



## OPEN ACCESS

## EDITED BY

John Archibald,  
University of Victoria, Canada

## REVIEWED BY

Mariko Nakayama,  
Tohoku University, Japan  
Christine Shea,  
The University of Iowa, United States

## \*CORRESPONDENCE

John H. G. Scott  
✉ jhgscott@umd.edu

RECEIVED 07 July 2023

ACCEPTED 02 September 2024

PUBLISHED 14 February 2025

## CITATION

Scott JHG (2025) Detargeting the target in phoneme detection: aiming the task at phonological representations rather than backgrounds. *Front. Lang. Sci.* 3:1254956. doi: 10.3389/flang.2024.1254956

## COPYRIGHT

© 2025 Scott. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Detargeting the target in phoneme detection: aiming the task at phonological representations rather than backgrounds

John H. G. Scott<sup>1,2\*</sup>

<sup>1</sup>Department of German Studies, School of Languages, Literatures and Cultures, University of Maryland, College Park, MD, United States, <sup>2</sup>L2+ Sound Learning Lab, Division of German, Russian, Arabic Language and Muslim Cultures, School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Calgary, AB, Canada

One challenge of learning a second or additional language (L2+) is learning to perceive and interpret its sounds. This includes acquiring the target language (TL) contrastive phonemic inventory, the sounds' systematic behavior in the TL phonology, and novel relationships between spelling and sound (GPCs; grapheme-phoneme correspondences). Many perception tasks require stipulation of written labels for target speech sounds (e.g., phoneme detection). Listening for this target is not necessarily, or even frequently, an equivalent cognitive task between participant groups. The incongruence of phonological and orthographic domains and their GPCs poses a methodological challenge for L2+ research. The author argues that phoneme detection tasks should avoid the phone of investigative interest ( $x$ ) as the direct target of listener attention and redirect focus to an adjacent listening target ( $y$ ). Ideally, this target should not trigger or otherwise be implicated in the phonological process or phonotactic constraint under investigation. The careful choice of listening target ( $y$ ) with both a familiar sound and a congruent orthographic label for both (or all) language groups of the experiment yields an equivalent task and better indicates implicit knowledge of the phenomenon under study. This approach opens up potential choices of phonological objects of interest ( $x$ ). The two phoneme detection experiments reported here employ this novel adjacent-congruent listening target approach, which the author calls the Persean approach. Experiment 1 establishes baseline performance in two assimilation types and replicates processing inhibition in first-language (L1) German speakers in response to violations of regressive nasal assimilation. It also uses [t] as the Persean listening target to test sensitivity to preceding violations of progressive dorsal fricative assimilation (DFA). Experiment 2 investigates sensitivity to violations of DFA in both L1 German speakers and L1 English L2+ German learners. Experiment 2 also uses the Persean method for the first phoneme detection investigation demonstrating sensitivity to violation of a prosodic/phonotactic constraint banning /h/ in syllable codas. The study demonstrates that phoneme detection with Persean listening targets is a viable instrument for investigating regressive and progressive assimilation, prosodic/phonotactic constraints, and prelexical perceptual repair strategies in different language background groups and proposes statistical best practices for future phoneme detection research.

## KEYWORDS

phoneme detection, L2+ perception, reaction times, phonological representation, grapheme-phoneme correspondence, underspecification, assimilation, German

