learner to perceive and, in consequence, difficult to produce as well. This results in erroneous vowel substitutions that can potentially lead to quasi-homophones, which may hamper understanding (Sobkowiak 2008: 133).

The difficulty is compounded by the fact that, together with vowel quality differences, vowels' inherent duration also varies across particular vowels in English (Roach 2009), which is not so much the case in Polish. On top of that, there is yet another layer of durational distinction, which is related to extrinsic (Gonet 2017: 139) or relative (Sobkowiak 2008: 193) vowel duration being dependent on the voicing status of the following consonant. Further features that make the two vowel systems disparate include the fact that Polish vowels, unlike English vowels, tend to be subject to heavy nasalisation and they tend not to be subject to reduction (Sobkowiak 2008: 130), and they exhibit glottalisation in the word-initial position (Gonet 2017: 91). Finally, there are vowel dynamics, which are said to differ across the two languages, with Polish exhibiting less formant movement throughout certain vowel portions (Schwartz 2020).

2.1.1. Vowel durational contrasts in English and Polish

As is known, the qualitative differences across English vowels are accompanied by the quantitative differences (Ashby and Maidment 2005: 75). According to (Roach 2009: 13-16), English RP vowels are traditionally divided into lax (short): /ɪ/, /e/, /æ/, /ʌ/, /ɒ/, /ʊ/, and tense (long) vowels: /iː/, /ɜː/, /ɑː/, /ɔː/, /uː/, which correspond to Wells' (1982) keywords: KIT, DRESS, TRAP, STRUT, LOT, FOOT and FLEECE, NURSE, PALM, THOUGHT, GOOSE. The duration contrast becomes an issue in the case of foreign language learners whose native language does not generally exhibit durational variability of vowels, such as is the case in Polish (Jassem 1962, reported in Rojczyk and Porzuczek 2012). Polish, not making a distinction between short and long vowels, compels learners of English from this linguistic background to insufficiently attend to such contrasts, as they tend to underdeliver in their distinguishing between English short and long vowels (Waniek-Klimczak 2005). Even the productions of advanced learners of English, although they show some competence in modifying vowel duration, are not error-free, by rendering words such as *good* as /gʊt/ (Nowacka 2010). At the same time, Polish learners of English may also overestimate vowel duration in differentiating English vowels, as was found by Rojczyk (2011). The study revealed that, because the TRAP vowel tends to be longer relative to the other lax vowels, the contrast in Poles' production of TRAP and STRUT is primarily brought about through the temporal, rather than the spectral cues. A similar pattern was found for the perception of these two vowels, with the longer stimuli being often identified as TRAP, and the shorter ones as STRUT, regardless of their spectral properties. Similarly, Bogacka (2004) found native Polish learners of English to not attend to spectral cues in their perceptual separation of FOOT and GOOSE, or of KIT and FLEECE. Instead, the participants relied on durational cues in both contrasts, although with a reversed identification pattern for the back-vowel contrast. The overreliance on vowel duration is in line with Bohn's (1995) Desensitisation Hypothesis, according to which, when L2 learners have not gained enough L2 experience to become sensitive to the spectral differences between L1 and L2 vowels, they will avail themselves of temporal contrasts between these vowels, even if such contrasts are absent in their L1.

Apart from the fact that English vowels manifest inherent temporal specification, vowels' duration is, independently, subject to variation depending on positional or contextual factors, resulting in extrinsic (Gonet 2017: 139) or relative (Sobkowiak 2008: 193) vowel duration. Apart from the rather universal tendency for open syllables to be longer than the closed ones, English, in particular, exhibits markedly longer stressed syllables, relative to the unstressed ones (Delattre 1966). Moreover, vowels in phrase-final syllables are found to be longer than those situated in other positions, and word-final syllables tend to be longer than word-initial or word-medial syllables (Klatt 1975). Additionally, a vowel in a two-syllable word exhibits shortening in the first syllable when it is followed by an unstressed syllable, but not in the second syllable when it is preceded by un unstressed syllable (Klatt 1973). More pertinently to the current investigation, English extrinsic vowel duration servers as a cue to signal the voicing status of the following coda consonant, in that vowels preceding voiceless consonants are shortened (Chen 1970), a fact observed as early as in 1950 by Daniel Jones (reported in Rojczyk 2010b).

Polish does not employ extrinsic vowel duration phonemically in distinguishing voice of the subsequent obstruents (Sobkowiak 2008; Jassem and Richter 1989), neutralising this contrast word-finally, although there is some evidence for marginal vowel duration differences in the two conditions (Slowiaczek and Dinnsen 1985, but see Slowiaczek and Szymanska 1989). Poles, not being accustomed to perceiving voice contrast stemming from vowel duration changes, fail to make use of this cue in English in both perception (Rojczyk 2010b) and production (Waniek-Klimczak 2005). Figure 2.1 illustrates the disparity in the implementation of vowel duration contrast as a cue to the voicing of the following consonant in the two languages.

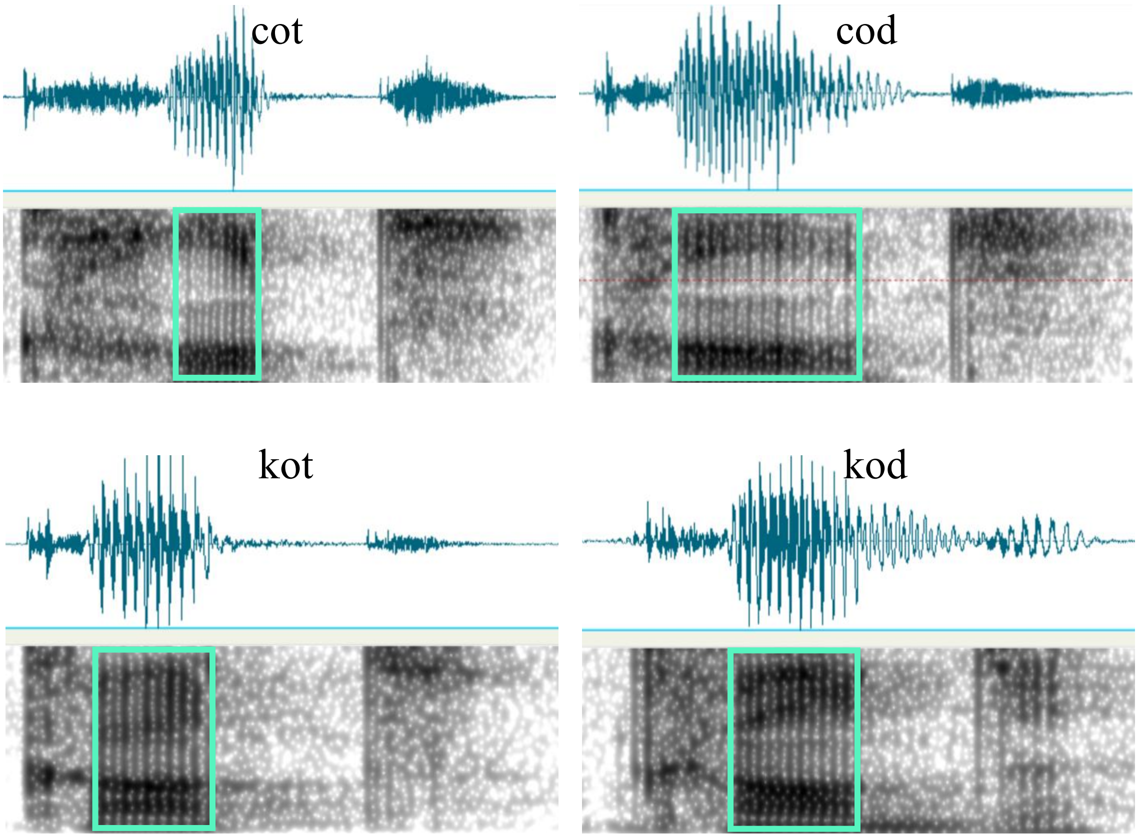

Figure 2.1. English *cot-cod* pair (top row) and Polish *kot-kod* (*cat-code*) pair (bottom row)

While in the case of English, vowel duration tends to be visibly longer in words such as *cod* relative to *cot* (all other things being equal), Polish word pairs such as *kot* and *kod* (*cat* and *code*) exhibit, at best, only a marginal duration difference in favour of the latter word. Ignoring both intrinsic and extrinsic vowel duration contrasts may result in a multitude of non-existent homophones, with such extreme examples as the words *bit*, *beat*, *bid*, and *bead* all being pronounced the same way by naïve Polish learners of English.

2.1.2. Vowel quality contrasts between English and Polish – a static approach

Apart from duration, another key factor contributing to the differences between the two vowel systems, and consequently to the difficulties in their separation in both perception and subsequent production, lies in vowel quality. Studies have demonstrated that Polish learners of English have difficulties when it comes to the formation of target vowel categories, which is evidenced by them resorting to their native vowel repertoire, or to L1-L2 merged vowel categories, when speaking English. Figure 2.2 illustrates the position of both Polish and English vowels in the vowel space, based on impressionistic (left and middle) and acoustic (right) data.

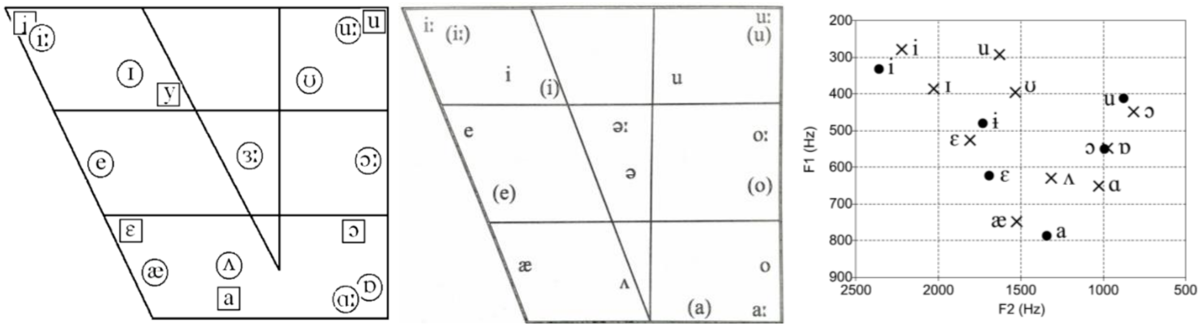

Figure 2.2. Comparisons of Polish and English vowel data based on different sources (two impressionistic and one acoustic):
left: Polish – squares, English – circles (Porzuczek et al. 2013),
middle: Polish – brackets, English – no brackets (Sobkowiak 2008),
right: Polish – dots, English – crosses (Weckwerth 2010, adopted from Bogacka et al. 2006; Ferragne and Pellegrino 2010)

As remarked earlier, with English exhibiting roughly twice as many vowels (oral monophthongs), situated approximately in between the Polish vowels, inevitable confusion is expected from the point of view of a Polish learner of English, especially in those cases where L1-L2 category distances, and consequent qualitative differences, are marginal. The following discussion is necessarily limited in terms of the number of vowels considered. The four front Standard Southern British English vowels, along with their closest Polish counterparts were chosen. The choice of these English vowels stems from their relatively frequently reported difficulty involved in their perception and realisation, but also from more pragmatic considerations, discussed in Chapter 4.

The first vowel to be considered is the FLEECE vowel, which in literature is described as one in which:

> … the front of the tongue is raised to a height slightly below and behind the front close position; the lips are spread; the tongue is tense, with the side rims making a firm contact with the upper molars … the vowel is often noticeably diphthongized, especially in final positions ... (Cruttenden 2008: 106)

The vowel is in very close vicinity to Polish /i/ in the vowel space, and based on impressionistic comparisons, the English vowel is reported as longer, with a lowered and/or retracted position relative to Polish /i/ (Porzuczek et al. 2013). Sobkowiak (2008) additionally describes FLEECE as more tense and peripheral. Due to the fact that FLEECE can easily by classified as very similar (using SLM terminology) to Polish /i/, it is predicted to be subsumed by the Polish vowel. The prediction is consistent with the observations made by Nowacka (2010), who found some of her native Polish advanced learners of English to mostly resort to their native /i/. However, Polish realisation of FLEECE has also been reported to have a KIT-like quality (Schwartz 2015: 211).

The English KIT vowel is also expected to pose a challenge to the Polish learner of English. In the literature, it is described as one that is:

> … pronounced with a part of the tongue nearer to the centre than to front raised just above the close-mid position; the lips are loosely spread; the tongue is lax (compared with the tension for /iː/), with the side rims making a light contact with the upper molars … the degree of closeness and centralization [of KIT] varies according to the accentual force falling upon the vowel and its position in the word … (Cruttenden 2008: 107-108)

Relative to its closest Polish neighbour /ɨ/, English KIT is fronted and less open, and resembles a sound that is intermediate between Polish /i/ and /ɨ/ (Porzuczek et al. 2013: 57). Incidentally, Cruttenden (2008: 108) also makes a note about the Polish vowel, describing it as a centralised form of KIT. Sobkowiak (2008: 163) adds that KIT exhibits a greater degree of tenseness than Polish /ɨ/. Even though Polish /ɨ/ appears to be in closer vicinity in the vowel space than Polish /i/, it is the latter Polish vowel that is more frequently resorted to in Poles' realisation of KIT (Gonet 2017: 111). In her longitudinal study, Nowacka (2010: 238) observed this pattern in both reading and speaking in advanced learners of English, despite their having undergone phonetic training. This tendency may be attributed to the strong influence of orthography on the Polish learner, who seems to rely on the fact that both English KIT and Polish /i/ are frequently similarly represented in orthography, i.e. by means of the letter <i> in similar contexts (Sobkowiak 2008: 162). At the same time, Polish advanced learners of English in Rojczyk (2010a) were found to successfully dissimilate English KIT from both Polish /i/ and /ɨ/, by having established a separate, although not entirely stable, category in the acoustic space.

One of the most deceptively similar (cross-linguistically) vowels that is found in the English repertoire is the DRESS vowel, in which case:

> … the front of the tongue is raised between the close-mid and open-mid positions; the lips are loosely spread and are slightly wider apart than for /ɪ/; the tongue may have more tension than in the case of /ɪ/, the side rims making a light contact with the upper molars. (Cruttenden 2008: 110-111)

Pronunciation textbooks describes English DRESS as less open when compared to Polish /ɛ/, with the sides of the tongue in contact with the upper teeth, and a greater degree of the spread of the lips (Porzuczek et al. 2013: 63). To achieve the desired vowel quality, the learner is instructed to make a compromise between KIT and Polish /ɛ/. In a similar vein, Polish /ɛ/ tends to be somewhat lower and requires much less muscle tension than English DRESS (Sobkowiak 2008: 146). The recommendation for the learner is slightly different, however, in that they should aim for the middle ground between two Polish vowels: /ɛ/ and /ɨ/. There is a narrow variety of Polish /ɛ/, found in the context of palatal consonants (Weckwerth and Balas 2019), which induce narrower mouth opening, and is therefore described as a better candidate for substitution for English DRESS (Gonet 2017: 114). The deceptive similarity between English DRESS and Polish /ɛ/ is predicted to lead even highly proficient learners to substitute the Polish vowel for the English one. This was confirmed by the already quoted studies of Nowacka (2010) and Rojczyk (2010a). In the latter study, the participants' attempted productions of DRESS exhibited an even lower tongue position relative to Polish /ɛ/. The author suggests that this finding may be ascribed to the phenomenon of hypercorrection, with Polish learners overusing the TRAP vowel, motivated by the desire to sound more like a native English speaker (Sobkowiak 2008: 146). Finally, the participants' DRESS in Weckwerth (2010) showed a considerable overlap with TRAP, but not with STRUT, with DRESS also exhibiting the least variability among the three vowels investigated.

The final English vowel to be considered is the TRAP vowel, in which, according to the literature:

> … the mouth is more open than for /e/; the front of the tongue is raised to a position midway just above open, with the side rims making a very slight contact with the back upper molars; the lips are neutrally open. (Cruttenden 2008: 112)

When compared to its closest neighbouring vowels in Polish, i.e. /a/ and /ɛ/, the TRAP vowel is generally described as falling half-way between the two Polish vowels (Sobkowiak 2008: 142). This means that Polish speakers are equally likely to resort to either Polish low central /a/ or front mid /ɛ/ (Rojczyk 2013). Nowacka (2010: 239) observed frequent use of both Polish vowels in advanced learners' instances of TRAP. Weckwerth (2010) noted TRAP's almost complete overlap with DRESS and STRUT in the acoustic space, at the same time showing considerable variability in his participants' production of TRAP, which may be posited to be the consequence of the Poles' indecision regarding the most optimal substitute for this vowel from among the two vocalic neighbours. Finally, as was reported earlier, due to the vowel's inherently greater duration, Polish learners of English may prioritise the temporal, rather than the spectral aspect of TRAP in both perception and production (Rojczyk 2011). The problem with the TRAP vowel might be exacerbated by such factors as its dialectal variation in American English (Nearey 2013) and generational variation (Hawkins and Midgley 2005), which may make learners of English, exposed to such varied input, confused about what the desired vowel quality should be.

In sum, the comparison of English front vowels with their closest Polish counterparts demonstrates that a Polish learner of English is faced with the discriminatory challenge involving not only the appreciation of the temporal aspects of the English vowels, but also recognising sometimes subtle qualitative differences between English and Polish vowels. In the case of FLEECE, the Polish learner is expected to equate it qualitatively with Polish /i/ or with English KIT-like quality, along with underappreciating its durational aspect. English KIT will either be assimilated by Polish /i/ or /ɨ/, while English DRESS by Polish /ɛ/. Finally, in the case of TRAP, Poles are expected to resort to its two closest Polish neighbours, i.e. Polish /ɛ/ and /a/. While these substitutions should be expected to manifest themselves in the case of inexperienced learners, research has shown that they can persist into more advanced stages of language learning.

2.2. Vowel inherent spectral change (VISC)

So far vowel quality has largely been discussed as if it were constant throughout the vowel's duration. While a static view of vowel quality is sometimes necessary, as Hillebrand (2013: 26) puts it, it is "a convenient simplification … with some important liabilities that are not always properly appreciated." In his excellent review of vowel dynamics in North American vowels, Hillebrand gives the example of vowel data from the classic study of Peterson and Barney (1952), where despite considerable vowel category overlap in the F_1-F_2 plane, the untrained listeners, tasked with the identification of these vowels, did so quite efficiently. One of the potential reasons for that was precisely because vowel overlap was only evident at a single time slice, and it quite clearly underrepresented the vowels' true quality throughout their duration, to which the listeners must have attended to.[5] These changes of spectral properties of particular vowels, even those traditionally referred to as monophthongs, have come to be known as vowel inherent spectral change or VISC (Nearey and Assmann 1986). Nearey and Assmann (1986), who coined the term, observed significant formant frequency changes not only over the course of (Canadian) English diphthongs, but also for nominal monophthongs /ɪ/, /ɛ/, and /æ/,[6] even without the presence of consonantal context. The unstable quality of even such vowels is illustrated in Figure 2.3, as measured between two points in the vowels' duration (24% and 65%).

[5] The reason mentioned aligns with dynamic specification theory of vowel perception (Strange et al. 1983: 696), according to which "dynamic articulatory events give rise to a dynamic acoustic pattern in which the changing spectro-tremporal configuration provides sufficient information for the identification of the phonetic units." Alternatively, target normalisation theories assume "the essential information for vowel identity is contained in the asymptotic spectral cross section within the syllabic nucleus, which most closely corresponds to the canonical (isolated) vowel targets. However, since these static spectral patterns are inherently ambiguous across speakers and contexts, the veridical perception of vowels requires complicated normalisation processes through which the variable acoustic input is recoded in some way to arrive at the invariant percept."

[6] For North American vowels, IPA symbols, instead of Wells's keywords, are used in the current dissertation.

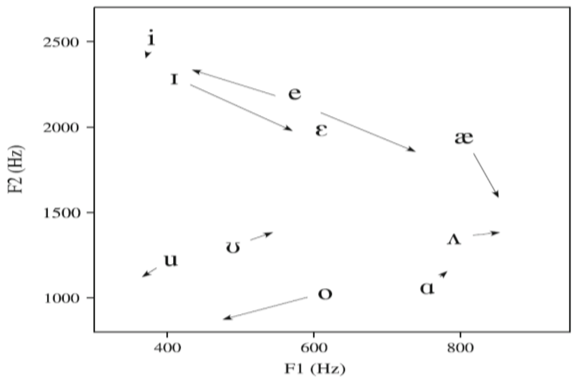

Figure 2.3. Formant trajectories of isolated Western Canadian English vowels, based on two-point measurements at 24% and 65% of vowel duration (Nearey and Assmann 1986, adapted from Nearey 2013)

Nearey (2013) distinguishes between the following possible patterns of formant movement in the F_1-F_2 space, based on the movements towards particular vowels, observed across North American English:

> 1. i-VISC or iota-VISC: movement toward [i] or [j] (/e/)
> 2. u-VISC or upsilon-VISC: movement toward [u] (/o/ and perhaps weakly /u/)
> 3. a-VISC or alpha-VISC: movement toward [a] (/ɛ/ and perhaps /æ/ and /ɔ/, also possibly /ɪ/ and /ʊ/)
> 4. ə-VISC or schwa-VISC: movement toward [ə], that is toward relatively neutral values of F_1 and F_2 (perhaps /ɪ/ and /ʊ/) (Nearey 2013: 55)

Based on previous findings, the author enumerates expected movement types for North American vowels, with the caveat that some of these may depend on the dialect or context considered. The tense vowel /e/ (narrowly [eɪ]) is expected to display iota-VISC, as is the vowel /i/, while the tense /o/ (narrowly [oʊ]), together with /u/ will tend to exhibit upsilon-VISC. The lax vowels /ɪ/, /ɛ/, /æ/ and /ʊ/ tend to show some degree of alpha-VISC, while /ɑ/ and /ʌ/ do not display visible VISC patterns.

The importance of vowel dynamics in the perception of vowels by North American speakers was demonstrated by experiments in which the Silent Centre paradigm was employed (e.g. Strange et al. 1983; Jenkins et al. 1983). In these experiments, which tested how removing various vowel sections would impair their identification rate, it was found that removing the centre section did not hinder vowel identification significantly. On the other hand, the removal of the extreme vowel sections, containing CV and VC formant transitions, had a negative impact on vowel identification. The role of vowel dynamics in vowel perception was proved yet in another away, by having listeners exposed to static spectral cues only. Hillenbrand and Gayvert (1993) used steady-state synthesised version of data from Peterson and Barney (1952), and found a much higher identification error rate for the flat-formant versions of vowel tokens when compared to the original error rate in Peterson and Barney (1952). Hillenbrand et al. (2001) exposed their listeners to vowels surrounded by a multitude of consonant pairings, and observed considerably improved category separability with a pattern classifier that incorporated spectral change information.

There is a general agreement that formant dynamics in the initial part of the vowel play an important role in its identification of North American vowels, but it is less clear what other cues are perceptually relevant (see Morrison 2013 for a review). In short, one hypothesis stresses the additional importance of formant values towards the end of a vowel, or as the change of formant values between the vowel's onset and offset. Alternatively, it is the slope of change that matters, reflecting the rate of change of formants over time, including the information whether the slope is upward or downward. Finally, the direction hypothesis states that the additional cue lies solely in the direction of formant movement.

Besides there being a number of factors affecting the patterns of vowel dynamics, such as the vowel-specific nature itself, consonantal context (Steven and House 1963), and broadly defined prosodic effects (Fox and Jacewicz 2009: 2603), studies have also shown that spectral dynamics vary with language dialect. For example, Fox and Jacewicz (2009) compared spectral dynamics in vowels in three regional varieties of American English (Western North Carolina, Central Ohio, Southern Wisconsin) and found dialect to be a strong source of variation in VISC in the case of all tested vowels (/ɪ, ɛ, e, æ, aɪ/). Farrington et al. (2018) investigated vowels among Southern varieties of American English (Tennessee, North Carolina, Virginia) and the authors confirmed the usefulness of non-static measures in better understanding not only vowel changes within these varieties, but also in sociophonetics generally. Williams and Escudero (2014) compared both monophthongs and diphthongs used in Standard Southern British English and in a Northern English dialect (Sheffield English). They found significant differences in terms of mean formant trajectories for phonologically back (and one central) monophthongs, while the magnitude and direction of formant trajectories were different mostly for the nominal diphthongs.

Vowel dynamics appear to vary systematically across languages, with languages such as Italian, French, and Polish being impressionistically judged as pure relative to languages such as English and Danish (Schwartz and Kaźmierski 2020: 231). The cross-linguistic view of VISC was inspected more explicitly by Williams et al. (2015), who demonstrated that VISC was more important for determining the quality of Standard Southern British English vowels than for Northern Standard Dutch. The cross-linguistic disparity between the degree of vowel dynamics was also investigated less directly, through the analysis of American vowels produced by native English, Korean, and Chinese speakers. The Chinese speakers, whose language has a sparse vowel system, turned out to exhibit the greatest VISC distances, even greater than those of native English speakers, although the spectral angle of formant movements did not appear to depend on the linguistic background of the participants.

2.2.1. Vowel quality contrasts between English and Polish – a dynamic approach

One of the richest contributions to our understanding of cross-linguistic VISC comes from the research done by Geoffrey Schwartz (and colleagues), who shed light on the specification of vowel dynamics from the point of view of a native Polish learner of English. Polish vowel system exhibits a relatively small repertoire of vowels, characterised by a relatively stable quality (Schwartz et al. 2016a: 61). In English, given the relatively high vowel density in the vowel space and consequent vowel overlap, it is expected that dynamic properties (together with duration) will be made use of in the differentiation between particular vowel categories to a higher degree than it should be the case in Polish, with its less crowded system (Schwartz et al. 2016a: 62). Schwartz et al. (2016b: 185) alludes to Święciński's (2004) comparison of articulatory settings in Polish and English, with the latter language exhibiting a more relaxed articulatory supralaryngeal setting, which the former authors hypothesise might lead to slower formant transitions, and thus to more robust formant dynamics in English relative to Polish. CV and VC transitions in Polish are realised more rapidly than in English, which means vowel targets are reached earlier and they take up more of the vowel's duration, resulting in a more stable overall quality relative to English, which exhibits slower transitions that are more spread over the vowel's duration (Schwartz et al. 2016b: 73-74).

As a consequence of the disparity between vowel dynamics in the two languages, Polish learners of English may be considered at a disadvantage in their attempt at the acquisition of the native-like vowels in the target language (Schwartz 2015: 206). This may be evident in perception, with Polish learners of English reportedly judging the quality of vowels based on their initial portion. For example, they may perceive such vowels as British English TRAP in terms of its onset quality, which resembles DRESS (easily confusable with Polish /ɛ/), as opposed to native English speakers, who tend to delay their vowel quality judgment as a result of greater reliance on later vowel portions (Schwartz 2015: 211; Schwartz et al. 2016b: 189). Similarly, Polish realisation of FLEECE as a vowel resembling KIT may be the consequence of the identification being made based on the vowel's initial portion, which may be described as KIT-like (Schwartz 2015: 211).

In their experiments, Schwartz and colleagues investigated experimentally the disparity between the two languages in terms of their vowel dynamics. Schwartz's (2015) pilot study explored the degree of vowel dynamics in the productions of native Polish beginner and advanced learners of English in the FLEECE and TRAP vowels. The study revealed that more proficient participants exhibited more robust formant dynamics, and their productions were in turn associated with better ratings on a scale of foreign accentedness by native English judges. Similar findings come from Schwartz et al. (2016b), who compared VISC in FLEECE and TRAP as realised by native Polish students and highly proficient university teachers, with the latter group manifesting greater dynamics.

Schwartz et al. (2016a) examined the role of vowel dynamics in Polish speakers' perception of both Polish and English vowels through the employment of the Silent Centre paradigm, mentioned earlier. In the case of the Polish stimuli, the type of stimuli, with different vowel portions presented and others silenced, did not affect identification accuracy for the most part, which confirmed the lesser role of vowel dynamics in the perception of Polish vowels. In the case of the English stimuli, the more advanced participants showed greater accuracy for SC tokens (first and last 20% of vowel duration preserved) and for initial tokens (first 35% of vowel duration preserved) than less proficient participants did, proving that more native-like attunement to dynamic vowel specification comes with linguistic experience and proficiency. However, the disparity in the reliance on vowel dynamics in English vowel perception was less evident across beginner and advanced groups in Schwartz and Dzierla (2018), who did a similar study for English stimuli only, but using a forced-choice rhyming task instead of an identification task, with both groups ultimately reacting rather similarly to VISC in the English stimuli, and with initial tokens proving the most difficult among all stimulus types.

Schwartz and Kaźmierski (2020) aimed to compare spectral dynamics of English front vowels (FLEECE, KIT, DRESS, TRAP) across different proficiency levels, taking into account baseline L1 (SSBE) English data, and L1 Polish data for the vowels that were deemed as closely corresponding to the English ones (Polish /i, ɨ, ɛ, a/). Out of five vowel intervals (the two extreme ones excluded), in the second interval (20%-40%), formant movement was expectedly smaller in baseline Polish than in baseline English, with L2 productions generally falling in between the baseline data for the two languages, but not much movement was observed in the third interval (40%-60%), for any of the data groups. Greater dynamics in the early portion of the English vowels, as reflected by larger Euclidean distances, which, incidentally, increased with proficiency, agreed with the authors' predictions resultant from the Onset Prominence representations in the two languages (discussed later). Interestingly, baseline Polish exhibited the greatest dynamics in the fourth interval (60%-80%), which the authors suggested could be due to the effects of the fortis coda consonants, potentially leading to less robust VC transitions in the English items. The authors observed that what primarily distinguished Polish and English vowel data was not so much the total amount of VISC over the course of the vowels, but rather in which vowel portion the formant movement was more evident, with it being housed earlier in English, but later in Polish.

Schwartz (2021) continued the VISC comparison between the four English vowels and their four closest Polish counterparts, using different metrics (e.g. F_1 and F_2 slopes) and larger vowel intervals (25% each, with the two extreme ones excluded). Somewhat consistently with the previous findings, the early portion of the vowels (25%-50%) manifested more dramatic formant dynamics (of F_1 in particular) in English relative to Polish, but the opposite found in the third interval (50%-75%), which was attributed to the more imminent effect of the coda consonant in Polish. In his second experiment, the author compared the participants' L1 Polish vowels with the same participants' L2 English vowels, the results of which largely mirrored those from the first experiment for L1 Polish and L1 English, respectively.

Overall, these studies show that Polish and English are different in terms of their implementation of and their reliance on vowel dynamics. However, despite the initial observations that Polish exhibits more stable quality than English, some later findings suggest that vowel dynamics can also be considerable in Polish, but towards the later portions of the vowel. It was also shown that in the perception of Polish vowels, spectral dynamics do not appear as relevant as they do for native English speakers. Moreover, when native-Polish learners of English are considered, their English vowel perception and production seem to be characterised by more robust formant patterns with increased L2 proficiency.

Schwartz proposes that the difference in vowel dynamics between the two languages may be phonological in nature. While the main source of VISC may be attributed to the coarticulatory effect of onset consonants (Schwartz et al. 2016b: 196), "what starts out as a co-articulatory effect of consonants shifts its affiliation to become a feature inherent to vowels" (Schwartz et al. 2016a: 73). When the CV transitions take more time, such as is the case in English, the co-articulatory effect of the consonant extends further into the vowel, and this effect may be considered as an integral characteristic of the vowel. The implementation of different degrees of phonologisation of VISC across the two languages is explained within the Onset Prominence representational framework or OP (e.g. Schwartz 2016). The essence of this framework lies in the ambiguity concerning the initial vowel portion, referred to as Vocalic Onset (VO), which necessarily carries the information about the preceding consonant in the form of formant transitions. Languages may vary depending on whether they assign this ambiguous initial portion to the representation of the consonant or the vowel itself. If VO is considered as embedded into the consonant, as it is predicted to be the case in English, more robust CV interaction is expected, and therefore more dynamic formant patterns ensue. In Polish, the ambiguous part is posited to belong with the representation of the vowel, hence the vowel targets are reached earlier. On the whole, the observed general confirmation of hypotheses based on the OP framework (e.g. in Schwartz and Kaźmierski 2020) lends support to its usefulness in cross-linguistic comparison of VISC patterns, including the distinction between Polish and English.

The relevance of phonological considerations of VISC becomes evident in the face of the issue of similarity, when addressing such questions as how vowel dynamics can be integrated into existing models of second language speech acquisition (Schwartz et al. 2018). According to the authors, in SLM and PAM, similarity between segments is understood rather generally, i.e. in terms of whole segments, without considering the segments' specific constituents. For example, if Polish learners base their vowel percept on the initial portions of the vowel, SLM's equivalence classification should apply only to the early portions of the vowel. Moreover, with increased L2 experience, the decision as to the identity of the vowel should be delayed until later in the vowel's duration, resulting in acquiring new more dynamic L2 vowel categories. Likewise, in the case of

PAM (PAM-L2) model, with experience, we should expect a refinement of the articulatory gestures that influence L2 formant dynamics.

2.3. Summary

The chapter focused on the differences between Polish and English vowels systems, and on the resultant difficulties that Polish learners of English may face in the acquisition of English vowels, stemming from cross-linguistic interference. The differences in the temporal organisation of vowels across the two languages are known to affect Poles' L2 pronunciation, as a result of their both underappreciation of durational cues in some cases, and overreliance on them in other contexts, with the accompanying apparent disregard for spectral cues. Because of the deceptive qualitative similarity of some of their native vowels to the target ones, the learners concerned will readily resort to their native repertoire, with some experienced learners establishing potentially new, although not necessarily native-like categories for the target vowels.

The discussion of vowel quality included the less holistic approach, one which takes into account dynamic patterns throughout the vowels' duration, yet another aspect differing the vocalic systems of the two languages. While Polish has been largely regarded as a language where vowel dynamics do not play a significant role in either perception or production, the opposite tendency is found in English, with native English speakers exhibiting considerable reliance on them when identifying vowels, which is accompanied by their dynamic vowel productions. Studies have shown that naïve Polish learners are likely to exhibit more stable formant patterns, at least in the initial portions of English vowels. They may also base their judgement of vowel identity on their initial impression of vowel quality, ignoring subsequent vowel portions, leading to erroneous percepts and subsequent productions. Finally, the phonological account of the cross-linguistic contrast was mentioned, specifically within the Onset Prominence representational framework, whose central idea lies in the ambiguity of the consonant-vowel transitions, potentially constituting an integral part of either the consonant or the vowel.

