



Phonetic imitation of t-glottaling by Czech speakers of English

Pavel Šturm (Charles University, Prague) –

Joanna Przedlacka (University of Oxford) –

Arkadiusz Rojczyk (University of Silesia in Katowice)

ABSTRACT

The paper focuses on the ability of Czech speakers to explicitly imitate native English realizations of the phoneme /t/ as [ʔ] (t-glottaling). In Czech, glottalization occurs as a boundary signal of word-initial vocalic onsets. We hypothesize that this allows for a better imitative performance in the intervocalic context as compared to non-prevocalic contexts. However, an alternative hypothesis based on language-external facts (frequency in the learners' English input) predicts the opposite pattern. Our experiment involves 30 participants in a shadowing task. In addition to words with /t/, words with /k/ are examined to establish if speakers can generalize to a phonologically similar category to which they have not been exposed. Speakers adapted their pronunciation after exposure to t-glottaling to some degree. Our hypothesis was confirmed for the shadowing task, while the alternative language-external hypothesis was confirmed for the post-test task, suggesting a different pattern of performance in terms of imitation versus learning.

KEYWORDS

adaptation, glottalization, glottal stop, phonetic imitation, t-glottaling

DOI

<https://doi.org/10.14712/18059635.2022.1.8>

1 INTRODUCTION

In this study we investigate the phenomenon of “t-glottaling”, i.e., the replacement of the English alveolar plosive [t] with a glottal stop [ʔ] in certain positions (often described as syllable-coda positions), resulting in pronunciations such as [ˈbeʔə] for *better* or [ˈhɔʔ] for *hot*. A related topic is that of “glottal reinforcement”, when the glottal gesture is produced along with the oral gesture rather than as its replacement (e.g. [ˈhɔʔt]). Considering that in many varieties we also find the alveolar flap [ɾ] in intervocalic positions, and most varieties have a strongly aspirated (and potentially affricated) plosive [tʰ] as well, the allophonic variation of the English /t/ phoneme is thus complex (cf. Skarnitzl and Rálišová 2022) and potentially challenging for learners.

We present the results of a phonetic shadowing experiment with Czech speakers of English in which the participants were explicitly instructed to imitate native speakers' productions. This language group was selected for structural reasons. In contrast to the English words presented above, Czech words involve neither glottal replacement nor reinforcement of [t]. As such, those variants are unfamiliar to the Czech speaker. However, although the Czech word *bota* /ˈbota/ ‘shoe’ would never be

pronounced as [ʔboʔa], the glottal stop is a common sound in the Czech inventory, appearing as a boundary marker in vowel-initial words or morphemes (see Section 1.3). Therefore, Czech speakers are quite familiar with the glottal stop in the prevocalic position, including the intervocalic position (*Potkal naopak Adama*. [ʔpotkal ʔnaʔopak ʔadama] ‘On the contrary, he ran into Adam.’).

This leads us to hypothesize that if Czech speakers are supposed to learn or phonetically imitate English words with t-glottaling, the task will be easier and the imitators more successful in the intervocalic context ([ʔbeʔə]) than in non-prevocalic contexts (before a pause, *hot* [ʔhɒʔ], or a consonant, *hot weather* [ʔhɒʔweðə]). However, an alternative hypothesis predicts that the intervocalic context should in fact be harder to imitate, since Czech learners encounter t-glottaling more frequently in preconsonantal or prepausal positions if we assume an extensive exposure to Standard Southern British English (see Jakšič and Šturm 2017) as opposed to varieties of English where t-glottaling occurs also intervocalically (see Sections 1.2 and 1.4). The experiment primarily aims to test these two conflicting hypotheses.

1.1 PHONETIC IMITATION AND LANGUAGE ADAPTATION

Phonetic imitation appears to be a fundamental human behaviour that plays a crucial role in language learning and acquisition. Infants imitate speech sounds in their ambient language to acquire new words (Kuhl and Meltzoff 1996). Later in life, phonetic imitation may be a source of language adaptation when speakers gradually pick up an accent of a new region (Chang 2012; Sancier and Fowler 1997). Finally, phonetic imitation may be a driving force in propagating sound changes when a speaker imitates and adopts a new sound and passes it to the next interacting speaker (Labov 2001; Lin et al. 2021; Siegel 2010).

Different research paths have attempted to investigate the nature of speech imitation and strived to identify the factors that shape its magnitude (Pardo et al. 2018). One approach concentrates on analyzing speech production in interacting partners in order to observe how the degree of imitation (in these studies referred to as *convergence* or *accommodation*) is modulated by social distance, the level of interaction, or the perception of a dialogue partner (Babel 2010, 2012; Babel et al. 2014; Gasiorek et al. 2015; Giles et al. 1991). Other studies look into the process of phonetic imitation using speech shadowing tasks in a laboratory setting (Goldinger 1998; Kwon 2019, 2021; Mitterer and Ernestus 2008; Mitterer and Müsseler 2013; Namy et al. 2002; Nielsen 2011; Shockley et al. 2004). In a typical paradigm of a shadowing task, the participant first reads words presented in an orthographic form in order to elicit their baseline productions. Next, they hear and repeat the same words after a model talker (shadowing) and finally re-read the words again in a post-test phase. The comparisons between the three conditions permit an insight into the degree of imitation changes and the level of post-exposure retention.

Considering that successful learning of speech sounds in a second language requires an effective interaction of perception and production, a number of studies have also directed their attention to the degree of imitation after exposure to second language (L2) speech. In such studies, shadowing after a model talker in L2 throws light on how the acoustic properties of an L1 sound inventory constrain the successful





attainment of L2 speech sounds (de Jong et al. 2009; Flege and Eeften 1988; Hao and de Jong 2016; Jia et al. 2006; Llompарт and Reinisch 2018; Podlipský and Šimáčková 2015; Rojczyk 2013; Rojczyk et al. 2013; Schouten 1977; Zajac and Rojczyk 2014). The results show that speakers are able to imitate phonetic features that are absent in their L1 relatively effectively. Compared to the baseline condition, imitated L2 productions after a model speaker tend to be more native-like. This effect has been found for voice onset time (Flege and Eeften 1988), vowel duration (Podlipský and Šimáčková 2015; Zajac and Rojczyk 2014), spectral properties of vowels (Jia et al. 2006; Llompарт and Reinisch 2018; Rojczyk 2013), the lack of release in stop consonants (Rojczyk et al. 2013), or tones in tone languages (Hao and de Jong 2016).

The results from these studies suggest that direct imitation may temporarily bypass phonological constraints emerging from cross-linguistic differences. For example, Rojczyk (2013) tested twenty-two Polish learners of English in how they imitated the quality of the non-native TRAP vowel /æ/.¹ This vowel is especially problematic for Polish learners because it is subsumed by two Polish neighbouring vowels, /ɛ/ and /a/. The results showed that the imitated productions dissimilated successfully from both vowels compared to a baseline reading task. In another study, Llompарт and Reinisch (2018) investigated the link between imitation and perception for two non-native contrasts differing in the level of difficulty: DRESS-TRAP (difficult) and FLEECE-KIT (easy). The analysis of the productions by German learners of English revealed that both imitation and perception were more successful for the easy contrast than for the difficult one. Moreover, the ability to imitate the DRESS-TRAP opposition was closely related to the perception of this contrast, which implies a strong impact of phonological representations on the magnitude of imitation. The authors concluded that imitation is linked with the level of attainment of non-native contrasts but does not need to reflect the learners' productive usage of such non-native distinctions.

As suggested by one of the reviewers, it is important to differentiate between *explicit* and *implicit* imitation in reviewing prior research on the role of imitation in L2. Although we agree that this is an important distinction to be made, it is sometimes difficult or even impossible to achieve since the methodological descriptions in previous studies frequently lack such information. For example, in Podlipský and Šimáčková (2015: 2) we are only informed that “in shadowing, they repeated each word right after they heard it”. Llompарт and Reinisch (2018: 603) do use the term “explicit”, but not referring to the type of imitation directly: “participants... were explicitly told that they had to wait until the native speaker finished talking before imitating”. In contrast, Zajac and Rojczyk (2014: 502) overtly suggested that their method relied on a precise distinction between implicit and explicit imitation by specifying that “twenty participants took part in the first session in which target-model words were presented without specific instructions inducing imitative behaviours: the participants were only instructed to wait until the recorded voice stopped producing the word and then read this from the screen. Another twenty participants took part in the second session in which they were instructed to imitate the words they heard as

1 The words in small capitals refer to John Wells' 24 lexical sets (representative keywords) for English vowels (Wells 1982).



faithfully as they could”. Interestingly, the authors reported that providing speakers with explicit instructions to imitate did not have a significant effect on the magnitude of convergence with a native model talker.