## 1 Introduction

Language users have different first-language (L1) and prior-language experience profiles, which poses an inherent methodological challenge for intergroup task parity when investigating cross-dialect or cross-language perception and second or additional language (L2+) phonology. Speakers of different languages may perceive a sound differently, have different familiarity with the sound (a familiar phone or allophone for one group is a novel phone for another), or may label it with different sets of letters (single sound-to-letter correspondence for one group and multiple spellings allowed for the sound in another group). When learners seek to acquire the phonology of L2+, one aspect that they must learn is the contrastive phonemic inventory (as well as predictable allophonic variants). Determining whether a particular learner has acquired a particular phoneme presents certain challenges to the psycholinguist. Because we cannot look directly at the phonological grammar, we must turn to a range of experimental tasks and then interpret the behavioral results to infer the relevant properties of the grammar. For example, imagine that we wonder if a learner has acquired a phonemic representation for front rounded vowels /y: ʏ/ in their L2+ German (represented orthographically as <ü><sup>1</sup>), particularly if the L1 inventory lacks the /y: ʏ/ pair. We commonly present tasks to see if they can reliably *identify* or *discriminate* sounds, such as the corresponding back rounded [u: ʊ] and front unrounded [i: ɪ] pairs, from front rounded [y: ʏ] in German words. Alternatively, we may want to see if participants can simply *detect* an [y:] or [ʏ] in a word or a phrase. Detecting the sound of interest (call it *x*) tells us something about the representation of *x* in the learner's interlanguage (IL). However, many factors can influence the behavioral results of this sort of *phoneme detection* task. Is *x* absent from or frequent in the L1? If present, is *x* a predictable allophone or a full phoneme? How is it represented featurally? Is it frequent or rare in the L2+ lexicon and usage? Is the phone reliably encoded in the orthography? All these factors have been shown to influence phonological identification and detection tasks (Bassetti et al., 2015b; Connine, 1994, p. 115–116; Connine and Titone, 1996, p. 639; Cutler and Otake, 1994; Darcy et al., 2007; Frauenfelder and Seguí, 1989; Otake et al., 1996; Scott and Darcy, 2023; Scott et al., 2022; Seguí and Frauenfelder, 1986). The literature presents a complex and, at times, contradictory picture of what the phoneme detection task can tell us.

The phoneme detection task measures accuracy and reaction time (RT) in response to detecting a specified listening target in the stimulus. As with many RT methods, behavioral responses to phoneme detection (accuracy, systematic changes in processing speed) are employed as proxy measures representing underlying grammatical knowledge (Hui and Jia, 2024). The phoneme detection task has the advantage that it does not require high target-language (TL) proficiency or lexical knowledge. As such, it is useful for investigating prelexical processing, even with pre-learner and early L2+ learner groups, as long as the listening target is viable and congruent between languages. In this article, I introduce a new variant of the phoneme detection task to shed light on some phonological representations in L1 and L2+. In this variant of the

phoneme detection task, participants do not focus on detecting the novel L2+ sound of interest *x* but rather attend to a sound that occurs *adjacent* to the object of interest (call it *y*). When *y* is not implicated in the phonological phenomenon of interest, I call this the Persean approach, in reference to how Perseus required the reflection of his shield to look on the Gorgon Medusa's face without being turned to stone by her direct gaze. I explain why having participants detect *x* directly can be as fatal an error as looking directly at Medusa, particularly when investigating multiple language-background groups or if the aim is to investigate *implicit*, or what may be called *optimum* or *automatized explicit*, knowledge in cross-language or IL phonological perception, all of which have theoretical and practical relevance for L2+ acquisition research (Bordag et al., 2021; Rebuschat, 2013; Strange, 2011; Suzuki, 2017). The experimental results of this Persean approach reveal that the detection of *y* can tell us something about the nature of the representation of *x*, adding an important tool to our methodological toolbox.

In Section 2, I highlight the difficulties that arise in task design for phoneme detection experiments. I focus on the problems found in choosing listening targets for L2+ learner experiments, especially regarding task parity for intergroup comparison. In Section 3, I review the sparse literature using phoneme detection tasks to investigate two place assimilation phenomena, right-to-left regressive nasal assimilation (RNA) and progressive (left-to-right) dorsal fricative assimilation (DFA), in German, and critically examine their choices of listening targets with this task. In Section 4, I briefly summarize the prosodic ban on /h/ in syllables codas in English and German to lay the groundwork for experiment 2, which conducts the first phoneme detection investigation of syllable structure constraints governing segment distribution. Then, in Section 5, I outline a strategic innovation to the phoneme detection task designed to avoid the potential methodological pitfalls described (adapted from Otake et al., 1996). The aim of this innovation is to thread the methodological needle of listening target labels in L2+ perception studies by focusing the listener's attention not on the actual object of interest (*x*) but rather on an *adjacent* listening target (*y*): a Persean approach to steal a glimpse of the Gorgon. This adjacent target should be (a) familiar to both L1 and TL phoneme inventories and (b) not directly implicated in the phonological process or phonotactic constraint under investigation (i.e., neither the trigger of the phonological process nor the phone to which the phonological process or constraint applies).