Chapter 3

The cross-linguistic interference of orthography

One aspect of cross-linguistic interference, one that may have a bearing on the current investigation, is the interference resulting of orthography. This chapter is intended to provide only a short review of some of the aspects related to the effects of spelling on L2 learners' pronunciation. The effects of orthography may be particularly relevant in those cases where the native and the target languages differ in terms of orthographic depth, such as is the case in the acquisition of English by Poles.

3.1. The interplay of spoken and written input

It seems well established that the perception of speech is inherently multimodal (Rosenblum 2008: 405; Erdener and Burnham 2005: 193), which prompts language users to make use of other modalities beyond the auditory one when such are available. Listeners may thus combine auditory cues with visual ones when forming speech percepts, the latter including the written representation of language. Goldinger and Azuma (2004), who tested the faithfulness of imitation of tokens read by the participants in two sessions (before and after being exposed to auditory training tokens), concluded that detailed traces of speech, which are preserved in memory, can be activated when encountering only orthographic representations of language. This is accounted for by considering printed words as probes that tap into long-term memory to be ultimately assembled into responses in the working memory (Goldinger and Azuma 2004: 720). Similar conclusions come from the shadowing experiments by Dufour and Nguyen (2013), who speculate that detailed traces formed during speech perception are subsequently used for speech production, and may be activated in the process of written word recognition (Dufour and Nguyen 2013: 5). This may be considered in line with the recurrent feedback model (Ziegler and Ferrand 1998: 683), according to which the presentation of a written word triggers grapheme nodes, which subsequently activate phoneme nodes, and vice versa in the case of the presentations of a spoken word, essentially resulting in a bidirectional flow of activation between orthography and phonology. This, however, should not obscure the fact there is evidence supporting the position that that access of orthographic lexemes and phonological lexemes occur independently (Caramazza 1997: 189).

3.2. The effect of orthography on L2 learners' speech

The vast majority of participants in studies on L2 phonological acquisition tend to be highly educated and literate learners, who are exposed to both written and auditory input right from the start and throughout the acquisition process (Colantoni et al. 2015: 23), with written input possibly constituting a large part of the overall input offered to L2 learners (Bassetti 2008: 191). While native speakers of a language have to cope with orthography-internal factors, L2 learners are impacted by the interplay between their L1 and L2 orthographies (Bassetti 2008: 198). More specifically, L2 orthographic input may be reinterpreted according to the L1 orthography-phonology conversion rules. Such distorted input may interact with L2 acoustic input, which itself is also reshaped by L1 phonology, resulting in non-target-like phonological representations of L2 phonemes, syllables, and words. L2 users may even strategically disregard the acoustically obtained percept in favour of orthographic forms (Bassetti 2024: 706). Bassetti (2008: 196-197) reports that the resulting orthography-induced non-target-like productions may take place not only at the moment of the L2 learner being exposed to L2 orthographic input, but also beyond, in such tasks as repeating spoken words.

The effect of orthographic input may vary across given L1-L2 pairs with respect to the degree to which their writing systems deviate from simple one-to-one letter-to-phone (or grapheme-to-phoneme) correspondence (Van den Bosch 1994: 1), ranging from relatively deep (or opaque) orthographies, such as in Hebrew or English, to shallow (or transparent) orthographies, as in Serbo-Croatian or Italian. Orthographic depth can be viewed not only from the point of view of complexity of print-to-speech correspondences, but also from the perspective of the unpredictability involved in deriving words' pronunciations based on their spelling (Schmalz 2015: 1614). It can thus be imagined that a learner, whose L1 orthography is transparent, may be at a disadvantage when learning a language with a more opaque writing system, where they are faced with the task of acquiring rules governing the print-to-speech correspondences, as well as exceptions to these rules. The problem may be exacerbated when the two given L1 and L2 languages display incongruent mappings between graphemes and phonemes. For example, <H> maps to /h/ in English, but to /n/ in Russian (Hayes-Harb and Barrios 2021:299).

The negative effect of orthography on L2 learners' performance is perhaps most evident in the case of epenthesis. Spelling conventions that did not keep up with historical sound changes or letter combinations that do not adequately reflect phonotactic constraints may naturally prompt the naïve L2 learners to pronounce sounds that are typically associated with such letters or letter combinations, especially if their L1 permits it. For example, Bassetti and Atkinson (2015) investigated the pronunciation of silent letters among native Italian learners of English in such words as *lamb* and *walk*. The participants pronounced the superfluous phone in 85% of the cases in a reading aloud task, and 56% in a word-repetition task. Word-final consonantal epenthesis was shown in the case of Brazilian learners of English, who tended to insert velar consonants in their realisation of the <ng> sequence (Cabanero and Alves 2008). Young-Scholten (reported in Young-Scholten 2002) tested the effect of access to orthographic input during learning and reading tasks of Polish items by English speakers, and observed more vowel epenthesis among the participants in the condition where spelling was available, relative to the picture-only condition. Detey and Nespoulous (2008) investigated the potential effect of orthography on how Japanese learners of French perform a perceptually-driven metaphonological task of syllable segmentation, and observed more cases of vowel epenthesis in the audiovisual and visual conditions than in the auditory condition.

As mentioned earlier, orthography-induced pronunciation errors also arise, perhaps more subtly, when L2 orthographic representations are reinterpreted according to the L1 orthography-phonology interface. Young-Scholten (2000, reported in Young-Scholten 2002) observed that word-final obstruents in German, which are devoiced in this position, were realised as voiced by L1 English speakers, which was ascribed to the likely effect of orthographic input provided (of <b>, <d>, and <g>). English learners of German were also driven by the letter <s> in their word-initial realisation of the voiced fricative /z/ in German as voiceless /s/ in that position, despite having been exposed to native German aural input over the period of twelve months (Young-Scholten and Langer 2015). Vokic (2011) investigated L1 Spanish speakers' productions of the flap sound in American English, a sound which exists in their native repertoire, and which can occur intervocalically in both languages. Despite this, the participants' access to the English flapping rule was blocked as a result of their reliance on Spanish orthography that obscured the English association between the sound with the graphs <t>, <d>, <tt>, and <dd>. Piske et al. (2002) suggested that some of the non-target-like productions of English vowels (/ɪ/, /ɛ/, /ʊ/) by Italian-English bilingual participants might be driven by how their native vowels are encoded in similar orthographic representations (<i>, <e>, <o>). However, the effect was observed only for late bilinguals (not early bilinguals) and for pseudowords (not real words). Studies have also suggested that the interplay between orthographic and phonological representations can emerge in the erroneous judgment of semantic relations between non-native words. Ota et al. (2009) observed that Japanese learners of English tended to semantically associate such words as *key–rock*, due to the confusability between *rock* and *lock*, which is related to Japanese learners' L1 phonology not accounting for the /ɹ/-/l/ contrast.

Despite the evidence suggesting the detrimental effect of orthography, some point to the positive impact of orthographic representations on L2 learners' perception and realisation of target linguistic units. On a general level, orthography can improve word learning performance. For example, Hu (2008) demonstrated that young Chinese learners of English can make use of orthographic information, even if they suffer from poor phonological awareness, possibly as a result of making them more certain of what they are hearing. Escudero et al. (2008) tested the ability of native Dutch proficient learners of English to distinguish between a confusable /ɛ/-/æ/ contrast, and found that while both were mistaken for one another in an auditory condition only, those provided with accompanied orthographic information tended not to confuse the former sound. Erdener and Burnham (2002) observed that the provision of orthographic forms helped Australian and Turkish participants in their Spanish and Irish productions being rated higher by native judges of two latter pair of languages. Escudero and Wanrooij (2010) revealed that Spanish learners of Dutch benefited from spelling when presented with the task of categorising vowels in the latter language. The effect was dependent on the vowel contrast considered, however, and orthography did not help in those cases in which vowels were correctly classified in the auditory-only task. Finally, Bürki et al. (2019) demonstrated some ambivalence of orthographic effects, in that while it improved accuracy and word processing speed during word learning phase among native French learners of English, what was ultimately found was that their formant values were more French-like when the access to orthographic input was provided.

It has also been demonstrated that the effect of orthography may depend on the degree of orthographic depth in the native and the target language. Erdener and Burnham (2005) examined how visual information (from the face) and the level of orthographic depth of the target language (transparent Spanish vs opaque Irish) affect error rate in the production of nonwords in those languages. The participants, who had not had any previous experience in the two languages, were themselves divided into two groups, depending on the orthographic depth as manifested in their native languages (transparent Turkish vs opaque Australian English). In general, the participants performed better in the presence of orthography. More specifically, the study revealed that the provision of orthographic input in the main task (a shadowing task) across all the conditions resulted in native Turkish (transparent) participants making less errors than English (opaque) participants when the target language was Spanish (transparent), but the opposite was true for Irish (opaque). In other words, the presence of spelling was beneficial for speakers of Turkish, but detrimental for Irish, while little difference was found for English speakers across the two foreign languages. One of the conclusions the authors drew was that the representatives of a transparent orthographic system (Turkish) are affected by orthographic information more than those with more opaque systems (English).

3.2.1. Orthographic influences on Polish learners' realisation of L2 vowels

With English ranking relatively high in terms of orthographic depth (Marjou 2021) and Polish manifesting generally transparent sound-to-spelling correspondence (Jassem 1981: 40-41),[7] native Polish learners of English may be considered at a disadvantage. Due to being often exposed to written form of the target language relatively early in their L2 English acquisition, their speech is bound to exhibit not only cross-linguistic interference from sound but also interference from spelling (Sobkowiak 2008: 25). The challenge seems most evident in the case of vowel phonemes, which are encoded in various vowel letter combination in English, to which the Polish speaker is not accustomed to. Still, the problem of interference from spelling remains in a more subtle form even in cases when particular vowels are represented by single letters in both languages. In such cases, Polish learners of English may be suspected of transferring their reliance on orthographical transparency, or more specifically, on given L1 graphemes unambiguously representing specific L1 phonemes, into English, with the consequence that corresponding vowel letters in English words will trigger their L1 phonemes.

For example, because they are so used to the grapheme <i> representing the phoneme /i/ in their native language, they will readily associate the same grapheme in English words with the Polish phoneme /i/. Consequently, in their realisation of the KIT vowel as in *rich*, inexperienced native Polish learners of English, guided by their native spelling conventions, will tend to resort to their native vowel /i/, despite the existence of a better substitute for the English vowel in the form of Polish /ɨ/, represented by different

[7] The relative predictability of pronunciation from spelling does not imply the predictability of spelling from pronunciation, as exemplified by the Polish phoneme /u/, which can be represented by both <u> and <ó>.

grapheme <y> (Sobkowiak 2008: 25; Gonet 2017: 111). Such L1 spelling-driven realisations are also expected in the DERSS vowel, in which case the already hardly perceptible acoustic difference between the English vowel and Polish /ɛ/ is obscured further by both being often represented by <e> in certain contexts. The problem also arises in the case of the schwa vowel, whose multitude of orthographic representations in English are known to guide Polish speakers' realisations of the reduced vowel (Sobkowiak 2008: 137-138). Although it did not concern English, an interesting study comes from Nimz and Khattab (2020), who investigated orthographic influence on Polish speakers' realisation of German vowels. The authors speculate that because both languages utilise the <o> grapheme to represent two different vowel qualities, i.e. /oː/ in German and /ɔ/ in Polish, the resultant German vowel production by Poles deviates from that of the native German speakers. Interestingly, the German vowel /eː/, which was expected to be assimilated into Polish /ɛ/ due to their identical mappings into <e>, exhibited diphthongisation. This was likely the consequence of the learners' integrating both orthographic and perceptual influences, with only the early portion of the vowel being assimilated into Polish /ɛ/, as suggested by <e>.

Another dimension of the reliance on native letters representing foreign sounds is also illustrated in the way some English borrowings are spelt in Polish. Weckwerth (2010: 546) notes that the choice of vowel letters in Polish words such as *skan* (*scan*) or *flesz* (*flash lamp*) stem from the ambiguity involved in the realisation of the TRAP vowel, which is situated roughly between Polish /ɛ/ and /a/ (usually mapped onto <e> and <a> in Polish, respectively). Another evidence of placing perhaps too much trust in orthographic information is found in some English-Polish dictionaries that make use of orthographic transcription, with such words as *ship* being transcribed as *szyp*, i.e. by means of Polish letters that stand for Polish sounds that are considered to be similar to the English ones. Bryła-Cruz (2022: 133) warns that such transcription may intensify the interference between and L1 and L2, and may prevent learners from recognising patterns and rules governing L2 phonology.

What should also be considered is whether or not orthography influences the temporal organisation of vowels in the second language. As mentioned in Chapter 2, vowel duration is not contrastive in Polish, while it is in languages such as English. While some less naïve learners may develop the intuition that the vowel coded by means of certain diagraphs such as <ea> (e.g. in *beat*)[8] tends to be longer relative to the one hidden behind just single graphs like <i> in similar contexts (e.g. in *bit*), on the whole, they are left with resorting to guesswork when it comes to decoding durational specification of vowels merely from spelling. The effect may evidently be different across languages, however, as Nimz and Khattab (2020) found that Polish learners of German were actually aided by orthography, more specifically by the lengthening effect of <h>, in their realisation of long vowels in German. When it comes to vowel duration contrast serving as a cue to the voicing of the following coda consonant, as in *beat*-*bead*, the contrast seems to be more predictable from spelling because of the unambiguous information in graphemic contrasts, such as <t> versus <d>. However, it appears that those learners that do recognise this durational distinction in the vowel do so only intuitively, without relying on that orthographic representation of the final coda in an explicit manner.

Finally, it remains an open question whether or not the perception or realisation of vowel dynamics is affected by orthography. It can only be speculated that spelling, due to its relative constancy, might contribute to the impression that vowels have uniform qualities, particularly when they are encoded by single graphs. This may become particularly evident in the case of a speaker of Polish, a language with reportedly rather stable formant patterns, who may transfer the associations between stable vowels and their corresponding vowel letters from L1 Polish into L2 English.

3.3. Summary

This brief chapter was intended to outline the issue of orthographic influence on the realisation of L2 speech, with particular focus on Polish L2 learners. Some researchers have suggested that detailed speech traces can be activated not only when language is heard, but also when it is seen, in the form of orthographic representations. Listeners may therefore integrate auditory and visual (orthographic) cues, stored as joint representations, resulting in a bidirectional flow of activation between orthography and phonology. Such an effect may not only be apparent in reading aloud, where the effect of orthography should be the strongest, but it may also spill over into spontaneous speech.

Second language learners coming from linguistic backgrounds where spelling-to-pronunciation correspondence is relatively straightforward may be negatively impacted

[8] Although the <ea> grapheme is not very consistent in terms of what vowel it encodes, which is exemplified by such words as *beat*, *bread*, and *break*.

in their acquisition of speech features of a foreign language that exhibits substantial orthographic depth. The challenge is multifaceted as it not only involves acquiring the rules behind the print-to-speech correspondences, but it also entails dealing with potential unpredictability in the cases which the rules fail to account for. What complicates matters further is that learners who are used to straightforward orthographic systems may transfer their grapheme-to-phoneme associations into their L2 speech. Such associations may be evoked in the minds of Polish learners of English when faced with English orthographic input, which in turn may lead to inadequate perception and production of L2 speech features, vowel qualities in particular. Despite all of the above, it is acknowledged that, while orthography can hinder L2 sound acquisition, some evidence points to its beneficial effects. The difficulties discussed here may have implications in the context of phonetic imitation, which is addressed in Chapter 4.

Chapter 4

Phonetic imitation of English vowels by Polish young and adult learners – an experimental study

The preceding chapters have demonstrated that imitation, including phonetic imitation or convergence, has received researchers' considerable attention. However, even though much research has been done, there remain a number of unexplored venues for investigation, particularly in the case of phonetic imitation in the context of L2 speech acquisition. The current study offers a humble but a noteworthy contribution to our understanding of the phenomenon, by accounting for factors which have either received little attention, or have been evaluated rather superficially, despite their reported importance.

4.1. Study design

4.1.1. Objectives and research questions

The main purpose of the current investigation is to assess the degree to which native Polish learners of English are able to imitate the characteristics of English vowels, including their quality, duration, and spectral variability. As has already been thoroughly discussed, the two languages differ considerably in terms of their vocalic systems to the extent that not a single Polish vowel could be described as having an exact equivalent in English. The main differences lie not only in the vowels' quality or timbre, as manifested by their formant structure, but also in terms of durational contrasts, which are employed in English both in the intrinsic and extrinsic sense, but which are almost non-existent in the Polish language. Polish and English vowels also differ as far as their dynamic specification is concerned, with the former language exhibiting more stable patterns, specifically in the initial vowel portions. In their attempt to speak English, Polish learners of English often resort to their native vowel repertoire or to some form of merged categories of L1-L2 vowels. They also underdeliver in their realisation of durational contrasts, and display less robust formant patterns than those that are typical for English. The two languages constitute a viable example of many pairs of languages, where one is relatively simple, and the other relatively complex in terms of their vowel systems, as described earlier. Consequently, the conclusions drawn from this study will not only be limited to these two languages.

The current study is intended to investigate the aforementioned differences in the context of phonetic imitation – a process that is generally understood as one in which one's speech becomes similar to another's, as a result of auditory exposure to the latter's speech, both in interactive and non-interactive settings. In the specific context concerned, i.e. the context of L2 speech acquisition, where the model speaker is a native L2 speaker, phonetic imitation can be additionally seen as a way of approximating native-like pronunciation, which is brought about by at least temporary abandonment of L1-like speech characteristics and the assumption of more L2-like ones. This of course necessitates perception mechanisms to pick up the model's relevant speech features and their at least partial regeneration by the L2 learner's speech apparatus. Therefore, what may stand in the way of successful imitation in this context, based on established models of L2 perception and on previous research, is the inability to perceive the subtle features differing L1 and L2 sounds, as well as the failure to engage one's articulatory faculties to reproduce the desired timing or quality. The learners' imitative performance is expected to be driven not solely by the instruction to repeat words after the model, but also, at least in some cases, by the desire to sound native-like or at least to improve their pronunciation skills. The effect is not expected to be a lasting one, as research has shown that the faithfulness with which the participants reproduce the models' speech is diminished in delayed shadowing or post-exposure tasks.

4.1.1.1. Vowel properties

As has been stated, the point of the current investigation is to establish whether, and to what degree, non-native speakers of a language can shift, at least temporarily, their native speech habits towards more native-like performance after hearing a native model. More specifically, what is of interest is whether Polish learners of English can assume more English-like vowel properties, absent in their native language, with regard to their quality, durational contrasts and vowel dynamics. With the preceding discussion in mind, the first research question is formulated as follows:

1. To what extent do native Polish learners of English imitate the properties of English FLEECE, KIT, DRESS, and TRAP vowels, specifically with regard to:
   a) vowel duration contrast as a cue to the voicing of the coda consonant,
   b) vowel quality from the static perspective, as manifested by the first and second formant values, located at vowel mid-points,
   c) vowel quality from the dynamic perspective, as manifested by the magnitude of total formant shifts across multiple vowel intervals, along with their rate of change?

As far as durational contrasts are concerned, the issue was already investigated by Zając (2013), Zając and Rojczyk (2014), and Rojczyk et al. (2022), as reported in Chapter 2. Convergence was only tentatively observed in Zając (2013), specifically in the case of the *bit-bid* pair, although the participants generally shortened their vowels after exposure. While, in Zając and Rojczyk (2014), the participants were confirmed to exhibit satisfactory durational contrasts, the picture was somewhat obscured by the fact that the participants already exhibited almost native-like performance in the baseline task. It should also be noted that the two studies considered absolute vowel durations, which may have left the potential effect of speech rate shift (across the baseline and the imitation tasks) unaccounted for. This was remedied in Rojczyk et al. (2022), where word duration shifts were included in the analysis, yielding significant results in favour of successful convergence. Therefore, the current investigation is partly intended to replicate these findings, by asking whether relative vowel duration undergoes phonetic imitation in the case of L1 Polish learners of L2 English. The predictions are in line with the results obtained by Rojczyk et al. (2022), but to varying extents depending on the additional factors incorporated in the current study, i.e. the factors of age (proficiency) and orthography, which are expanded on in the second and the third research questions, respectively. No predictions are made with regard to which particular vowels should elicit the greatest baseline-to-imitation shifts in the realisation of durational contrasts, but the participants are likely to follow the patterns exhibited by the model speaker.

Investigation of the imitation of spectral properties of vowels has largely been conducted outside the scope of second language acquisition. One of the few studies that did address the issue in the context of L2 speech acquisition was by Rojczyk (2013), who confirmed Polish learners' ability to imitate English TRAP. However, our understanding remains incomplete, considering that only one English vowel was investigated, one that has also been found to be one of the most prone to the imitative effect in the context of L1 speech imitation (Babel 2012). What is more, in view of the premises of the Speech Learning Model, the TRAP vowel may be more subject to imitation due to its relatively distinct quality when compared to its closest Polish counterparts. This is in line with the observation that greater initial phonetic distance elicits greater convergence, as a result of there simply being a greater gap for the interactants to bridge (Nycz and Mooney 2017). The question, therefore, is whether other vowels, particularly vowels such as English FLEECE, and DRESS, whose quality is deceptively more similar to the Polish vowels (/i/ and /ɛ/ respectively), would also be prone to phonetic imitation. In line with the previous findings, the TRAP vowel is expected to exhibit the greatest imitative effect, with Polish learners potentially shifting away from the baseline L1 /ɛ/-like or /a/-like productions towards more native TRAP-like values. A significant effect is also expected in the case of KIT, whose baseline realisation is expected to be similar to that of FLEECE, leaving a good distance for the participants to close. The FLEECE vowel itself, as well as DRESS, due to their considerable respective similarity to Polish /i/ and /ɛ/, are not expected to be imitated, but the effect may depend on the specific formant values of the model.

Previous studies on vowel imitation restricted their analyses by assuming a static perspective of vowel quality. The static view of vowel quality posits that almost all information needed to specify vowel quality lies within a short steady portion of the vowel, typically extracted around the vowel's mid-point. However, as has been discussed in Chapter 2, English vowel-phonemes, including those traditionally labelled as monophthongs, have been shown to undergo changes in their spectral characteristics throughout their duration. This dynamic view of vowel quality has been found to play an important role in the recognition of vowel identity among speakers of English. Polish and English represent a pair of languages that differs in terms of how much vowel spectral change is observed, in that Polish vowels are purer (more stable) in terms of their formant structure (Schwartz 2020), particularly in the vowel's initial portions (Schwartz and Kaźmierski 2020). Nor do vowel dynamics seem relevant for vowel perception in Polish (Schwartz et al. 2016b). In light of the above, the current study is meant to answer the question of whether or not greater amount of spectral change, typical for English, undergoes imitation in the case of Polish learners. What is taken into account are both total formant movements over the course of the major portion of the vowel, as well as the movements in particular vowel sections. The study also considers the rate at which these movement take place. It is tentatively assumed that the participants will shift their more stable patterns towards more dynamic ones in the imitation task, at least in the initial portion of the vowel. It seems uncertain if there are any particular vowels that may elicit relatively greater imitative effect in this respect. However, it seems reasonable to expect the greatest effect to take place in those cases where the participant-to-model distances are most pronounced.

4.1.1.2. The effect of orthography

The current investigation also aims to determine the effect of orthography in the process of phonetic imitation. Imitation experiments tend to use orthographic representations of the stimuli, and no study seems to have assessed the potential effect of this way of stimuli presentation on the degree of imitation. As mentioned in Chapter 3, detailed L1 traces formed through speech perception may be activated in the process of written word recognition (Dufour and Nguyen 2013), and word presentation may co-activate both graphemic and phonemic representations (Ziegler and Ferrand 1998). Moreover, L2 learners may strategically disregard the acoustic layer of the perceived input, by relaying largely on its orthographic form (Bassetti 2024). It may be suspected, therefore, that the presentation of the L2 stimuli in their orthographic forms in the imitation task may induce the learners to persist in relying on their L1 grapheme-to-phoneme associations, leading to L1-like productions. These associations are expected to be triggered once a given written word is seen, right before the acoustic input from the model speaker is heard, resulting in the former overriding the potential effect of latter. With this in mind, the second research question is:

2. Is the degree of imitation affected by the presence versus absence of orthographic input in the imitation task?

As was reported in Chapter 3, there is evidence that a number of pronunciation errors in L2 speech are orthography-induced, particularly when the target language exhibits greater orthographic depth than the learners' L1. The influence of spelling may be subtle and interfere with the way second language learners perceive and produce foreign consonants and vowels, i.e. according to their native language spelling conventions. Because Polish is characterised by a relatively strong correspondence between its spoken and its written form, native Polish learners of English may be inclined to transfer this correspondence when pronouncing English vowels, especially when the speech is read out. For example, because they are naturally used to the fact that the letter <i> represents the sound /i/ in their native language, Poles will also associate the same letter in English words, such as in *live*, with the Polish vowel /i/, which is similar to FLEECE in terms of quality, resulting in the vowel KIT being replaced with Polish /i/. In an imitation task, therefore, it is predicted that learners with access to orthographic input will be more prone to the retention of Polish /i/-like quality in the case of KIT tokens. Similarly, in cases where English DRESS is encoded by a single graph <e>, as is often the case in Polish /ɛ/, access to orthographic input of words such as *bet* is likely to trigger more Polish /ɛ/-like productions in the imitation task. As for TRAP, although it is generally assimilated by either Polish /ɛ/ or /a/, the presence of spelling in the form of the <a> graph, might induce the latter option to be more common, since the latter Polish vowel is encoded by <a>. No clear predictions are made as to FLEECE, whose written representation, such as <ee> or <ea>, although initially misleading for the Polish learners, tend not be realised as sequences of vowels /ɛ/-/ɛ/ or /ɛ/-/a/, respectively, but rather by means of the Polish /i/ vowel.

The predictions are also less certain when it comes to how access to orthographic input will influence imitative performance in the case of vowel duration contrasts and vowel dynamics. As regards durational contrasts, such as those found in *bet*-*bed* pairs, there does not appear to be a strong reason to suspect printed word of suppressing imitative performance, as the relevant temporal information seems concealed enough anyway, apart from <t> and <d> graphs indirectly pointing to the contrast. However, it can be tentatively conjectured that the most naïve Polish learners will be more likely to ignore such contrasts if they see the words in the imitation task. Having been made thus (visually) certain what words are the ones to be imitated, the participants may discard the acoustic layer of the input, and realise words such as *bed* in a Polish-like manner, i.e. without noticeable durational contrast relative to words such as *bet*. When it comes to vowel dynamics, it can only be conjectured that orthographic input may induce more Polish-like, i.e. stable formant patterns. As shown earlier, vowel letters may guide the realisation of vowel quality by L2 learners, particularly when their L1's orthography is more transparent than that of the L2's. Perhaps, it is also the case that letters, because of their rather unchanging nature, give the impression that sounds, vowels in particular, that are represented by those letters are also constant.

In sum, while imitation is expected to take place regardless of whether the learners see the words in the imitation task, it is predicted that access to orthographic input in the task is likely to induce more Polish-like vowel realisation due the their rather strong reliance on grapheme-to-phoneme straightforward correspondences being transferred from Polish into English. On the other hand, those participants who will have to rely solely on the auditory modality are expected to be more likely to deliver more model-like productions, particularly in the case of vowel quality. However, what should not be ignored is that orthographic input may have its benefits, and for some learners, the reliance only on what they hear may actually impede their performance by making them less certain what it actually is that they hear and what they are supposed to repeat. On the other hand, it is also to be noted that all participants are exposed to the written representations of the words in the baseline task, but this presents another potential limitation, which renders the potential orthographic influence not entirely avoidable.

4.1.1.3. The factor of age and L2 proficiency

Finally, to gain a richer perspective, the factor of age, together with L2 proficiency, is taken into account by including both child (teenage) and adult learners of English. The majority of studies on phonetic imitation have focused on adult speakers, not infrequently university students, neglecting younger age groups. Even fewer studies have considered juxtaposing the imitative performance between the two age groups, and those that have have yielded variable results, as was discussed in Chapter 1. Indeed, when we narrow the scope of investigation to the context of L2 speech imitation, the scarcity of research becomes even more evident, which is surprising, considering the undisputed effect of age on speech perception and production in the second language. With what little research that has been done on the effect of age on imitative performance, there do not emerge definite conclusions, with some authors pointing to the advantageous position of younger learners, and others to that of adults. While the intention behind the current study is not to explicitly compare the performance of two groups, it is hoped that it will contribute to our understanding of age difference with regard to imitation, by posing the third and final research question:

3. How does the factor of age, together with L2 proficiency, affect the degree of imitative performance?

Some of the studies reviewed in Chapter 1 seem to suggest that the potential imitative advantage of young learners may lie in their lesser predilection for categorisation, as a result of having acquired a fewer number of exemplars relative to adults (Nielsen 2014), while some others point the deterioration of articulatory flexibility in the case of older learners (Cochrane and Sachs 1979). Conversely, adults have been judged as superior imitators, potentially due to their imitation skill having been sharpened with age and phonetic experience (Alivuotila et al. 2007). As remarked in Chapter 1, both age and immersion may jointly influence imitative performance in L2 (Jia et al. 2006), and there may be a slight imitative advantage in the case of more advanced learners (Burin 2018). The issue of L2 experience and proficiency is difficult to disentangle from the factor of age, with there not being many truly naïve adult learners of English, particularly among Poles.[9] In the current investigation, therefore, the factor of age is rather inextricably considered together with the factor of L2 proficiency or L2 experience.

In the current investigation, it is predicted that younger (and less proficient) learners will exhibit seemingly greater imitative performance not so much because of their potentially greater perceptual and productional flexibility relative to adults, but primarily because of their lesser L2 proficiency, likely to lead to worse pre-exposure performance. Initially, this reason seems to run contrary to what was mentioned earlier, that is that more proficiency should lead to more robust imitation. This observation, however, does not seem to align with the fact that proficient L2 learners, whose L2 speech should, presumably, already exhibit at least a certain degree of native-likeness, resulting in there not being so much left to imitate. In other words, if imitation is understood as bridging the participant-to-model gap,[10] which is expected to be greater in the case of naïve learners, then it seems natural that greater imitation will be observed in such learners, given that greater initial phonetic distance elicits greater convergence (Nycz and Mooney 2017). However, it should be acknowledged that it may be considered controversial whether to consider imitation solely in terms of the degree of bridging the said distance, and that imitation may still take place even when the initial distance in small (Macleod 2021).

In summary, the final factor that is loosely accounted for in the current investigation is that of age, which is inextricably connected with L2 experience or proficiency. It is naturally expected that, because more experienced adults will likely exhibit more native-like baseline vowel properties, the initial participant-to-model distances will be shorter than those of less experienced (younger) learners, leaving more room for imitation to the latter group. It does not seem reasonable at this stage, however, to consider such a scenario as evidence pointing to a lesser degree of imitation in the case of adults. Therefore, the primary purpose behind the inclusion of the factors of age (and proficiency) is not so much to explicitly compare the imitative performance between those groups, but rather to broaden the perspective of the current research, and to not limit the study's conclusions to just one group of participants.

[9] Unless we consider late adult learners, who may have had limited exposure to either formal or naturalistic opportunity to acquire English.
[10] Where the model is a native speaker of the target language.

In sum, the current investigation aims to enrich our understanding of the process of phonetic convergence by answering the following questions:

1. To what extent do native Polish learners of English imitate the properties of English FLEECE, KIT, DRESS, and TRAP vowels, specifically with regard to:
   a) vowel duration contrast as a cue to the voicing of the coda consonant,
   b) vowel quality from the static perspective, as manifested by the first and second formant values, located at vowel mid-points,
   c) vowel quality from the dynamic perspective, as manifested by the magnitude of total formant shifts across multiple vowel intervals, along with their rate of change?
2. Is the degree of imitation affected by the presence versus absence of orthographic input in the imitation task?
3. How does the factor of age, together with L2 proficiency, affect the degree of imitative performance?

4.1.2. Stimuli

The stimuli used in the experiments that are the core of data analysis included eight English monosyllable words listed in Table 4.1. The words were divided into two equally sized groups, the first of which included a voiceless consonant /t/, while the second a voiced consonant /d/ as codas. The words were selected so that they included the four Standard British English front vowels: FLEECE, KIT, DRESS, and TRAP. The reason for focusing on front vowels lay in the fact that the first two formants extracted from them can be reliably distinguished,[11] which may be particularly important for the extraction of formant values at multiple points throughout the vowels' duration. Moreover, as discussed in Chapter 2, although the front vowels are by no means the only ones posing problems to Polish learners of English, they have been frequently reported as such, making their selection doubly warranted.

vowel	voiceless coda	voiced coda
FLEECE /i:/	beat	bead
KIT /ɪ/	bit	bid
DRESS /e/	bet	bed
TRAP /æ/	bat	bad

Table 4.1. The stimuli used in the experiment

As can be seen, the words were paired so that each of the four vowels was followed by both fortis and lenis consonants. To limit the coarticulatory effect of neighbouring

[11] As Gonet (2017: 109) remarks, "spectrograms constitute a good analytic tool for studying vowel qualities in front and central vowels, but fail to do so in back vowels."

consonants on vowel quality (Cole et al. 2010), as well as to obtain more reliable duration measures (Rojczyk and Porzuczek 2012), the vowels were not flanked by continuants, such as approximants, liquids or nasals. The onset consonant /h/, commonly used for experimental stimuli (Peterson and Barney 1952; Hillenbrand 1995) due to its relatively neutral realisation in English (Porzuczek et al. 2013), was avoided due to its more consonantal nature in the Polish language, which many participants would undoubtedly have adopted. With such a restricted choice, following Rojczyk (2010a), the context /b_t/ was chosen, along with its voiced coda counterpart /b_d/. Other stop consonants were not included in order to preserve a stable consonantal environment for the extraction of formant patterns. To make up for the scarcity of words meant for the imitation experiment, each of them was meant to be produced four times, leaving thirty-two data points coming from a single participant in one task. In addition, seven distractors were included: *pet*, *tell*, *cake*, *desk*, *mother*, *yes*, *book*, which were also to be imitated multiple times in one task, although not an equal number of times each. Another factor determining the word selection was their relative ease of pronunciation, which was particularly important in the case of the young (less proficient) participants. The words were not surrounded by carrier phrases in order to avoid imposing undue burden on the participants, and extending the recording sessions to the point where the measurements obtained might have become compromised due to the participants' fatigue.