To our knowledge, no previous studies have investigated the magnitude of phonetic imitation of English glottal articulations by non-native speakers. In the following two sections, we focus on glottalization in English and Czech and present the hypotheses that emerge from comparing the two language systems and other factors.

1.2 GLOTTALIZATION AND T-GLOTTALING IN ENGLISH

The phenomenon of t-glottaling refers to the realization of the English phoneme /t/ as a glottal stop [ʔ] rather than a voiceless alveolar plosive or similar sounds in words like *cat* or *city*. T-glottaling occurs in many British English varieties, having spread surprisingly quickly in the latter part of the twentieth century, eventually losing its negative connotations (Fabricius 2002; Hughes, Trudgill and Watts 2013). T-glottaling, at least in some phonological contexts, is attested not only in current standard and non-standard varieties of British English (Gavaldà 2016; Schleeff 2021), but also in American English (Eddington and Channer, 2010; Seyfarth and Garellek 2020). Importantly, non-prevocalic environments seem to be more favoured in terms of t-glottaling than intervocalic ones (Cruttenden 2014; Fabricius 2002).

The term “glottal stop” is a cover term for a range of glottal gestures that can be subsumed under a broader term, “glottalization”. The phonetic realization of these glottal events varies, from glottal plosives to various lenited variants, creaky voice or laryngealization (e.g. Ashby and Przedlacka 2014; Keating, Garellek and Kreiman 2015; Redi and Shattuck-Hufnagel 2001). The canonical glottal plosives are produced by a complete closure of the vocal folds, obstructing the airflow into supralaryngeal cavities. As a result, the subglottal pressure increases and is subsequently released by a rapid parting of the folds. Cruttenden (2014: 182) comments on the auditory impressions of the glottal plosive as “its presence being perceived auditorily by the sudden cessation of the preceding sound or by the sudden onset (often with an accompanying strong breath effort) of the following sound”. However, other glottal variants seem to be more prevalent. According to Ashby and Przedlacka (2011: 50–51), “examples of ‘glottal stops’ with silent hold phases are hardly to be found in natural speech at all. Real glottal events are in fact themselves almost invariably ‘lenited’, consisting chiefly of a period of disturbed vocal fold vibration [...] still serving as syllable margins”.²

T-glottaling is closely related to another phenomenon affecting the pronunciation of words like *cat*. T-glottaling is a **glottal replacement** of the alveolar segment, so that no trace of the oral articulation of [t] is left. However, an alternative strategy is a **glottal reinforcement** of [t], which can be seen in segmental terms as a (partially overlapping) sequence of a glottal and an alveolar stop [ʔt] (Cruttenden 2014: 184). The word *cat* could thus be pronounced as [kæʔ] or [kæʔt]. As these strategies are in some respect equivalent, we can describe both types of pronunciation as “glottal articula-

2 See Docherty and Foulkes (1999) and Ashby and Przedlacka (2014) for illustrations of glottal events in varieties of British English. For glottalization in American English, see for instance Seyfarth and Garellek (2020) or Kaźmierski (2020).



tions". Although glottal articulations of /t/ are a common feature of British English, they are not represented in pronunciation dictionaries (Sturiale 2012), which typically contain only phonemic transcription.

The phonological contexts where t-glottaling occurs are summarized under (1). In one environment, the /t/ phoneme is word-final (Vt#, *right*) or followed by a consonant (VtC, *football*). Both represent a syllable-final, non-prevocalic position and thus can be treated together. The contexts (1b) are intervocalic (VtV), either within a word (*city*) or across a word boundary (*a lot of*). They can be considered syllable-final only if we accept Wells' syllabification, which assigns the segment to the syllable coda (Wells 1990). The location of stress is not relevant to the present study, as the /t/ can follow the stressed vowel immediately or at a distance (*senator*). Finally, the context of (1c) is a special case, as it is often not clear whether the following segment is a genuine syllabic consonant (=1a), or a schwa nucleus intervenes (=1b). However, contexts (1c) are not the focus of our study.

(1) Examples of glottalization in English /t/-glottaling

- (a) Non-prevocalic contexts
eat [ˈiːʔ], *right* [ˈraɪʔ], *football* [ˈfʊʔbɔːl], *get down* [geʔdʌʊn]
- (b) Intervocalic contexts
city [ˈsɪʔi], *letter* [ˈleʔə], *senator* [ˈsenəʔə], *a lot of* [ˈlʌʔəv]
- (c) Preceding a syllabic consonant
bottle [ˈbɒʔl], *button* [ˈbʌʔŋ]

Glottalization in English is also connected to the phenomenon of **linking (liaison)** in connected speech. Typically, there is a smooth transition between the word-final and word-initial segment, and no break is perceived between the words. Consonant-to-vowel linking would be the norm in *an hour* [ən_ɑʊə], vowel-to-vowel linking in *two hours* [tuː_ɑʊz]. Transient glides are used after high vowels, whereas liaison /r/³ is used after non-high vowels (Cruttenden 2014: 315–317). However, initial vowels can be pronounced with glottalization when the prosodic context or the pragmatic situation necessitates it, as in emphasis: *I haven't seen* [ʔ]ANYBODY. *She's* [ʔ]AWFULLY good. *Finally, the law* [ʔ]ACTED (cf. Cruttenden 2014: 183). This can occur also within words, where glottalization may function as a morpheme boundary marker. The morpheme starts with a vowel, following a vowel or consonant, as in *reaction* [riːˈrækʃən], *co-operate* [kəʊˈɒpəreɪt] or *post-empiricism* [pəʊstˈemˌpɪrɪsɪzəm]. Unlike in the examples under (1), the glottal stop here is not a realization of a particular segment, but a means of vowel hiatus resolution when the elements are not linked.

3 In their study of the speech of BBC newsreaders, Mompeán and Gómez (2011) report that laryngeal gestures were the most common hiatus breaking strategy in the potential r-liaison sites where no rhotic was used, with the prevailing realization being creaky voice and the canonical stops only occurring in a small minority of cases.



1.3 GLOTTALIZATION IN CZECH

In the Czech language, glottalization has a demarcative function. The glottal plosive or its lenited variants subsumed under glottalization (see previous section) appear at some lexical and morphemic boundaries, cuing the beginning of a vowel-initial word or a morpheme (prefix or stem, but not suffix). In effect, the glottal gesture is a marker of vowel-initial “phonological words”, as exemplified under (2). For instance, *mimoevropský* ‘non-European’ consists of two phonological words: *mimo* ‘outside (of)’ and *evropský* ‘European’. Although words in (2a) are all stressed on the first syllable, words in (2b) demonstrate that the presence of glottal articulations does not depend on stress location. Example (2c) represents a special category, as the words are preceded by non-syllabic prepositions that form one phonetic syllable with the first syllable of the word itself. It is the only context in which the usage of glottalization is mandatory in standard Czech pronunciation (Hála, 1967; Palková, 1994). The pronunciation [‘kʰakʃɪ] or [‘gʰakʃɪ] instead of [‘kʰakʃɪ] is thus a non-standard variant.

(2) Examples of glottalization in Czech

(a) Word-initial contexts

akce [‘ʔakʃɛ] ‘action’, *Evropa* [‘ʔɛvropa] ‘Europe’, *obsah* [‘ʔopsax] ‘content’, *útok* [‘ʔu:tok] ‘attack (noun)’

(b) Word-medial, prefix- or stem-initial contexts

protiakce [‘prociʔakʃɛ] ‘counteraction’, *mimoevropský* [‘mimoʔevropski:] ‘non-European’, *bezobsažný* [‘besʔopsažni:] ‘contentless, without content’, *zaútočit* [‘zaʔu:toʃit] ‘attack (verb)’

(c) Contexts with non-syllabic prepositions

k akci [‘kʰakʃɪ] ‘to action’, *s Evropou* [‘sʔɛvropou] ‘with Europe’, *v okně* [‘fʰokɲɛ] ‘in (the) window’

The contexts (2a) and (2b) provide a choice in standard Czech between a pronunciation with or without glottalization. The likelihood of glottalization is affected strongly by the need for speech clarity, as the presence of glottalization cues the word- or morpheme-initial parses of the stream of speech. However, only 12% of Czech words begin with a vowel (Šturm and Bičan 2021), so this function should not be overrated; on the other hand, many of these words have high frequency, for instance *a* ‘and’, *ale* ‘but’, *aby* ‘so that’, *už* ‘yet, already’, or *on* ‘he’. The rate of occurrence of glottalization in Czech was studied by Volín (2012). He compared a formal speaking style (news-reading on the Czech radio) with spontaneous conversations. Overall, glottalization was present in 95% of potential contexts in the former, whereas it was only 65% of potential contexts in the latter. Women glottalized more often than men regardless of the style.