I present the research questions in Section 6, and in Sections 7, 8, I report on two experiments that serve as test cases for the modified phoneme detection method, based on studies originally reported by Scott (2019a,b). The first tests the modified phoneme detection method in L1, investigating German RNA and DFA in L1 German speakers. This experiment is a replication and expansion of studies by Otake et al. (1996) and Weber (2001a,b, 2002). The nasal data may offer insight into theories of phonological feature (under-)specification and variation as they relate to place assimilation. The second experiment investigates German DFA with L1 German and L1 American English L2+ German learner groups, following Weber (2001a,b, 2002) and Lindsey (2013; unpublished thesis, Indiana University, Bloomington), whose studies investigated L1 Dutch and L1 American English groups, respectively. It additionally investigates the phonotactic/prosodic

1 ... or rarely <y>. Angle brackets < > indicate an orthographic representation.

ban on /h/ in syllable codas. Crucially, both experiments in this study avoid listening targets with unfamiliar or incongruent orthographic representations, unfamiliar phonetic transcriptions (e.g., Thomson, 2018), or other symbol types (e.g., Thomson, 2011) for listening targets that may be subphonemic (intra-category) variants or that lack graphemic or phonemic congruence between L1 and L2+ phoneme inventories.

## 2 The problems of labeling listening targets: facing a Gorgon

### 2.1 Phonological knowledge: more than phones

Phonemes are multifaceted knowledge structures that include the categorization and distribution of phones and, for most adults, orthographic labels. The *categorization* of phones sorts acoustically similar speech sounds into discrete categories according to articulatory features or acoustic cues with various manifestations along several continua (e.g., place of articulation and voice onset time). The *distribution* of phones describes where in a word a language permits a particular phone or allophonic variant to occur. This is phonotactics, a statistical type of well-formedness knowledge that may derive from categorical constraints or probabilistic knowledge based on the lexical (in)frequency of a particular form (Steinberg, 2014, p. 11–17). For reading populations, *orthographic* labels are conventionally used to denote a particular phone or phoneme. These three types of representational knowledge are necessarily connected by mappings that function to associate a specific label to a specific phonetic category in a specific context. Such context may be determined by phonotactic distribution, morphological structure, lexical-semantic content, and other factors. We know very little about how L2+ learners acquire these aspects of phonemes in relation to each other (see Ontogenesis Model: Bordag et al., 2021).

### 2.2 The underacknowledged problem of orthography

In designing experiments to reveal the properties of phonological representation, one must also take certain orthographic factors into account. Alphabetic literacy—that is, the knowledge of how labels are applied to individual speech sounds by orthographic convention—influences phonological awareness in undeniable but still poorly understood ways. Adult non-readers without alphabetic literacy exhibit a reduced capacity for phonological tasks that require manipulation at the segmental level (e.g., phoneme deletion or detection) relative to former non-readers who have later learned to read. This effect is less pronounced when syllables or rhymes are the unit of focus (Morais et al., 1986). For those with literacy of a script that encodes syllables rather than segments, the development of L1 phonemic awareness may proceed along different paths (Mann, 1986). Investigating the connection between alphabetical literacy and L1 phonemic awareness has a long tradition in reading and cognition research

(see Bertelson, 1986b, special issue articles in Bertelson, 1986a, and Castro-Caldas, 2004 for helpful reviews). More recently, research connecting this vein with L2+ phonology is rapidly emerging (e.g., Bassetti et al., 2015a, special edition of *Applied Psycholinguistics*, including a state-of-the-art review by Bassetti et al., 2015b). In addition, qualitative evidence suggests that groups from different L1 orthography backgrounds, despite similar quantitative performance results on the same phonemic awareness tasks, may employ different phonological processing procedures in L2+ scenarios (e.g., Korean vs. Chinese; Koda, 1998).