All fifteen words (including seven distractors) were pre-recorded by four professional independent voiceover artists, who were native British English speakers. The model speakers were instructed to read out the words in isolation, as naturally as possible, using the falling intonation pattern, much like how the recordings for foreign language course books are made. The choice of the British English variety stems from the fact that it is frequently recognised as standard in EFL courses in Poland. From among the four models, only one was chosen for the final model stimuli (a female actor), so as not to overcomplicate data analysis, which, it might be supposed, would have otherwise yielded more undesired variability in the participants' productions. Choosing more than one model would have also required some form of normalisation of the model tokens, which would have been likely possible only for the temporal aspect of the stimuli. What is more, the intention was to expose the participants to the stimuli that were produced relatively naturally, without having to resort to any form of acoustic manipulation. The reasoning behind choosing that particular model was two-fold. First, this model displayed rather intermediate values in terms of vowel duration contrasts and vowel dynamics relative to the other models, and it was desired not to expose the participants to too extreme values. While too small model values might not have elicited any imitative effect, too high values might have been too conspicuous, potentially rendering the imitative effect almost certain, and it was clearly not the intention to obtain significant results that would stem solely from exorbitant model values. Figure 4.1 illustrates two sample vowel durational contrasts of the chosen model for the *bet-bed* (top) and *bat-bad* (bottom) pairs, showing that although some contrasts were rather conspicuous, as in *bat-bad*, they were not so obvious in other cases, as in *bet-bed*.

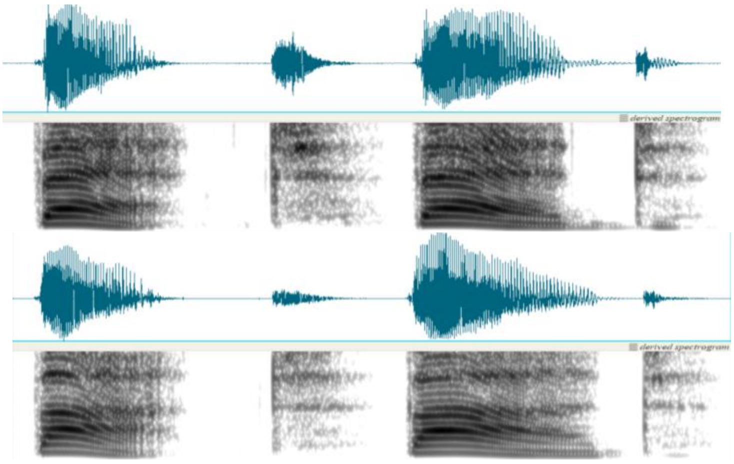

Figure 4.1. Sample vowel durational contrasts of the chosen model: *bet-bed* (top) and *bat-bad* (bottom)

At the same time, the model was also subjectively judged to sound quite friendly relative to the others, which may be important in the face of evidence that voices that induce positive attitude are more likely to be imitated (e.g. Yu et al. 2013). The chosen model's values for each vowel and for each tested parameter are included in Appendices 1–3, together with the participants' mean (and SD) values.

4.1.3. Participants

A total number of 60 native Polish participants (6 male, 54 female) agreed to take part in the experiment, half of whom involved primary school students, and the other half included university students. The first group were 30 young learners (4 male, 26 female), aged 12-13 (M = 12.7; SD = 0.45), recruited and recorded at Primary School Nr 5 with Sports Sections in Chorzów, Poland. Being in their seventh school year, all of them had learnt English for at least six years, and they displayed a rather uniform proficiency level in English corresponding to the A2 level according to the Common European Framework of Reference for Languages, as evidenced by their grades obtained through the administration of both oral and written tests. The second group consisted of 30 second-year students (2 male, 28 female), aged 20-23 (M = 21.1; SD = 0.78), at the University of Silesia in Katowice, Poland. Their estimated proficiency was approximately C1 or above on the CEFR scale, and, having attended classes at the Institute of English, they had completed a three-semester course in practical English phonetics, and had therefore acquired some rudimentary knowledge of the differences between Polish and English speech sounds. Clearly therefore, the two groups of participants differed not only in terms of age, but also in terms of their language proficiency. It was ascertained that none of the participants showed any signs of speech or hearing disorders.

4.1.4. Procedure

The recording sessions took place in a quiet room at the respective premises where the participants had been recruited, and the procedure was exactly the same for both groups. Each participant was recorded individually and each recording session took approximately twenty minutes. They were instructed on what they were supposed to do in Polish (to ensure understanding in the case of the younger group), and were comfortably seated in front of a laptop computer. They were assured that their performance would not be graded, and that, regardless of their performance, they would provide a valuable scientific input. This was particularly important from the perspective of ethical concerns in the case of the young participants, who tended to be more concerned with how well they would do in the experiment. The participants were asked a few basic questions in English and were then asked to read eight simple English training words, in order to activate the desired language mode and to better familiarise them with the procedure. Upon this, they were presented with two tasks: a baseline word reading task and an imitation task (see Figure 4.2).

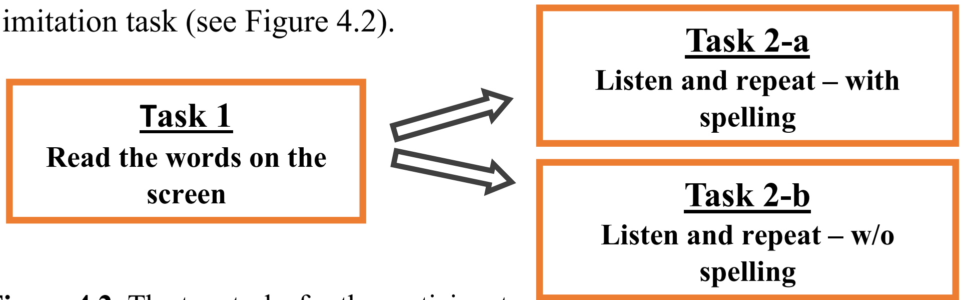

Figure 4.2. The two tasks for the participants

The point of the first task was to establish the participants' reference values for the tested parameters. They were asked to read orthographic representations of 54 English words described earlier (32 experimental stimuli and 22 distractors). They were told to read them the way that seemed natural to them. Each word appeared on a separate Power Point slide, in a white Times New Roman font, size 96, centred on a black background (see Figure 4.3). The slides had been randomly shuffled for each participant prior to the experiment. The procedure was self-paced and the experimenter (the author) proceeded to the next slide when the participant finished reading a word, making sure not to rush them. In rare cases when the participants, particularly the young ones, grossly mispronounced a word, they were asked to give it another try.[12]

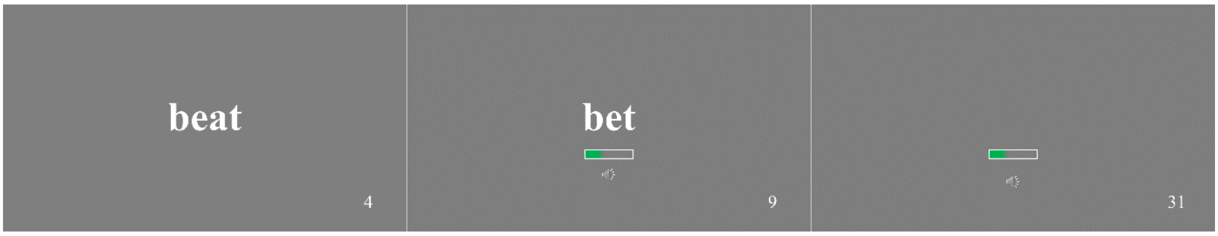

Figure 4.3. The presentation of the stimuli in the baseline task (on the left), the imitation task with spelling (in the middle), and the imitation task without spelling (on the right)[13]

The second task was an imitation task, which had two versions. For this reason, the participants from both age groups were further divided into two equal subgroups, resulting in four subgroups of 15 participants in each subgroup. In one of the versions of the imitation task, 15 young and 15 adult participants were presented with the same words appearing separately on the screen in a different (random) order. To prepare the participants for the moment when the model stimuli would be played, a green progress bar marking the interval of 750 ms for each slide had been introduced. Once the progress bar was filled, the participants heard the model produce the word, upon which they were instructed to wait for the loudspeaker icon to disappear, which happened 500 ms after the model was done saying the word, and to simply repeat what they heard. In the second version of the task, two other subgroups of 15 young and 15 adult participants underwent a very similar procedure with the exception that the words to be imitated were not displayed on the screen. See Figure 4.3 for sample presentation of the stimuli for the two versions of the imitation task. Similarly to the first task, the order of the presentation of the words was randomised for each participant, and the interstimulus interval was self paced, but it did not deviate significantly across participants.

[12] The only problems that recurred were with the vowel in the words *beat* and *bead*, and it was deemed justifiable to help the participants by telling them that the target words rhymed with the words they clearly knew, i.e. *feet* and *read*, respectively.

[13] The background colour of the slides was originally black.

4.1.5. Recording equipment

The participants' productions were captured using Samson Q2U dynamic microphone at the sampling rate of 48 kHz and 16-bit quantization. The choice of a dynamic microphone was motivated by the need to filter out occasional background noises, which are clearly unavoidable on the premises of a primary school even during class time. The participants were instructed to keep relatively still, leaving an approximately 10-centimetre distance between the microphone and the mouth. The microphone was placed to the side of the participant and was fitted with a pop-filter. The stimuli to be imitated were presented aurally through PreSonus Eris E3.5 studio monitors. The studio monitors were chosen instead of ordinary loud speakers for the sake of preserving the naturalness of auditory stimuli, by delivering relatively flat frequency response. The stimuli were played at a comfortable, roughly identical loudness levels. Headphones were not used because it would have been more difficult to ensure the stability of both loudness and comfort to the participants. Both audio devices were connected to the 3rd Focusrite Scarlett audio interface.

4.1.6. Data Analysis

A total number of 3840 relevant words (60 participants × 2 tasks × 8 words × 4 repetitions) were obtained from the participants. All acoustic measurements were made based on waveform and spectrogram displays in the Praat software (version 6.4.11, Boersma and Weenink 2024). To create TextGrid files, two interval tiers were created: one for vowel and the other for word duration measurements. All recorded items of interest were inspected both visually and auditorily, and both vowel duration and word duration were manually marked on the TextGrid tiers. Based on the inspection, 62 items were excluded from subsequent analysis. These included items that were either grossly mispronounced or items where vowel duration was impossible to measure due to excessively fuzzy transitions between the vowels and the following coda consonants. Additionally, where appropriate, extreme outliers were removed.[14] Further discarded data points are

[14] The removal of extreme outliers was motivated by the need to preserve the reliability of the statistical models used. These might be attributed to the participants' unnatural (accidental) productions. The removed items were not judged to convey any relevant information about the variability inherent in the study area, and their inclusion (these being just a handful of cases) would not have altered the obtained results. A modified (stricter) Interquartile Range (IQR) method was used to detect extreme outliers, i.e. data points that were less than (Q1 – 2.5 × IQR) or greater than (Q3 + 2.5 × IQR).

mentioned in the subsequent subsections, devoted to the specific parameters tested, therefore the number of words included in the analyses varied slightly depending on the parameter.[15] Having marked all vowel and word duration intervals, a Praat script, written based on the guide by Stanley and Lipani (2019), was run to extract the appropriate temporal data (vowel and word duration), and spectral data (F_1 and F_2 values at 5 points of the vowels' duration). All extracted acoustic measures underwent statistical analyses to determine potential differences across the two tasks in the two age groups, who are henceforth referred to as the YL (Young Learners) group and the AL (Adult Learners) group. Although the model values are provided, the tested parameters are not evaluated in terms of participant-to-model distances, but rather in terms of baseline-to-imitation distances for given participants.[16] Data cleaning, exploration, visualisation and analysis, as well as all statistical computations were conducted using the R programming language (version R-4.4.0, R Core Team 2024) in the RStudio environment (version 2024.04.2+764, RStudio Team 2024), supported primarily by the tidyverse package (Wickham et al. 2019).

The initial evaluation of the effects of the tested parameters was made based on various plots, most of which show arithmetic mean values (with standard deviations) calculated for given predictors, separately for the two age groups. To ascertain statistical significance of the predictors, linear mixed-effects models (LME, Winter 2020) were used for each of the two age groups separately,[17] and for each of the tested parameters as response variables, related to: extrinsic vowel duration, static vowel quality, and VISC. The assumptions[18] for mixed-effects linear models were verified using the `check_model()` function from the Performance package (Lüdecke et al. 2021). The effect of `task`[19] on a given parameter was evaluated through a series of likelihood ratio tests (LRT, from lme4 package, Bates et al. 2015), which compare reduced models (with

[15] For example, the cases where vowel formants were not measurable were still included in the analysis of vowel duration for the sake of preserving the highest number of data points suitable for analysis of a given parameter.
[16] This is mainly motivated by the risk of obtaining questionable results, especially when evaluating formant values. It seems much safer to compare speakers with themselves before and after imitative exposure to ascertain potential shifts, if only to preserve variation due to anatomical factors. The issue is further discussed in the sections related to vowel quality.
[17] The statistical models were fitted for the YL and AL groups separately in order not to overfit the models, which might have undermined their reliability and interpretation.
[18] The assumptions included: the independence of observations, linear relationship between the fixed effects and the dependent variable, normality of residuals, homogeneity of variance of the residuals (homoscedasticity), lack of multicollinearity, and normality of random effects.
[19] When the predictors are mentioned in the context of LME, Courier New font is used to highlight them. The same font is used for all elements that represent either the input or the output of the R code.

fewer or no predictors) to more specified models (with more predictors). Additionally, `compare_performance()` function from the Performance package was used to compare the models' performance metrics, such as AIC and BIC. In the current study, the sequentially added predictors (fixed effects), included `task`, `vowel` type, and `spelling` condition, as well as their interaction. To account for by-participant variability, the random effects structure included random intercepts and slopes for `subject`.[20] The overall structure of the models is provided in Table 4.2. Detailed model summaries[21] are provided in Appendices 4–8. Pairwise comparisons were performed, using the emmeans package (Lenth 2024), with the Bonferroni correction method applied.

Response variables
`RVDD` (relative vowel duration difference)
`f1`, `f2` (formant values at vowel mid-point)
`TL` (trajectory length)
`roc` (spectral rate of change)

Fixed effects with levels	
`task`	baseline, imitation
`vowel`	FLEECE, KIT, DRESS, TRAP
`spelling`	yes, no

Random effects
`subject` (intercept)
`subject` (slope)

Table 4.2. Summary of response variables, fixed effects (with levels), and random effects

4.2. Tested parameter 1: Relative vowel duration difference

As described in more detail in Chapter 2, the two languages differ in terms of the temporal organisation of their vowels, with Polish not showing much variability in that regard, as opposed to English. Polish learners of English, therefore, are expected to underdeliver in their realisation of vowel duration contrasts, particularly when such contrasts signal the voicing status of the final coda consonant. The current investigation is intended to answer the question whether exposure to a native English model in an imitation task leads to an improvement in the realisation of this durational distinction.

[20] The random effects structure does not include `word` or `word pair`, since the by-item variation is already captured by `vowel`.
[21] These were generated using `report_table()` function from the report package (Makowski et al. 2023).

4.2.1. Measurements

Vowel duration measurements were done manually, based on waveform and spectrographic displays. Standard measurement criteria were used, in that vowel duration encompassed the interval between the onset and the offset of periodicity, which was accompanied by clear formant structure in the spectrogram (e.g. Fox and Jacewicz 2009). Because absolute vowel duration values are clearly dependent on speech rate (e.g. Gopal 1990), comparing such values across the two tasks, with the participants potentially shifting their speech rate, if only as a result of exposure to the model speaker, might produce unreliable results. To at least partially remedy this, word duration was also measured, from the onset of the release of /b/[22] to the cessation of noise accompanying /t/ or /d/. See Figure 4.4 for sample vowel (green) and word duration (red) intervals.

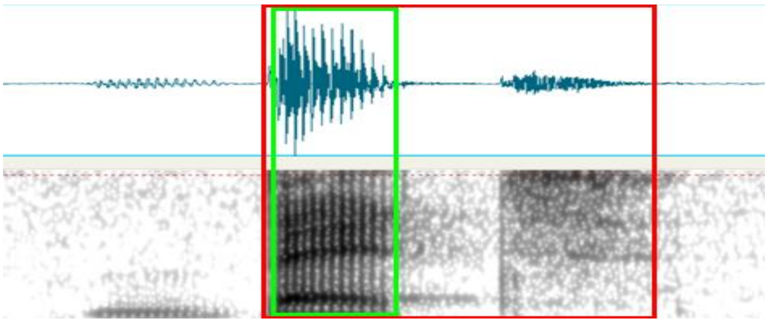

Figure 4.4. Sample vowel duration (green) and word duration (red) intervals

Based on the two measurements, relative vowel duration was calculated by dividing absolute vowel duration by absolute word duration for each word, which resulted in a percentage or a proportion of a word that its vowel occupied. Afterward, a given participant's words (in a given task) were paired so that the vowels in them matched but coda voicing differed, resulting in the following pairings: *beat-bead, bit-bid, bet-bed,* and *bat-bad*. Since the participants had been asked to produce four instances of each word, the words were paired in the order that they were produced by them, so, for example, the first *beat* a given participant produced was matched with their first *bead* in a given task. Finally, relative vowel duration difference (henceforth referred to as RVDD), the first tested parameter, was calculated by simply subtracting the relative vowel duration in a word with a voiceless coda from the relative vowel duration of the matched word with a voiced coda. For example, Figure 4.5 shows relative vowel durations of a sample word

[22] Closure phase for /b/ was not included in the measurement of word duration for the sake of measurement consistency across the tokens, as they varied with respect to the presence versus absence of prevoicing.

pair *bet-bed*, amounting to roughly 40% and 60%, respectively. RVDD for this pair, therefore, equals to 0.60 – 0.40 = 0.20 (or 20 percentage points). It is to be observed that this single metric attempts to capture the participants' vowel duration contrast (stemming from the coda voicing contrast), without ignoring the potential factor of speech rate. The core of this analysis lay in whether this parameter differed significantly across the two tasks.

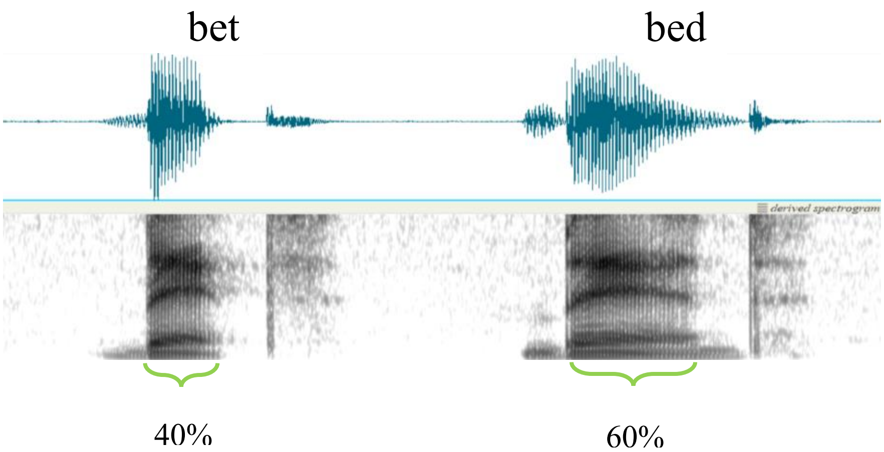

Figure 4.5. Sample word pair *bet-bet*, with their approximate relative vowel durations marked

4.2.2. Analysis and results

Because 62 tokens were initially excluded from the analysis, some tokens missed their pairs, in which cases the corresponding tokens were discarded as well. Additionally, 5 pairs in the YL group and 3 pairs in the AL group[23] were flagged as extreme outliers, and were therefore removed, as justified earlier, leaving a total of 1853 word pairs (938 for the YL and 915 for the AL group) for the analysis of durational contrasts. Figure 4.6 shows the YL (left)[24] and AL (right) groups' distributions of RVDD across the two tasks, with orange curves indicating baseline responses and blue ones indicating imitated responses, and the black vertical line marking the point at which the vowel length was not differentiated.

[23] Recall: YL – Young Learners, AL – Adult Learners.
[24] From now on, the pairs of plots comparing the two groups are shown next to each other, with the plot referring to the YL group always being on the left.

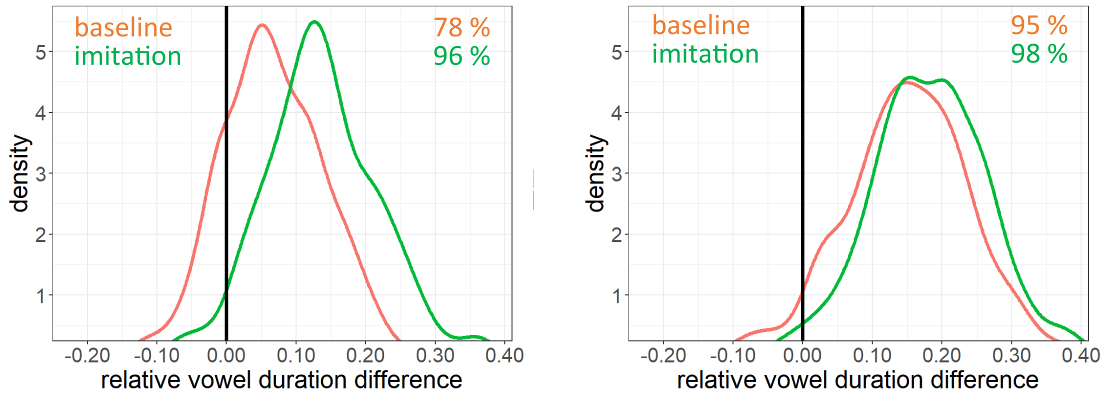

Figure 4.6. Distributions of the RVDD data points produced by the YL group (right) and the AL group (left) in the two tasks: baseline (orange) and imitation (blue)

It can be observed that the YL group's distribution shifted visibly after imitative exposure, while the AL group's shift was more modest. However, the ALs' baseline curve was already shifted relative to the YLs' baseline curve. It is also to be observed that the major part of the baseline curve was on the right of the borderline (black vertical line), showing clear vowel length differentiation prior to exposure, particularly in the case of the AL group. The distributions also show that YL's productions were more concentrated in both tasks than those of the latter group. The position of the curves agrees with the participants' high proportion of positive responses, i.e. cases where relative vowel duration in <bVd> words exceeded that in <bVt> words, with YL showing an increase from 78% to 96%, and AL group from 95% to 98%, across the two tasks.

The above observations are confirmed by mean RVDD values, shown in Figure 4.7.

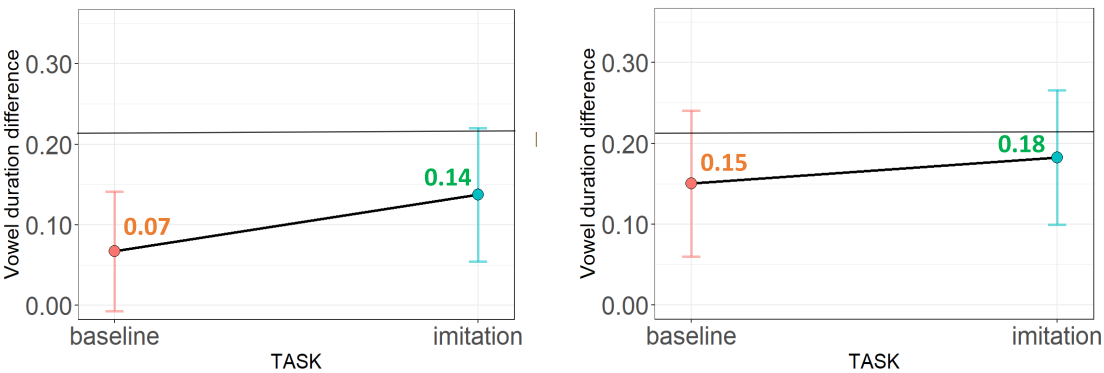

Figure 4.7. Mean values of (relative) vowel duration differences (with standard deviations) for the YL group (left) and the AL group (right) in the two tasks, with the model value (horizontal line)

YLs approximately doubled their mean vowel duration contrast, from 0.07 to 0.14,[25] while ALs' already high baseline value of 0.15 increased to 0.18. The dark grey horizontal line represents the model speaker's mean value of approximately 0.22, which left both groups, the YL group in particular, with sufficient participant-to-model distance to elicit imitative effect.

Figure 4.8 shows the participants' relative vowel durations (not differences), dependent on the voicing status of the coda consonant and the task. This helps visualise the fact that, predictably, the increases in vowel duration differences described above were due to positive vowel duration shifts in <bVd> words across the two tasks (blues lines), from 0.48[26] to 0.55 for YL, and from 0.54 to 0.58 for AL. Vowel duration in <bVt> remained virtually unchanged in the case of both groups, with an average of 0.40.

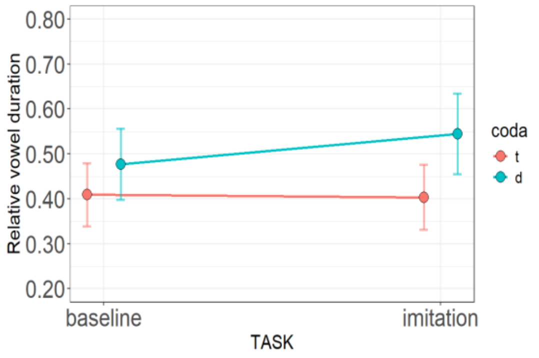

Figure 4.8. Mean relative vowel duration (with SD) for the two coda consonants /t/ (orange) and /d/ (blue) in the two tasks for the YL group (left) and the AL group (right)

Figure 4.9 reveals deeper patterns present in the data, by considering the shifts of vowel duration contrasts across the two tasks for each of the four vowels.

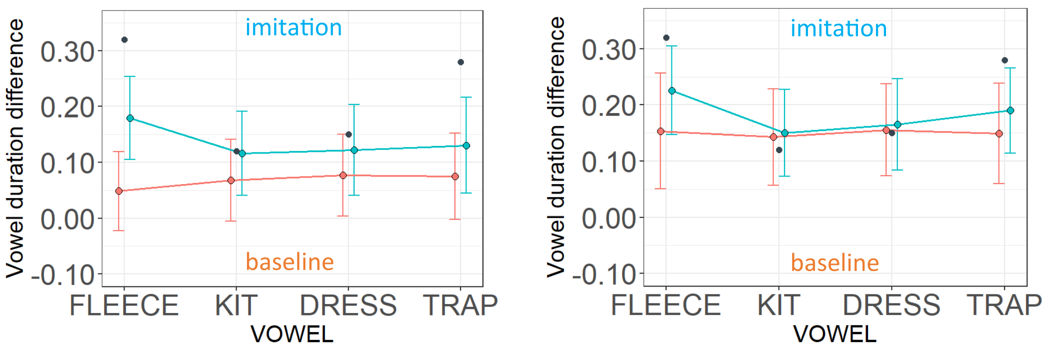

Figure 4.9. Mean values of RVDD (with SD) for the four vowels, for YLs (left) and the ALs (right) in the two tasks, with model values (dark points)

[25] That is from 7 to 14 percentage points (pp).

[26] Recall: 0.48 means that, on average, a given vowel occupied almost half of the word's duration.

The YL group displayed RVDD increases from the baseline (orange) to the imitation (blue) task for all vowels, while the AL group's already relatively high baseline values made observable shifts only in the case of FLEECE and TRAP vowels. The degree of the shifts seems to correspond to the model speaker's mean values (represented by dark points), which were the highest exactly for these two vowels. In the YL group, the model values were consistently higher than participants' baseline values, and the participants seemed to approximately match them for KIT and DRESS after exposure, but they were still far from reaching the model in the case of FLEECE and TRAP. Surprisingly, the young participants' TRAP shifted rather modestly, considering the baseline-to-model distance. In the AL group, the model values, being approximately the same as those of the participants', elicited no shifts for KIT and DRESS in the imitation task.

Figure 4.10 includes the factor of presence versus absence of orthographic representations of the stimuli presented to the participants in the imitation task.

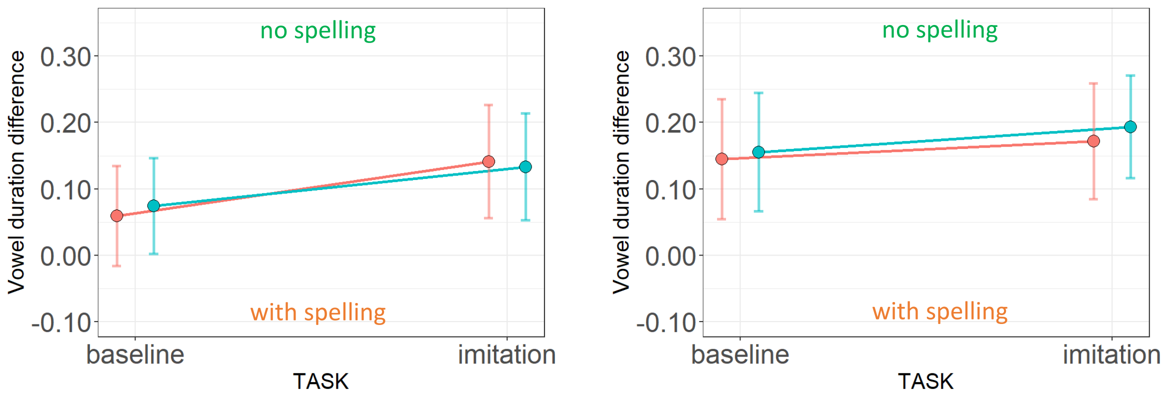

Figure 4.10. Mean values of RVDD (with SD) for the YL group (left) and the AL group (right) in the two tasks, with the effect of spelling

In the case of both groups, the shifts in RVDD did not seem to depend on whether the participants belonged to the either of the spelling subgroups. The two lines, with the orange one representing the spelling subgroup and the blue the one without spelling, seemed almost parallel in both age groups, which corresponds to comparable mean RVDD shifts from 0.06 to 0.14 (spelling) and 0.07 to 0.13 (no spelling) for YLs and from 0.15 to 0.17 (spelling) and 0.16 to 0.19 (no spelling) for ALs.

Finally, Figure 4.11 additionally splits the picture above by accounting for mean RVDD for each of the four vowels under analysis and the two spelling conditions.

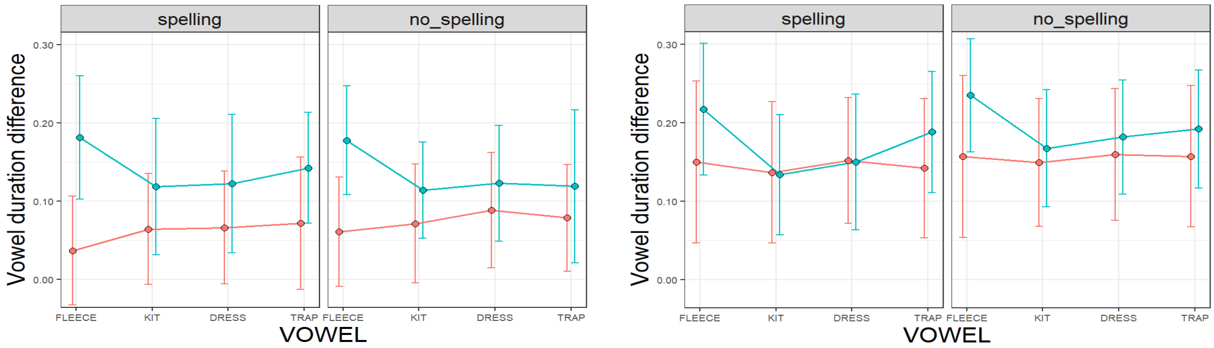

Figure 4.11. Mean values of RVDD (with SD) for the YL group (left) and the AL group (right) in the two tasks (baseline: orange, imitation: blue), with the effect of spelling and vowel type

While the overall patterns seem similar, what emerges from this closer examination is that the adult participants' mean RVDD showed somewhat clearer shifts for KIT and DRESS vowels, relative to the ones observed in Figure 4.9, where the two spelling conditions were unaccounted for. Figure 4.11 shows that, while adult participants in the spelling condition did not increase their RVDD in the imitation task, those in the condition without spelling seem to have marginally done so: from 0.15 to 0.17 for KIT and from 0.16 to 0.18 for DRESS, despite having higher baseline values in the latter condition. Conversely, spelling seems to have somewhat aided both groups in the imitation of TRAP's durational contrast, as indicated by slightly greater shifts across the two tasks in the condition with orthography provided. See Appendix 1 for all mean and SD values for RVDD, for different vowels, spelling conditions, tasks, and groups.

4.2.2.1. Statistical significance

As described earlier, to assess the effect of task on `RVDD`, a series of linear mixed effects models were fitted to the data, which were then compared using likelihood ratio tests. The models ranged from the null or reference model to gradually more specified models, with constant random effects:

```
model_0: RVDD ~ 1 + (1+task|subject)
model_1: RVDD ~ task + (1+task|subject)
model_2: RVDD ~ task + vowel + (1+task|subject)
model_3: RVDD ~ task × vowel + (1+task|subject)
model_4: RVDD ~ task × vowel × spelling + (1+task|subject)
```
[27]

[27] The effect of `spelling` is only relevant within the context of `task`. Therefore, `spelling` was included in the model as part of a three-way interaction with `task` and `vowel` straight away, without first considering its effect separately as an additive term.

The five models listed above were fitted for each of the two age groups separately. The likelihood ratio test indicated that `model_1` provided a significantly better fit to the data relative to the null model `model_0` for both the YL group ($\chi^2(1) = 45.05$, $p < 0.001$) and the AL group ($\chi^2(1) = 14.75$, $p < 0.001$), confirming the significance of the effect of `task`, especially for the younger group (higher Chi-squared value for YLs). Introducing the factor of `vowel`, particularly in the interaction with `block`, led to a significant improvement in model fit, with `model_3` considerably surpassing `model_1` (more than `model_2` did) for both YLs ($\chi^2(6) = 84.26$, $p < 0.001$) and ALs ($\chi^2(6) = 74.039$, $p < 0.001$).[28] This proves that the effect of `task` on `RVDD` was highly moderated by `vowel` type. Further expanding the interactive term of `task` × `vowel` by including the effect of `spelling` condition, did not improve the model fit for either of the groups ($\chi^2(8) = 6.3671$, $p = 0.61$; $\chi^2(8) = 7.31$, $p = 0.5$). Overall, the statistical tests support the initial observations made earlier, in that RVDD increased in the imitation task in both age (proficiency) groups, and that it did so to various degrees depending on the vowel, as Figure 4.9 suggested. It was also confirmed that whether the participants saw the stimuli in the imitation task or not did not significantly affect overall vowel duration contrasts, as was shown in Figure 4.10.