Importantly, although glottalization in the sense above is a natural part of Czech utterances, t-glottaling does not occur in Czech.



1.4 RESEARCH QUESTIONS AND HYPOTHESES

There are several research questions considered in this study, reflected in the hypotheses under (3). We must distinguish two types of items — **shadowed** and **non-shadowed** — according to whether or not they are presented auditorily during the exposure phase. If shadowing triggers learning of the glottal gesture, the post-test reading task should show a higher rate of glottalization not only in the shadowed items (3a), but also in non-shadowed items (3b). In other words, the expectation is that the process will be applied productively to words of similar structure, yet not previously heard. These are either from the same category (e.g., *meet*, corresponding to shadowed *feet*) or from a new but phonetically and distributionally similar category (/k/ in *week*). Furthermore, the strength of adaptation after exposure should depend on the type of reaction that is required in the task and on memory. According to the hypothesis in (3c), immediate explicit phonetic repetition will elicit higher rates of glottalization than the delayed post-test reading task (performed when the auditory memory of the stimulus has already faded away). The two tasks are otherwise comparable as they both include orthographic intervention on the screen. If the data do not support hypotheses (3a) and (3b) for the post-test performance but do support (3c) for the shadowing, it might mean that there is no actual learning involved, only explicit imitation without an attempt at learning the glottal articulations.

The remaining two hypotheses are related to the comparison of the interacting languages. Hypothesis (3d) predicts that Czech speakers of English will adapt their speech towards glottalization more readily in the intervocalic rather than in the non-prevocalic contexts. Where Czech speakers might expect a [t] or [r] (as in *city*), native speakers of some English varieties produce t-glottaling instead ([ˈsɪʔɪ], [ˈbeʔ], [ˈfʊʔbɔːl]). Importantly, only the first of these — **intervocalic [ʔ]** — has a corresponding segmental structure in the Czech language (e.g., [ˈnaʔopak]), whereas the other two contexts — pre-pausal and pre-consonantal — are unfamiliar in Czech. We argue that it is the familiar structure that should be more easily shadowed and retained.

Alternatively, however, **the non-prevocalic context** might in fact prove easier to imitate, given the frequency of t-glottaling in various word positions in the learners' input. As Fabricius (2002) showed, t-glottaling is a consistent feature of Standard Southern British English only in pre-consonantal environments (and less consistently in pre-pausal environments); furthermore, her data from pre-vocalic environments suggest that the intervocalic position is prone to resist t-glottaling in this variety. As a result, the Czech speakers would be more ready to accept VʔC or Vʔ# forms that they encounter quite frequently, as compared to the relatively rare VʔV forms that might thus be less familiar. The experiment should resolve which of the two aspects — language internal or language external — plays a more crucial role.

(3) Hypotheses regarding the effect of phonetic shadowing

- (a) Speakers will glottalize more in both post-exposure tasks than in the baseline task. (Shadowing task triggers learning.)



- (b) Non-shadowed items with similar characteristics will also show a higher rate of glottalization in the post-test as compared to the baseline. (Shadowing task triggers learning and the process is generalized based on analogy.)
- (c) Speakers will glottalize more in the immediate repetition task than in the delayed post-exposure task. (Imitation is present, learning not necessarily or to a lower degree.)
- (d) Speakers will glottalize more in VtV contexts than in VtC or Vt# contexts. (Language structure is a key factor.)
- (e) Speakers will glottalize less in VtV contexts than in VtC or Vt# contexts. (Language input frequency is a key factor.)

2 METHOD

2.1 MATERIAL

The material includes several types of items. Some appeared in all three tasks (shadowed items, 4a-c), others only in the baseline and post-test tasks (non-shadowed items, 4d-e). In order to facilitate fluency in reading, difficult or less familiar words were not used. Crucially, the /t/ segments appeared either in an **intervocalic (VtV)** position, or in a **non-prevocalic (VtC or Vt#)** position, where # marks a word-final position before a pause. There were 16 items in each category. Filler items comprised 16 additional words that did not involve a /t/ or /k/ segment at all, with the aim of concealing the object of investigation to some degree. Finally, the material contained 18 more words to test whether participants generalize beyond the auditorily presented words. Nine had a /t/ segment in the same positions, nine a /k/ segment in the non-prevocalic position (note that in *fact*, *picture*, and *practice* only the velar plosive is expected to be glottalized). These **non-shadowed tokens** will be referred to as from the **same** or **different category**, respectively.

(4) Experimental material

- (a) VtV: *beautiful, better, butter, cutting, daughter, eat up, energetic, forty, getting, hotter, it all, lot of, patriotic, relative, sort of, water.*
- (b) VtC or Vt#: *a lot, bat, bit, cat, cut, eight, feet, fit, football, hot, hot weather, sit down, start, straight, what, white.*
- (c) Filler items: *ago, always, blue, brother, deal, eyebrows, floor, girls, home, children, money, mouse, nothing, roof, shoes, walls.*
- (d) Non-shadowed tokens of /t/ (same category): *city, later, letter, putting, a lot more, meet, nightlife, nut, rot.*
- (e) Non-shadowed tokens of /k/ (different category): *background, book, fact, joke, picture, practice, sack, shock, week.*

Instead of selecting and recording a single speaker, we opted for a multi-speaker approach that increases the variability of voices and idiolects. The recordings of words with the target segments were obtained from two sources. First, most of the re-



cordings come from a corpus of Southern British English collected for an earlier sociophonetic study (Przedlacka 2002), where realizations of /t/ intervocally and non-prevocally was one of the phonetic variables. The words or short phrases were responses to a spoken lexical questionnaire (e.g. *What happens to water in 100° C?*) in one-to-one interviews in the subjects' schools (aged 14 to 16). The questionnaire was preceded by an informal chat about their interests and plans. The speakers appeared relaxed during the questionnaire and often made informal asides. Those factors indicate that the data represents a relatively casual speaking style in spite of the short, often single-word utterances (note especially that it was not a word-list reading task). The original analogue field recordings were digitized at a sampling rate of 16 kHz.

The accent of the informants was Estuary English and Cockney. Phonologically, their pronunciation was fairly close to Standard Southern British English (the accent variety described in standard textbooks on British English pronunciation). A variety of /t/ tokens were produced. For the present experiments, words with t-glottaling, i.e., with glottal (stop, creak or other) realizations of /t/ (all transcribed as [ʔ]) were purposely selected. That is, glottally reinforced voiceless alveolar stops ([t̟]) were excluded, so that the type of material could be restricted to a single category.

The second source of recordings are YouTube videos. Since the number of tokens with glottal replacement in the corpus mentioned above was not sufficient or included the target segments in other contexts, several more speakers and words were added. Care was taken to make the stimuli as similar to the previous recordings as possible. Namely, single-word utterances were selected. The speakers' age or background was often unknown, but they appeared in their twenties or thirties. The quality of the recordings was comparable; we also converted the recordings to mono files with a sampling rate of 16 kHz to fit in with the previous recordings. All stimuli were normalized to an RMS of 70 dB in *Praat* (Boersma and Weenink 2021) and saved as wav files, with an additional short silence at the beginning and end of the stimuli.

2.2 PARTICIPANTS

In total, 30 participants were recorded for this study (15 male and 15 female). Their age ranged from 19 to 39 years, with a mean of 25.3 (SD = 5.0). They were all native speakers of Czech without any reported hearing or speaking disorders. They were compensated financially for their time. Most of the participants studied at the Faculty of Arts, Charles University, Prague, a minority worked at the university library or had no connection to the university. In effect, there were three groups of participants: (i) students of English studies, i.e., expert users of the English language familiar with accents variation (2 participants), (ii) students of phonetics as a full programme (not an introductory class), i.e., expert listeners (5 participants), and (iii) other participants ($n = 23$). Given their **expert** knowledge, the seven participants from groups (i) and (ii) will therefore be analyzed separately. In addition, an attempt was made to take into account the approximate level of the speakers' English. The participants were asked for their CEFR level (e.g., B1, C2 etc.) and the onset of learning English. There was one participant with level A (length of learning



English = 15 years), 16 participants with level B (mean length of learning English = 15.8 years), and 13 participants with level C (mean length of learning = 18.9 years). In the analyses, the levels will be merged into two broad categories: **intermediate** (A or B) and **advanced** (C).