Just as phonology and lexical items are language-specific, so is orthography. Not only must contrastive phoneme distinctions of the language be represented (*chip* vs. *ship*), but each grapheme–phoneme correspondence (GPC) also has its own phonotactic and morphological distribution in the lexicon (e.g., <sh> and <ch> vs. <ci> and <ti> as labels for English /ʃ/ in *ship*, *fish*, *shanty/chantey*, *chute* vs. *commercial*, *navigation*). Invented L2+ spellings based on L1 GPCs illustrate the specificity of orthography nicely. Consider examples such as <JUELLULIB> and <GUARIYUSEI> (“Where do you live?” and “Waddayasay?”), both attempts by Mexican migrant workers to write down helpful phrases of English spoken in rural southern Illinois with Spanish spellings (Kalmar, 2015, p. 19, 51). On the Ontogenesis Model of lexical representations in L2+ (Bordag et al., 2021), such examples illustrate *fuzziness* in the phonolexical and lexico-semantic representations of the migrant workers’ L2+ English (Cook et al., 2016; Darcy et al., 2013). For these speakers, the phonological and orthographic domains, and the mappings (GPCs) between them, lag behind the semantic domain that allows them to use these phrases communicatively.

The problems of incongruent GPCs are similar for phoneme detection tasks, where the label of the listening target may represent different phonological information between groups (e.g., <N> with L1 Japanese vs. L1 Dutch; Otake et al., 1996; <CH> or <G> for L1 German vs. L1 Dutch or L1 English; Lindsey, 2013; Weber, 2001a,b, 2002; [u] for L1 French vs. [u] in [.Cu.] but not [.u.] for L1 Japanese; Dupoux et al., 1999, p. 1,570). In such cases, the intended label represents a phoneme or allophonic variant in one language but not the other (e.g., [x] and [ç] in German vs. American English).

### 2.3 Why labels are a problem for L2+ learners

As L2+ learners learn the sound system of a new language, they gradually acquire what the sounds are, where the sounds go, how they are written, how they are combined to label lexemes, and the relationships between these components. Just as we do not expect early, intermediate, and even advanced learners to have native-like production, vocabulary, and semantics, we should also not expect learners to have native-like, also called *optimal*, phonological perception and representations in the TL (*optimal encoding, optimum range*; Bordag et al., 2021). It is crucial for the design of laboratory phonology studies of L2+ acquisition to take into account that the components of phonological representations may not all be fully optimized in cross-language and subsequent L2+ perception. Indeed, they most likely are neither optimized nor closely yoked. Learners’ IL phonological categories, orthographic

knowledge, and phonolexical representations both are unevenly optimized and may remain divergent from and less precise than the representations of L1 speakers of the TL (Best and Tyler, 2007; Cook et al., 2016; Darcy et al., 2013). If we drop any assumption that L2+ learners' phonological, phonotactic, and orthographic knowledge is fully optimized, then labels for speech sounds become a problem, as we cannot assume congruent meaning for the label between languages or stages of IL development. The two experiments of this study serve to demonstrate the benefits of employing a phoneme detection task in cross-language and L2+ studies if one can avoid certain methodological concerns that may arise from stipulating listening targets by means of labels that differ in their phonological status between L1 and L2+ groups.