While the inclusion of the `vowel` predictor and its interaction with `task` significantly improved the model, meaning that the effect of `task` on `RVDD` varied depending on `vowel`, the statistical picture remains incomplete without pinpointing the particular pairs of `task` and `vowel` combinations that actually contributed to the interaction effect. By means of emmeans package for R, post-hoc pairwise comparisons of the estimated marginal means of `RVDD`, across different levels of `task` within each `vowel` level, were conducted, with the Bonferroni correction applied. Figure 4.12 illustrates the contrast estimates, i.e. estimated `RVDD` changes from the baseline to the imitation `task` for each `vowel` level, with 95% confidence intervals. The grey dashed lines mark the hypothetical points at which there were no `RVDD` changes across the two `task` levels. Negative contrast estimates indicate that `RVDD` was greater in the imitation relative to the baseline `task`.

[28] See Appendix 4 for the summaries of `model_3` for both age groups.

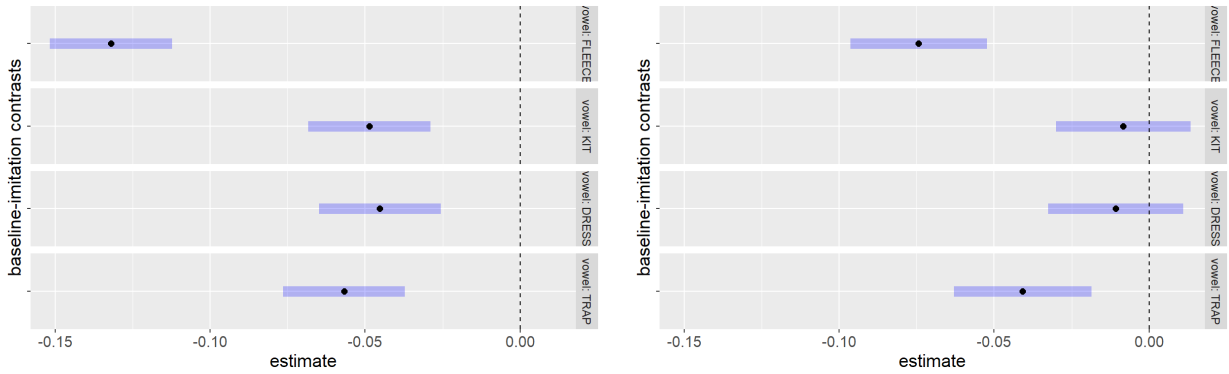

Figure 4.12. Pairwise contrasts of `task` levels (baseline vs. imitation) for each level of `vowel` (FLEECE, KIT, DRESS, TRAP) for the YL (left) and the AL (right) group

For the YL group, pairwise comparisons between the baseline and imitation tasks were significant for all `vowel` levels: -0.132 for FLEECE, -0.049 for KIT, -0.045 for DRESS, and -0.057 for TRAP ($p < 0.001$ in all cases), confirming the previous observations of FLEECE eliciting the greatest imitative effect. In the case of the AL group, significant contrast estimates were found for FLEECE (-0.074, $p < 0.001$) and TRAP (-0.041, $p < 0.001$), but not for KIT (-0.008, $p = 0.448$) and DRESS (-0.011, $p < 0.332$), as indicated by confidence intervals straddling the line for the two middle vowels in Figure 4.12. Overall, the post-hoc tests validate the previously made observations based on Figure 4.9, in that young participants successfully imitated all vowel duration contrasts, while adults did so only in the case of FLEECE and TRAP, where the model values sufficiently different from the adults' baseline values.

4.2.3. Discussion

The purpose of this section was to contribute to our understanding of vowel imitation, more specifically, imitation of vowel durational contrasts signalling the voicing of the coda consonant, by L1 Polish learners of L2 English. In line with the results of previous studies, the participants were expected to increase durational contrasts, as a result of exposure to the native-speaker. Overall, the prediction was confirmed, in that Polish learners of English were found to follow the model speaker's greater contrasts. They significantly shifted their relative vowel duration differences across the two tasks: baseline word reading, and an immediate imitation task. Expectedly, the contrast increases were almost exclusively due to vowels in /bVd/ words being pronounced longer, rather than the vowels being pronounced shorter in /bVt/ words.

Significant imitation was ascertained in the case of both young and adult participants. While it was not the purpose of the study to explicitly (and statistically) compare the imitative performance between the two age (and proficiency) groups, a tentative descriptive comparison seems warranted. As expected, the young group, who was also less proficient than the adult group, displayed much less pronounced durational contrasts in the baseline task than the latter group. This meant there was a considerable participant-to-model distance to be bridged in the case of the younger group, and indeed, they virtually doubled the vowel durational contrasts in the imitation task. While adults also exhibited more pronounced contrasts in the imitation task than in the baseline task, their relatively much higher baseline values (than the young learners' baseline values), inevitably stemming from their higher L2 proficiency, meant that these increases were not considerable, due to there not being much distance between their contrasts and those of the model speaker. Therefore, although the shifts by the younger group were much greater across the two tasks, it does not seem justifiable to declare them as better imitators relative to adults based on this data, especially considering their mean imitated value was slightly smaller than those of the adults' mean baseline value (0.14 vs. 0.15). Therefore, the adults' high baseline values can be considered as ones that were already shifted in the first task, as a result of the participants being instructed to produce words in English, almost exhausting the potential for further shifts in the subsequent task. While it must be acknowledged that the young group made seemingly good use of the opportunity to become more native-like, at least temporarily, as far as vowel durational contrasts are concerned, it is the adult group that ultimately reached more model-like values.

Although no differences were expected as regards the effect of vowel type on the degree of imitation, some vowels elicited significantly greater imitative effect than others. On the whole, most successfully imitated duration contrasts were found in the *beat-bead* pair and in the *bat-bad* pair. There does not appear to be anything inherent about the vowels, or the words they were in, that would make them more imitable than others. It seems the participants more readily imitated the contrasts in the case of those vowels that exhibited the greatest baseline-to-model distances, and indeed, the model's durational contrasts were most pronounced for FLEECE (*beat-bead*) and TRAP (*bat-bad*). The young group imitated all vowels, and they even matched the model values for KIT. Surprisingly, although TRAP was generally imitated more than KIT (*bit-bid*) and DRESS (*bet-bed*), the younger group did not cover much of the distance that the model left for them. The adult group's KIT and DRESS contrasts remained rather unchanged across the two tasks, as their baseline values virtually converged with the model values.

Finally, it was assessed whether or not seeing the imitated words would impact the extent to which durational contrasts were imitated. While no clear predictions were offered, it was speculated that the participants would be more resistant to durational shifts if the words' written representations were displayed in the imitation task, as all the information needed for the words' production might already be considered enough upon visual presentation of the words, potentially rendering the incoming acoustic signal superfluous. The study revealed that, overall, access to orthographic input in the imitation task (or lack thereof) did not significantly affect the participants' performance. Although not statistically significant, initial observations hinted at a feeble advantage that the adults displayed in the no-spelling condition in the case of KIT and DRESS, but, as noted earlier, durational contrasts in these vowels did not improve because of short participant-to-model distances. Conversely, there appeared to be a marginal advantage of orthographic access in the case of TRAP contrasts, but at this point no clear explanation presents itself for these observations.

In sum, Polish learners of English successfully improved their English vowel duration contrasts as a result of exposure to a native-English speaker, particularly by lengthening their vowels in words with voiced word-final codas. Although both young and adult learners followed the model speaker, the young group made much more considerable improvement than adults, due to the former group's inherently worse baseline performance. By contrast, the adult participants, having much less participant-to-model distance to bridge, displayed more modest shifts, but eventually reached more model-like values than the young ones. As for particular vowels, the durational contrasts in FLEECE and TRAP were most robustly imitated, but the most likely explanation for this lies in nothing else but the considerable contrasts of the model speaker for these two vowels. Finally, overall, the participants' imitative performance was not clearly affected by the presence versus absence of orthographic input in the imitation task.

4.3. Tested parameters 2: F_1 and F_2 mid-point values

As discussed in Chapter 2, Polish and English vowels system differ in many respects, and the main contrasts lie not only in the temporal specification of vowels but also in their qualitative aspects. The point of this section is to verify whether Polish learners of English can, at least temporarily, abandon their expected Polish-like realisations of English vowels, and assume more native-like, or more specifically, more model-like vowel productions in terms of their quality. Although we know that vowel quality is subject to convergence in the context of L2 speech acquisition, the issue remains vague as regards a wider range of vowels, particularly those that are considered similar to the closest L1 equivalents.

4.3.1. Measurements

Traditionally or impressionistically, the quality of vowel monophthongs has been regarded from the static point of view, as if the vowel had one uniform quality. However, since acoustically vowels are not constant, to capture the vowel's qualitative representation, static formant measurements are made, most commonly at the steady vowel portion. This is the perspective assumed at this stage of analysis of the current data. The first and second formants were extracted, which are expected to approximately correspond to vowel height (inversely) and frontness (directly), respectively. The F_1 and F_2 formant values were identified at the vowel mid-points, at exactly half of the vowels' duration (see Figure 4.13), using Praat's spectrographic displays and with formant tracker enabled.[29] Prior to formant extraction,[30] all spuriously estimated formants by the formant tracker were marked for exclusion or, if possible, for manual correction at the stage of data cleaning.

[29] Default formant settings were used with the exception of 'Formant ceiling', referring to the frequency range within which a particular number of formants (5 by default) are calculated. It was set individually between 5000 Hz and 6000 Hz for each participant, based on whether the majority of a given participant's estimated (LPC method) formant tracks aligned with the dark bands representing formants in wideband spectrograms.
[30] The formants were extracted by means of the same script used earlier.

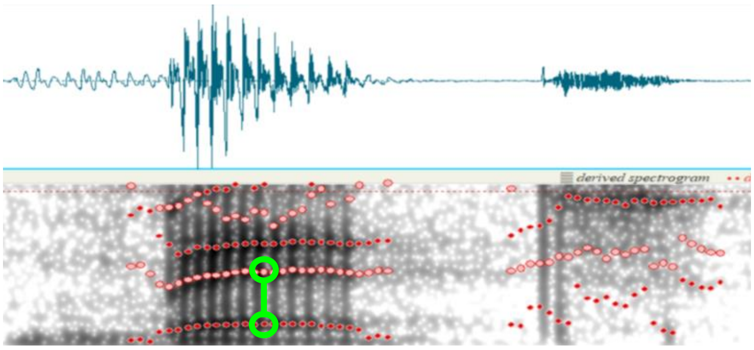

Figure 4.13. Sample measurement of F_1 and F_2 based on both the formant tracker and formant bands on the spectrogram

All subsequent formant analyses are based on unnormalised (raw) formant values. The reason for this lies in the author's conviction that, when it is justified to avoid normalisation, it should be avoided.[31] In the context of the current study, what is compared are essentially the formant values of a group of speakers, along with their inherent characteristics, with the same group of speakers, across two conditions (tasks), making the normalisation process redundant. Loose comparisons across the two age groups can still be made when we compare not their formant values in the absolute sense, but how their formant values change across the two tasks. The only concern may arise when comparing two subgroups within an age group (with spelling and without spelling), but, again, because we are interested in formant changes, and not in formant values themselves, the subsequent findings are still considered to be valid.

4.3.2. Analysis and results

The main focus of this analysis was to determine whether given participants' F_1 and F_2 values showed a significant increase or decrease between the two tasks. The two formants were analysed separately for two reasons. The primary motivation lay in the need to not overfit the statistical models, which could have compromised the reliability of their results.[32] Second, it was desired to know which of the two formants would contribute more to potential imitative effect in the context of the current study. The ensuing analyses

[31] By their very nature, normalisation processes involve (often non-linear) transformations of data, which, while useful in revealing patterns in the data, casts a certain degree of abstraction and uncertainty on it, by distancing it from its original form, and in consequence, potentially from the objective truth that it initially represented. Moreover, the maximisation of vowel-space similarity, that normalisation methods often prioritise, may actually erase legitimate phonetic variation inherent to the data, and therefore obscure linguistic facts (see Barreda 2021 for a review).

[32] The MANOVA method was considered, which allows the assessment of multiple dependent variables simultaneously, but it was not chosen due to the difficulty in incorporating (important) random effects and for the second reason mentioned.

always incorporate the factor of vowel type from the outset, to account for formant patterns inherent to each vowel. Apart from the 62 data points discarded in the initial data cleaning, 121 more tokens were excluded due to immeasurable formants, and additional 21 flagged as extreme outliers,[33] leaving a total of 3636 tokens for analysis (1850 for YLs; 1786 for ALs).

Figure 4.14 shows scatter plots of raw formant values for young participants (left) and adults (right). The ellipses can be understood as ideally encompassing roughly 67% of the data points for each vowel and for each task, under the assumption that the data follow a bivariate normal distribution, which the current case can be loosely judged as such. The solid ellipses represent data points obtained in the baseline task, while the dashed ones encompass formant values from the imitation task.

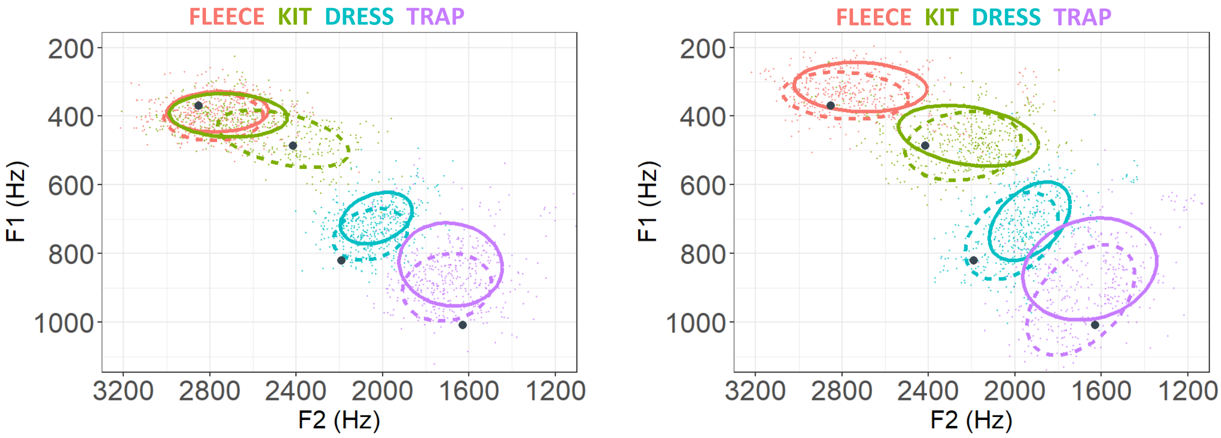

Figure 4.14. The participants' raw formant values (in Hz) for the four vowels, with their concentrations delineated by solid ellipses (baseline data points) and dashed ellipses (imitation data points), along with model speaker values (black points)

As can be seen, in the majority of cases, modest but observable mid-point formant concentration shifts took place, and these shifts generally followed the model speaker values (black points).[34] Reasonably, the shifts seem greater in those cases where the model values lay outside of the participants' baseline ellipses for a given vowel. The smallest formant changes seem to have occurred in the case of FLEECE, particularly for the YL group, which agrees with the fact that the model value (top-right black point) was the closest one to the centre of the corresponding baseline (orange) ellipse. The formants underwent visibly the greatest baseline-to-imitation formant changes for the KIT vowel in the case of the YL group. Overall, both age groups exhibited similar patterns with the

[33] Before excluding the outliers, the data were grouped by vowel type.
[34] Even though the plots represent unnormalised values.

exception of the AL group's baseline values for KIT, which were already much more convergent with the model values in the first task.

Figure 4.15 shows average F_1 values (top) and average F_2 values (bottom) for both the YL (left) and AL (right) groups, with bars representing baseline (orange) and imitation (blue) mean values, paired for each vowel, together with model values (black points).

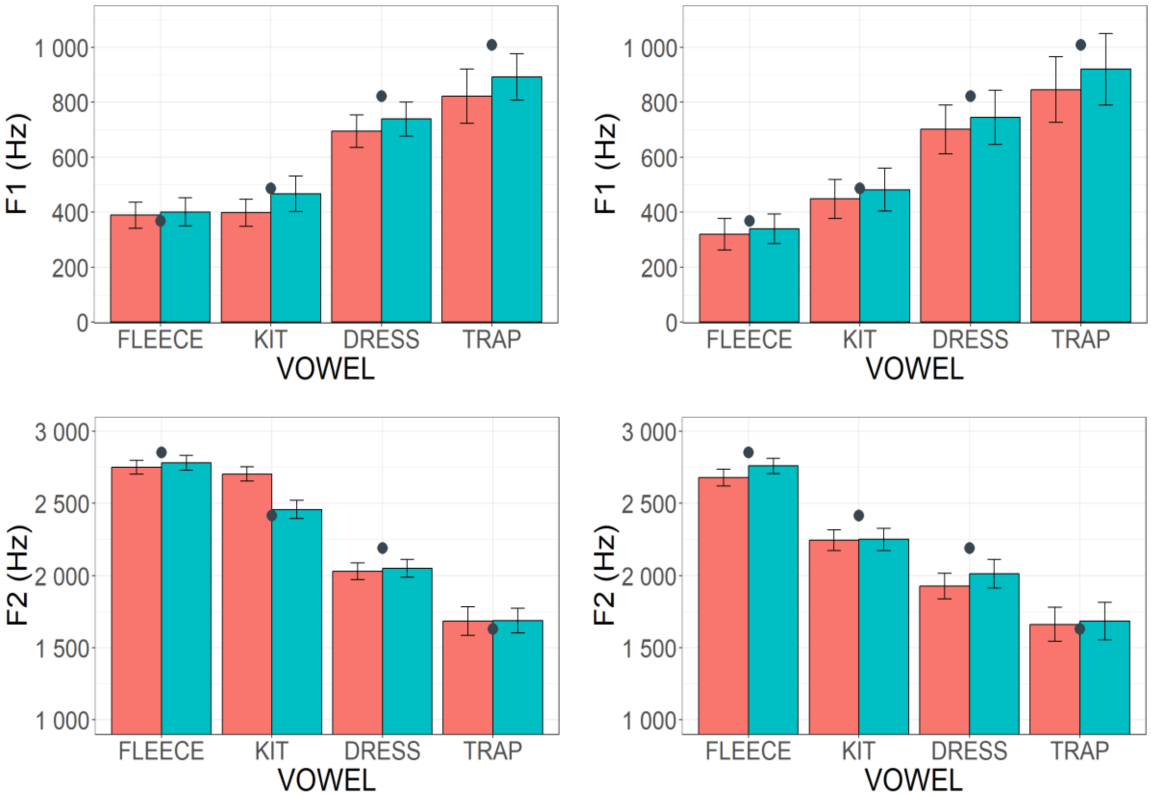

Figure 4.15. Mean F_1 values (top) and mean F_2 values (bottom) for YLs (left) and ALs (right), grouped by the task (baseline – orange, imitation – blue) and the vowel, together with model values (black points).

The bar charts generally seem to confirm the initial observations based on the scatter plots above. Both groups increased their mean F_1 values for all vowels, particularly for YLs' KIT (from 398 Hz to 467 Hz) and TRAP (from 822 Hz to 892 Hz), and ALs' TRAP (from 846 Hz to 920 Hz). Again, the degree of baseline-to-imitation shifts seemed correlated with the distances between the participants' baselines and the model values (FLEECE: 369 Hz, KIT: 486 Hz, DRESS: 821 Hz, TRAP: 1008 Hz). As the bottom pair of bar charts in Figure 4.15 demonstrates, the changes in F_2 were generally more modest, with virtually no changes for YLs, except for the considerable decrease in KIT (from 2703 Hz to 2458 Hz), and only relatively subtle upward shifts for ALs' FLEECE (from 2679 Hz to 2759 Hz) and DRESS (from 1926 Hz to 2012 Hz). Overall, the participants seem to have followed the model's F_2 values (FLEECE: 2852 Hz, KIT: 2414 Hz, DRESS: 2190 Hz, TRAP: 1629 Hz), although no clear shifts towards the respective model values appear to have happened in YLs' DRESS and ALs' KIT.

Both the scatter plots in Figure 4.14 and the bar charts in Figure 4.15 were further split to account for the effect of orthography on baseline-to-imitation formant shifts. However, on the whole, they showed no sharp contrasts either for F_1 or F_2 changes, so the split charts are not shown here. The bars charts for both F_1 and F_2 were virtually identical for the YL group across the two spelling conditions. As for adults, somewhat greater degree of imitation was observed in F_1 for FLEECE and KIT when no spelling was provided, but their performance was slightly better with spelling for F_2 in FLEECE. See Appendix 2 for all mean and SD values for F_1 and F_2, for different vowels, spelling conditions, tasks, and groups.

Due to the formant data not having been normalised, it seems appropriate to consider the participants' variability in terms of their formant shifts, which is visualised in a series of scatter plots in Figure 4.16. Each of the four rows corresponds to a given vowel of interest for both age groups. The arrows, one for each participant, can be considered vectors, whose initial points are fixed at a given participant's baseline F_1 and F_2 values, and the terminal points at their imitated F_1 and F_2 values, with the magnitude of the vectors representing the degree of baseline-to-imitation formant shifts. The vectors' colours were provided purely for visual clarity, with each colour representing a particular quadrant at which the vectors point (upper-left, upper-right, bottom-left, bottom-right).

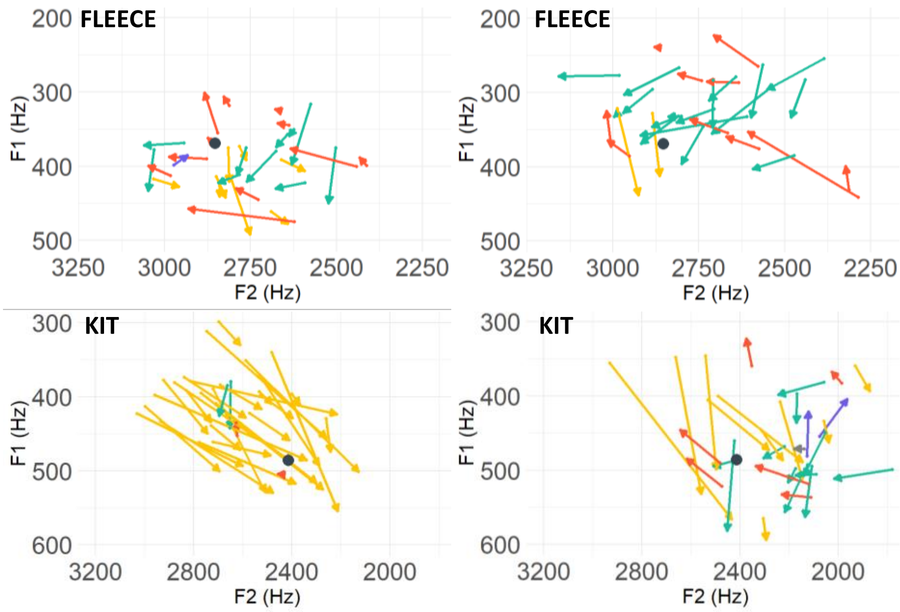

Figure 4.16. Scatter plots with vectors representing individual participants' baseline-to-imitation F_1 and F_2 shifts for the four vowels, for the YL groups (left) and AL group (right), with model values (black points)

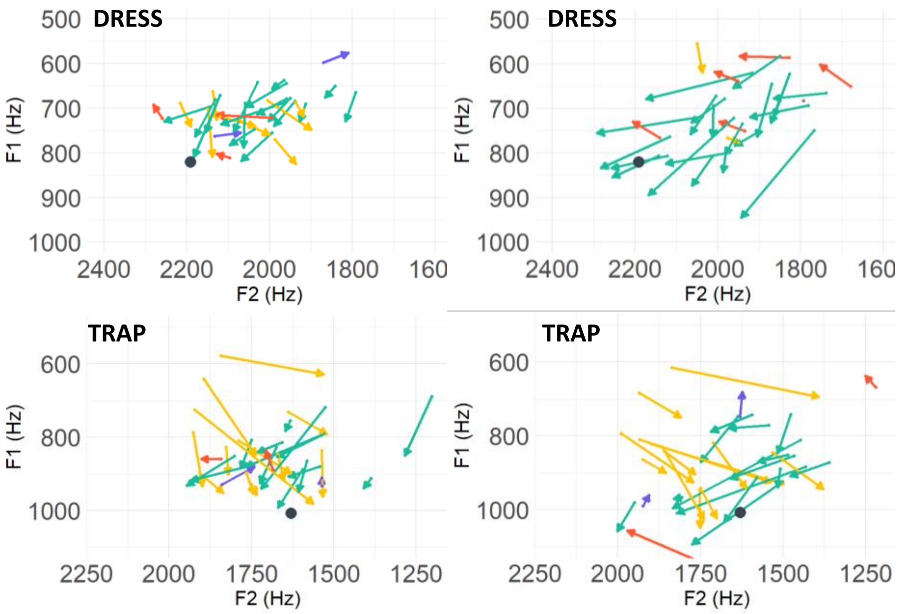

Figure 4.16. (continued)

As can be observed, there is a lot of expected variation both in terms of the direction at which the vectors point, as well as their magnitude. Despite this, overall, the trends observed earlier agree with these more nuanced patterns, and what seems particularly evident is that the majority of vectors approximately point in the direction of model values (black points). These plots also show what averaged values in earlier figures did not show, namely that while some participants within a given group shifted their values towards a particular direction, others neutralised this pattern by exhibiting divergent shifts, which is perhaps most evident in the case of F_2 values for TRAP, where many vectors point in the opposite directions along the F_2 axis, accounting for meagre mean F_2 changes in the bar charts. In congruence with the earlier observations, the vectors manifested the most decisive changes both in terms of direction and magnitude in the case of KIT for YLs, as well as in FLEECE and DRESS for ALs.

4.3.2.1. Statistical significance

As was done previously, to ascertain statistical significance of the potential effect of the predictors on the tested parameter, a series of linear mixed effects models were fitted and compared by means of likelihood ratio tests. The following models were fit separately for the two formants and the two age groups:

```
model_0: f1/f2[35] ~ vowel[36] + (1+task|subject)
model_1: f1/f2 ~ vowel + task + (1+task|subject)
model_2: f1/f2 ~ vowel × task + (1+task|subject)
model_3: f1/f2 ~ vowel × task × spelling + (1+task|subject)
------------------------------------------------------------------------
model_3r: f1/f2 ~ (vowel × task) + (vowel × spelling) + (1+task|subject)
```

The first likelihood ratio comparison, between `model_0` and `model_1`, yielded significant results, but, as was the case previously, including the interaction term `vowel × task` improved the null model even better than `model_1` did. The comparisons were significant for both formants and for both groups:

YLs' `f1`: $\chi^2(4) = 119.6$, $p < 0.001$	ALs' `f1`: $\chi^2(4) = 59.26$, $p < 0.001$
YLs' `f2`: $\chi^2(4) = 317.93$, $p < 0.001$	ALs' `f2`: $\chi^2(4) = 42.95$, $p < 0.001$

The much higher Chi-squared values for the YL group, especially for F_2, suggest greater effect of `task` in the case of young participants relative to adults. Although at the stage of inspecting the scatter plots and bar charts, the factor of `spelling` did not seem to have an effect, its inclusion in `model_3` led to significant differences in all four cases:

YLs' `f1`: $\chi^2(8) = 50.62$, $p < 0.001$	ALs' `f1`: $\chi^2(8) = 30.99$, $p < 0.001$
YLs' `f2`: $\chi^2(8) = 59.19$, $p < 0.001$	ALs' `f2`: $\chi^2(8) = 64.95$, $p < 0.001$

However, the three-way interaction term `(vowel × task × spelling)` inherently includes three two-way interaction terms: `(vowel × task)`, `(task × spelling)` and `(vowel × spelling)`, the last of which might be responsible for the significant result, even though it is hardly of interest in the current study. After all, formant values for given vowels are expected to be different across the two spelling subgroups for a given age group, as these subgroups represent different participants, with their own inherent vowel characteristics. To isolate the effect of `(task × spelling)` on `f1` and `f2`, which is clearly of interest, `model_3` was compared to its reduced version `model_3r`, with the term `(task × spelling)` removed, which yielded insignificant results in all cases:

YLs' `f1`: $\chi^2(4) = 0.63$, $p = 0.96$	ALs' `f1`: $\chi^2(4) = 8.84$, $p = 0.07$
YLs' `f2`: $\chi^2(4) = 5.16$, $p = 0.27$	ALs' `f2`: $\chi^2(4) = 6.92$, $p = 0.14$

[35] The same series of models were fit separately for `f1` and `f2` in order not to overfit the models.
[36] The null model already contains the predictor `vowel`, as this predictor naturally has an effect on both formants, which was not so evident in the case of vowel duration contrasts.

In other words, the full model model_3 does not fit the data significantly better than the reduced model model_3r, suggesting the (task × spelling) interaction term does not add much explanatory power to the model, and consequently does not improve its ability to predict either f1 or f2. Because the significant term (vowel × spelling) in model_3r adds nothing of value, model_2 is resolved to be the final model.[37]

However, for the sake of clarity and certainty, it seemed warranted to perform post-hoc pairwise comparisons of the estimated marginal means of f1 and f2, taking into account various effect combinations of task, vowel (model_2) and spelling (model_3), with the Bonferroni correction method applied. Table 4.3 shows estimated f1 changes from the baseline to the imitation task for each vowel, where zero would denote no such changes. Contrast estimates below zero indicate formant increases, while positive estimates indicate formant decreases.

FLEECE	f1	12 Hz	$p = 0.09$
KIT	f1	-69 Hz	$p < 0.001$
DRESS	f1	-45 Hz	$p < 0.001$
TRAP	f1	-71 Hz	$p < 0.001$

FLEECE	f1	-16 Hz	$p = 0.07$
KIT	f1	-35 Hz	$p < 0.001$
DRESS	f1	-44 Hz	$p < 0.001$
TRAP	f1	-71 Hz	$p < 0.001$

FLEECE	f2	-35 Hz	$p = 0.05$
KIT	f2	244 Hz	$p < 0.001$
DRESS	f2	-18 Hz	$p = 0.29$
TRAP	f2	-5 Hz	$p = 0.78$

FLEECE	f2	-82 Hz	$p < 0.001$
KIT	f2	-6 Hz	$p = 0.74$
DRESS	f2	-92 Hz	$p < 0.001$
TRAP	f2	-12 Hz	$p = 0.52$

Table 4.3. Estimated f1 and f2 baseline-to-imitation contrasts for YLs' (left) and ALs' (right) for the four vowels, with p-values

As can be seen, f1 increased significantly in all cases except for FLEECE in both groups, while f2 changed significantly only in the case of YLs' KIT and ALs' FLEECE and DRESS. In Figures 4.17 and 4.18, the pairwise comparisons were expanded to include the predictor of spelling.

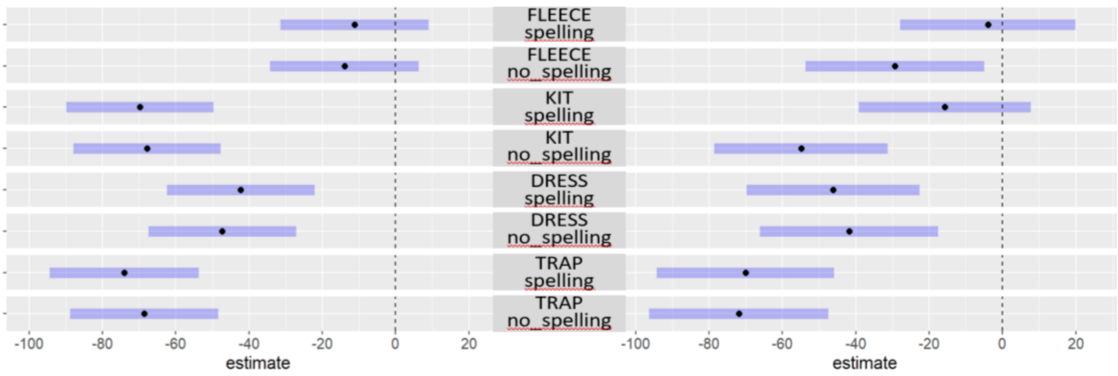

Figure 4.17. f1 pairwise contrasts of task levels (baseline, imitation) for each vowel (FLEECE, KIT, DRESS, TRAP) for YLs (left) and ALs (right)

37 See Appendices 5–6 for model summaries.

Figure 4.18. `f2` pairwise contrasts of task levels (baseline, imitation) for each vowel (FLEECE, KIT, DRESS, TRAP) for YLs (left) and ALs (right)

The contrast estimates for both `f1` and `f2`, with their 95% confidence intervals, seem quite closely aligned for the opposite `spelling` conditions in the case of young learners (left). More observable contrasts are in the adult group (right), where the contrast estimates seem more divergent, particularly for `f1` in FLEECE and KIT, which seem to have increased more in the subgroup without spelling provided.

As was the case with vowel duration contrasts, statistical tests confirm the trends observed earlier, in that all participants tended to shift their F_1 values more than F_2 values, although both exhibited significant shifts. Overall, the young participants seem to have made bigger formant shifts than adults did, which was particularly evident in the case of the KIT vowel. Again, the inclusion of the factor of spelling did not significantly affect the degree of baseline-to-imitation formant shifts.

4.3.3. Discussion

One of the key questions asked in the current investigation was whether L2 learners, specifically Polish learners of English, could imitate spectral characteristics of vowels in the target language. Proper vowel quality poses a challenge for L2 learners, as they are often left to rely on impressionistic judgments of how the desired vowel qualities compare to those in their native repertoire. Previous research has shown that vowel quality generally undergoes significant changes as a result of exposure to auditory cues in immediate shadowing tasks. What little similar research has been done in the context of L2 speech also suggests that non-native speakers are capable of imitating native speaker's spectral properties of vowels, although only distinct vowel qualities seem to have been considered. What was of interest in the current investigation was whether L2 vowels that resemble L1 vowels were also subject to imitation, as well as to determine the potential effect of spelling on its degree. Finally, the factor of age, together with L2 proficiency, was considered so as to at least partially account for the variability among L2 learners. Overall, the current investigation confirmed that Polish speakers imitate English vowel quality, as reflected by significant mid-point formant shifts across the baseline word reading and the imitation tasks. The general tendency was that the participants followed the model's higher formant values,[38] which usually pointed toward more fronted and open vowel realisation.