2.3 PROCEDURE

The experiment was prepared in *DMDX* (Forster and Forster 2003), a software for visual and auditory stimuli presentation. There were three blocks separated by a short break. The first was a **pre-test (baseline) reading task**: participants read the words that appeared on the screen with their usual English pronunciation. No other action was necessary, as the items were run continuously (the word in upper-case letters was displayed for 2.8 sec). The second block was a **shadowing (imitation) task**: participants saw the words again on the screen (for 4 sec) and simultaneously heard the auditory stimulus over headphones. They were asked to imitate the pronunciation of the native speaker as closely as possible. As before, they were allowed only one attempt and no repetition. The block was followed by a break and a **post-test reading task** (identical to the baseline pre-test). The three blocks were in a fixed order, but the order of items within each block was randomized for each participant.

Apart from the presence of an auditory stimulus during shadowing, the blocks also differed in the number of items. There were 16 intervocalic, 16 non-prevocalic and 16 filler items in each block (= 48 shadowed items). However, the pre-test and post-test included 18 additional items that were not heard during shadowing (= non-shadowed items). A total of 180 items were presented to the participants, 12 minutes in total. Additional 3–5 minutes were needed for instructions, training and breaks. Three filler items that did not appear in the experimental material and that did not include a /t/ segment were used for a training session, in which the participants familiarized themselves with the procedure and stimuli, particularly with the pacing (how much time they had for a response). Each participant gave an informed consent before the experiment and filled a questionnaire about their background (see Section 2.2).

2.4 CODING

Excluding fillers, there were 3,960 observations (30 speakers × 132 items). Shadowed items were produced three times by each speaker, while non-shadowed items were produced only twice (in the pre-test and post-test conditions).

Each token was examined in *Praat* (Boersma and Weenink 2021) and analyzed with respect to the phonetic realization of the target segments. A combination of spectrographic and waveform cues together with auditory analysis resulted in a transcription of the sound as one of the following: alveolar or velar plosives ([t], [d], [k], [g]), alveolar or velar plosives with no audible release ([t̚], [d̚], [k̚], [g̚]), alveolar flap ([ɾ]), glottal plosive ([ʔ]), glottally reinforced plosives ([t̪], [k̪]), elision ([∅]). Aspiration, affrication or spirantization were not distinguished (so [t^h], [t̪s] and [t̪s] were all transcribed as [t]). The “no audible release” category has formant transitions towards the end of the preceding vowel, and a /t/ or other respective segments (/d/, /k/, /g/)



are thus perceived in careful listening. In contrast, “elision” differs from no audible release in that the word *straight* sounds like *stray* and *white* like *why*, leaving no trace of formant transitions in the offset of the vowel.

The most relevant distinction was between tokens with and without the glottal gesture. In non-glottal [t], there is a sudden cessation of sound at the offset of the preceding vowel, and the mostly silent phase is followed by an oral release gesture producing a short burst with energy concentrated around 4 kHz. In a glottally reinforced [ʔt], the alveolar gesture is accompanied — often preceded — by glottalization in the form of a complete glottal stop or irregular, creaky phonation. Importantly, the glottal gesture is localized in the vicinity of the /t/ segment. Namely, some speakers may produce creaky phonation in all vowels as a habit, but this was not considered an instance of a glottal segment unless other, localized cues were also present (e.g. sudden loss of formant structure at the V-C transition, prolonged duration of the silent interval). Although these characteristics are relevant for the glottal replacement of [t] (t-glottaling) as well, in that case the alveolar gesture is completely missing.

2.5 STATISTICAL ANALYSIS

For the purposes of this study, it is sufficient to categorize the transcriptions described above into two types only: (i) **glottalized tokens** ([ʔ], [ʔt], [ʔk]) and (ii) **non-glottalized tokens**. The second type might be labelled “oral”, but it also includes elided segments. The binary variable of glottalization was statistically analyzed using mixed-effects logistic regression (Bates et al. 2015; Jaeger 2008; Winter 2020). The regression output is in terms of logit (log-odds ratio) values, but is displayed in effect graphs as probability, using the libraries *emmeans* (Lenth 2020) and *ggplot2* (Wickham 2009). All data processing and statistics was performed in R (R Core Team 2020). The significance of an effect was evaluated by likelihood ratio tests (LRT), where a model with an effect or interaction term is compared to a reduced model without that parameter. When multiple comparisons are reported, the *p*-values have been adjusted with Bonferroni correction and are subsequently compared to $\alpha = 0.05$.

The regression models include fixed effects (TASK, POSITION, SEX, EXPERTNESS, LEVEL) and random effects (ITEM and PARTICIPANT). The slope of the relevant fixed effect is in addition allowed to vary across participants, which makes it possible to include interactions between for instance PARTICIPANT and POSITION (in other words, the model will be able to capture the data more accurately since some participants might show an effect of position, while others might not show it or show it to different degrees). Biological sex, expertness and proficiency level in English are included as control variables.

3 RESULTS

3.1 EFFECTS OF TASK AND POSITION (SHADOWED ITEMS)

The data in Table 1 and Figure 1 suggest that very few tokens were glottalized in the baseline reading task. Any such cases were part of the non-prevocalic con-

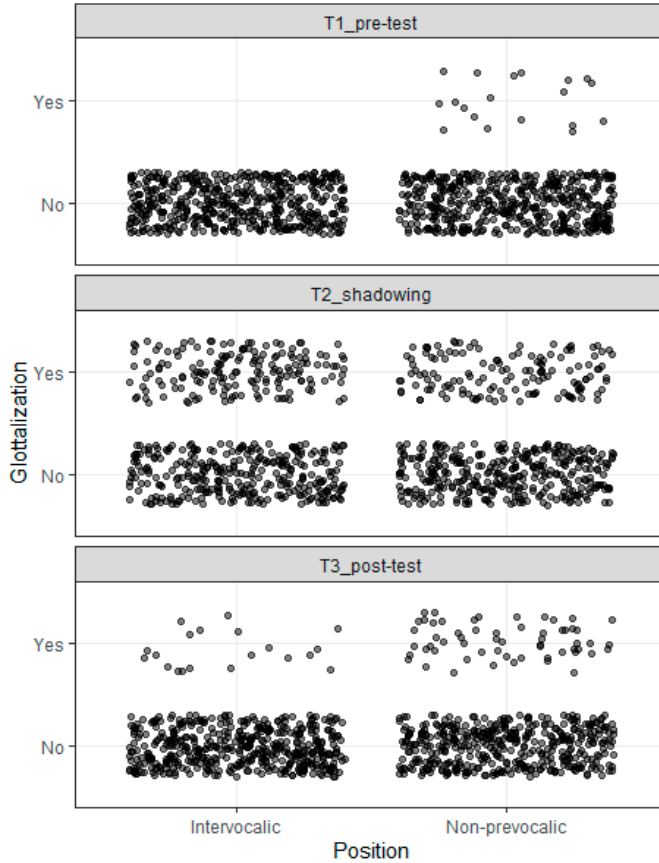


FIGURE 1: Scatterplot of the shadowed data in the three tasks depending on the position of /t/. Glottalized tokens at the top, non-glottalized at the bottom of each plot.

texts.⁴ A much larger proportion of glottalized tokens was produced in the shadowing task. Although the figure does not seem to suggest a difference between the two positions, Table 1 shows that the intervocalic position was associated with somewhat higher rates of glottalization. Finally, the post-test reading task yielded lower rates of glottalization, and the pattern with respect to positions was reversed.

To evaluate the two variables and their interaction, a logistic model was fitted to the data after the intervocalic baseline category had been replaced with NA values due to zero observations of glottalized items. Varying by-participant random slopes for the effects of position and task were allowed. The TASK was a significant predic-

⁴ A note concerning the items *eat up*, *sort of*, *lot of* and *it all* (of the VtV type): they were all ($n = 120$) pronounced as non-glottalized in the pre-test, i.e., with an alveolar consonant. 34% of the tokens were pronounced with [r] or [d] and 35% with [t], indicating linking (*eatup* etc.). The remaining 30% were pronounced with final [t] plus word-initial [ʔ], indicating the words were not treated as a group (*eat ʔup* etc.).



tor ($\chi^2(2) = 46.96, p < 0.001$) but POSITION was not ($\chi^2(1) = 0.02, p = 0.889$). However, the interaction term was evaluated as significant ($\chi^2(1) = 28.19, p < 0.001$). Post-hoc pairwise comparisons revealed that the non-prevocalic position in the post-test was associated with significantly higher rates of glottalization as compared to VtV ($p = 0.0013$), unlike in the shadowing task where there was no significant contrast ($p = 0.586$). Further comparisons confirmed that the shadowing task was indeed significantly different from both the pre- and post-test tasks ($p < 0.001$) and that, for non-prevocalic items, the post-test yielded significantly higher rates of glottalization than the pre-test ($p = 0.018$).