## 2.4 Why labels are a problem for intergroup comparisons

A necessary condition for experimental research in cross-language and L2+ phonology is that we investigate groups with different profiles of L1 and prior language experience. Yet this also poses a serious methodological challenge for ensuring task parity between groups. We routinely control for such factors as age of acquisition, proficiency, and literacy, among others, but we should also control for the comparability of task demands for the different groups. Many perception tasks require using labels for speech sounds, such as those motivating the Perceptual Assimilation Model (Best, 1995; Best and Tyler, 2007) or a phoneme detection task (aka phoneme monitoring; Foss, 1969). An example of the former would be something along the lines of "When you hear the sound [r] does it sound more like a type of /t/ or a type of /r/?" For an English speaker, a [r] might be thought of as a kind of /t/ or <t> (as in words like *city*), while for a Spanish speaker, a [r] might be thought of as a kind of /r/ or <r> (as in words like *pero*). An example of the latter would be "Press the red button if you hear a [θ] in the following sentence." The nature and cognitive load of such tasks, due to the representation of the listening target itself ([r] and [θ] in the preceding examples), may differ between language groups (Otake et al., 1996, p. 3,838–3,840). This depends on factors such as the label's status in the listeners' phonemic inventory, exposure to subphonemic variants, phonological or orthographic nativeness, or the phonological status of certain features of the stimuli for each group (e.g., cue weightings). A specific listening target might be an L1 phoneme to one participant group, an allophonic variant to another, and a novel non-phoneme to a third. For example, to an L1 Japanese speaker, [r] is the straightforward realization of the phoneme /r/, whereas an L1 English speaker may perceive it as a positional allophonic variant of the phoneme /t/, and an L1 Mandarin speaker may perceive it as a novel non-phoneme (or perceptually assimilate it to /l/ or /t/). Thus, listening for a target can constitute cognitively distinct tasks between groups. Similarly, listening for the target <u> likely would not be equivalent for L1 French listeners and L1 Japanese listeners. In French, the letter <u> typically represents the front rounded vowel phoneme /y/. The back rounded vowel [u] (phonemic /u/) is typically represented in French orthography by a digraph (e.g., <ou>). In contrast, L1 Japanese listeners may interpret <u> as either the syllable "u" represented in *rōmaji* (Roman script) as <u> and in *hirigana* by its own glyph

<う>, or as the nuclear constituent of another canonical syllable (e.g., <bu>/<ぶ>, <pu>/<ぷ>, <mu>/<む>; Dupoux et al., 1999, p. 1,570). For them, <u> represents a close back unrounded [u] or compressed vowel [u<sup>β</sup>]. Such examples are common in cross-language and L2+ phonology, with methodological implications for perception research on many language pairings. How should we determine if Japanese or French speakers hear an [u]? Should we ask the French speakers if they hear <ou> but ask the Japanese speakers if they hear <う>? Would this allow us to compare the results across groups? Would such a comparison still be confounded due to fundamentally different vowel qualities or because a French speaker cued to <ou> has one phonological unit to consider, whereas a Japanese speaker has both a full vowel (V) syllable and the nuclei of several consonant vowel (CV) syllables to listen for? Most likely, French and Japanese speakers face tasks with different cognitive processes and different cognitive loads in this case. In addition to creating congruency problems between languages, orthography also can be misleading due to inconsistency within a language when one label does not reliably indicate one sound (e.g., <CH> represents [tʃ] or [ʃ] in English *chant*, *chute*). This sort of confound can be compounded when using an L2+ label to focus attention on an L2+ category listening target that has no analog in the L1 (e.g., using <CH> for [ç] or [x] in German). For such reasons, it is important for ensuring study validity that perception tasks do not rely on listening targets with divergent phonological status between participant groups. To ensure parity between groups for phonological awareness tasks such as phoneme detection, listening targets and the GPCs used to stipulate them to participants should be selected for their congruence, as much as possible, given each group's language background. This study adopts this standard.

## 3 Phoneme detection for investigating place assimilation

### 3.1 The classic phoneme detection paradigm

The phoneme detection task was introduced to psycholinguistics by Foss (1969). As characterized by Weber (2001b, p. 12), it is a dual task, entailing the detection of a predetermined target sound in speech presented aurally and then a timed response. Participants indicate their detection (or not) of the listening target in the stimulus by pressing a response key as quickly as possible after the target sound is heard. Like other reaction time methods, accuracy and reaction time (RT) are the dependent variables of interest (Hui and Jia, 2024). To avoid response bias, the target items and distractors are counterbalanced by *fillers* that do not contain the listening target. If real words are used, then semantic priming (and other lexical factors) may affect RT (Frauenfelder and Seguí, 1989; Seguí and Frauenfelder, 1986). If non-words are used, it has been shown that for items that phonologically resemble a real word, RT is relatively faster than for those differing more from real words. Such similarity to extant words or sequences of phones may motivate facilitation even for non-words. This is a potential source of variation in RT for phoneme detection and similar tasks (Connine, 1994, p. 115–116; Connine and Titone, 1996, p. 639). Despite some observed effects of the lexicon on