The prediction that some English vowels would be more prone to imitation, based on whether the particular vowels were deemed similar to the closest Polish counterparts or not, was partially validated. Predictably, the participants' formants remained relatively unchanged in the case of FLEECE, with only adult participants making noticeable shifts along the F_2 dimension, making their FLEECE more fronted. As opposed to FLEECE, the model left considerable participant-to-model distance in the case of KIT, particularly for the younger participants, and they made full use of it by virtually matching both formants of the model speaker, resulting in a more open and retracted KIT, relative to their baseline realisation, which was much more Polish /i/-like (and FLEECE-like). Adults' baseline productions of KIT were already more convergent with the model values, but they managed to significantly reduce the vowel's height in the imitation task. Surprisingly, even though there was a relatively considerable distance between their baseline values and the model's values along the F_2 dimension, adults persisted in their initial degree of frontness for KIT.

Similarly to FLEECE, the DRESS vowel was not expected to exhibit significant shifts across the tasks, due to its deceptive similarity to Polish /ɛ/. However, the model's rather outstanding values did not reflect this similarity, as there was enough participant-to-model distance for both groups to close. The model's more open realisation of DRESS induced both groups to shift their productions in that direction. The model also exhibited more fronted realisation, which attracted only the adult group, who followed the model quite unanimously, as confirmed by a number of vectors pointing towards the model values. Finally, substantial shifts were expected in the case of TRAP, which is ambiguously replaced with either Polish /a/ or /ɛ/, the two closest Polish neighbours of TRAP. The participants readily modified their baseline production of TRAP, which was on the whole less open and apparently more Polish /ɛ/-like, towards the model's more

[38] Even though the formant data did not undergo normalisation, both groups' formant values were drawn towards the model's values.

open realisations. At first glance, no shifts seem to have occurred along the frontness dimension, with both groups of participants retaining their mean F_2 values. However, a closer inspection of individual variability revealed that F_2 shifts did take place, as illustrated by particular participants' vectors at the bottom of Figure 4.16. While the vectors' starting points seemed polarised, they ultimately appeared to converge towards the centre, where the approximate position of the model's values lay. This suggests that both groups displayed similar ambivalence in their baseline realisations of the vowel, some of which were more fronted, potentially Polish /ɛ/-like, while others more retracted, suggesting a more /a/-like quality.

As far as the potential effect of orthography on imitative performance is concerned, it was predicted that imitators who had no access to words' orthographic representations in the imitation task would approximate the model vowel qualities to a greater extent than those who did see the words. Based on previous research investigating orthography-induced errors, Polish learners who saw the words on the screen right before and while imitating them were expected to be guided by their L1 associations between vowel letters and vowel sounds, and to rely less on the auditory aspect of the stimuli. Surprisingly, the current data did not support this position, with the young participants exhibiting virtually equivalent performance in both conditions. Although, overall, not significantly different, the adults' degree of imitation may have varied slightly depending on the condition, the vowel and the formant, with marginally better no-spelling performance for F_1 in FLEECE and KIT, but worse for F_2 in FLEECE. One of the reasons for the lack of expected effect may lie in that the effect was not entirely isolated, as all of the participants necessarily did see the words in the word reading task, which may have influenced their vowel productions in the subsequent imitation task. In line with this, as was mentioned earlier, the effect of orthographic input may be more global and may spill over into speech that is not explicitly read out, meaning that, even when the words are not seen, the learners' stored representations and their subsequent reproductions may, in a way, be considered multimodal, as shaped by both written and spoken input. It may also be that the presence of spelling helped some participants in imitating the model's vowel qualities by making them more certain what words they were supposed to imitate.

To conclude, on the whole, Polish learners of English successfully imitated more L2-like English vowel quality, as manifested by the first and second mid-point formant shifts, after exposure to a native English model. The participants followed the model's formant values, which generally pointed toward more fronted and open realisations. The shifts were more evident along the F_1 dimension than in the F_2 dimension, which tentatively suggests vowel height induces greater changes than the degree of frontness. As for particular vowels, KIT was most readily imitated, particularly in the case of children, as well as TRAP, as these two front vowels appear to be most distinct relative to their closest Polish neighbours. The FLEECE vowel, being quite similar to Polish /i/ in terms of quality, exhibited meagre shifts, but the DRESS vowel, also similar to Polish /ɛ/, induced rather remarkable formant shifts, particularly for F_2 in adults. This could be ascribed to participant-to-model distances, which in the case of DRESS were long enough to elicit imitative performance. The distances were also likely responsible for the apparent imitative advantage of the younger group in the case of KIT, which was much more Polish /i/-like in the baseline task than the adults' baseline productions of KIT. No effect of spelling was ascertained as far as vowel quality imitation is concerned, but it may be speculated that the current experimental paradigm simply did not capture the effect of orthography on L2 speech due to its potentially underestimated complexity and subtlety.

4.4. Tested parameters 3: Trajectory length (TL) and spectral rate of change (roc)

As discussed earlier, the static view of vowel quality, while traditionally considered informative, is rather simplistic, in that it fails to capture vowels' spectral variability. To the author's knowledge no study has attempted to verify explicitly whether or not vowel dynamics are subject to phonetic imitation. The current investigation is intended to supplement our understanding of vowel imitation by accounting for the dynamic view of vowel quality. Polish learners of English, whose native language displays more stable patterns, particularly in initial vowel portions, were tested for the imitation of VISC, as a result of exposure to a native English model, whose language is reported to exhibit more robust formant patterns.

4.4.1. Measurements

The metrics used for the measurement of vowel dynamics, based on (Fox and Jacewicz 2009; Farrington et al. 2018),[39] were derived from F_1 and F_2 formant values taken at the four equidistant points throughout the vowels' duration, these being at 20%, 40%, 60%,

[39] Vector length (VL), which marks formant trajectory only between the two extreme points, was not used in the current study, because, as the former authors remark themselves, it is likely to underestimate the amount of spectral change and may lead to erroneous conclusions as to the nature of VISC of the vowels under analysis. However, VLs were essentially used, but for particular vowel sections, as discussed later.

and 80% of the vowels' duration, resulting in the division into five sections, each representing 20% of the total vowel duration (see Figure 4.19). The formant measurement methods and criteria were equivalent to the ones described for mid-point formant measurements.[40]

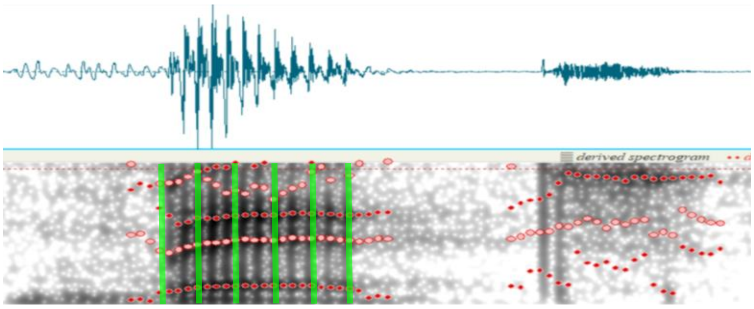

Figure 4.19. Sample division of vowel duration into five sections

Out of the five sections, the two most extreme ones were excluded from further analysis in order to reduce the coarticulatory effect of the surrounding consonants, leaving three vowel sections of interest: 20-40%, 40-60%, and 60-80%. For each of the three vowel sections, Euclidean distances in the F_1 by F_2 plane were calculated, with greater distances (in Hz) reflecting greater magnitude of combined F_1 and F_2 movement throughout a given vowel section. These distances are referred to as vowel section lengths (VSLs):

$$VSL_1 = \sqrt{(F1_{20} - F1_{40})^2 + (F2_{20} - F2_{40})^2}$$

$$VSL_2 = \sqrt{(F1_{40} - F1_{60})^2 + (F2_{40} - F2_{60})^2}$$

$$VSL_3 = \sqrt{(F1_{60} - F1_{80})^2 + (F2_{60} - F2_{80})^2}$$

Figure 4.20 illustrates a sample vowel with its formant trajectories that encompass the middle 60% of its duration, with the four vowel points and three VSLs marked between successive points. As can be seen, the division of a vowel into separate sections allows closer inspection of formant frequency change over the course of a vowel, which a simple vector between the two extreme points would have missed.

---

[40] Formant values were not normalised for the reasons outlined earlier. Additionally, in the case of VISC, there seems to be even less of a reason to normalise as we are interested in changes in formant values rather then formant values themselves, although it should be acknowledged that one's inherently greater formant values will potentially also lead to greater vowel dynamics. Fox and Jacewicz (2009), from whom the matric used here were adopted, worked on unnormalized values, although Farrington et al. (2018), who used mostly the same metrics, did use normalisation.

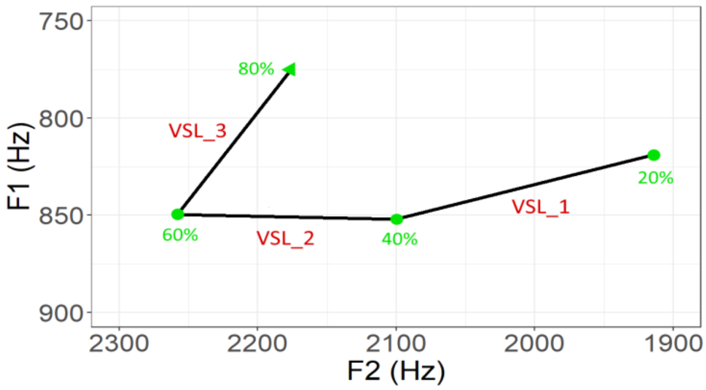

Figure 4.20. The three vowel section lengths

From the three VSLs, the first tested parameter emerges, defined as the total trajectory length (TL) in Hz, over the course of the three sections:

$$TL = VSL_1 + VSL_2 + VSL_3$$

While the magnitude of shifts, such as those in the figure above, give us a sense of formant dynamics, they only consider the aspect of time in relative terms, meaning that each of the three shifts took the same amount of time. To gain a more complete understanding of formant dynamics, it is necessary to supplement the magnitude of formant change with the rate at which this change happens, which can be accomplished by incorporating vowel duration over which the formant shifts take place. Therefore, we divide total trajectory length (TL) by the 60% of the vowel's duration (VD), to obtain the spectral rate of change (roc) in Hz/ms:

$$roc = \frac{TL}{0.6 \times VD}$$

Therefore, for example, while two people may show similar magnitude of formant shifts, it may also turn out that one of them accomplishes these shifts faster.[41] Apart from TL and roc, what is also inspected in the subsequent analysis, although not tested statistically, is the magnitude of formant shifts for particular sections themselves and their individual spectral roc. Moreover, it is also attempted to account for how particular formants contribute to the overall formant trajectories, and to inspect how their direction potentially change relative to the model values.

[41] Fox and Jacewicz (2009: 2604) note that "the higher spectral roc is related to faster articulatory movements typically invoked to reach the formant target."

4.4.2. Analysis and results

Because both formants had to measured at four vowel points, leaving more room for erroneous measurements, 45 more tokens had to be excluded from analysis relative to the number of tokens analysed in the case of vowel mid-points, which resulted in a total of 3612 tokens (3840 – 228). Further tokens, those that represented extreme outliers, were discarded.[42] This left a total of 3562 tokens for the analysis of TL (YL: 1834, AL: 1728), and 3556 for the analysis of roc (YL: 1832, AL: 1724).

Figure 4.21 illustrates a broad view of the participants' vowel dynamics as reflected by the three vowel section lengths (VSLs) for each of the four vowels placed in formant scatter plots. The solid trajectories represent VSLs obtained in the baseline tasks, while the semi-transparent ones stand for VSLs from the imitation tasks. The thin grey trajectories belong to the model speaker.

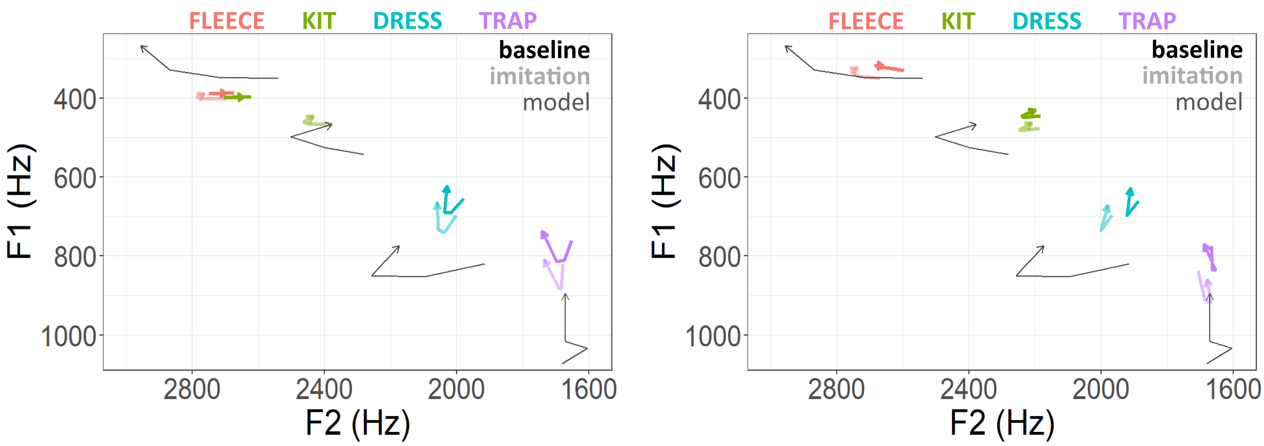

Figure 4.21. Scatter plots with YLs' (left) and ALs' (right) vowel section lengths for particular vowels in the baseline task (solid) and the imitation task (semi-transparent) for the four vowels, with model values (thin grey)

The baseline-to-imitation overall shifts in vowel trajectories were due to changes of absolute formant values at particular vowel points, which is congruent with midpoint formants shifts observed earlier. From this global picture, there do not emerge any distinct patterns as regards the participants' across-the-task changes in either the magnitude or the

[42] Although there were not many outliers, they significantly skewed the distributions, which could have invalidated the statistical models had they been included. The outliers were identified directly for the final derived metrics separately (TL and roc), which were then subjected to statistical tests. The outliers were not removed at the stage of formant extraction for the derived metrics, because it would have resulted in a considerable loss of data points for the derived metrics (due to the reliance on many F_1 and F_2 values, at four vowel points, in their calculation). However, in the case of most of the plots below, those based on underived metrics (F_1 and F_2), or on partially derived metrics (VSLs), outliers were excluded at the stage of calculation of these respective metrics.

direction of vectors pertaining to particular vowel sections. What appears quite evident, however, is the fact that the participants' vectors were nowhere near the magnitude of the model vectors, meaning their VSLs were much lower, even in the imitation task. What seems worthy of noting is the overall U-turn shape of the participants' trajectories. Their FLEECE and KIT vowels tended to become somewhat fronted towards the middle section of the vowels, but became partly retracted again towards the end, while their DRESS and TRAP became more open in the middle, only to later assume a comparable degree of height to that observed at the onset.

To offer a more nuanced analysis of potential baseline-to-imitation changes in the participants' vowel dynamics, what is considered are their trajectory length (TL), i.e. the sum of three VSLs for each vowel. Figure 4.22 illustrates the participants' mean TL (with SD) in the baseline task (orange) and the imitation task (blue) for all four vowels, with mean model values for reference (black points).

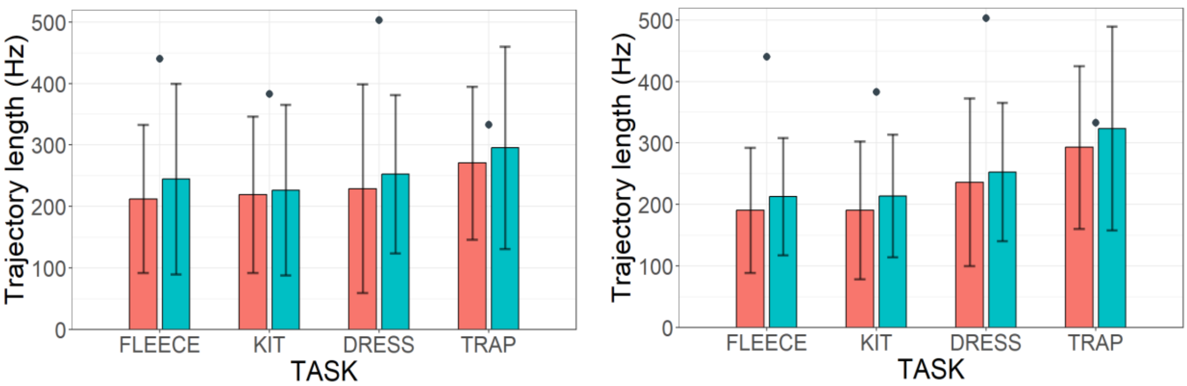

Figure 4.22. Mean values of trajectory length (with standard deviation) for each vowel, in baseline (orange) and imitation (blue), with model values (black points), for both YLs (left) and ALs (right)

From the above bar charts, what seems clear is that TL increased in the second task in virtually all cases, with the global mean value rising from 224 Hz (SD = 101 Hz) to 239 Hz (SD = 100 Hz) for the YL group and from 219 Hz (SD = 105 Hz) to 242 Hz (SD = 111 Hz) in the AL group. However, the baseline-to-imitation increases do not appear considerable, given the relatively high mean model value of 415 Hz (SD = 124 Hz). The participants' vowels manifested similar TL with the exception of the TRAP vowel, which reached the highest mean values, even though the model had the lowest TL for this vowel. The factor of spelling was inspected, but no differences in TL were discerned across the two spelling conditions.[43] No striking differences appear to emerge between the two age groups, but the former group, interestingly, attained higher baseline values for FLEECE and KIT, and the latter group increased their TL more for KIT. It should not escape one's attention that the participants' TL was highly variable, with SD reaching almost half of their mean values. The participants' distributions shown in Figure 4.23 exhibit a certain degree of positive skewness due to there still remaining some outlying values along the right tails, which contributed to the considerable variability.

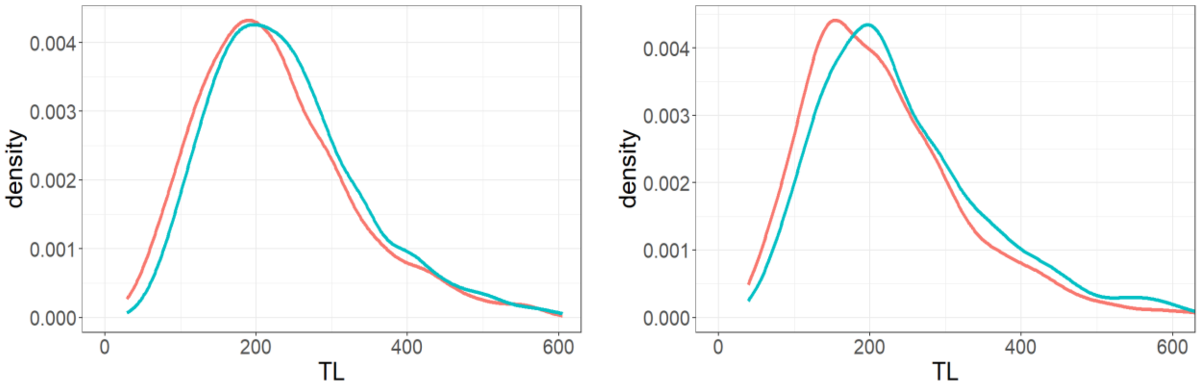

Figure 4.23. Distributions of the TL data points produced by the YL group (left) and the AL group (right) in the two tasks: baseline (orange) and imitation (blue)

Figure 4.24 breaks down the mean TL (for all vowels combined) into separate VSLs, from which it was derived, with baseline bars (orange) and imitation bars (blue) representing the magnitude of formant movement for the three vowel sections: 20-40%, 40-60%, and 60-80%.

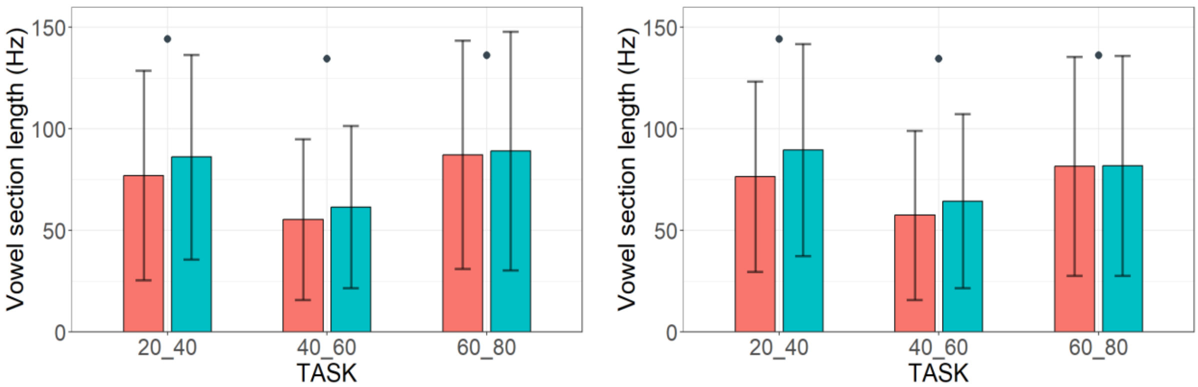

Figure 4.24. Mean values of vowel section length (with standard deviation) for each section, in baseline (orange) and imitation (blue), with model values (black points), for both YLs (left) and ALs (right)

[43] For the sake of saving space, the plots for the two spelling conditions are not included, them being virtually identical.

The figure reveals that it was the first and the second vowel sections which contributed the most to TL changes across the two tasks. Both groups showed comparable patterns, in that the middle section proved more stable (had lower VSL) than the more peripheral sections in terms formant movement, and both increased their dynamics in the first two sections in the imitation task, with no change in the third section, which already manifested rather high baseline values, particularly for the YL group. As regards the model speaker, her mean VSL values were rather consistent across the three sections, with somewhat greater movement in the first section. Evidently, the model surpassed the participants' both baseline and imitation mean values. The data proved again to be highly variable, as indicated by quite extended error bars (high SD).

While Figure 4.22 showed mean trajectory lengths for the four vowels, Figure 4.25 incorporates absolute vowel duration into the measure, resulting in spectral rate of change for each vowel and task. Spectral roc helps better understand vowel dynamics by showing how fast formants change over time.

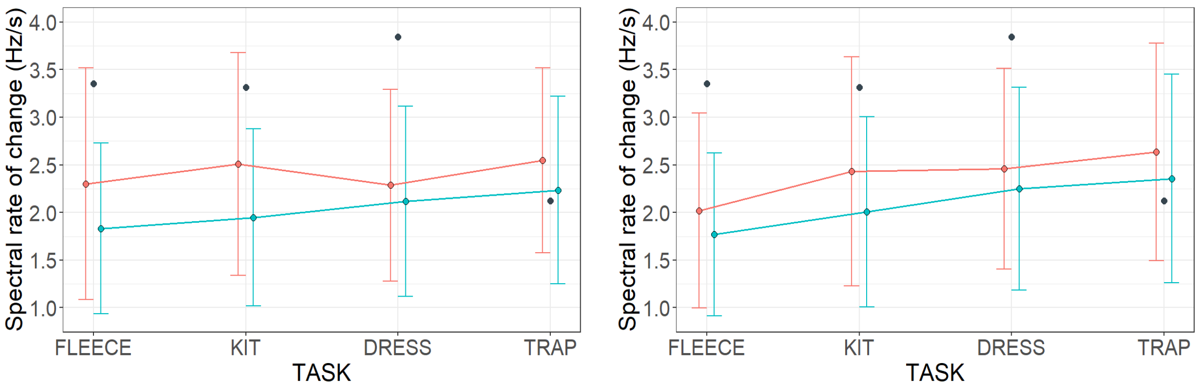

Figure 4.25. Mean values of spectral rate of change (with standard deviation) for each vowel, in baseline (orange) and imitation (blue), with model values (black points), for both YLs (left) and ALs (right)

Contrary to trajectory length, which increased across the two tasks for all vowels, mean spectral roc was lower in the second task (blue) than in the baseline task (orange) in both groups. The participants' global mean of spectral roc decreased from 2.41 Hz/ms (SD = 1.10 Hz/ms) to 2.03 Hz/ms (SD = 0.96 Hz/ms) for the YL group and from 2.38 Hz/ms (SD = 1.13 Hz/ms) to 2.09 Hz/ms (SD = 1.03 Hz/ms) in the AL group. The participants lagged behind the model speaker (M = 3.16 Hz/ms, SD = 1.4 Hz/ms) in terms of formant movement rate in all vowels except for TRAP, which agrees with the model's relatively low TL for this vowel. Again, it did not seem that whether the participants saw the words or not while imitating affected roc.[44] Figure 4.26 shows spectral roc for particular vowel sections (20-40%, 40-60%, 60-80%), across the two tasks.

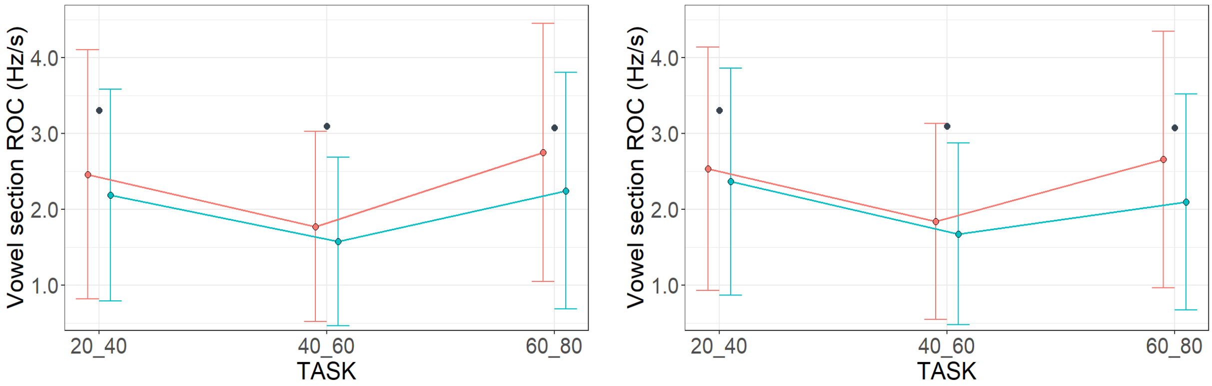

Figure 4.26. Mean values of spectral rate of change (with standard deviation) for each vowel section, in baseline (orange) and imitation (blue), with model values (black points), for both YLs (left) and ALs (right)

As shown earlier, the two peripheral sections (20%-40% and 60%-80%) had the greatest overall formant movement, and naturally, roc for particular vowel sections displayed the same patterns. When it comes to the differences across the two tasks, the largest drops in the rate at which formants shifted was observed in the third section, which was a consequence of VSL showing no positive change for that section which could make up for longer vowel (section) duration. Similarly to TL, the participants' overall spectral roc was marked by considerable variability and an even greater degree of distributional skewness.

To better understand lower spectral roc after imitative exposure, what should be considered are mean baseline-to-imitation changes in TL, spectral roc, and absolute vowel duration, grouped by coda type (Table 4.4).

coda	task	TL	roc	VD
t	baseline	222	2.51	152
t	imitation	238	2.25	180
d	baseline	226	2.31	168
d	imitation	241	1.82	240

coda	task	TL	roc	VD
t	baseline	214	2.56	139
t	imitation	240	2.38	169
d	baseline	223	2.21	176
d	imitation	244	1.80	238

Table 4.4. Mean trajectory length, spectral rate of change, and absolute vowel duration, depending on the voicing status of coda, in the YL group (left) and AL group (right)

We can clearly see that the drops in spectral roc were accompanied by the participants' longer absolute vowel durations in the imitation task. Even though TLs were higher in

[44] See Appendix 3 for all mean and SD values for TL and roc, for different vowels, spelling conditions, tasks, and groups.

the second task, the increases in vowel duration were more considerable, resulting in slower roc, meaning the increase in total formant trajectory did not keep up with the more considerable increase in vowel duration. This is, perhaps, most evident in the case of <bVd> words, whose longer vowels contributed greatly to increases in RVDD across the tasks, as was shown earlier. To better see how the increases in TL accompanied the increases in absolute vowel durations, Pearson correlation coefficients were calculated separately for the two tasks. In the baseline task, TL and VD appeared quite correlated (YLs: $r = 0.195$, $p < 0.001$; ALs: $r = 0.212$, $p < 0.001$), but the correlation, while still significant, was weaker in the imitation task (YLs: $r = 0.100$, $p = 0.002$; ALs: $r = 0.142$, $p < 0.001$).

Finally, what has been missing in the current analysis is how particular formants contributed to the overall vowel dynamics, as well the formants' direction of change. Below are a series of figures illustrating the participants' mean F_1 and F_2 values at four vowel points (20%, 40%, 60%, 80%) for each vowel, in the two tasks, together with the model speaker's values.

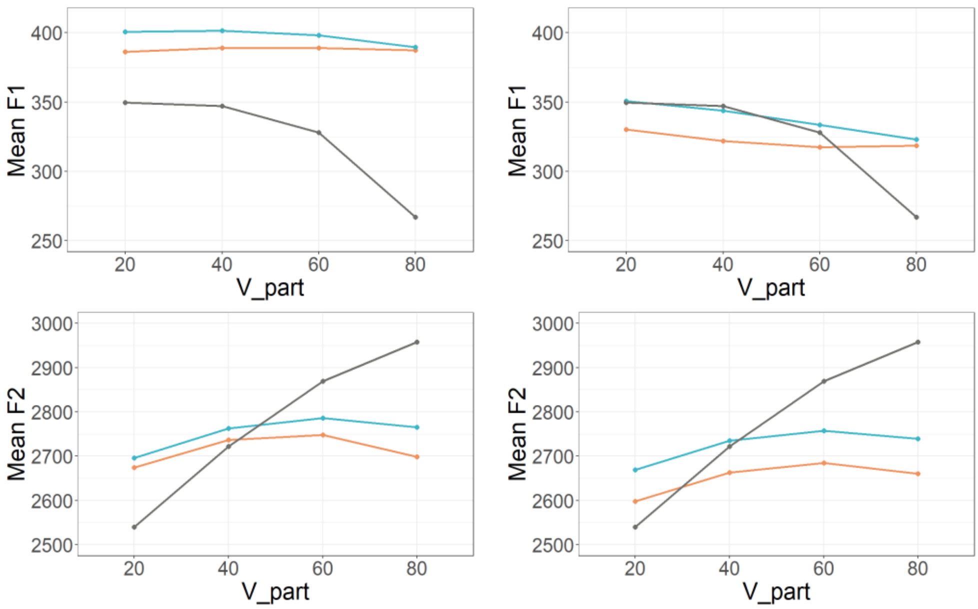

Figure 4.27. Mean F_1 (top) and F_2 (bottom) changes across four vowel points in FLEECE in baseline (orange), imitation (blue), for YLs (left) and ALs (right), with model speaker (grey)

The figure above shows that the participants' F_1 and F_2 dynamics for the FLEECE vowel were not visibly affected by the model's much more volatile formant shifts. The only discernible baseline-to-imitation change was observed in the case of F_1 in the third section, where there was a slight shift in the formant's direction across the two tasks, following the model's (much greater) downward movement. The steep upward movement of the model's F_2 throughout the vowel did not seem to result in the participants' following that trend, with the exception of YLs, whose F_2 decline in the last section became somewhat less pronounced.

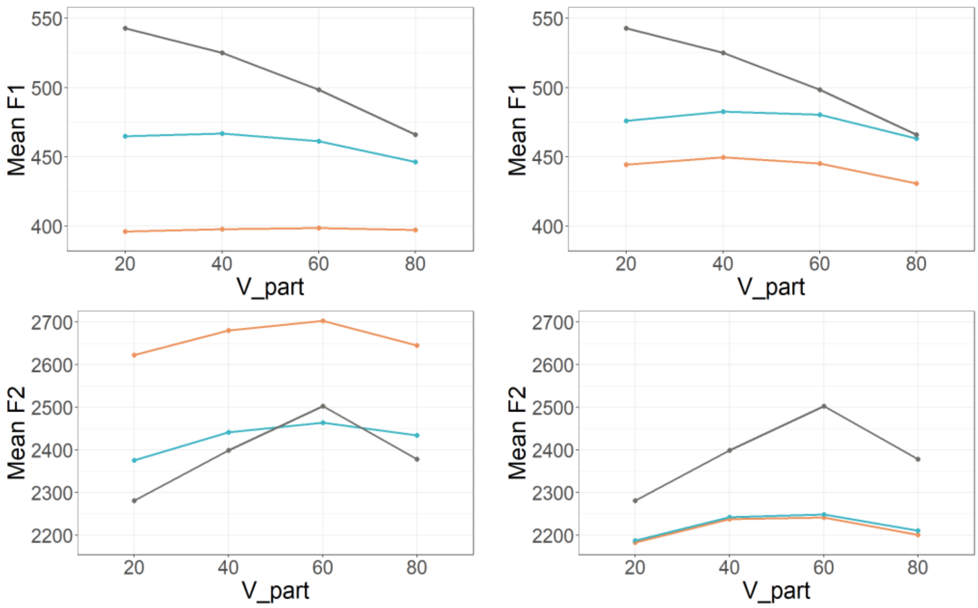

Figure 4.28. Mean F_1 (top) and F_2 (bottom) changes across four vowel points in KIT in baseline (orange), imitation (blue), for YLs (left) and ALs (right), with model speaker (grey)

While the pattern of formant movements in KIT, shown in Figure 4.28, remained rather unchanged for the AL group, the YL group displayed a greater imitative effect in this regard. Their relatively flat F_1 trajectory in the baseline task turned into a slightly downward slope, which conformed with the model's (steeper) falling trajectory. In accordance with previously observed F_2 mid-point shifts in KIT for this group, F_2 values at all vowel points shifted considerably towards those of the model. Their positions, however, relative to each other, remained hardly changed across the two tasks, although their trajectory's shape did not deviate considerably from that of the model.