Task	Position	Glottalization		Percent glottalized
		No	Yes	
Pre-test	Intervocalic	480	0	0% [0% – 0.8%]
	Non-prevocalic	461	19	4% [2.4% – 6.1%]
Shadowing	Intervocalic	311	169	35% [30.9% – 39.7%]
	Non-prevocalic	341	139	29% [24.9% – 33.2%]
Post-test	Intervocalic	460	20	4% [2.6% – 6.4%]
	Non-prevocalic	421	59	12% [9.5% – 15.6%]

TABLE 1: The number of glottalized and non-glottalized tokens according to task and position. The 95% confidence intervals were computed from a binomial test.

3.2 EFFECTS OF PARTICIPANT VARIABLES (SHADOWED ITEMS)

Figure 2 shows the percentage of glottalized items in the three tasks and two positions, comparing the relevant characteristics of participants. Biological sex of the speakers (top panel) does not seem to be a major factor, as the confidence intervals overlap. It is probably not involved in interactions either. As for expertness of the speakers (middle panel), experts (phoneticians or students of English as a major) generally showed considerably higher rates of glottalization, especially in the shadowing task. English proficiency level (bottom panel) seems to have a similar effect, favoring glottalization in advanced speakers.

A logistic model was created for the *shadowing task*. POSITION was not a significant predictor ($\chi^2(1) = 1.60, p = 0.206$); the factors SEX and LEVEL were not significant either ($p > 0.05$), but EXPERTNESS reached statistical significance ($\chi^2(1) = 4.23, p = 0.040$), with expert participants yielding a higher probability of glottalization. Moreover, there was a significant interaction of POSITION with EXPERTNESS ($\chi^2(1) = 4.52, p = 0.034$), but not with LEVEL ($p = 0.962$). In other words, whereas both advanced and intermediate speakers behaved in the two positions in a similar way, the expert participants not only glottalized more than the non-experts, but they also glottalized more in the intervocalic position than in the non-prevocalic position. This is captured in Figure 3.

Another logistic model was fit to the data from the *post-test task*. This time, no significant interactions emerged (EXPERTNESS: $\chi^2(1) = 0.03, p = 0.860$; LEVEL: $\chi^2(1) = 0.61$,

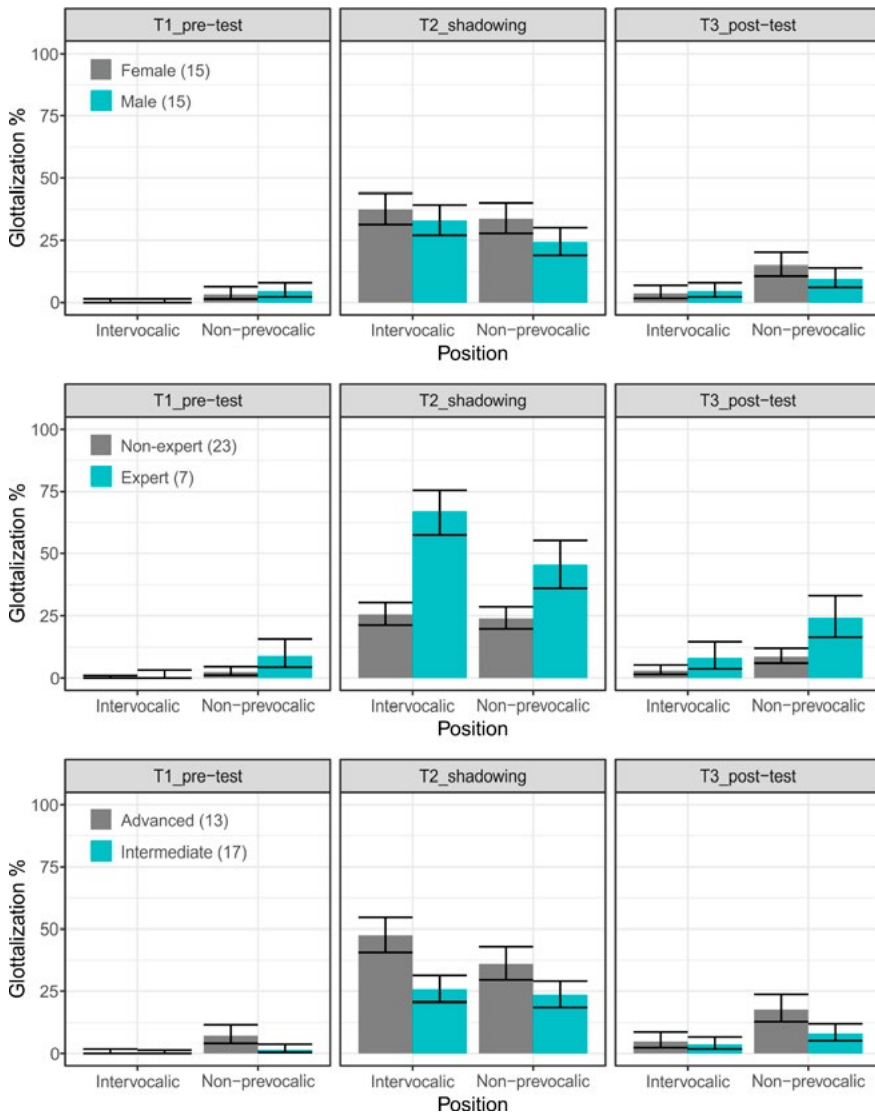


FIGURE 2: The rate of glottalization according to task and position. From top to bottom: biological sex, expertness, level. The 95% confidence intervals were computed from a binomial test.

$p = 0.434$), and a model without these terms seems more warranted, as the interactions have nearly no impact. Individual factors were not significant either ($p > 0.05$), with the exception of POSITION ($\chi^2(1) = 6.97, p = 0.008$). As Figure 4 indicates, it was the non-prevocalic position that was associated with a higher probability of glottalization. However, it is clear from the figures (and the previous section) that the rate of glottalization in the post-test task was generally low.

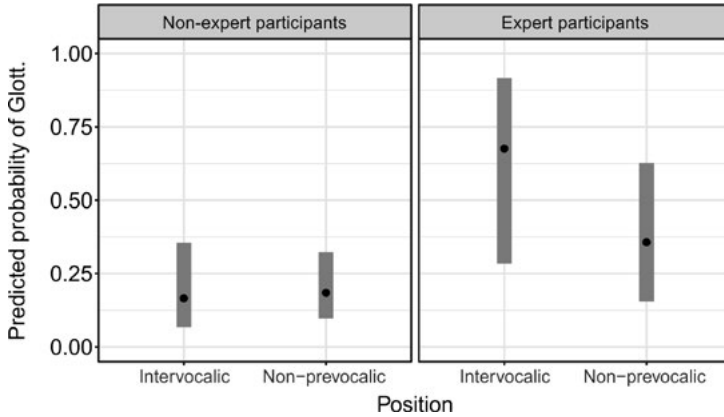


FIGURE 3: Interaction of position with expertness in a logistic model of the shadowing task.

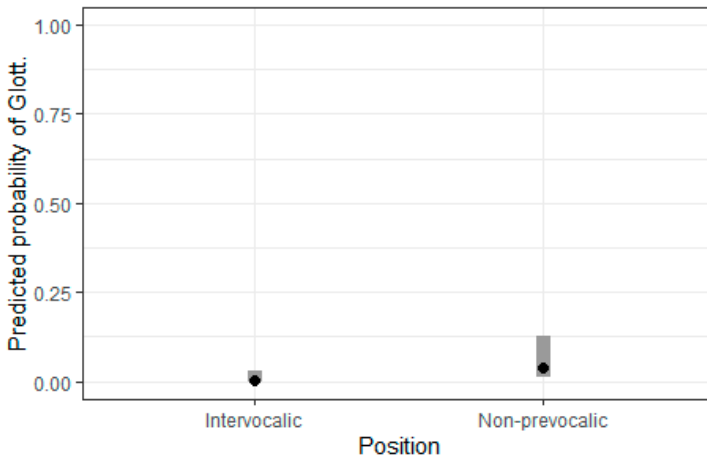


FIGURE 4: The effect of position in a logistic model of the post-test task.