tasks using non-word stimuli, numerous studies support some degree of abstraction from the input before accessing the lexicon, thus placing phonotactic processing (e.g., reinterpretation of raw percepts according to well-formedness conditions; Selkirk, 1984, p. 114) at a prelexical stage of processing, as early as 200 ms after stimulus onset (Dehaene-Lambertz et al., 2000; Steinberg, 2014; Steinberg et al., 2010a,b, 2011; Whalen, 1991). The influence of real-word phonolexical representations—even on non-words—and the early (prelexical) timing of phonotactic processing will be especially relevant for interpreting the results of experiment 2, in which L1 German speakers selectively compensate for the illicit occurrence of [h] in syllable codas but only in vocalic contexts where [x] would be licit.

One way to determine if a participant has acquired a particular process or constraint is to see if they react differently when a given string violates that TL pattern. Like other reaction time procedures such as phonetic decision, repetition, and lexical decision, phoneme detection typically indicates a phoneme's goodness of fit to its context through overall *slower* RT for a mismatching context (Whalen, 1984) and, specifically, violations of obligatory assimilation (Weber, 2001a, p. 96). Following established usage for this experimental paradigm, I refer to these slower RTs as processing *inhibition* (Marslen-Wilson and Warren, 1994; Martin and Bunnell, 1981, 1982; Otake et al., 1996; Streeter and Nigro, 1979; Whalen, 1984, 1991). However, violation of listener expectations under certain conditions may also yield a *faster* RT, which I refer to as processing *facilitation* (Cutler et al., 1987; Mills, 1980; Swinney and Prather, 1980). Because phoneme detection can yield either inhibitory or facilitative RT effects, it may be necessary to analyze different phonological conditions in separate statistical models rather than as levels of the same factor in a unified model, as experiments 1 and 2 demonstrate.

## 3.2 Place assimilation: nasals and dorsal fricatives

Two of the three phonological phenomena investigated here are types of place assimilation. Place assimilation often yields allophonic variants by context. For example, in English, underlying velar stops typically surface as velar before back vowels but palatal before front vowels (*cougar* ['ku:ɡɪ] vs. *keener* ['ci:nɪ]). This phonetic effect results from regressive place assimilation, where the place of the vowel influences the place of the preceding stop. See Winters (2003) for a historical review of developments in theoretical, typological, and experimental research on place assimilation and an account of its articulatory motivation and Hura et al. (1992) regarding perceptual motivations for assimilation.

Feature geometric approaches to autosegmental phonology have used hierarchical relations between distinctive features to describe rules of place assimilation.<sup>2</sup> Figure 1 displays example assimilation and default rules that specify place for specified and underspecified segments. Feature geometry approaches typically analyze place assimilation as the application of a single *spreading rule*. This might spread Place wholesale, like the example rule in Figure 1A, by

which RNA supplies Place specification to an underspecified nasal. Or it might spread a lower-tier feature, such as the example rule in Figure 1C, which adds [CORONAL] to a specified [DORSAL] fricative, a possible analysis of Standard German DFA. Such analyses may describe both regressive/leftward (Figure 1A; e.g., English [k]ougar ~ [c]eener) or progressive/rightward assimilation (Figure 1C). Figure 1B depicts a typical *default rule* to supply [CORONAL] to a Place-underspecified segment. Feature underspecification theories avoid delinking and favor *structure-filling* rules like those in Figures 1A, B, or *structure-changing* rules like the one in Figure 1C. Feature underspecification arguments are not uncommon in the German phonology literature (e.g., Glover, 2014; Hall, 1995, 2010). Experiment 1 probes whether specified mismatches behave differently from underspecified mismatches in L2+ learners.