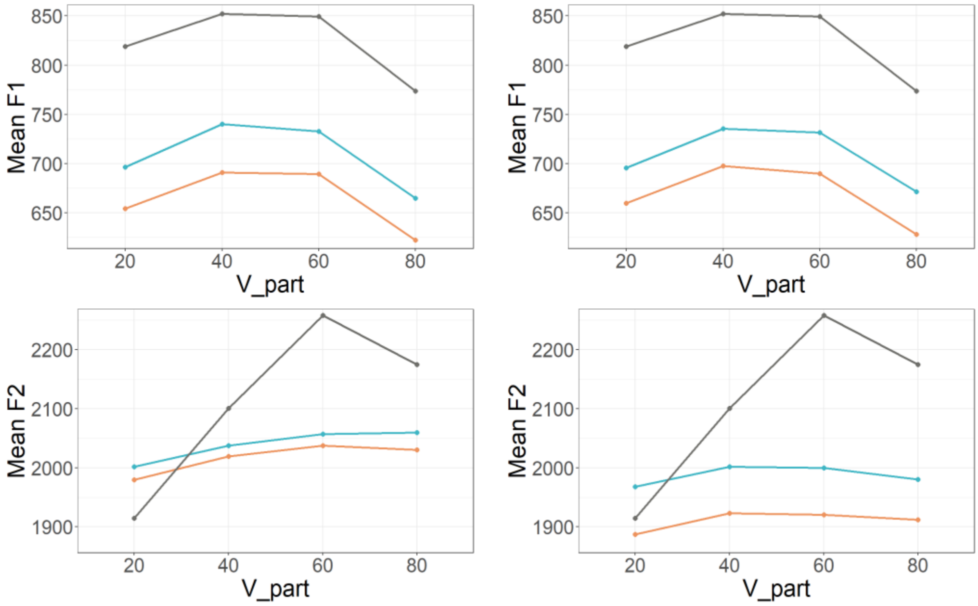

Figure 4.29. Mean F_1 (top) and F_2 (bottom) changes across four vowel points in DRESS in baseline (orange), imitation (blue), for YLs (left) and ALs (right), with model speaker (grey)

When it comes to F_1 trajectories of the DRESS vowel of both groups, shown in Figure 4.29, the participants' patterns were very similar to those of the model speaker, and unsurprisingly, showed no changes across the two tasks. As for F_2 values for the same vowel, both groups retained their somewhat flat F_2 trajectories, despite the model's very distinct upward movement in the first two vowel sections, followed by a rapid decline in the last section.

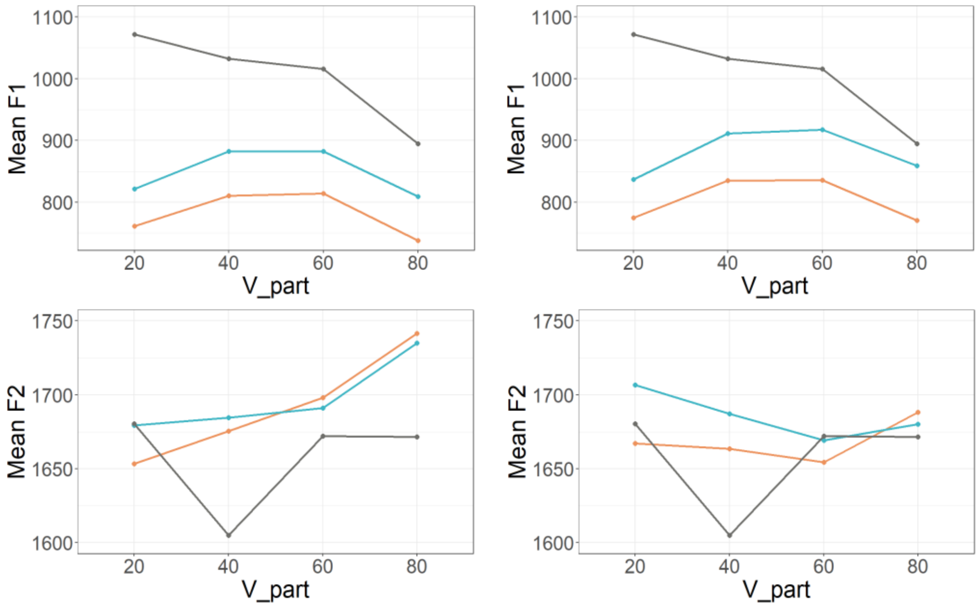

Figure 4.30. Mean F_1 (top) and F_2 (bottom) changes across four vowel points in TRAP in baseline (orange), imitation (blue), for YLs (left) and ALs (right), with model speaker (grey)

Finally, Figure 4.30 provides mean formant values at the TRAP vowel's four points. As for F_1, both groups seem to have remained rather unaffected by the model's shifts, particularly in the first section, where she displayed a downward movement, while the participants persisted in their upward movement. More distinct baseline-to-imitation changes emerge in TRAP's F_2 values, which appear somewhat distorted after exposure to the model speaker, but in an unexpected way. The model displayed a sudden drop from 20% to 40% of the vowel's duration and an almost equally rapid upward recovery between 40% and 60%. The participants, however, increased their F_2 values at the model's bottom at 40%, relative to the baseline. At the same time, following the model speaker, their F_2 values at 20% increased visibly, particularly in the case of adults, whose F_2 decline in the first two sections became steeper after imitative exposure.

4.4.2.1. Statistical significance

To determine whether the participants' vowel dynamics changed significantly across the two tasks, two most relevant metrics were selected as response variables for the statistical models, i.e. trajectory length `TL` and spectral rate of change `roc`. As remarked earlier, both variables manifested positive skewness in both groups, particularly in the case of adults' `roc`. To assure the satisfaction of the statistical models' assumptions, the data were square-root transformed, which greatly improved their resemblance to the normal distribution. A series of linear mixed effects models were fitted, separately for `TL` and `roc`, and separately for the two age groups:

```
model_0: TL/roc ~ 1 + (1+task|subject)
model_1: TL/roc ~ task + (1+task|subject)
model_2: TL/roc ~ task + vowel + (1+task|subject)
model_3: TL/roc ~ task × vowel + (1+task|subject)
model_4a: TL/roc ~ task × spelling + vowel + (1+task|subject)
model_4b: roc ~ task × vowel × spelling + (1+task|subject)
```

The comparison between `model_0` and `model_1` showed that adding the predictor of `task` significantly improved the model's explanatory power and overall fit to the data for both for `TL` and `roc`:

YLs' `TL`: $\chi^2(1) = 7.53$, $p = 0.006$

YLs' `roc`: $\chi^2(1) = 23.85$, $p < 0.001$

ALs' `TL`: $\chi^2(1) = 12.96$, $p < 0.001$

ALs' `roc`: $\chi^2(1) = 10.16$, $p = 0.001$

This means that both young and adult participants' vowel dynamics, expressed as the combined magnitude of formants shifts across the three vowel sections, were significantly higher in the imitation task, while the rate at which these shifts happened was significantly lower in that task. Adding the predictor of `vowel` in `model_2` improved the model further, suggesting significantly different degrees of vowel dynamics across the four vowels:

YLs' `TL`: $\chi^2(3) = 98.02$, $p < 0.001$
YLs' `roc`: $\chi^2(3) = 35.15$, $p < 0.001$

ALs' `TL`: $\chi^2(3) = 286.58$, $p < 0.001$
ALs' `roc`: $\chi^2(3) = 107.47$, $p < 0.001$

To determine whether the effect of `task` on `TL` and `roc` depended on `vowel`, the interaction term `(task × vowel)` was included in `model_3`, which produced the following results:

YLs' `TL`: $\chi^2(3) = 4.98$, $p = 0.174$
YLs' `roc`: $\chi^2(3) = 11.54$, $p = 0.009$

ALs' `TL`: $\chi^2(3) = 1.31$, $p = 0.728$
ALs' `roc`: $\chi^2(3) = 3.64$, $p = 0.303$

When it comes to the effect on `TL`, the inclusion of the interaction term did not significantly improve the model fit when compared to the model without the interaction, suggesting that both `task` and `vowel` had an independent effect on `TL`. As for `roc`, the effect of `task` on `TL` depended on `vowel` type in the case of the YL group, but not in the AL group.

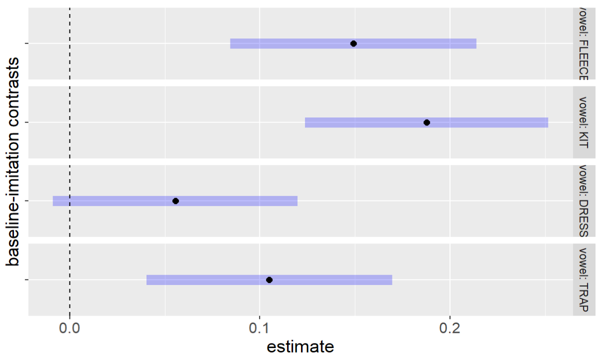

Figure 4.31. Spectral `roc` pairwise contrasts of `task` levels (baseline vs. imitation) for each level of `vowel` (FLEECE, KIT, DRESS, TRAP) for the YL group

The post-hoc pairwise comparisons were performed only for YLs' `roc`, which provides more insight into the significant interaction found between `task` and `vowel`. As can be seen in Figure 4.31, positive contrast estimates,[45] related to decreases in roc across the two task levels, were evident in all but one vowel, that being DRESS.

The inclusion of the factor of spelling in model_4a (for YLs' TL, ALs' TL, ALs' roc) and in model_4b (for YLs' roc) did not yield significant results:

YLs' TL: $\chi^2(2) = 0.13$, $p = 0.937$ ALs' TL: $\chi^2(2) = 1.13$, $p = 0.570$
YLs' roc: $\chi^2(8) = 7.43$, $p = 0.491$ ALs' roc: $\chi^2(2) = 0.54$, $p = 0.763$

The tests confirmed the significance of the main observations made earlier. The total magnitude of formant movement, expressed as trajectory length (TL), increased significantly across the two tasks in both groups, although this was not immediately clear from the scatter plots in Figure 4.21. While, when considered independently, TL differed significantly depending on the specific vowel, the baseline-to-imitation increases in TL did not vary significantly across different vowels. This is in line with Figure 4.22, where the across-the-task TL increases seemed consistent for all vowels in both groups. As for the rate at which formant shifts happened (spectral roc), it decreased significantly in both groups, which is consistent with the observations made based on Figure 4.25. While roc varied significantly across the vowels, only in the case of young participants was there a significant impact of vowel type on the degree of baseline-to-imitation decrease in roc. Yet, again, in congruence with initial impressions, there were no significant differences in the degree of either TL or roc shifts across the two spelling conditions.

4.4.3. Discussion

Previous research on phonetic imitation does not appear to have accounted for the varying nature of vowels, which they tend to exhibit throughout their duration, particularly in languages such as English. The current investigation sought to fill that gap, specifically in the context of second language acquisition, where the native language of the potential imitators (Polish) differs from the model speaker's language (English) in terms of vowel dynamic specification. Polish learners of English, particularly the less experienced ones, were expected to show more Polish-like productions in the baseline word reading task, i.e. with more stable formant patterns, especially at the vowel's onset. Having been exposed to more dynamically rich vowel realisation of the model, it was anticipated that the participants would assume more English-like dynamic properties of four English

[45] Recall that these represent square-root transformed estimates of roc contrasts.

monophthongs: FLEECE, KIT, DRESS and TRAP. Overall, both young and adult Polish learners of English were found to increase the amount of formant movement after imitative exposure, although this increase was often accompanied by a decrease in the rate at which these movements were exercised.

To ascertain convergence, the first metric considered was trajectory length (TL), which refers to the sum of combined F_1 and F_2 movements throughout three vowel sections (20%-40%, 40%-60%, 60%-80%). Both young and adult participants were found to exhibit significant TL increases across the baseline and the imitation tasks to roughly similar extents across all vowels. While both groups increased the total magnitude of formant movement over the major vowel portion (20%-80%) in the imitation task, there remained a considerable participant-to-model gap in terms of TL in all vowels except for TRAP, where, ironically, the participants showed the greatest TL in both tasks. Participant-to-model distances did not seem to affect the degree of imitation, as illustrated by both FLEECE and DRESS eliciting average baseline-to-imitation shifts in TL, despite the model's outstanding TL values for these two vowels. It should also be noted that TL of both groups exhibited a considerable degree of variability.

To account for whether the rate at which formant movement takes place also undergoes imitation, spectral roc (i.e. TL divided by 60% of the vowel's absolute duration) was included in the analysis. Generally, higher spectral roc is related to faster articulatory movements employed to reach formant targets. As opposed to TL, the participants showed consistently lower spectral roc in the imitation task, despite the considerably higher model values than the participants' baselines. This apparent divergence was due to the participants' increases in TL not being high enough to compensate for the relatively higher increases in vowel duration across the two tasks. As it was revealed in Section 4.2, the participants increased vowel duration contrasts, particularly by lengthening their vowels in /bVd/ words (see also Table 4.4). It seems to align with the results of Fox and Jacewicz (2009), who observed that spectral roc is higher in the case of vowels preceding voiceless consonants than when preceding voiced ones. This seems to suggest that while Polish learners of English can imitate English-like overall magnitude of formant movement, their ability to do so is accompanied by the extension of vowel duration, leaving in question whether they would be able to increase spectral dynamics within similar temporal constraints. However, it appears that the participants displayed a much lesser correlation between trajectory length and absolute vowel duration in the imitation task than they did in the baseline task, which indicates that they managed to increase vowel dynamics not merely through the manipulation of vowel duration.

To obtain a more detailed picture showing which of the three vowel sections contributed the most to the overall dynamics, the magnitude of formant movement for each of the three vowel sections (vowel section length or VSL) was examined. While the model manifested rather comparable VSL for particular sections, with slightly greater magnitude in the first section, both groups showed more variability, favouring more spectral movement in the first and the last vowel section. As a result of the imitation task, the participants' movement was higher in the first and the second section, but not in the third section. The fact that Polish speakers already displayed relatively substantial baseline dynamics in the first vowel section (20%-40%) is surprising, as according to previous research (e.g. Schwartz and Kaźmierski 2020), more Polish-like habits should encourage more stable patterns in the initial vowel portions. However, the fact that, in the imitation task, they showed enhanced dynamics in the two initial portions (20%-40% and 40%-60%), but not in the final portion (60%-80%), seems to suggest that there was more room for improvement in the early vowel sections. On the other hand, their already dynamic last section, following the typically more robust Polish-like spectral movements in that section, remained unchanged after imitative exposure, possibly as a result of there not being any further perceived distance between the participants and the models.

The inspection of overall trajectories for particular formants for given vowels revealed that FLEECE and KIT remained relatively flat along the F_1 dimension, but they exhibited a somewhat U-turn shape in terms of F_2. The opposite pattern was observed for DRESS and TRAP, where F_1 tended to increase towards the middle section and then return to its starting point, whereas F_2 shifts were not so pronounced, specifically for DRESS. However, although it was not statistically tested, it did not seem that a given formant contributed more to the overall vowel dynamics more than the other. Nor did it appear that the trajectories of one of them tended to become more model-like after exposure. Still, an interesting case is that of the TRAP vowel, which exhibited considerable overall TL relative to the other vowels, and it improved further in the imitation task, despite the model's unremarkable values. TRAP's overall duration was comparable to that of FLEECE, which had the lowest TL, so the factor of vowel duration, found previously to correlate with TL, was unlikely to be responsible. Only a tentative explanation can be offered to account for this, which may lie in the participants' indecision as to which Polish vocalic quality the target English vowel corresponds to.

Since, as remarked earlier, the English vowel tends to be substituted for by either Polish /ɛ/ or /a/, it may be possible that this ambiguity emerged throughout the vowel’s production, contributing to its varying quality, and hence greater formant dynamics. It is also worthy of attention that their overall F_1 trajectories for TRAP were close to their F_1 trajectories for DRESS, although the absolute F_1 values at given vowel points were higher for TRAP, suggesting a greater degree of openness. Interestingly, the same observation was not made for F_2, whose patterns did not resemble those of any other vowels analysed.

It was also of interest whether the factor of age (and L2 proficiency) visibly impacted the degree of imitation of vowel dynamics. Both young (less proficient) and adult (more proficient) learners showed remarkably similar both overall dynamic patterns and imitative performance. Previous studies (Schwartz 2015; Schwartz et al. 2016a; Schwartz and Kaźmierski 2020) have consistently shown that more proficient Polish learners of English exhibit more robust formant dynamics than less proficient ones. It was therefore expected that the (more proficient) adult participants in the current study would show greater overall dynamics at least in the initial vowel portions, but the current results do not conform with the previous evidence. There does not appear to be a satisfactory explanation for the current results. Judging by the younger group’s worse baseline performance as regards vowel durational contrast and static vowel quality, as established earlier, it does not seem likely that they should suddenly exhibit comparable performance with that of more proficient adults in their realisation of VISC. It could be speculated that the adults’ advantage, stemming from their greater L2 proficiency and experience, was offset by the potential advantage of younger age of the other group. Nittrouer (2007) found that native English children rely more heavily on dynamic spectral structure to recognise English vowels than adults. Children were also found to exhibit more variable formant trajectories than adults (Lee et al. 2004). However, although Assmann et al. (2013) observed a general tendency for absolute formant frequencies to decrease with age, the patterns of formant movements were fairly consistent between children and adults. In light of this, while it does not seem very likely that younger age favours a more robust realisation of VISC, especially in the case of native Polish children, the fact that more proficient adult learners are superior to less proficient learners of roughly the same age (as the studies of Schwartz and colleagues have established) hints at a tentative advantage of young participants in terms of dynamic vowel productions.

In sum, Polish learners of English were shown to be capable of modifying spectral dynamics in their realisation of the four front English vowels as a result of exposure to a native English speaker. This modification took the form of increased trajectory length, denoting the total magnitude of formant movement over the three middle vowel sections (20%-40%, 40%-60%, 60%-80%). This increase was accompanied by a significant increase in vowel duration across the two tasks, and a consequent reduction of spectral rate of change, which means that while the participants exhibited greater vowel dynamics, they apparently needed to lengthen their vowels to accomplish that. Still, the across-the-task increases in the first and the second section, but not in the third section, and a lesser correlation between trajectory length and vowel duration in the imitation task both indicate that changes in dynamics were not only due to vowel lengthening. Both groups displayed surprisingly similar performance in their overall realisation of vowel dynamics, which does not fully align with previous findings. Their overall performance was particularly surprising in the initial vowel section (20%-40%), which was comparable to the last vowel section considered (60%-80%), although only the former one (along with the middle one) underwent significant improvement across the tasks. Again, the factor of spelling, although considered, did not seem to play the slightest role in the participants' performance in terms of VISC imitation. Contrary to the two previous tested parameters, the participant-to-model distances appear to have been less relevant here. Despite the model having similarly high values for the three vowel sections, the participants' response, particularly in the form of only slight increases in the second section, left much of the participant-to-model distance unbridged. Surprisingly, the TRAP vowel, manifesting the lowest dynamics in the model's productions, displayed the highest baseline dynamics, which further increased in the imitation task, potentially due to the vowel's ambivalent nature from the point of view of the Polish learner. Finally, it should be recognised that the data exhibited a notable degree of skewness to the right, resulting from a number of outlying high TL values, which suggests that vowel dynamics, operationalised as trajectory length, are not firmly established across the participants' productions.

Chapter 5

General discussion and conclusions

The purpose of the current investigation was to ascertain whether or not, and to what degree, Polish learners of English can imitate durational and qualitative properties of English vowels. Because the two languages concerned differ in these two respects, it was expected that, after auditory exposure to a native English model, the learners would temporarily shift away from their more Polish-like realisations of these properties towards more English-like ones. It was also of interest whether age, together with second language proficiency and experience, affects imitative performance, which was predicted to be greater in the case of young (less proficient) learners, due to potentially greater initial participant-to-model distances. Finally, because of the known influence of orthography on the pronunciation of second language learners, it was anticipated that those learners who saw the words in the imitation task would be more likely to exhibit more Polish-like vowel characteristics.

To answer these questions, both children (teenage) and adult learners were enlisted to participate in the experiment, in which they first read, and then repeated English words spoken by a native English model. Having compared the participants' production in the two tasks, it was revealed that, on the whole, Polish learners of English can imitate both durational and qualitative vowel properties of the latter language, following the English model's speech. While the younger (less proficient) group exhibited overall greater degrees of shifts across the baseline and the imitation tasks, the adult (more proficient) group could be considered as equally successful in imitating the features concerned, considering they already displayed more native-like performance in the baseline task. Contrary to what was expected, the role of orthography in the imitation task was found to be insignificant, potentially due to its much more subtle effect, which the current study failed to capture.

5.1. Tested parameters

In an attempt to ascertain imitative effect on the participants' production of vowels, the main research question was posed:

To what extent do native Polish learners of English imitate the properties of English FLEECE, KIT, DRESS, and TRAP vowels, specifically with regard to:

a) vowel duration contrast as a cue to the voicing of the coda consonant,
b) vowel quality from the static perspective, as manifested by the first and second formant values, located at vowel mid-points,
c) vowel quality from the dynamic perspective, as manifested by the magnitude of total formant shifts across multiple vowel intervals, along with their rate of change?

In line with previous findings of Zając (2013), Zając and Rojczyk (2014), and Rojczyk et al. (2022), vowel durational contrasts, which were expected to be less sharp in the word reading task, exhibited significant increases in the imitation task. These results agree particularly with those of Rojczyk et al. (2022), where the contrasts were considered in relative terms, i.e. by indirectly accounting for speech rate, word duration in particular, which itself varied across the tasks. The realisation of sharper contrasts was largely due to the increases in relative vowel durations in /bVd/ words, confirming that the learners' baseline productions of these words were more Polish-like, i.e. the coda consonants in them were subject to devoicing, without much regard for the duration of the preceding vowels. Of the four front vowels considered, FLEECE and TRAP were the ones that elicited the most profound baseline-to-imitation shifts, although it does not appear that there was anything inherently special about the vowels that led to this. Rather, it was the relatively high participant-to-model baseline distances for the two vowels that led to these particular vowels eliciting the greatest changes across the two tasks.

The significant increases in durational contrasts were accompanied by overall shifts in vowel qualitative properties, as manifested by their F_1 and F_2 mid-point value changes. The results agree with previous findings (Rojczyk 2013) that vowels such as TRAP, whose quality can be considered as somewhat distinct from its two closest Polish neighbours, undergo convergence. The same was observed for KIT, particularly in the case of children, since their baseline realisations tended to be more FLEECE-like, leaving a good distance to cover in the imitation task. Predictably, the quality of FLEECE itself, which is quite similar to Polish /i/, remained largely unchanged with the exception of F_2 for adults. Contrary to initial predictions, the DRESS vowel, which generally tends to be replaced with a very similar Polish /ε/, elicited significant shifts, particularly in the case of adults, who increased both formants. Overall, the learners followed the model mid-point values, shifting their baseline vowel quality realisations towards more fronted and open productions.

Vowel quality was also considered from a dynamic point of view, by accounting for formant changes across three vowel sections (20%-40%, 40%-60%, 60%-80%). Based on the evidence indicating that Polish is characterised by more stable formant patterns than it is the case in English (Schwartz 2020; Schwartz and Kaźmierski 2020), particularly in the initial sections, it was predicted that Polish learners would demonstrate more robust vowel dynamics in the imitation task, following the native English speaker's patterns. Although the participants productions were highly variable, they significantly increased their trajectory length, indicating the total formant movement over the three vowel sections. While they assumed more model-like values for trajectory length, it was achieved at the expense of slower rate at which formant movements were executed (lower spectral roc), as evidenced by the accompanied extension of vowel durations. Interestingly, the participants' first vowel section (20%-40%) was already dynamically rich, together with the third one (60%-80%), but only the former was found to show even greater dynamics in the imitation task, suggesting there was room for improvement in this respect. The middle section (40%-60%), which turned out to be the most stable one, exhibited significant, albeit still mediocre, increases, considering the model's high mean values. The baseline-to-imitation increases in trajectory length were comparable across all four vowels, but the TRAP vowel proved to be rather exceptional, as it had the greatest overall dynamics, which were convergent with the model values from the outset.

Overall, the current results confirm that second language learners, whose native language is rather simplistic in terms of its vocalic system relative to the one of the target language, are capable of assuming more target-like vowel realisations, as a result of successful perception and reproduction of the features concerned. The imitative effect was not only evidenced by the learners' ability to adjust their temporal parameters, which may be considered rather conspicuous, but also more nuanced qualitative aspects of vowels. Although qualitative convergence was sometimes selective and it was almost never complete, the study demonstrated that Polish learners of English can detect vowel timbre in the target language, and modify their more Polish-like productions to approximate the English-like ones. The current investigation enhanced our understanding of imitation of L2 vowels, by showing that more vowels are subject to imitation, beyond those established in previous research. It also showed that more subtle qualitative aspects are worthy of being considered, as was demonstrated by the participants' overall successful improvement of vowel dynamics in the imitation task.

5.2. Factors influencing phonetic imitation

The current study adds to the body of evidence showing that phonetic imitation takes place not only in socially-rich contexts, but also in non-interactive settings, where the participants repeat what another person, separated from them both in time and space, has said. While there may not be an apparent motive to exhibit imitation in such a context, other than being instructed to do so, there are a number of factors that may affect whether or not, and to what degree, convergence is observed, as was discussed in Chapter 1, in Section 1.3. Despite the non-interactive nature of the tasks, most second language learners are expected to be driven to imitate not only by merely being told to do so, but also by the desire to improve their English pronunciation. Also, since the current participants were accustomed to this way of pronunciation practice throughout their English classes, i.e. though imitation, it may be supposed that this habit beneficially affected their imitative performance. What is more, studies have shown that people with established authority may be more prone to being imitated (Giles 1973), and native speakers of the target language, by virtue of having an obviously greater command of the target language, may be loosely considered as such. From among less socially-mediated factors, word frequency or the number of repetitions, found to affect the degree of convergence by some authors (Goldinger 1998; but see Black 2012; Shockley et al. 2004), do not seem to have been crucial factors in the current investigation, although these were not explicitly tested, given a small number of words. The tendency for low vowels to be imitated more than others (Babel 2012) seems partially confirmed by the current data, particularly in the case of the TRAP vowel, which exhibited mostly clear baseline-to-imitation shifts in all three tested parameters.

5.2.1. Initial phonetic distance

One of the factors that was predicted to play a role in the current investigation was one related to initial phonetic distance between the participant and the model (Nycz and Mooney 2017), which may also be connected with perceptual salience of features to be imitated (MacLeod 2014). This may be particularly relevant in the context of L2 speech, where, according to one of the models accounting for the difficulties in the perception of

L2 speech, the success in the acquisition of target sounds is dependent on there being enough acoustic distance between a given L2 sound and its closest counterpart in L1 (Flege 1995; Flege and Bohn 2021). It was therefore expected that, in cases where the baseline participant-to-model distances were long, a seemingly greater imitative effect would be observed, as opposed to the situations in which the participants already converged to the model speaker. Overall, the current data confirmed this tendency, with some exceptions, but the model values, perhaps expectedly, were hardly ever reached by the participants, unless they were already approximately matched initially.

What should be considered is how the issue of initial phonetic distance affected the participants' performance for given parameters. In particular, vowel duration contrasts were most robustly imitated in the case of FLEECE, which coincided with high model durational contrasts for this vowel, but this was not observed for TRAP, which, despite high model values, did not lead to impressive baseline-to-imitation shifts, particularly in the case of the younger group, for which no clear explanation presents itself. It could only be speculated that, because of the inherently long duration of TRAP relative to the higher vowels, further TRAP lengthening in *bad* was impeded.

With regard to vowel quality, the analysis of the current data revealed a rather clear tendency of the participants to be attracted by the model's higher mid-point formant values. Small initial distances, for example in the case of the young group's FLEECE, resulting from the vowel's close qualitative proximity to Polish /i/, did not exhibit visible shifts across the two tasks, whereas their F_2 values for KIT underwent considerable changes due to their baseline Polish /i/-like realisations of KIT. An interesting case was that of DRESS, which was not expected to undergo significant qualitative changes, because of its similarity to Polish /ɛ/. The relatively distinct model values, however, allowed the participants to detect the qualitative contrast, and to shift their mid-point formant values, with the exception of F_2 for the younger group. Expectedly, the TRAP vowel displayed quite observable shifts, specifically in terms of F_1, where the initial distance was large, but F_2 was also found to be affected, although in a less straightforward manner, due to the ambiguity in the participants' baseline realisations.

Finally, participant-to-model distances appeared to be less relevant with regard to the imitation of vowel dynamics. Polish learners increased their trajectory length rather indiscriminately across all vowels, regardless of the model's varying levels. Despite considerable distances for DRESS, its baseline-to-imitation shifts were comparable to those of TRAP, the least dynamically rich vowel of the model speaker. It could be argued that the marked degree of variability in the participants' responses contributed to the obtained results, in that, while some learners may have adequately followed the mean model values, the pattern was obscured by other participants actually not showing significant dynamics, which would be in line with the skewed shape of the distributions for the overall realisation of trajectory length.

The issue of initial phonetic distance can lead to problems when comparing imitative performance across different groups or vowels. The often-used difference-in-distance metric in imitation studies can sometimes encourage forming premature conclusions, as greater starting distance can give rise to apparent greater convergence, whereas when the distance is short, seemingly no convergence may be ascertained to have occurred.[46] For example, it would not be reasonable to conclude that the vowel durational contrast for the pair *beat-bead* is more imitable due to a potentially unique nature of the FLEECE vowel, as the current data may seem to have suggested. The considerable baseline-to-imitation shifts for this vowel should be ascribed to profound initial participant-to-model distance for FLEECE, resulting from high model values. Conversely, when the participants' values happen to already approximate the model's values in the baseline condition, where there appears to be no further possibility for convergence to occur, as was sometimes the case in the current study, rather than stating that no convergence has taken place, it seems more reasonable to leave the question of whether it has occurred or not unanswered. Convergence, or lack thereof, becomes only clear when there is sufficient initial participant-to-model distance, which is or is not significantly shortened in the imitation task. In light of this, the current results should be interpreted according to the following scenarios:

- if the participants significantly reduced the initial participant-to-model distance, convergence is interpreted to have occurred,[47]
- if, despite sufficient initial participant-to-model distance, the participants did not significantly shift their baseline values, no convergence is said to have occurred,
- if the participant-to-model distance was initially very short, and the participants did not significantly reduce the distance, convergence (or lack thereof) cannot be determined,[48]
- if two vowels or two participant groups had different degrees of initial participant-to-model distance, and both vowels or both groups were found to significantly reduce

[46] See MacLeod (2021) for a review of the problem of Starting Distance Bias.
[47] Imitation is judged to have equally occurred regardless of whether the initial participant-to-model distance was short or long, as long as the reduction of the distance was significant.
[48] Alternatively, if the short distance remains unchanged across the tasks, maintenance could be argued to have occurred, whereas, if the distance increased, divergence seems to be the most intuitive assessment.

the distance, but to varying extents, caution should be exercised in judging the relative imitative performance for the two vowels or the two participant groups.

The fact that features exhibiting shorter initial participant-to-model distance are less likely to show clear imitation patterns seems in line with the predictions of the SLM model, in the context of L2 speech acquisition. If the distance is small, the learner is unlikely to notice the subtle difference, and is therefore less likely to reproduce it. Interestingly, however, the mid-point quality of the DRESS vowel in the current data constitutes a conflicting case. On the one hand, the vowel is perceptually quite similar to Polish /ε/, but, on the other hand, the model's DRESS showed rather distinct formant values from the participants' baselines.[49] Despite the impressionistic similarity of the two vowels, the participants, on the whole, were able to make use of the available participant-to-model mid-point formant distance, which suggest that the factor of initial distance may supersede other factors affecting imitative performance. Finally, the issue of similarity poses another challenge in the context of VISC. Schwartz and Kaźmierski (2020: 248) observe that the evaluation of similarity between vowels should take into account qualitative aspects across various vowel portions, which may have perceptual implications. For example, because of the similarity between FLEECE and KIT or DRESS and TRAP in the early vowel portions, Polish learners of English, who tend to base their qualitative percept on these portions, may confuse the two vowels in the pairs mentioned for this very reason, meaning a more fine-grained approach to the issue of similarity is warranted.

5.2.2. The effect of age and L2 proficiency

To offer a more comprehensive view on the issue of phonetic imitation in L2 speech, both children (early teenagers) and adults (early adults) were included in the experiment. As discussed in Chapter 1, in Section 1.5, previous research has yielded inconclusive results with regard to the factor of age in imitative performance. What compounds the difficulty, particularly in the context of imitation in L2 speech, is the issue of L2 proficiency or experience, which is not infrequently intertwined with age. Another problem that emerges is how to reliably compare imitative performance across two groups, where one group

[49] The model's different formant values should, perhaps, not be surprising, considering the values were not normalised, as justified earlier. What should be more surprising, however, is that the participants were attracted to these values, which may suggest that raw formant values carry important information that learners attend to, which may have remained undetected had the formant values been normalised from the outset.

may display better baseline performance, and thus shorter initial phonetic distance, while the other could demonstrate much greater initial participant-to-model distance. In light of this, it was not the intention for the current investigation to answer the question which group performed better, but rather to broaden the research perspective and make the conclusions more generalisable. The only prediction afforded was that the younger, and, simultaneously, the less proficient group, would show more Polish-like realisations in the baseline task than the more proficient adult group, which would result in greater initial phonetic distance for the former group to be potentially bridged in the imitation task.

The prediction was mostly confirmed, depending on the tested parameter under consideration. With regard to vowel duration contrasts, the younger group displayed much worse baseline durational distinction than the adults did, which resulted in much greater shifts across the two tasks by the former group, by virtue of having more room for improvement in the imitation task. Notably, despite the fact that the adult group showed overall better durational contrasts, they still made significant baseline-to-imitation shifts across the two tasks. One vowel that stood out in terms of between-group differences was TRAP, which exhibited rather modest durational changes for the young group, despite high model values, but showed more improvement in the case of adults.

With respect to baseline-to-imitation shifts in vowel quality, at first glance, there do not appear any striking differences, particularly along the F_1 dimension, where both groups exhibited comparable changes in KIT, DRESS, and TRAP. Greater disparity emerged along the F_2 dimension, specifically for KIT, which was much more FLEECE-like or Polish /i/-like in the case of the younger group's baseline values, which shifted considerably towards more native values after exposure. Despite clearly higher F_2 model values for KIT, the adults did not exhibit further fronting in the imitation task. An interesting case was that of FLEECE and DRESS, the two front vowels that are considered most similar to their closest Polish equivalents. While the younger group made no observable F_2 adjustments for these two vowels, despite the distance available, the adults capitalised on the distance, and produced more fronted versions of FLEECE and DRESS in the imitation task. This hints at the possibility that adults, or more proficient and experienced L2 learners, may be more attuned to finer phonetic differences, and they may be more likely to reduce the disparity between their default qualitative vowel realisations and those of the model.