3.3 GENERALIZATION TO NON-SHADOWED TOKENS

Let us move now to the items not presented as auditory stimuli. Table 2 shows that in the baseline reading, shadowed and non-shadowed items did not differ much in the rate of glottalization, as predicted (it was *before* the phonetic imitation phase). There were only two tokens of glottalized /k/ and five tokens of glottalized /t/ (approx. 1–2% of the observations), all in the non-prevocalic context.⁵ The shadowing data from the previous section are presented for comparison (32% glottalized). The crucial result concerns the post-test reading task. Table 2 suggests a descending order of glottalization rates from shadowed items to non-shadowed /t/ (same category) and to non-

⁵ For /k/: words *shock* (M11) and *fact* (F05). For /t/: words *rot* (M02 + F05), *nut* (M03) and *a lot more* (M13 + F05).



shadowed /k/ (different category).⁶ However, the differences between the categories were not substantial; rather, they indicate interesting tendencies. A comparison can also be made between the pre-test and post-test results. For all types of items, there was a minor increase of glottalized tokens in the post-test task.

Task	Set	Glottalization		Percent glottalized
		No	Yes	
Pre-test	Shadowed	941	19	2% [1.2% - 3.1%]
	Non-shadowed: /t/	265	5	2% [0.6% - 4.3%]
	Non-shadowed: /k/	268	2	1% [0.1% - 2.7%]
Shadowing	Shadowed	652	308	32% [29.1% - 35.1%]
Post-test	Shadowed	881	79	8% [6.6% - 10.2%]
	Non-shadowed: /t/	257	13	5% [2.6% - 8.1%]
	Non-shadowed: /k/	264	6	2% [0.8% - 4.8%]

TABLE 2: The number of glottalized and non-glottalized tokens according to task and set of items. The 95% confidence intervals were computed from a binomial test.

To evaluate these results, a logistic model was fitted to the pre- and post-test data. The effect of TASK was significant ($\chi^2(1) = 7.89, p = 0.005$), but none of the other fixed effects reached significance, including SET ($\chi^2(2) = 4.45, p = 0.108$). The interaction between TASK and SET was not significant either ($\chi^2(2) = 1.59, p = 0.452$), suggesting that there is a higher probability of glottalization in the post-test phase regardless of the type of items. Therefore, when individual variation and the other control variables are considered in the model, the interesting differences apparent in Table 2 are not supported in a generalized model.

3.4 REALIZATION OF GLOTTALIZATION

It would be useful to differentiate between glottal replacement and glottal reinforcement in the data (it should be noted that the audio stimuli contained only glottal replacement). Table 3 shows that reinforcement occurred mainly in the non-prevocalic position, and that it was almost absent in the shadowing task. The speakers thus imitated the t-glottaling accurately, i.e., as a glottal gesture and not as a combination of a glottal and oral gesture. In the post-test phase, the contribution of reinforcement to the glottalized tokens was quite substantial. Moreover, Table 3 reveals that alveolar flaps [r] were produced — quite expectedly — in the intervocalic position, unlike plosives with no audible release or deleted segments, which were associated mostly with the non-prevocalic position.

⁶ For /k/: words *sack* (3×), *shock*, *joke*, *background*, produced by three different speakers (M01, F03, F06). For /t/: words *rot* (5×), *nut* (2×), *meet*, *a lot more* (2×), *nightlife*, *letter* (2×), produced by eight different speakers (F02, F03, F05, F06, F08, M10, M11, M13). Furthermore, there was another item with a word-final plosive, [p] (*eat up*), outside our analysis; it was replaced by a glottal stop once (M07, non-expert).



Task	Position	Glottalized		Non-glottalized					
		ʔ	ʔ̥	t	d	r	t̥	d̥	ø
Pre-test	Intervocalic	0	0	468	25	107	0	0	0
	Non-prevocalic	5	19	543	27	0	18	16	2
Shadowing	Intervocalic	168	1	176	6	77	0	0	52
	Non-prevocalic	134	5	141	12	0	36	27	125
Post-test	Intervocalic	19	3	403	9	155	1	0	10
	Non-prevocalic	39	33	421	20	0	27	25	70

TABLE 3. The number of tokens from each category depending on task and position.

4 DISCUSSION

4.1 IMITATION AND LEARNING

This study examined the process of explicit phonetic imitation of English words with t-glottaling by Czech participants. In the set of hypotheses assembled under (3), the first three concern the form and degree of learning, i.e., what an increase in glottal articulations of /t/ would reflect. Hypothesis (3a) was confirmed, as there was an increase in post-test glottalization as a response to the shadowing task. Hence, some degree of learning must have taken place, although it is open to question what its nature and extent is (see below) and whether it is a short-term or a long-term effect. The small size of the effect hinders any serious interpretation, apart from saying that long-term learning is probably not triggered, and the effect would disappear within several days.

Importantly, there were higher rates of glottalization in the post-test not only in the shadowed items, but also in the non-shadowed items (3b). These are novel items in the sense that they have not been presented as sound stimuli to the participants. The Czech participants are expected to productively apply t-glottaling to words that are in some respect analogous to the shadowed items. Presumably, position of the target segment is a crucial conditioning of the process (e.g. /t/ in *talk* would never be glottalized). Since there were no instances of glottal /k/ or /p/ in the shadowing stimuli, two scenarios may follow. On the one hand, the glottal gesture learned from the shadowed items may be extended to all phonetically similar segments (voiceless plosives) in the relevant *positions* (intervocalic or non-prevocalic). In that case, both types of non-shadowed items (/t/ and /k/) should show increased glottalizations in the post-test, and to a similar degree. On the other hand, if the basis for analogy is *category membership*, we would predict glottalization of the non-shadowed /t/ items to the exclusion of the /k/ items. A difference between the two phonemes should ensue.

The participants were able to generalize the imitation of t-glottaling to novel items of both types, although we must keep in mind that the number of glottalized tokens in the post-test was generally low. Statistically speaking, the interaction of task with item type was not significant, leading to a post-test increase in glottalization for all types of items. This seems to point to the latter conclusion: generalization



to [k] in *week* stems from the fact that [k] occupies the same position as in *feet* (and the similarity of [k] to [t] phonetically/phonologically is just a prerequisite). However, the patterns were more nuanced. The three sets of items yielded a gradual decrease in the differences between the pre-test to post-test results: the departure from the control condition was highest for shadowed items (change from 2% to 8%), lower for non-shadowed items from the same category (= /t/, from 2% to 5%), and lowest for non-shadowed items from the new category (= /k/, from 1% to 2%). This suggests that (i) there are in fact some differences between /t/ and /k/, and (ii) the generalization ability is not so robust.⁷

Proceeding to the hypothesis in (3c), a comparison of the two post-exposure tasks again suggests that proper learning does not occur. Speakers glottalized much more extensively after immediate exposure in the shadowing task (32% of tokens glottalized) than in the delayed post-test reading task (8% glottalized). Taken together, the findings so far indicate that participants are quite responsive to glottal articulations in immediate phonetic shadowing, but do not retain these productions in a delayed task. In fact, the process at play seems to be imitation — which was explicitly called for in the instructions — rather than learning.

Nevertheless, we cannot preclude the possibility of a combination with learning as there was at least some increase from pre-test to post-test.⁸ Moreover, if elided /t/s or plosives with no audible release (Table 3) are counted as (imperfect) imitation (an intermediate category between [t] and [ʔ]), the imitative behaviour of our speakers might actually be better than what we reported based on glottalized items only, and the effect of exposure to t-glottaling would stand out more.

4.2 LANGUAGE INTERNAL VS. EXTERNAL FACTORS

Hypotheses (3d) and (3e) involved a comparison between two phonological environments. Structural differences between English and Czech led us to the prediction that imitating words with the glottal stop [ʔ] in intervocalic position (VtV) would be an easier task for the participants than imitating words with the target in non-prevocalic position (VtC or Vt#). The rationale was the presence of glottal stops in Czech intervocalically, but not before a consonant or a pause. Alternatively, a higher frequency of t-glottaling in VtC/Vt# than in VtV in the English input, especially in standard varieties, would predict the opposite pattern of results. Facilitation of t-glottal-

7 For instance, Nielsen (2011) examined the amount of aspiration after exposing her participants to lengthened VOT tokens of /p/. The imitated production was generalized to new instances of the target phoneme /p/, and also to the novel phoneme /k/, albeit to a lesser degree. Nevertheless, the generalization effect was much stronger in her experiment than in ours. The reason may of course relate to the different target feature (t-glottaling vs. VOT), or to a difference in method (also, see the following footnote).