## 3.3 Regressive Nasal Assimilation (RNA)

### 3.3.1 RNA in English and German

RNA is typologically widespread but manifests differently between languages (Speeter Beddor and Evans-Romaine, 1995). In English and German, examples of tautomorphic homorganic nasal-obstruent sequences, commonly argued to arise through place assimilation (e.g., Wiese, 1996), are plentiful (e.g., *ramp* [ræmp], *rant* [rænt], and *rank* [ræŋk] and German *Kampf* [kampf] “struggle, combat,” *Land* [lant] “country,” *Bank* [baŋk] “bank”). In English, RNA does not apply to morphologically derived nasal-obstruent sequences (e.g., *dreamt* [dɹæm-t], *ashamed* [ə.ʃeɪm-d]) or word-internally across morpheme boundaries (e.g., *confess* [kən.ˈfɛs], *infinite* [ˈɪn.fɪ.nɪt], *kingpin* [ˈkɪŋ.pɪn]). The application of RNA is further limited in German, which allows labial nasals before alveolar stops (e.g., *Amt* [amt] “office, agency,” *Hemd* [hɛmt] “shirt,” *Samstag* [ˈzams.tak] “Saturday”), as well as rarely before velar stops across syllable boundaries (e.g., *Lemgo* [ˈlɛm.go:] “city of Lemgo, North Rhine-Westphalia,” *Imker* [ˈɪm.kɐ] “beekeeper,” *Irmgard* [ˈɪr̩m.gɑrd] “proper name (fem.);” examples from Wiese, 1996, p. 218; Wiese, 2011, p. 105). These facts, and a phoneme detection experiment by Weber (2001a,b), suggest that some nasals of English and German are specified for Place and thus resist RNA. Assuming a theory of underspecification and a rule like Figure 1A, English and German nasals may only undergo RNA when not blocked by a prior place specification (e.g., labial /m/). Any nasals still lacking Place undergo repair by a default rule like in Figure 1B. For an additional discussion of the susceptibility of nasals to place assimilation, see Winters (2003).

### 3.3.2 Phoneme detection investigations of RNA

Otake et al.'s (1996) and Weber's (2001a) phoneme detection studies of RNA crucially inform the approach to listening target stipulation in the present task design. Following Otake et al.'s (1996) investigation of Japanese RNA, Weber (2001a) used a similar task to confirm that violations of RNA yield a similar RT effect in L1 German listeners. Together, these two studies provide the basis for the prediction that violation of RNA will yield slower RT (processing inhibition). Additionally, Otake et al. set

<sup>2</sup> See Hura et al. (1992) for relevant theoretical discussion.

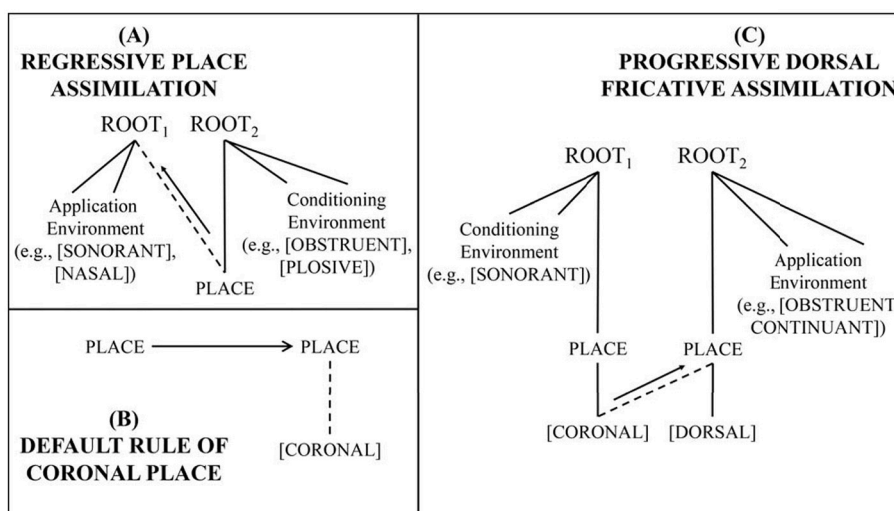


FIGURE 1 Example feature geometry rules: (A) Regressive nasal assimilation, (B) coronal default, and (C) progressive dorsal fricative assimilation.

the stage for using phoneme detection methodology to investigate underlying phonological representation using processing data from a behavioral task.