When it comes to the final parameter tested, the two groups showed remarkably similar performance, both in their overall vowel dynamics, as well as in terms of baseline-to-imitation shifts. Previous studies on VISC among Polish learners of English (Schwartz 2015; Schwartz et al. 2016a; Schwartz and Kaźmierski 2020) revealed, rather consistently, that more proficient (adult) learners exhibit more robust formant dynamics than less proficient (also adult) learners. No effect of potential advantage of higher proficiency seems to have been confirmed by the current data, with one tentative exception as regards the spectral rate of change, which manifested somewhat lesser baseline-to-imitation drops for FLEECE and KIT for the more proficient group. It may also be possible that the apparent similarity is due to greater variability reported for children's realisation of formant trajectories (Lee et al. 2004), and indeed, the current data was somewhat more variable in the case of the younger and less proficient group.

In sum, on the whole, both young (less proficient) and adult (more proficient) learners of English were found to imitate vowel duration contrasts and their qualitative properties. In some cases, expectedly, the younger participants showed worse baseline performance than adults did, which resulted in overall greater baseline-to-imitation shifts across the two tasks. These greater shifts should not, however, be considered as evidence of better imitative performance of younger learners, since greater initial phonetic distance is naturally expected to lead to greater convergence, as discussed earlier. The adults' high baseline values can be considered as ones that already partly shifted in the first task, as a result of being instructed to produce words in English, potentially limiting the possibility for further adjustments in the imitation task. Notably, despite their higher baseline values (with the exception of VISC), the adults generally still exhibited significant baseline-to-imitation adjustments. They appeared to be more attuned to fine phonetic features of the model speaker, particularly in the case of greater lengthening of TRAP and higher second formant values for FLEECE and DRESS, which the younger participants failed to appreciate. Overall, the seemingly better imitative performance of young learners (Nielsen 2014), as manifested by greater reduction in participant-to-model distance, can be judged to have been matched by the performance of adult learners (Alivuotila et al. 2007), by showing greater attunement to subtle phonetic differences.

5.2.3. The effect of orthography

The current investigation predicted that those Polish learners of English who were exposed to both the spoken and printed representations of the stimuli would be more likely to retain their L1-like realisation of vowel properties, by being driven by the supposed strong associations between vowel letters and Polish vowels. The prediction was based on the possibility that speech perception, L1 speech specifically, may leave traces that are activated during written word recognition (Dufour and Nguyen 2013) and that seeing a word may activate both graphemic and phonemic representations (Ziegler and Ferrand 1998). The question was also warranted in view of the ample evidence indicating that orthography interacts with speech perception and production in the second language, as was discussed in Chapter 3, particularly when the two languages concerned differ in terms of orthographic depth.

The current findings consistently showed that the presence versus absence of written word representations in the imitation task does not affect the degree of convergence, regardless of which acoustic parameters the participants were tested on. While no definite predictions were offered with regard to vowel durational contrasts or vowel dynamics, the participants' realisation of vocalic mid-point quality was anticipated to be more Polish-like in the multimodal condition. The effect was expected to be particularly profound for the KIT vowel, encoded by <i>, as this letter is associated with more FLEECE-like quality in Polish, in similar contexts. Although some subtle differences were noted for adults' productions of the FLEECE and KIT vowels across the two conditions, overall, it cannot be stated that spelling visibly affected imitative performance of any of the two groups.

A few, not mutually exclusive, explanations are proposed as to the lack of orthographic effect on the degree of convergence. One of the reasons for the participants' apparently comparable performance across the two spelling conditions may lie in the study's failure to isolate the effect of orthography, since all participants were exposed to the written forms of the stimuli in the baseline task. Thus, the participants who only heard the words in the imitation task may be suspected of still having been affected by the prior activation of graphemic traces in the reading task.

It could also be argued that the affect of orthographic input was more subtle and global, in the sense that orthography-induced non-target-like productions may manifest themselves not only when the speech is read out, but they may constitute an inherent feature of L2 speech, even when it is spontaneous (Bassetti 2008). Bassetti et al. (2022: 683) conclude that, once established, the effect of orthography may be impervious to change, despite naturalistic exposure or instructional intervention. In the current study, the effect was particularly observable in the younger group's realisation of the KIT vowel, which was much more FLEECE-like in the baseline task, due to the strong association between <i> and /i/ in Polish. However, as remarked earlier, this association was overridden by the acoustic input in the subsequent task, regardless of the spelling condition.

Finally, while it could be argued that additional access to orthographic input, which naturally makes the participants more certain as to what they are supposed to imitate, may prompt them to rely less on the auditory layer of the input, it should be acknowledged that such certainty may also favourably affect the degree of imitation. In the face of highly variable acoustic layer of the L2 input across dialects and speakers, printed forms, due to their relatively immutable nature, may be considered as fixed reference points, at which the learners can orient themselves. Bearing in mind that orthographic input may have a beneficial influence on L2 learners' pronunciation, as was discussed in Chapter 3, the potentially negative effect of orthography that was predicted may have been thus neutralised in the current study. Therefore, although the current results did not reveal a clear effect of orthography on the degree of imitation, it may be that the investigation failed to capture potential subtleties and complexities of the issue.

5.3. Pedagogical implications

Imitation has been considered an important factor in the acquisition of second-language speech (Rojczyk et al. 2013), and it has played a crucial role in pronunciation teaching (Dziubalska-Kołaczyk et al. 2014). It may be considered as particularly useful when it comes to the acquisition of features pertaining to vowels, in which case the role of explicit instruction, as to how exactly these sounds should be produced, is particularly limited. However, the teacher's ability to provide adequate feedback on whether imitation has actually taken place is rather minimal when students repeat words chorally. Even when the students' productions are assessed individually, the teacher is often left to rely on nothing else but on their own subjective sense of judgement, which may not be seen as an accurate assessment of reality.

The current study confirmed, in an objective manner, that learners are capable of temporarily adjusting their vowel parameters to approximate the vocalic features of the model speaker. Convergence, however, was, quite expectedly, almost never complete, in that the participant-to-model distances were rarely bridged entirely. As the current results have generally confirmed, participant-to-model distances are important for the learners to perceive the contrast between their baseline and the target realisations of the features concerned. What may, therefore, help in their acquisition is the exaggeration of the model feature that is intended for practice, which may be particularly useful in the case of durational contrasts. Moreover, although it was not explicitly studied, previous studies, as well as the current data coming from the initial four model speakers, suggests that vowel dynamics tend to exhibit inter-speaker variability. Consequently, it may be advisable to ascertain a satisfactory, if not an exaggerated, level of vowel dynamics in the model's productions, prior to exposing learners to such productions, in order to make this elusive feature more salient for the learners.

The current investigation also showed that it is not only the young learners who may benefit from imitative practice. Explicit phonetic instruction among primary school students is not expected to be commonplace, with the exception of some basic articulatory explanation, as well as rather intuition-based sound comparisons across the L1 and the L2. In consequence, repetition is arguably the most optimal method of teaching pronunciation in the case of such learners. When it comes to adults, especially those who are proficient, imitation also proves to be a way of improving one's pronunciation, despite satisfactory baseline performance. Although they showed overall better performance in the word reading task, the current adult participants still managed to reduce the participant-to-model distances in the imitation task. Some of the current results suggest that more proficient adults are capable of noticing and bridging subtle qualitative differences between their baseline and target realisations, proving further that such tasks are beneficial to even proficient learners.

Finally, the current data did not seem to confirm that the reportedly detrimental effects of orthographic input, which lead to orthography-induced pronunciation errors in L2, are explicitly applicable in the context of imitation practice. The degree to which Polish learners imitated English vowel properties was not significantly affected by the presence versus absence of access to the words' spelt forms in the imitation task. While the overall effect of orthography is believed to be at work in an inconspicuous manner, it is acknowledged that learners may actually benefit from seeing the words to be repeated by making them certain what the words actually are. At the same time, it is recommended that learners are encouraged to realise that the orthographic input, to which they may be frequently exposed to in their formal L2 instruction, does not adequately represent the true form of the language they are attempting to acquire.

5.4. Limitations and recommendations for future research

It is acknowledged that, because of the rather broad scope of the current research, the investigation may not have captured all intricacies potentially present in the data. First, it should be noted that the stimuli included only two words for each of the four vowels, and while this was attempted to be compensated for by the inclusion of multiple repetitions, such a limited repertoire of lexical items casts some doubt on the definiteness of the conclusions drawn. Preferably, studies should include a wider range of vocabulary items to ensure broader relevance of the results, by accounting for such factors as the potential effect of word frequency on imitative performance. What further limits the generalisability of the current results is the fact that only one native model speaker was ultimately engaged in the study, although four models were initial considered. To ascertain the significance of the results in imitation experiments, more models should be involved in order to prevent model-specific effects, potentially inhibiting or overly facilitating imitative effect. However, this would also necessitate the consideration of how to account for the model speakers in statistical analyses without overfitting the statistical models.

It is also recognised that the imitative effect was only ascertained in an immediate shadowing task, which informs us only about the learners' ability to tap into the acoustic details of recently heard speech. This fails to account for how long-lasting the imitative effect is, although, based on previous research, it can be assumed that improvement in the realisation of vowel properties would diminish in a post-task, as found by Rojczyk et al. (2022). Likewise, the participants' imitative performance was reported to be worse in a delayed or a distracted imitation task, relative to an immediate imitation task (Rojczyk et al. 2013). Despite the fact that a reduction in imitative effect in a post-task should be expected, the inclusion of such a task, while potentially more taxing for the participants, is believed to offer more insight into the nature of imitation, by helping to account for its potentially variable degrees of impermanence under different conditions or speech features meant to be imitated.

As was already discussed, the current investigation offers no definite answers with regard to age or orthographic influence on imitative performance. Ideally, futures studies may wish to at least partly disentangle the effect of age and L2 proficiency, for example, by involving various proficiency levels for given age groups. Additionally, the young group in the current study, aged 12-13, obviously did not exhaust the spectrum of young participants, leaving the question of imitative ability of even younger learners unanswered. There does not emerge a satisfactory remedy for the problem of orthographic input, whose effect is not easily isolated because of the baseline task. What could be introduced instead of a reading task is a picture-naming task, but that would make the range of vocabulary items limited, especially in the case of less proficient participants. Moreover, the additional cognitive load imposed on the participants in such a task may itself have a confounding impact on the results.

Finally, it should be noted that the way of examining vowel dynamics in the current study may not have captured all the intricacies potentially present in the current data. First, the already mentioned limited number of stimuli may be particularly detrimental to a thorough evaluation of spectral patterns, which are known to vary across consonantal contexts (Schwartz 2021), accounting for which may have naturally provided a more comprehensive view of the imitation of vowel dynamics. Second, the vowel metrics used in the current study may not have revealed certain subtleties in the data, which may have yielded the seemingly comparable dynamic patterns shown by the two age/proficiency groups. Therefore, other ways of capturing spectral patterns may be advisable, such as the formant slope and stability measures (Schwartz et al. 2016b), the vector angle (Farrington et al. 2018), or parametric curves (Morrison 2013). Moreover, although studies have variably considered the dynamics in the most extreme vowel sections, due to the immediate influence of consonantal context, the inclusion of such sections may provide more insight with regard to cross-linguistic differences, which the inner sections may fail to reveal.

5.5. Conclusions

Despite certain limitations inherent to the study design, such as a narrow range of stimuli used, the current investigation is believed to have contributed to our understanding of phonetic imitation, specifically in the context of second language acquisition. It was confirmed that both young (less proficient) and adult (more proficient) Polish learners of English are capable of approximating native English vowel properties. Their ability was not only demonstrated by their extension of vowel duration contrasts, which may be considered rather conspicuous, but also by shifts in qualitative features of vowels, viewed both from a static and a dynamic perspective. The static quality was imitated selectively, however, mostly along the F_1 dimension, which is related to the degree of vowel height. Interestingly, even vowels considered to be similar to the Polish ones exhibited some convergence, particularly in the case of adult learners. Vowel dynamics, which displayed considerable variability among the learners, showed visible increases, particularly in the initial vowel sections, as a result of imitative exposure. However, while the overall magnitude of formant movement was observed, it should be noted that it came at a cost of slower spectral rate of change, which was a consequence of vowel lengthening. Overall, the imitative performance of both age (and proficiency) groups can be judged as comparable. While the young learners showed overall greater parameter increases in the imitation task, the adult participants managed to shift their productions towards more model-like values, despite better baseline performance. The latter group also demonstrated greater attunement to subtle qualitative differences in the case of similar vowels. No apparent influence of orthographic input was revealed as having an effect on imitative performance, although it is believed that overall orthographic effects in the learners' productions were at work, but in a more subtle and global way. On the whole, the current investigation has demonstrated that subtle qualitative vowel properties are also subject to phonetic imitation. It was also confirmed that initial participant-to-model distance impacts the degree of imitation, and caution is advised when drawing conclusions as to various degrees of imitation across different features or groups. What is more, the presence of orthography in imitation practice does not appear to have a detrimental effect on the degree of imitation. Finally, the study showed that imitation practice is an effective way of improving pronunciation in the second language, even in the case of highly proficient learners.

References

Adank, P., Hagoort, P., & Bekkering, H. (2010). Imitation improves language comprehension. *Psychological Science*, 21(12), 1903–1909.

Alivuotila, L., Hakokari, J., Savela, J., Happonen, R.-P., & Aaltonen, O. K. (2007). Perception and imitation of Finnish open vowels among children, naive adults, and trained phoneticians. In *Proceedings of the 16th ICPhS* (pp. 361–364). Saarbrucken, Germany.

Ashby, M., & Maidment, J. (2005). *Introducing phonetic scienc*e. Cambridge: Cambridge University Press.

Assmann, P. F., Nearey, T. M., & Bharadwaj, S. V. (2013). Developmental patterns in children's speech: Patterns of spectral change in vowels. In G. Morrison, & P. Assmann (Eds.), *Vowel Inherent Spectral Change*. *Modern Acoustics and Signal* Processing (pp. 199–230). Berlin, Heidelberg: Springer.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society*, 39(4), 437–456.

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, 40(1), 177–189.

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology*, 5(1), 123–150.

Bassetti, B., & Atkinson, N. (2015). Effects of orthographic forms on pronunciation in experienced instructed second language learners. In: R. Hayes-Harb, B. Bassetti, & P. Escudero (Eds.), *Orthographic effects in second language phonology* (pp. 67–91). Applied Psycholinguistics, 36(1), special issue.

Bassetti, B. (2008). Orthographic input and second language phonology. In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 191-206). Clevedon, UK: Multilingual Matters.

Bassetti, B., Cerni, T., & Masterson, J. (2022). The efficacy of grapheme-phoneme correspondence instruction in reducing the effect of orthographic forms on second language phonology. *Applied Psycholinguistics*, 43(3), 683–705.

Bassetti, B. (2024). Orthographic effects in the phonetics and phonology of second language learners and users. In M. Amengual (Ed.), *Cambridge Handbook of Bilingual Phonetics and Phonology* (pp. 699-720). Cambridge: Cambridge University Press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204.

Beňuš, Š. (2014). Social aspects of entrainment in spoken interaction. *Cognitive Computation*, 6(4), 802–813.

Berger, S. M., & Hadley, S. W. (1975). Some effects of a model's performance on an observer's electromyographic activity. *American Journal of Psychology*, 88, 263–276.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 171–204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In J. Munro, & O.-S. Bohn (Eds.), *Language experience in second language speech learning. In honor of James Emil Flege* (pp. 13–34). Amsterdam/Philadelphia: John Benjamins.

Bilous, F. R., & Krauss, R. M. (1988). Dominance and accommodation in the conversational behaviours of same and mixed-gender dyads. *Language & Communication*, 8(3), 183–194.

Black, A. (2012). Acoustic and social parameters on phonetic imitation: gender, emotion, and feature saliency. *University of British Columbia Working Papers in Linguistics*, 32, 16–33.

Boersma, P., & Weenink, D. (2024). *Praat: Doing phonetics by computer* (Version 6.4.11) [Computer program]. http://www.praat.org/

Bogacka, A. (2004). On the perception of English high vowels by Polish learners of English. In E. Daskalaki, N. Katsos, M. Mavrogiorgos, & M. Reeve (Eds.), *CamLing 2004: Proceedings of the University of Cambridge second postgraduate conference in language research* (pp. 43–50). Cambridge Institute of Language Research. Cambridge: Cambridge University Press.

Bogacka, A., Schwartz, G., Zydorowicz, P., Połczyńska, M., & Orzechowska, P. (2006). The production and perception of schwa in second language acquisition: The case of Polish learners of English. In K. Dziubalska-Kołaczyk (Ed.), *IFAtuation: A life in IFA. A festschrift for Professor Jacek Fisiak on the occasion of his 70th birthday* (pp. 71–84.). Poznań: Wydawnictwo Naukowe UAM.

Bohn, O.-S. (1995). Cross language speech production in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Timonium: York Press.

Bourgeois, P., & Hess, U. (2007). The impact of social context on mimicry. *Biological Psychology,* 77, 343–352.

Brady, S., Shankweiler, D., & Mann, V. A. (1983). Speech perception and memory coding in relation to reading ability. *Journal of Experimental Child Psychology*, 35(2), 345–367.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, 75(2), B13–B25.

Brouwer, S., Mitterer, H., & Huettig, F. (2010). Shadowing reduced speech and alignment. *Journal of the Acoustical Society of America*, 1–14.

Brown, C. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Language Research*, 14, 136–193.

Bryła-Cruz, A. (2022). More harm than good: Why dictionaries using orthographic transcription instead of the IPA should be handled with care. *Research in Language*, 20(2), 133–152.

Bürki, A., Welby, P., Clément, M., & Spinelli, E. (2019). Orthography and second language word learning: Moving beyond “friend or foe?”. *Journal of the Acoustical Society of America*, 145, EL265–271.

Cabanero, M. B., & Alves, U. K. (2008). A transferência grafo-fônico-fonológica na produção de seqüências ortográficas ‘ng’ do inglês (L2): uma abordagem conexionista. *ReVEL*, 6(11).

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*, 14, 177–208.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893–910.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22(3), 129–159.

Cochrane, R., & Sachs, J. (1979). Phonological learning by children and adults in a laboratory setting. *Language and Speech*, 22, 145–149.

Colantoni, L., Steele, J., & Escudero, P. (2015). Second language speech: Theory and practice. Cambridge: Cambridge University Press.

Colder, B. (2011). Emulation as an integrating principle for cognition. *Frontiers in Human Neuroscience*, 5, Article 54.

Cole, J. S., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, 38(2), 167–184.

Cruttenden, A. (2008). *Gimson's pronunciation of English* (7th ed.). London: Hodder Education.

D’Imperio, M., Cavone, R., & Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology*, 5, 1226.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics*, 4, 183–198.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica,* 64, 145–173.

Detey, S., & Nespoulous, J. L. (2008). Can orthography influence second language syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua*, 118(1), 66–81.

Dias, J. W., Rosenblum, L. D., & Cook, T. C. (2013). Unconscious imitation as a metric for crossmodal influences: Visibility of the mouth can enhance alignment to audiovisual speech. *Journal of Vision*, 13.

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science*, 11(1), 86–89.

Dufour, S., & Nguyen, N. (2013). How much imitation is there in a shadowing task? *Frontiers in Psychology,* 4, Article 346.

Dukiewicz, L., & Sawicka, I. (1995). *Fonetyka i fonologia*. Kraków: Wydawnictwo Instytutu Języka Polskiego PAN.

Dziubalska-Kołaczyk, K., Balas, A., Schwartz, G., Rojczyk, A., & Wrembel, M. (2014). Teaching to suppress Polglish Processes. In E. Waniek-Klimczak, & M. Pawlak (Eds.), *Teaching and Researching the Pronunciation of English Studies in Honour of Włodzimierz Sobkowiak* (pp. 235–247). Switzerland: Springer International Publishing.

Erdener, V. D., & Burnham, D. K. (2002). The effect of auditory-visual information and orthographic background in L2 acquisition. In *Proceedings of the International Conference on Spoken Language Processing (ICSLP)* (pp. 1929–1932). Bonn, Germany: International Speech Communication Association.

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, 55(2), 191–228.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, 26(4), 551–585.

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, 53, 343–65.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, 36(2), 345–360.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of* America, 121(6), 3814–3826.

Farrington, C., Kendall, T., & Fridland, V. (2018). Vowel dynamics in the southern vowel shift. *American Speech*, 93(2), 186–222.

Ferragne, E., & Pellegrino, F. (2010). Formant frequencies of vowels in 13 accents of the British Isles. *Journal of the International Phonetic Association*, 40, 1–34.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 233–277). Timonium, MD: York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge: Cambridge University Press.

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, 83(2), 729–740.

Flege, J. E., & Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, 5(1), 1–16.

Flege, J. E., Bohn, O. S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470

Floccia, C., Delle Luche, C., Durrant, S., Butler, J., & Goslin, J. (2012). Parent or community: Where do 20-month-olds exposed to two accents acquire their representation of words? *Cognition*, 124(1), 95–100.

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America*, 99(3), 1730–1741.

Fronda, M. P. (2012). Imitation (mimesis, imitatio). In R. S. Bagnall, K. Brodersen, C. B. Champion, A. Erskine, & S. R. Huebner (Eds.), *The encyclopedia of ancient history* (pp. 3416–3417). Wiley-Blackwell.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: quantifying the benefits of linguistic coordination. *Psychological Science*, 23(8), 931–939.

Galef, B. G., Jr. (1988). Imitation in animals: History, definition and interpretation of the data from the psychological laboratory. In T. Zentall, & B. G. Galef Jr. (Eds.), *Social learning: Psychological and biological perspectives (*pp. 3–28). Hillsdale, NJ: Erlbaum.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics*, 15(2), 87–105.

Giles, H., & Ogay, T. (2007). Communication Accommodation Theory. In B. B. Whaley, & W. Samter (Eds.), *Explaining communication: Contemporary theories and exemplars* (pp. 293–310). Lawrence Erlbaum Associates Publishers.

Giles, H., Coupland, N., & Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. In H. Giles, J. Coupland, & N. Coupland (Eds.), *Contexts of accommodation: Developments in applied sociolinguistics* (pp. 1–68). Cambridge University Press; Editions de la Maison des Sciences de l'Homme.

Giles, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, 32(5), 790–804.

Girard, F., Floccia, C., & Goslin, J. (2010). Perception and awareness of accents in young children. *British Journal of Developmental Psychology*, 26(3), 409–433.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review,* 105, 251–279.

Goldinger, S. D. (2013). The cognitive basis of spontaneous imitation: Evidence from the visual world. In *Proceedings of Meetings on Acoustics* (Vol. 19, Article 060136).

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review*, 11(4), 716–722.

Gonet, W. (2017). *Explorations in the acoustics of English sounds*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.

Gopal, H. S. (1990). Effects of speaking rate on the behavior of tense and lax vowel durations. *Journal of Phonetics, 18*(4), 497–518.

Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, 70, 1231–1240.

Gregory, S. W., Dagan, K., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental frequencies to perceptions of communication quality. *Journal of Nonverbal Behavior*, 21(1), 23–43.

Hao, Y. C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, 54, 151–168.

Hauser, I., Graham, E., & Zhang, X. (2023). Competing targets in English sibilant imitation. *Journal of the Acoustical Society of America Express Letters*, 3(7), 075201.

Hawkins, S., & J. Midgley. (2005). Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association*, 35(2), 183–199.

Hayes-Harb, R., & Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, 54(3), 297–326.

Heath, J. (2017). How automatic is convergence? Evidence from working memory. *Linguistic Society of America Proceedings,* 2(35), 1–10.

Heyes, C. (2011). Automatic imitation. *Psychological Bulletin*, 137(3), 463–483.

Hillenbrand, J. M., & Gayvert, R. T. (1993). Identification of steady-state vowels synthesized from the Peterson-Barney measurements. *Journal of Acoustical Society of America*, 94, 668–674.

Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America*, 105, 3509–3523.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111.

Hillenbrand, J., Clark, M. J., & Nearey, T. M. (2001). Effects of consonant environment on vowel formant patterns. *Journal of Acoustical Society of America*, 109, 748–763.

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review,* 93, 411–428.

Honorof, D. N., Weihing, J., & Fowler, C. A. (2011). Articulatory events are imitated under rapid shadowing. *Journal of Phonetics*, 39, 18–38.

Hu, C.-F. (2008). Use orthography in L2 auditory word learning: Who benefits? *Reading and Writing*, 21(8), 823–841.

Huang, B. H. (2014). The effects of age on second language grammar and speech production. *Journal of Psycholinguistic Research,* 43(4), 397–420.

Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. *Annual Review of Psychology*, 60, 653–670.

Iverson, P., & Evans, B. G. (2007). Auditory training of English vowels for first-language speakers of Spanish and German. In *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1625–1628). Saarbrüken, Germany.

Jassem W., & Richter L. (1989). Neutralization of voicing in Polish obstruents. *Journal of Phonetics,* 17, 317–325.

Jassem, W. (1981). *Podręcznik wymowy angielskiej*. Warszawa: Państwowe Wydawnictwo Naukowe.

Jenkins, J. J., Strange, W., & Edman, T. R. (1983). Identification of vowels in 'vowelless' syllables. *Perception & Psychophysics*, 34, 441–450.

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, 119(2), 1118–1130.

Jiang, F., & Kennison, S. (2022). The impact of L2 English learners' belief about and interlocutor's English proficiency on L2 phonetic accommodation. *Journal of Psycholinguistic Research,* 51, 217–234.

Jin, S.-H., & Liu, C. (2013). The vowel inherent spectral change of English vowels spoken by native and non-native speakers. *Journal of the Acoustical Society of America*, 133(5), 363–369.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.

Joos, M. A. (1948). Acoustic phonetics. *Language Supplement*. 24, 1–136.

Kappes, J., Baumgaertner, A., Peschke, C., & Ziegler, W. (2009). Unintended imitation in nonword repetition. *Brain & Language*, 111, 140–151.

Kent, R. D. (1979). Imitation of synthesized English and nonEnglish vowels by children and adults. *Journal of Psycholinguistic Research*, 8, 43–60.

Kim, M., Horton, W., & Bradlow, A. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, 2, 125–156.

Klatt, D. H. (1973). Vowel duration as a function of the syllabic structure of a word. *Journal of the Acoustical Society of America*, 54, 312–313.

Klatt, D. H. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics*, 3, 129–140.

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America*, 100, 425–438.

Kwon, H. (2019). The role of native phonology in spontaneous imitation: Evidence from Seoul Korean. *Laboratory Phonology*, 10(1), 10.

Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: University of Michigan Press.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27(3), 145–162.

Laland, K. N., & Bateson, P. (2001). The mechanisms of imitation. *Cybernetics and Systems,* 32, 195–224.

Lee, S., Narayanan, S., & Byrd, D. (2004). A developmental acoustic characterization of English diphthongs. *Journal of the Acoustical Society of America*, 115, 2628 (A).

Lehiste, I., & Shockey, L. (1980). Labeling, discrimination and repetition of stimuli with level and changing fundamental frequency. *Journal of Phonetics*, 8, 469-474.

Lenth, R. (2024). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.10.4.900003) [R package]. https://rvlenth.github.io/emmeans/

Leonard, L., & Kaplan, L. (1976). A note on imitation and lexical acquisition. *Journal of Child Language*, 3, 449–456.

Lewandowski, N. (2012). *Talent in nonnative phonetic convergence*. Doctoral Dissertation, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Llompart, M., & Reinisch, E. (2018). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech,* 62(3), 594–622.

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6(60), 3139.

MacLeod, B. (2014). Investigating the effects of salience and regional dialect on phonetic convergence in Spanish. In M.-H. Côté, & E. Mathieu (Eds.), *Variation within and across Romance Languages: Selected papers from the 41st Linguistic Symposium on Romance Languages (LSRL), Ottawa, 5–7 May 2011* (pp. 351–378). John Benjamins.

Makowski, D., Lüdecke, D., Patil, I., Thériault, R., Ben-Shachar, M., & Wiernik, B. (2023). *Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption* (Version 0.5.9) [R package]. CRAN. https://easystats.github.io/report/

Markham, D. (1997). P*honetic imitation, accent, and the learner*. Lund: Lund University Press.

Marjou, X. (2021). OTEANN: Estimating the transparency of orthographies with an artificial neural network. In *Proceedings of the Third Workshop on Computational Typology and Multilingual NLP* (pp. 1–9). Association for Computational Linguistics.

Meltzoff, A., & Moore, M. (1999). Persons and representation: Why infant imitation is important for theories of human development. In J. Nadel, & G. Butterworth (Eds.), *Imitation in Infancy* (pp. 9–35). Cambridge: Cambridge University Press.

Meltzoff, A., Kuhl, P., Movellan, J., & Sejnowski, T. (2009). Foundations for a New Science of Learning. *Science*, 325, 284–288.

Miller, R. M., Sanchez, K., & Rosenblum, L. D. (2010). Alignment to visual speech information. *Attention, Perception, & Psychophysics*, 72(6), 1614–1625.

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, 109, 168–173.

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception, & Psychophysics*, 75(3), 557–575.

Morrison, G. (2013). Theories of vowel inherent spectral change. In G. Morrison, & P. Assmann (Eds.), *Vowel Inherent Spectral Change. Modern Acoustics and Signal Processing* (pp. 31-48). Berlin, Heidelberg: Springer.

Nearey, T. M. (2013). Vowel inherent spectral change in the vowels of North American English. In Morrison, G., & Assmann, P. (Eds.), *Vowel Inherent Spectral Change. Modern Acoustics and Signal Processing* (pp. 49–86). Berlin, Heidelberg: Springer.

Nearey, T. M., & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80, 1297–1308.

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics*, 39(2), 132–142.

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language and Hearing Research*, 57(6), 2065–2075.

Nimz, K., & Khattab, G. (2020). On the role of orthography in L2 vowel production: The case of Polish learners of German. *Second Language Research*, 36(4), 623–652.

Nittrouer, S. (2007). Dynamic spectral structure specifies vowels for children and adults. *Journal of the Acoustical Society of America*, 122, 2328–2339.

Nowacka, M. (2010). The ultimate attainment of English pronunciation by Polish college students: A longitudinal study. In E. Waniek-Klimczak (Ed.), *Issues in Accents of English: Variability and* Norm (pp. 233–257). Newcastle upon Tyne: Cambridge Scholars Publishing.

Nycz, J., & Mooney, S. (2017). Variable vowel convergence in a novel cooperative task. *The Journal of the Acoustical Society of America*, 142, 2583.

Ostaszewska, D., & Tambor, J. (2012). *Fonetyka i fonologia współczesnego języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, 111(2), 263–269.

Paquette-Smith, M., Schertz, J., & Johnson, E. K. (2022). Comparing phonetic convergence in children and adults. *Language and Speech*, 65(1), 240–260.

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America*, 119, 2382–2393.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America,* 24, 175–184.

Phillips, S. C., & Clopper, C. G. (2011). Perceived imitation of regional dialects. Paper presented at the 181st Meeting of the Acoustical Society of America, Seattle, WA, United States.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(2), 169–190.

Piske, T., Flege, J. E., MacKay, I. R., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, 59(1), 49–71.

Podlipský, V. J., & Šimáčková, Š. (2015). Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, United Kingdom.

Porzuczek, A., Rojczyk, A., & Arabski, J. (2013). *Praktyczny kurs wymowy angielskiej dla Polaków*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

R Core Team. (2024). R: *A language and environment for statistical computing* (Version 4.4.0). R Foundation for Statistical Computing. https://www.R-project.org/

Repp, B. H., & Williams, D. R. (1985). Categorical trends in vowel imitation: Preliminary observations from a replication experiment. *Speech Communication*, 4, 105–120.

Repp, B. H., & Williams, D. R. (1987). Categorical tendencies in imitating self-produced isolated vowels. *Speech Communication*, 6, 1–14.

Riverin-Coutlée, J., Roy, J.-P., & Gubian, M. (2023). Using Mahalanobis distances to investigate second dialect acquisition: A study on Quebec French, *Language and Speech*, 66(2), 291–321.

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: A basic principle of brain function. *Nature Reviews Neuroscience*, 17(12), 757–765.

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge: Cambridge University Press.

Rogers, C. L., Glasbrenner, M. M., DeMasi, T. M., & Bianchi, M. (2013). Vowel inherent spectral change and the second-language learner. In G. Morrison, & P. Assmann (Eds.), *Vowel Inherent Spectral Change. Modern Acoustics and Signal Processing*. Springer, Berlin, Heidelberg.

Rojczyk, A. (2010a). Forming new vowel categories in second language speech: The case of Polish learners' production of English /e/ and /ɪ/. *Research in Language*, 8, 85–97.

Rojczyk, A. (2010b). *Temporal and spectral parameters in perception of the voicing contrast in English and Polish.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Rojczyk, A. (2011). Overreliance on duration in nonnative vowel production and perception: The within lax vowel category contrast. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in SLA of speech: New Sounds 2010* (Vol. 2, pp. 239–249). Bern: Peter Lang.

Rojczyk, A. (2012). Phonetic and phonological mode in second-language speech: VOT imitation. Paper presented at EUROSLA 2012, Poznań, Poland.

Rojczyk, A. (2013). Phonetic imitation of L2 vowels in a rapid shadowing task. In J. Levis, & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 66–76). Ames IA: Iowa State University.

Rojczyk, A., Porzuczek, A. & Bergier, M. (2013). Immediate and distracted imitation in second-language speech: Unreleased plosives in English. *Research in Language,* 11, 3–18.

Rojczyk, A., Šturm, P., Przedlacka, J., & Wieczorek, B. (2022). How is phonetic imitation conditioned by instructions? Explicit versus implicit imitation. Paper presented at the international conference Accents, Łódź, Poland.

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, 17(6), 405–409.

RStudio Team. (2024). *RStudio: Integrated development for R* (Version 2024.04.2+764) [Computer software]. RStudio, PBC. http://www.rstudio.com/

Schertz, J., & Johnson, E. K. (2022). Voice onset time imitation in teens versus adults. *Journal of Speech, Language, and Hearing Research*, 65, 1839–1850.

Schertz, J., Adil, F., & Kravchuk, A. (2023). Underpinnings of explicit phonetic imitation: perception, production, and variability. *Glossa Psycholinguistics*, 2(1).

Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review, 22*(6), 1614–1629.

Schouten, M. E. H. (1977). Imitation of synthetic vowels by bilinguals. *Journal of Phonetics*, 5, 273–283.

Schwartz, G. (2015). Vowel dynamics for Polish learners of English. In E. Waniek-Klimczak, & M. Pawlak (Eds.), *Teaching and Researching the Pronunciation of English - Studies in Honour of Włodzimierz Sobkowiak* (pp. 205–217). Berlin: Springer.

Schwartz, G. (2016). On the evolution of prosodic boundaries – paremeter settings for Polish and English. *Lingua*, 171, 37–74.

Schwartz, G. (2020). Refining representations for L2 phonology. *Second Language Research*, 36, 691–707.

Schwartz, G., & Dzierla, J. (2018). Polish listeners' perception of vowel inherent spectral change in L2 English. *Poznan Studies in Contemporary Linguistics*, 54(2), 307–332.

Schwartz, G., & Kaźmierski, K. (2020). Vowel dynamics in the acquisition of L2 English – an acoustic study of Polish learners. *Language Acquisition – a Journal of Developmental Linguistics,* 27(3), 227–254.

Schwartz, G. (2021). The phonology of vowel VISC-osity – acoustic evidence and representational implications. *Glossa: a journal of general linguistics*, 6(1), 26.

Schwartz, G., Aperliński, G., Jekiel, M., & Malarski, K. (2016a). Spectral dynamics in L1 and L2 vowel perception. *Research in Language*, 14(1), 61–77.

Schwartz, G., Aperliński, G., Kaźmierski, K., & Weckwerth, J. (2016b). Dynamic targets in the acquisition of L2 English vowels. *Research in Language*, 14(2), 181–202.

Schwartz, G., Kaźmierski, K., Weckwerth, J., Jekiel, M., & Malarski, K. (2018). Vowel dynamics in the acquisition of L2 English – a cross-sectional and longitudinal study of L1 Polish learners. *ResearchGate*. https://doi.org/10.13140/RG.2.2.24905.65122

Shankweiler, D., Strange, W., & Verbrugge, R. (1975). Speech and the problem of perceptual constancy. In R. Shaw, & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.

Shepard, C. A., Giles, H., & Le Poire, B. A. (2001). Communication accommodation theory. In W. P. Robinson, & H. Giles (Eds.), *The new handbook of language and social psychology* (pp. 33–56). New York: John Wile.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics,* 66(3), 422–429.

Slowiaczek, L. M., & Dinnsen, D. A. (1985). On the neutralizing status of Polish word-final devoicing. *Journal of Phonetics*, 13, 325–341.

Slowiaczek, L. M., & Szymanska, H. J. (1989). Perception of word-final devoicing in Polish. *Journal of Phonetics*, 17, 205–212.

Snow, C. E. (1981). The uses of imitation. *Journal of Child Language*, 8(1), 205–212

Snow, C. E., & Hoefnagel-Höhle, M. (1977). Age differences in the pronunciation of foreign sounds, *Language and Speech*, 20, 357–365.

Snow, C. E. (1989). Imitativeness: A trait or a skill? In G. E. Speidel, & K. E. Nelson (Eds.), *The many faces of imitation in children*. New York: Springer-Verlag.

Sobkowiak, W. (2008). *English phonetics for Poles. A resource book for learners and teachers*. Poznań: Wydawnictwo Poznańskie.

Stanley, J. A., & Lipani, L. (2019). Automatic formant extraction in Praat. *Faculty Publications*, (6104). https://scholarsarchive.byu.edu/facpub/6104

Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulations by consonantal context: An acoustical study. *Journal of Speech and Hearing Research*, 6, 111–128.

Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*, 74, 695–705.

Street, R. L., Brady, R. M., & Putman, W. B. (1983). The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. *Journal of Language and Social Psychology*, 2(1), 37–56.

Street, R. L., Jr. (1983). Noncontent speech convergence in adult-child interactions. *Communication Yearbook*, 7, 369–395.

Studdert-Kennedy, M., & Terrace, H. S. (2017). In the beginning: a review of Robert C. Berwick and Noam Chomsky's why only us. *Journal of Language Evolution*, 2, 114–125.

Święciński, R. (2004). Articulatory setting in Polish and its implications for teaching English pronunciation to Poles. In W. Sobkowiak, & E. Waniek-Klimczak (Eds.), *Dydaktyka fonetyki języka obcego. Zeszyt Naukowy Instytutu Neofilologii Państwowej Wyższej Szkoły Zawodowej w Koninie nr 3* (pp. 141–150). Konin: Wydawnictwo PWSZ w Koninie.

Tahta, S., Wood, M., & Loewenthal, K. (1981). Age changes in the ability to replicate foreign pronunciation and intonation. *Language and Speech*, 24, 363–372.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review, Monograph Supplements*, 2, 4–160.

Tilsen, S. (2009). Subphonemic and cross-phonemic priming in vowel shadowing: Evidence for the involvement of exemplars in production. *Journal of Phonetics*, 37, 276–296.

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society*, 37(2), 241–254.

Ulbrich, C. (2021). Phonetic accommodation on the segmental and the suprasegmental level of speech in native-nonnative collaborative tasks. In B. Gili Fivela, C. Avesani, & M. Savino (Eds.), *PaPE 2019 Special Issue: Language contact and speaker accommodation*. *Language and Speech*.

Vaughan, K. B., & Lanzetta, J. T. (1980). Vicarious instigation and conditioning of facial expressive and autonomic responses to a model's expressive display of pain. *Journal of Personality and Social Psychology*, 38(6), 909–923.

Vokic, G. (2011). When alphabets collide: Alphabetic first-language speakers' approach to speech production in an alphabetic second language. *Second Language Research*, 27(3), 391–417.

Waniek-Klimczak, E. (2005). *Temporal Parameters in Second Language Speech: An Applied Linguistic Phonetics Approach*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Weckwerth, J. (2010). Vowel variation in advanced Polish learners of English. In K. Dziubalska-Kolaczyk, M. Wrembel, & M. Kul (Eds.), *Proceedings of the 6th*

International Symposium on the Acquisition of Second Language Speech (New Sounds 2010) (pp 542–547). Poznań: AMU Publishing House.

Weckwerth, J., & Balas, A. (2019). Formant frequencies of Polish raised /ɛ/ and /a/ in palatal contexts. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS 2019), Melbourn*e. Australia, 4–10 August.

Wells, J. (1982). *Accents of English: an introduction*. Cambridge: Cambridge University Press.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686.

Wieczorek, B., & Rojczyk, A. (2021). Phonetic imitation of English VOTs by Polish young learners: Investigating the effect of a model speaker gender. Paper presented at Accents, Łódź, Poland.

Wiese, R. (1996). *The phonology of German*. Oxford: Oxford University Press.

Williams, D., Van Leussen, J. W., & Escudero, P. (2015). Beyond North American English: Modelling vowel inherent spectral change in British English and Dutch. In *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: University of Glasgow.

Winter, B. (2020). *Statistics for linguists: an introduction using R*. New York: Routledge.

Young-Scholten, M., & Langer, M. (2015). The role of orthographic input in second language German: Evidence from naturalistic adult learners' production. *Applied Psycholinguistics*, 36(1), 93–114.

Young-Scholten, M. (2000). New sounds in old spellings. Paper presented at New Sounds 2000. Amsterdam, 5 September.

Young-Scholten, M. (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honour of Henning Wode (*pp. 263–279). Wissenschaftlicher Verlag Trier.

Yu, A. C. L., Abrego-Collier, C., & Sonderegger, M. (2013). Phonetic imitation from an individual-difference perspective: Subjective attitude, personality, and "autistic" traits. *PLoS ONE*, 8(9), e74746.

Zając, M. (2013). Phonetic Imitation of Vowel Duration in L2 Speech. *Research in Language*, 11(1), 19–29.

Zając, M., & Rojczyk, A. (2014). Imitation of English vowel duration upon exposure to native and non-native speech. *Poznan Studies in Contemporary Linguistics*, 50(4), 495–514.

Zellou, G., Dahan, D., & Embick, D. (2017). Imitation of coarticulatory vowel nasality across words and time. *Language, Cognition and Neuroscience,* 32(6), 776–791.

Zellou, G., Scarborough, R., & Nielsen, K. (2016). Phonetic imitation of coarticulatory vowel nasalization. *The Journal of the Acoustical Society of America*, 140(5), 3560–3575.

Zentall, T. R. (2006). Imitation: Definitions, evidence, and mechanisms. *Animal Cognition*, 9(4), 335–353.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review*, 5, 683–689.

Summary

The purpose of the investigation is to assess the degree to which native Polish learners of English can imitate acoustic properties of English vowels, including their quality durational contrasts, and spectral variability. Since Polish and English differ in terms of how these properties are specified in their vocalic inventories, the speech of Polish learners of English is bound to exhibit cross-linguistic interference. This means that their realisation of English vowels is likely to exhibit Polish characteristics, which contributes to the perception of foreign accent, and possibly to miscommunication. The main question asked in the current investigation is whether Polish learners of English can assume more English-like properties in their speech, after an immediate exposure to a native English speaker. What is also of interest is how age, language proficiency and orthography affect the degree of imitative performance.

To answer the research questions, two groups of participants, young (less proficient) and adult (more proficient) learners of English, were engaged in an experiment, where they were asked to first read English words, and then repeat the same words, as produced by a recorded voice of a native English speaker. To determine the effect of orthography, only half of the participants had access to the words' spelt forms in the imitation task. The participants' productions from both tasks were recorded and subjected to acoustic and statistical analyses. The measures included: relative vowel duration difference, mid-point values of formant frequencies (F_1 and F_2), trajectory length, and spectral rate of change.

The results confirmed that both young and adult Polish learners of English are capable of approximating native English vowel properties, related to durational and qualitative (static and dynamic) properties. Both groups showed observable imitative performance, which was primarily driven by initial participant-to-model phonetic distance. No effect of orthography was ascertained on the degree of imitation. On the whole, the current investigation demonstrated that subtle qualitative vowel properties are subject to imitation. It was also concluded that caution should be exercised when judging varying degrees of imitation across different features or groups. Finally, the study showed that imitation practice is an effective way of improving pronunciation in the second language, even in the case of highly proficient learners.

The dissertation is divided into five chapters. In Chapter 1, what is offered is a general discussion of the process of imitation, which then focuses specifically on phonetic

imitation, with various factors affecting it, which is culminated in a review of studies within the context of second language speech imitation. Chapter 2 explores the differences related to the vocalic systems in Polish and English, with a special focus on vowel dynamics. Chapter 3 constitutes a brief overview of studies dealing with the impact of orthography on the speech of second language learners. The design of the current study is detailed in Chapter 4, which is followed by a discussion of the results for each of the three groups of tested parameters, related to vowel durational and qualitative contrasts. Finally, Chapter 5 captures the overall conclusions drawn from the findings, with regard to the research questions posed.

Streszczenie

Celem niniejszej rozprawy jest oszacowanie, do jakiego stopnia polscy uczniowie języka angielskiego są w stanie odwzorować w swojej wymowie cechy akustyczne angielskich samogłosek, a konkretnie ich barwę, długość oraz dynamikę. Ponieważ język polski i język angielski różnią się pod tymi względami, wymowę angielskich samogłosek przez Polaków nierzadko cechuje wpływ języka pierwszego, co przyczynia się do obcego akcentu – i potencjalnie – do problemów ze zrozumieniem. Podstawowym pytaniem badawczym postawionym w rozprawie jest, czy polscy uczniowie języka angielskiego potrafią przenieść anglojęzyczne cechy samogłosek do swojej wymowy pod wpływem zasłyszanych słów wypowiedzianych przez anglojęzycznego modela. W badaniu brane są pod uwagę takie czynniki jak wiek i poziom zaawansowania uczniów, a także potencjalny wpływ ortografii na zakres imitacji.

Aby udzielić odpowiedzi na pytania badawcze, w badanie zostały zaangażowane dwie grupy uczestników: młodzi uczniowie (początkujący) oraz dorośli (zaawansowani). Pierwsze zadanie postawione przed uczestnikami polegało na przeczytaniu angielskich słów, zaś drugie – na ich powtórzeniu za anglojęzycznym modelem. Aby oszacować potencjalny wpływ ortografii na zakres imitacji, jedynie połowa uczestników widziała słowa podczas ich powtarzania. Uczniowie zostali nagrani, a ich wymowę poddano analizie akustycznej oraz statystycznej. Zmierzono następujące parametry akustyczne: względną różnicę czasu trwania samogłosek, wartości środkowe częstotliwości formantów (F_1 i F_2), długość trajektorii formantów oraz tempo zmian spektralnych.

Wyniki eksperymentu potwierdziły, że zarówno młodzi, jak i dorośli uczniowie potrafią w przybliżeniu upodobnić cechy swoich samogłosek do tych, które charakteryzowały wymowę modela. Zaobserwowano to nie tylko jako zmianę długości samogłosek, ale też zmianę ich barwy oraz dynamiki. Zjawisko imitacji zaobserwowano u obydwu grup. Głównym czynnikiem sprzyjającym imitacji była początkowa różnica w wymowie poszczególnych cech między uczestnikami a modelem. Nie stwierdzono, aby wpływ ortografii na zakres imitacji był znaczący. Badanie pokazało, że również cechy akustyczne samogłosek, które nie są łatwo zauważalne, podlegają imitacji. Wyniki pozwoliły stwierdzić, że formułowania wniosków co do zakresu imitacji pomiędzy poszczególnymi grupami uczestników lub dźwiękami nie należy opierać jedynie na różnicy w zmianie poszczególnych parametrów akustycznych. Badanie stanowi również

dowód na to, że imitacja może być skuteczną metodą umożliwiającą poprawę obcojęzycznej wymowy nawet w przypadku dorosłych i zaawansowanych uczniów.

Rozprawa składa się z pięciu rozdziałów. Rozdział pierwszy poświęcony jest ogólnej dyskusji na temat procesu imitacji, w szczególności imitacji fonetycznej, z uwzględnieniem różnych czynników na nią wpływających, kończąc na imitacji w kontekście przyswajania drugiego języka. W rozdziale drugim nakreślone zostały różnice w systemach samogłoskowych języka polskiego i angielskiego, koncentrując się na aspektach dynamicznych samogłosek. Rozdział trzeci stanowi zwięzły przegląd badań poświęconych wpływowi ortografii na wymowę w drugim języku. W rozdziale czwartym szczegółowo opisano plan badań, ich wyniki oraz zawarto dyskusję odnośnie do wszystkich trzech grup cech akustycznych podlegających badaniu. Rozdział piąty stanowi podsumowanie rozprawy, gdzie wyciągnięte zostały najważniejsze wnioski w świetle zadanych wcześniej pytań badawczych.

List of figures

List of tables

Appendices

Appendix 1: Mean and SD values for RVDD

VOWEL	CONDITION	TASK	MEAN RVDD YL	SD	MEAN RVDD AL	SD	MEAN RVDD MODEL
FLEECE	spelling	baseline	0.04	0.07	0.15	0.10	0.32
FLEECE	spelling	imitation	0.18	0.08	0.22	0.08	0.32
FLEECE	no_spelling	baseline	0.06	0.07	0.16	0.10	0.32
FLEECE	no_spelling	imitation	0.18	0.07	0.24	0.07	0.32
KIT	spelling	baseline	0.06	0.07	0.14	0.09	0.12
KIT	spelling	imitation	0.12	0.09	0.13	0.08	0.12
KIT	no_spelling	baseline	0.07	0.08	0.15	0.08	0.12
KIT	no_spelling	imitation	0.11	0.06	0.17	0.07	0.12
DRESS	spelling	baseline	0.07	0.07	0.15	0.08	0.15
DRESS	spelling	imitation	0.12	0.09	0.15	0.09	0.15
DRESS	no_spelling	baseline	0.09	0.07	0.16	0.08	0.15
DRESS	no_spelling	imitation	0.12	0.07	0.18	0.07	0.15
TRAP	spelling	baseline	0.07	0.08	0.14	0.09	0.28
TRAP	spelling	imitation	0.14	0.07	0.19	0.08	0.28
TRAP	no_spelling	baseline	0.08	0.07	0.16	0.09	0.28
TRAP	no_spelling	imitation	0.12	0.10	0.19	0.08	0.28

Appendix 2: Mean and SD values for F_1 (top) and F_2 (bottom)

VOWEL	CONDITION	TASK	MEAN F1 YL	SD	MEAN F1 AL	SD	MEAN F1 MODEL
FLEECE	spelling	baseline	386	50	313	55	369
FLEECE	spelling	imitation	396	51	344	60	369
FLEECE	no_spelling	baseline	392	45	473	53	369
FLEECE	no_spelling	imitation	405	52	489	67	369
KIT	spelling	baseline	386	50	422	77	486
KIT	spelling	imitation	455	68	476	88	486
KIT	no_spelling	baseline	410	44	695	89	486
KIT	no_spelling	imitation	478	59	741	100	486
DRESS	spelling	baseline	689	63	707	89	821
DRESS	spelling	imitation	731	68	751	98	821
DRESS	no_spelling	baseline	702	54	856	90	821
DRESS	no_spelling	imitation	748	54	927	118	821
TRAP	spelling	baseline	794	108	835	142	1008
TRAP	spelling	imitation	866	95	914	141	1008
TRAP	no_spelling	baseline	848	78	313	55	1008
TRAP	no_spelling	imitation	917	63	344	60	1008

VOWEL	CONDITION	TASK	MEAN F2 YL	SD	MEAN F2 AL	SD	MEAN F2 MODEL
FLEECE	spelling	baseline	2726	208	2700	243	2852
FLEECE	spelling	imitation	2744	192	2807	225	2852
FLEECE	no_spelling	baseline	2776	145	2656	267	2852
FLEECE	no_spelling	imitation	2819	141	2706	260	2852
KIT	spelling	baseline	2732	200	2204	163	2414
KIT	spelling	imitation	2458	230	2232	189	2414
KIT	no_spelling	baseline	2673	208	2288	330	2414
KIT	no_spelling	imitation	2459	226	2269	227	2414
DRESS	spelling	baseline	1983	122	1942	129	2190
DRESS	spelling	imitation	2004	133	2020	143	2190
DRESS	no_spelling	baseline	2076	111	1910	169	2190
DRESS	no_spelling	imitation	2094	120	2002	203	2190
TRAP	spelling	baseline	1652	191	1685	213	1629
TRAP	spelling	imitation	1664	149	1703	167	1629
TRAP	no_spelling	baseline	1715	184	1638	235	1629
TRAP	no_spelling	imitation	1711	170	1665	217	1629

Appendix 3: Mean and SD values for TL (top) and roc (bottom)

VOWEL	CONDITION	TASK	MEAN TL YL	SD	MEAN TL AL	SD	MEAN TL MODEL
FLEECE	spelling	baseline	210	102	178	80	440
FLEECE	spelling	imitation	232	113	222	102	440
FLEECE	no_spelling	baseline	203	107	198	101	440
FLEECE	no_spelling	imitation	231	102	202	87	440
KIT	spelling	baseline	223	104	182	85	383
KIT	spelling	imitation	223	108	206	90	383
KIT	no_spelling	baseline	205	99	186	96	383
KIT	no_spelling	imitation	209	77	217	96	383
DRESS	spelling	baseline	210	80	215	74	503
DRESS	spelling	imitation	238	86	243	100	503
DRESS	no_spelling	baseline	217	88	240	111	503
DRESS	no_spelling	imitation	238	98	260	117	503
TRAP	spelling	baseline	254	103	267	104	332
TRAP	spelling	imitation	279	106	297	125	332
TRAP	no_spelling	baseline	270	105	299	122	332
TRAP	no_spelling	imitation	270	86	304	126	332

VOWEL	CONDITION	TASK	MEAN roc AL	SD	MEAN roc AL	SD	MEAN roc MODEL
FLEECE	spelling	baseline	2.44	1.19	1.92	0.97	3.35
FLEECE	spelling	imitation	1.86	0.95	1.78	0.88	3.35
FLEECE	no_spelling	baseline	2.17	1.23	2.12	1.07	3.35
FLEECE	no_spelling	imitation	1.80	0.84	1.75	0.83	3.35
KIT	spelling	baseline	2.66	1.15	2.41	1.19	3.32
KIT	spelling	imitation	2.06	1.02	1.95	0.91	3.32
KIT	no_spelling	baseline	2.35	1.17	2.45	1.23	3.32
KIT	no_spelling	imitation	1.84	0.82	2.07	1.09	3.32
DRESS	spelling	baseline	2.34	1.06	2.37	0.98	3.84
DRESS	spelling	imitation	2.15	1.00	2.17	0.98	3.84
DRESS	no_spelling	baseline	2.24	0.96	2.55	1.13	3.84
DRESS	no_spelling	imitation	2.08	1.00	2.34	1.15	3.84
TRAP	spelling	baseline	2.56	1.00	2.48	1.07	2.12
TRAP	spelling	imitation	2.34	1.09	2.29	1.03	2.12
TRAP	no_spelling	baseline	2.53	0.94	2.80	1.20	2.12
TRAP	no_spelling	imitation	2.14	0.86	2.42	1.16	2.12

Appendix 4: Summaries of LME models with best fit for RVDD for YLs (left) and ALs (right)

Parameter	Coefficient	95% CI	t(926)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	0.05	[0.03, 0.07]	5.11	< .001	fixed		-0.64	[-0.85, -0.43]	
block [imitation]	0.13	[0.11, 0.15]	13.27	< .001	fixed		1.53	[1.31, 1.76]	
vowel [KIT]	0.02	[0.00, 0.04]	2.50	0.013	fixed		0.24	[0.05, 0.43]	
vowel [DRESS]	0.03	[0.01, 0.05]	3.67	< .001	fixed		0.35	[0.16, 0.54]	
vowel [TRAP]	0.03	[0.01, 0.04]	3.37	< .001	fixed		0.32	[0.13, 0.51]	
block [imitation] × vowel [KIT]	-0.08	[-0.11, -0.06]	-7.16	< .001	fixed		-0.97	[-1.23, -0.70]	
block [imitation] × vowel [DRESS]	-0.09	[-0.11, -0.06]	-7.46	< .001	fixed		-1.01	[-1.27, -0.74]	
block [imitation] × vowel [TRAP]	-0.08	[-0.10, -0.05]	-6.45	< .001	fixed		-0.87	[-1.14, -0.61]	
	0.04				random	sbj			
	0.03				random	sbj			
	-0.15				random	sbj			
	0.06				random	Residual			
AIC									-2328.79
AICc									-2328.46
BIC									-2270.67
R2 (conditional)									0.47
R2 (marginal)									0.22
Sigma									0.06

Parameter	Coefficient	95% CI	t(903)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	0.15	[0.13, 0.18]	12.76	< .001	fixed		-0.15	[-0.42, 0.11]	
block [imitation]	0.07	[0.05, 0.10]	6.68	< .001	fixed		0.84	[0.60, 1.09]	
vowel [KIT]	-9.22e-03	[-0.03, 0.01]	-1.00	0.317	fixed		-0.10	[-0.31, 0.10]	
vowel [DRESS]	3.01e-03	[-0.02, 0.02]	0.33	0.744	fixed		0.03	[-0.17, 0.24]	
vowel [TRAP]	-1.23e-03	[-0.02, 0.02]	-0.13	0.895	fixed		-0.01	[-0.22, 0.19]	
block [imitation] × vowel [KIT]	-0.07	[-0.09, -0.04]	-5.05	< .001	fixed		-0.75	[-1.04, -0.46]	
block [imitation] × vowel [DRESS]	-0.06	[-0.09, -0.04]	-4.86	< .001	fixed		-0.72	[-1.01, -0.43]	
block [imitation] × vowel [TRAP]	-0.03	[-0.06, -0.01]	-2.53	0.011	fixed		-0.38	[-0.68, -0.09]	
	0.05				random	sbj			
	0.03				random	sbj			
	-0.70				random	sbj			
	0.07				random	Residual			
AIC									-2078.46
AICc									-2078.11
BIC									-2020.63
R2 (conditional)									0.38
R2 (marginal)									0.09
Sigma									0.07

Appendix 5: Summaries of LME models with best fit for F_1 for YLs (left) and ALs (right)

Parameter	Coefficient	95% CI	t(1838)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	388.58	[374.41, 402.76]	53.77	< .001	fixed		-1.02	[-1.09, -0.95]	
vowel [KIT]	9.35	[-0.79, 19.50]	1.81	0.071	fixed		0.05	[0.00, 0.09]	
vowel [DRESS]	306.28	[296.09, 316.48]	58.91	< .001	fixed		1.49	[1.44, 1.54]	
vowel [TRAP]	432.91	[422.70, 443.12]	83.16	< .001	fixed		2.10	[2.05, 2.15]	
block [imitation]	12.43	[-1.76, 26.61]	1.72	0.086	fixed		0.06	[-0.01, 0.13]	
vowel [KIT] × block [imitation]	56.40	[42.12, 70.67]	7.75	< .001	fixed		0.27	[0.20, 0.34]	
vowel [DRESS] × block [imitation]	32.43	[18.09, 46.77]	4.44	< .001	fixed		0.16	[0.09, 0.23]	
vowel [TRAP] × block [imitation]	58.75	[44.35, 73.15]	8.00	< .001	fixed		0.29	[0.22, 0.36]	
	33.99				random	sbj			
	27.60				random	sbj			
	-0.25				random	sbj			
	55.45				random	Residual			
AIC									20232.42
AICc									20232.59
BIC									20298.70
R2 (conditional)									0.93
R2 (marginal)									0.90
Sigma									55.45

Parameter	Coefficient	95% CI	t(1774)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	318.62	[298.98, 338.25]	31.83	< .001	fixed		-1.19	[-1.28, -1.11]	
vowel [KIT]	129.64	[116.86, 142.42]	19.90	< .001	fixed		0.56	[0.50, 0.61]	
vowel [DRESS]	381.09	[368.26, 393.91]	58.30	< .001	fixed		1.64	[1.58, 1.69]	
vowel [TRAP]	525.84	[512.74, 538.93]	78.75	< .001	fixed		2.26	[2.20, 2.32]	
block [imitation]	16.27	[-0.87, 33.41]	1.86	0.063	fixed		0.07	[0.00, 0.14]	
vowel [KIT] × block [imitation]	18.76	[0.57, 36.95]	2.02	0.043	fixed		0.08	[0.00, 0.16]	
vowel [DRESS] × block [imitation]	27.36	[8.95, 45.78]	2.91	0.004	fixed		0.12	[0.04, 0.20]	
vowel [TRAP] × block [imitation]	54.78	[36.10, 73.46]	5.75	< .001	fixed		0.24	[0.16, 0.32]	
	48.50				random	sbj			
	30.75				random	sbj			
	0.35				random	sbj			
	69.80				random	Residual			
AIC									20355.42
AICc									20355.59
BIC									20421.27
R2 (conditional)									0.91
R2 (marginal)									0.85
Sigma									69.80

Appendix 6: Summaries of LME models with best fit for F_2 for YLs (left) and ALs (right)

Parameter	Coefficient	95% CI	t(1838)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	2747.84	[2702.54, 2793.15]	118.95	< .001	fixed		1.02	[0.92, 1.12]	
vowel [KIT]	-45.52	[-70.31, -20.72]	-3.60	< .001	fixed		-0.10	[-0.15, -0.04]	
vowel [DRESS]	-716.08	[-741.00, -691.15]	-56.34	< .001	fixed		-1.53	[-1.59, -1.48]	
vowel [TRAP]	-1064.66	[-1089.62, -1039.70]	-83.66	< .001	fixed		-2.28	[-2.33, -2.22]	
block [imitation]	34.57	[0.77, 68.36]	2.01	0.045	fixed		0.07	[0.00, 0.15]	
vowel [KIT] × block [imitation]	-278.93	[-313.83, -244.03]	-15.68	< .001	fixed		-0.60	[-0.67, -0.52]	
vowel [DRESS] × block [imitation]	-16.18	[-51.24, 18.87]	-0.91	0.365	fixed		-0.03	[-0.11, 0.04]	
vowel [TRAP] × block [imitation]	-29.64	[-64.85, 5.57]	-1.65	0.099	fixed		-0.06	[-0.14, 0.01]	
	116.41				random	sbj			
	63.88				random	sbj			
	-0.34				random	sbj			
	135.56				random	Residual			
AIC									23540.21
AICc									23540.38
BIC									23606.49
R2 (conditional)									0.92
R2 (marginal)									0.86
Sigma									135.56

Parameter	Coefficient	95% CI	t(1774)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	2678.35	[2616.92, 2739.79]	85.50	< .001	fixed		1.19	[1.05, 1.33]	
vowel [KIT]	-432.74	[-458.51, -406.97]	-32.94	< .001	fixed		-0.99	[-1.05, -0.93]	
vowel [DRESS]	-759.07	[-784.93, -733.21]	-57.57	< .001	fixed		-1.73	[-1.79, -1.67]	
vowel [TRAP]	-1011.53	[-1037.94, -985.11]	-75.11	< .001	fixed		-2.31	[-2.37, -2.25]	
block [imitation]	81.54	[44.81, 118.27]	4.35	< .001	fixed		0.19	[0.10, 0.27]	
vowel [KIT] × block [imitation]	-75.43	[-112.13, -38.74]	-4.03	< .001	fixed		-0.17	[-0.26, -0.09]	
vowel [DRESS] × block [imitation]	10.32	[-26.82, 47.47]	0.55	0.586	fixed		0.02	[-0.06, 0.11]	
vowel [TRAP] × block [imitation]	-69.29	[-106.97, -31.62]	-3.61	< .001	fixed		-0.16	[-0.24, -0.07]	
	163.64				random	sbj			
	71.08				random	sbj			
	-0.18				random	sbj			
	140.75				random	Residual			
AIC									22884.31
AICc									22884.49
BIC									22950.16
R2 (conditional)									0.90
R2 (marginal)									0.76
Sigma									140.75

Appendix 7: Summaries of LME models with best fit for TL for YLs (left) and ALs (right)

Parameter	Coefficient	95% CI	t(1825)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	14.12	[13.67, 14.58]	61.02	< .001	fixed		-0.23	[-0.37, -0.09]	
block [imitation]	0.56	[0.20, 0.93]	3.03	0.003	fixed		0.17	[0.06, 0.28]	
vowel [KIT]	-0.10	[-0.50, 0.29]	-0.52	0.606	fixed		-0.03	[-0.15, 0.09]	
vowel [DRESS]	0.34	[-0.06, 0.73]	1.66	0.097	fixed		0.10	[-0.02, 0.22]	
vowel [TRAP]	1.70	[1.30, 2.10]	8.35	< .001	fixed		0.52	[0.40, 0.64]	
	3.06				random	Residual			
	0.91				random	sbj			
	0.65				random	sbj			
	-0.51				random	sbj			
AIC									9396.76
AICc									9396.86
BIC									9446.39
R2 (conditional)									0.12
R2 (marginal)									0.06
Sigma									3.06

Parameter	Coefficient	95% CI	t(1719)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	13.45	[12.84, 14.06]	43.17	< .001	fixed		-0.38	[-0.55, -0.20]	
block [imitation]	0.88	[0.48, 1.28]	4.30	< .001	fixed		0.25	[0.14, 0.37]	
vowel [KIT]	-0.18	[-0.56, 0.20]	-0.93	0.355	fixed		-0.05	[-0.16, 0.06]	
vowel [DRESS]	1.30	[0.91, 1.68]	6.62	< .001	fixed		0.37	[0.26, 0.49]	
vowel [TRAP]	2.97	[2.57, 3.36]	14.61	< .001	fixed		0.86	[0.74, 0.97]	
	2.88				random	Residual			
	1.48				random	sbj			
	0.82				random	sbj			
	-0.27				random	sbj			
AIC									8693.82
AICc									8693.92
BIC									8742.91
R2 (conditional)									0.32
R2 (marginal)									0.14
Sigma									2.88

Appendix 8: Summaries of LME models with best fit for roc for YLs (left) and ALs (right)

Parameter	Coefficient	95% CI	t(1820)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	1.47	[1.41, 1.53]	48.17	< .001	fixed		0.05	[-0.12, 0.22]	
block [imitation]	-0.15	[-0.21, -0.08]	-4.55	< .001	fixed		-0.43	[-0.62, -0.25]	
vowel [KIT]	0.08	[0.02, 0.14]	2.67	0.008	fixed		0.23	[0.06, 0.40]	
vowel [DRESS]	9.63e-03	[-0.05, 0.07]	0.33	0.743	fixed		0.03	[-0.14, 0.20]	
vowel [TRAP]	0.10	[0.04, 0.16]	3.45	< .001	fixed		0.30	[0.13, 0.46]	
block [imitation] × vowel [KIT]	-0.04	[-0.12, 0.04]	-0.94	0.348	fixed		-0.11	[-0.35, 0.12]	
block [imitation] × vowel [DRESS]	0.09	[0.01, 0.17]	2.27	0.023	fixed		0.27	[0.04, 0.51]	
block [imitation] × vowel [TRAP]	0.04	[-0.04, 0.13]	1.06	0.287	fixed		0.13	[-0.11, 0.37]	
	0.12				random	sbj			
	0.08				random	sbj			
	-0.35				random	sbj			
	0.31				random	Residual			
AIC									1082.86
AICc									1083.03
BIC									1149.01
R2 (conditional)									0.18
R2 (marginal)									0.05
Sigma									0.31

Parameter	Coefficient	95% CI	t(1715)	p	Effects	Group	Std. Coef.	Std. Coef. 95% CI	Fit
(Intercept)	1.39	[1.32, 1.47]	37.50	< .001	fixed		-0.17	[-0.38, 0.03]	
block [imitation]	-0.09	[-0.15, -0.04]	-3.52	< .001	fixed		-0.27	[-0.42, -0.12]	
vowel [KIT]	0.10	[0.06, 0.14]	5.04	< .001	fixed		0.29	[0.18, 0.40]	
vowel [DRESS]	0.16	[0.12, 0.20]	7.69	< .001	fixed		0.44	[0.33, 0.56]	
vowel [TRAP]	0.21	[0.17, 0.25]	9.98	< .001	fixed		0.59	[0.48, 0.71]	
	0.30				random	Residual			
	0.18				random	sbj			
	0.12				random	sbj			
	-0.60				random	sbj			
AIC									899.95
AICc									900.06
BIC									949.02
R2 (conditional)									0.28
R2 (marginal)									0.06
Sigma									0.30