8 The low retention of glottal articulations could also be linked to the following explanation. Three words in the material included a final plosive that was not glottalized in the stimuli, which runs counter to the expectations based on /t/. The distracting presence of non-glottalized word-final [p] (*eat up*) and [k] (*patriotic, energetic*) could have lowered post-test generalization to /k/.



ing in one of the two positions was expected for both tasks (immediate shadowing and the post-test reading).

In the baseline condition — before exposure — it is easy to explain why there were no glottal tokens intervocalically and a somewhat higher incidence of glottalization word-finally or before a consonant. Czech learners of English are likely to be targeting either Standard Southern British English or General American pronunciation (Jakšič and Šturm 2017). In the former, glottal reinforcement or replacement is widespread in the non-prevocalic position, whereas t-glottaling in the intervocalic position is a characteristic of non-standard varieties and might be perceived as stigmatizing. In the latter, intervocalic /t/s are usually realized as alveolar flaps (this was indeed the pronunciation of many of our speakers). So Czech learners have no impetus to glottalize VtV in their own speech, and some might glottalize non-prevocalic /t/.

In the *shadowing task*, there was indeed a larger number of glottalized tokens in the intervocalic condition (35% vs. 29%), but the difference was not significant. The predicted effect appeared only when the participants were split into experts, who produced significantly higher rates of glottalization in the VtV position, and non-experts (without such a difference). This could not be extended simply to the English proficiency level, since advanced and intermediate speakers behaved in the two positions in a similar way. The hypothesis (3d) was thus supported only for some participants. What seems to be the case considering all the data is that the higher rate of glottalization in the intervocalic position reflects a conscious copying of the glottal realization of /t/, which is especially salient intervocalically. Naturally, the expert participants are likely to be attentive to such details of articulation as t-glottaling in following the instruction to imitate the recordings “as closely as possible”.

Moreover, in the *post-test task*, we found evidence of the opposite direction of the effect, namely, the non-prevocalic position being associated with higher rates of glottalization (12%) compared to VtV (4%). This supports hypothesis (3e). As in the pre-test, Czech learners do not have the motivation to glottalize intervocalic contexts due to the absence of t-glottaling in standard British and American accents, a sort of anti-t-glottaling VtV constraint. Without the immediacy of phonetic shadowing, the minimal learning that might have occurred as a response to the exposure of t-glottaling is thus reflected mainly in the non-prevocalic position.

4.3 PARTICIPANT VARIABLES

Biological sex of the participants was not a significant predictor. It is well known that women usually glottalize more than men in various speech corpora (e.g., Seyfarth and Garellek 2020; Volín, 2012), but it is not clear whether we should also predict a difference in terms of the imitation ability. More important participant aspects in our study were those already mentioned: the proficiency level in English and expertness. Since their effects were similar (for more proficient participants, the shadowing intervention led to somewhat higher rates of glottalization both immediately and with delay), one might expect a strong correlation between the two variables. However, a variance inflation factor (VIF) analysis did not reveal any substantial multicollinearity in the predictors. Moreover, there were 13 advanced participants, but only 7 expert participants (with substantial overlap). It would be necessary to exam-

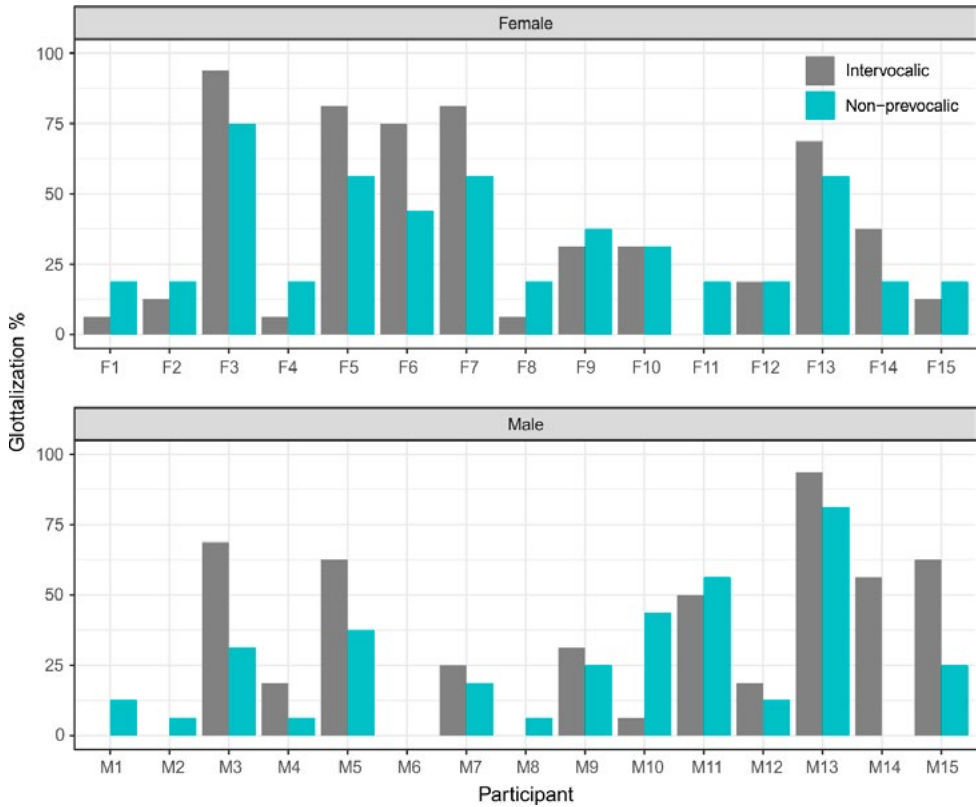


FIGURE 5: The effect of position in the shadowing task for individual participants.

ine a balanced group of participants in a factorial design to evaluate the contribution of these effects reliably. Finally, a future experiment should provide exclusively auditory stimuli (without orthography) during the shadowing task, as especially the non-expert participants may have been reluctant to replace the canonical alveolar [t] with a glottal stop when presented with the words in their orthographic form on the screen.

There is unfortunately no space to discuss individual variation, although it is clear that various dispositions of the speaker or other speaker variables can affect the degree and conditions of phonetic imitation. Different speakers showed different patterns in terms of (i) the general rate of glottalization, (ii) the strength and direction of the position effect, (iii) the strength of retention of the imitated glottal gestures in the post-test, (iv) the particulars of the articulated target sounds (see Table 3). To illustrate some of these, Figure 5 is presented below with the data from the shadowing task. For instance, speaker M14 produced all glottalized tokens in the intervocalic position, whereas M01 or F11 in the non-prevocalic position. One of the positions could be favored by some speakers, or both positions could be perceived as more or less equivalent by others.



5 CONCLUSIONS

This paper has focused on the imitation and learning of glottal realizations of English /t/ by native speakers of Czech. The glottal variant, a well-established feature of both standard and non-standard varieties of British English, is rarely explicitly taught to users of English as L2. We have provided evidence that imitation is facilitated if the target sound appears in the same position (VtV) in the imitators' target and source language. However, the structural similarities do not positively impact learning, as the successfully imitated native productions tend not to be retained (given the current level of experimental exposure). Other factors seem to play a role in long-term retention, such as previous exposure to glottal allophones of /t/ in classroom and informal learning. Furthermore, expert listening skills and expert knowledge of L2 seem to facilitate imitation.

Acknowledgements

Research supported by the National Science Centre Poland grant *Phonetic imitation in a native and non-native language* (UMO-2019/35/B/HS2/02767) and by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

REFERENCES

- Ashby, M. and J. Przedlacka (2011) The stops that aren't. *English Phonetics 14-15. Festschrift commemorating the retirement and the 70th birthday of Professor John C. Wells*. The English Phonetic Society of Japan.
- Ashby, M. and J. Przedlacka (2014) Measuring incompleteness: Acoustic correlates of glottal articulations. *Journal of the International Phonetic Association* 44, 283-296.
- Babel, M. (2010) Dialect convergence and divergence in New Zealand English. *Language in Society* 39, 437-456.
- Babel, M. (2012) Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40, 177-189.
- Babel, M., G. McGuire, S. Walters and A. Nicholls (2014) Novelty and social preference in phonetic accommodation. *Laboratory Phonology* 5, 123-150.
- Bates, D., M. Mächler, B. Bolker and S. Walker (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67, 1-48.
- Boersma, P. and D. Weenink (2021) *Praat: Doing phonetics by computer (version 6.1.42)*. Retrieved from <http://www.praat.org/>.
- Chang, C. B. (2012) Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40/2, 249-268.
- Cruttenden, A. (2014) *Gimson's Pronunciation of English*, 8th ed. London: Routledge.
- de Jong, K., Y. C. Hao and H. Park (2009) Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent. *Journal of Phonetics* 37/4, 357-373.
- Docherty, G. and P. Foulkes (1999) Derby and Newcastle: Instrumental phonetics and variationist studies. In: Foulkes, P. and G. Docherty (eds) *Urban voices*, 47-71. London: Arnold.
- Eddington, D. and C. Channer (2010) American English has go? a lo? of glottal stops: Social diffusion and linguistic motivation. *American Speech* 85, 338-351.

- Fabricius, A. (2002) Ongoing change in modern RP: Evidence for the disappearing stigma of t-glottalling. *English World-Wide* 23/1, 115–136.
- Flege, J. E. and W. Eefting (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.
- Forster, K. I. and J. C. Forster (2003) DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers* 35, 116–124.
- Gasiorek, J., H. Giles and J. Soliz (2015) Accommodating new vistas. *Language and Communication* 41, 1–5.
- Gavaldà, N. (2016) Individual variation in allophonic processes of /t/ in Standard Southern British English. *The International Journal of Speech, Language and the Law* 23/1, 43–69.
- Giles, H., J. Coupland and N. Coupland (1991) *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press.
- Goldinger, S. D. (1998) Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105/2, 251–279.
- Hála, B. (1967) *Výslovnost češtiny I (výslovnost slov českých) [The pronunciation of Czech I (pronunciation of Czech words)]*. Prague: Academia.
- Hao, Y. C. and K. de Jong (2016) Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics* 54, 151–168.
- Hughes, A., P. Trudgill and D. Watt (2013) *English Accents & Dialects: An Introduction to Social and Regional Varieties of English in the British Isles*. London: Routledge.
- Jaeger, T. F. (2008) Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59, 434–46.
- Jakšič, J. and P. Šturm (2017) Accents of English at Czech schools: Students' attitudes and recognition skills. *Research in Language* 15, 353–369.
- Jia, G., W. Strange, Y. Wu, J. Collado and Q. Guan (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.
- Kaźmierski, K. (2020) Prevocalic t-glottalling across word boundaries in Midland American English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11/1, Article 13.
- Keating P., M. Garellek and J. Kreiman (2015) Acoustic properties of different kinds of creaky voice. In: *Proceedings of the 18th International Congress of Phonetic Sciences*, paper 821. Glasgow: University of Glasgow.
- Kuhl, P. K. and A. N. Meltzoff (1996) Infant vocalizations in response to speech: Vocal imitation and developmental change. *Journal of the Acoustical Society of America* 100, 2425–2438.
- Kwon, H. (2019) The role of native phonology in spontaneous imitation: Evidence from Seoul Korean. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10/1, Article 10.
- Kwon, H. (2021) A non-contrastive cue in spontaneous imitation: Comparing mono- and bilingual imitators. *Journal of Phonetics* 88, 101083.
- Labov, W. (2001) *Principles of Linguistic Change, Volume 2: Social Factors*. Oxford: Blackwell.
- Lenth, R. (2020) emmeans: Estimated Marginal Means, aka Least-Squares Means. (1.5.2-1) [Computer software]. Available at: <https://CRAN.R-project.org/package=emmeans>.
- Lin, Y., Y. Yao and J. Luo (2021) Phonetic accommodation of tone: Reversing a tone merger-in-progress via imitation. *Journal of Phonetics* 87, 101060.
- Llompert, M. and E. Reinisch (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.
- Mitterer, H. and M. Ernestus (2008) The link between speech perception and production is





- phonological and abstract: Evidence from the shadowing task. *Cognition* 109/1, 168–173.
- Mitterer, H. and J. Müsseler (2013) Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception & Psychophysics* 75, 557–575.
- Mompeán, J. A. and F. A. Gómez (2011) Hiatus resolution strategies in non-rhotic English: The case of /r/-liaison. In: *Proceedings of the 17th International Congress of Phonetic Sciences*, 1414–1417.
- Namy, L. L., L. C. Nygaard and D. Sauerterig (2002) Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology* 21, 422–432.
- Nielsen, K. (2011) Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39, 132–140.
- Palková, Z. (1994) *Fonetika a fonologie češtiny*. Praha: Karolinum.
- Pardo, J. S., A. Urmanche, S. Wilman, J. Wiener, N. Mason, K. Francis and M. Ward (2018) A comparison of phonetic convergence in conversational interaction and speech shadowing. *Journal of Phonetics* 69, 1–11.
- Podlipský, V. J. and Š. Šimáčková (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.
- Przedlacka, J. (2002) *Estuary English? A Sociophonetic Study of Teenage Speech in the Home Counties*. Frankfurt: Peter Lang.
- R Core Team. (2020) R: A language and environment for statistical computing: Version 4.0.3 [software]. Vienna: R Foundation for Statistical Computing. Available at: <https://www.r-project.org>.
- Redi, L. and S. Shattuck-Hufnagel (2001) Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, 29/4, 407–429.
- Rojczyk, A. (2013) Phonetic imitation of L2 vowels in a rapid shadowing task. In: Levis, J. and K. LeVelle (eds) *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*, 66–76. Ames IA: Iowa State University.
- Rojczyk, A., A. Porzuczek and M. Bergier (2013) Immediate and distracted imitation in second-language speech: Unreleased plosives in English. *Research in Language* 11, 3–18.
- Sancier, M. L. and C. A. Fowler (1997) Gestural drift in bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25, 421–436.
- Schleef, E. (2021) Individual differences in intra-speaker variation: T-glottalling in England and Scotland. *Linguistics Vanguard* 7(s2), 20200033.
- Schouten, M. E. H. (1977) Imitation of synthetic vowels by bilinguals. *Journal of Phonetics* 5, 273–283.
- Seyfarth, S. and M. Garellek, M. (2020) Physical and phonological causes of coda /t/ glottalization in the mainstream American English of central Ohio. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11/1, Article 24.
- Shockley, K., L. Sabadini and C. A. Fowler (2004) Imitation in shadowing words. *Perception and Psychophysics* 66, 422–429.
- Siegel, J. (2010) *Second Dialect Acquisition*. Cambridge: Cambridge University Press.
- Skarnitzl, R. and D. Rálišová (2022) Phonetic variation of Irish English /t/ in the syllabic coda. *Journal of the International Phonetic Association*, online first version. DOI: <https://doi.org/10.1017/S0025100321000347>.
- Sturiale, M. (2012) No bot'le no party: T-glottalling and pronouncing dictionaries. *Language and History* 55/1, 63–74.
- Šturm, P. and A. Bičan (2021) *Slabika a její hranice v češtině [The Syllable and its Boundaries in Czech]*. Praha: Karolinum.
- Volín, J. (2012) Jak se v Čechách “rázuje” [How the glottal stop is realized in Czechia]. *Naše řeč* 95, 51–54.
- Wells, J. (1982) *Accents of English. Volume 1: An Introduction*. Cambridge: Cambridge University Press.
- Wells, J. (1990) Syllabification and allophony. In: Ramsaran, S. (ed) *Studies in the Pronunciation of English: A Commemorative*

- Volume in Honour of A. C. Gimson, 77–86.
London: Routledge.
- Wickham H. (2009) *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.
- Winter, B. (2020) *Statistics for Linguists: An Introduction Using R*. London: Routledge.
- Zajac, M. and A. Rojczyk (2014) Imitation of English vowel duration upon exposure to native and non-native speech. *Poznan Studies in Contemporary Linguistics* 50/4, 495–514.

**Pavel Šturm**

Institute of Phonetics
Faculty of Arts, Charles University
Address: nám. Jana Palacha 2, 116 38, Prague, Czech Republic
ORCID ID: 0000-0001-5521-029X
Pavel.Sturm@ff.cuni.cz

Joanna Przedlacka

Phonetics Laboratory
Faculty of Linguistics, Philology and Phonetics, University of Oxford
Address: Clarendon Institute, Walton Street, OX1 2HG, Oxford, UK
ORCID ID: 0000-0001-5016-0043
joanna.przedlacka@phon.ox.ac.uk

Arkadiusz Rojczyk

Institute of Linguistics
University of Silesia in Katowice
ul. Grota-Roweckiego 5, 41-205 Sosnowiec, Poland
ORCID ID: 0000-0002-7328-5911
arkadiusz.rojczyk@us.edu.pl