Otake et al. (1996) report six experiments (summarized in Table 1) investigating the phonemic representation of Japanese moraic nasals in real words based on responses to aurally presented phonetic realizations. In Japanese, moraic nasals—that is, consonant vowel nasal (CVN) syllable codas represented by the final nasal monograph <ㇺ>—undergo complete RNA to the following consonant (including sonorants) obligatorily (Vance, 1987), suggesting a lack of underlying Place specification (Hura et al., 1992, p. 68), which they gain only through RNA. This yields homorganic clusters in the medial position (e.g., *kanpa* [m.p], *tento* [ŋ.t], *konro* [n.r], *denki* [ŋ.k]). Their experiment 1 (L1 Japanese) and experiment 2 (L1 Dutch comparison) employed naturally produced stimuli. Based on previous research supporting the use of *rōmaji* for phoneme detection with L1 Japanese speakers (Cutler and Otake, 1994), these experiments used the letter <N> to label the listening target (i.e., the nasal undergoing assimilation) for both groups (Otake et al., 1996, p. 3,832). Their experiment 4 (L1 Japanese) investigated sensitivity to the moraic nasal under valid application of RNA (i.e., homorganic clusters) and RT inhibition in response to violations of RNA causing invalid (place mismatched) clusters, for example, \**to*[ŋ]bo, \**ko*[ŋ]to, \**ko*[m]to \**ro*[m]go, \**ro*[n]go (p. 3,836), still using <N> as the listening target. Their experiment 5 was a replication of experiment 4 (L1 Japanese, cross-spliced stimuli) with an important design modification: Rather than using the nasal as the listening target, it used the following consonants (<P B D T R K G>) as the listening targets (see Figure 1A, conditioning environment). They presented each listening target visually before each sequence (p. 3,838). Their experiment 6 employed the same procedure for an L1 Dutch comparison group. Overall, Otake et al. (1996) showed three key findings. First, L1 Japanese speakers can rapidly recover an abstract, unitary archiphonemic representation of the moraic nasal from its wide variety of phonetic realizations (see also Darcy et al., 2007, on the recoverability of the underlying phonological

representations from assimilated stimuli). Second, the phonetic realization of the moraic nasal creates a high expectation that allows L1 Japanese speakers to anticipate the following consonant that conditions RNA (cf. Key, 2014). Third, this language-specific knowledge is not shared by L1 Dutch speakers, who have a fundamentally different phonemic representation of nasals and for whom all stimuli were non-words. Their study also instructively highlights the methodological importance of the listening target for phoneme detection in two aspects—namely, (a) the nature of the label itself and (b) the choice of focus on either the application environment (object of interest *x*) or the conditioning environment of assimilation, as depicted in Figure 1A. I take up both again in Section 5.1. Otake et al.’s (1996) methodological change in experiment 5—to use the following obstruent as the listening target rather than the nasal target of place assimilation—crucially informs the innovation of the [t]-detection condition introduced here in experiment 1 and extended in experiment 2 (see Table 1).

### 3.4 Dorsal Fricative Assimilation (DFA)

#### 3.4.1 Phonological accounts of DFA in German

Like RNA and the English *cougar-keener* example, DFA manifests as coarticulation of adjacent obstruents and sonorants. In contrast to these, Standard German DFA is progressive rather than regressive. The German dorsal fricatives, commonly called the *ich-* and *ach-* sounds or “front and back *ch*,” are the voiceless palatal [ç] and velar [x], respectively (Dollenmayer et al., 2014, p. 192; Valaczkai, 1998, p. 112–114). Supplementary material A summarizes the phonetic characteristics of voiceless fricatives to contextualize these phones and their acoustic cues in the contrastive inventory of German. Spectral quality will be relevant for analyzing the L1 German results in experiment 2.

In Standard German and many regional dialects, palatal [ç] and velar [x] comprise a non-contrastive front-back allophonic pair

TABLE 1 Listening target alignments summary of phoneme detection task designs.

References	Location and identity of listening target					
	Nasal assimilation (regressive)			Dorsal fricative assimilation (progressive)		
	<u>V</u> NC	V <u>N</u> C Target = x	VNC	<u>X</u> FC	X <u>F</u> C Target = x	XFC Target = y
Otake et al. (1996)	–	Exp 1, 2, 4	Exp 5, 6	–	–	–
Weber (2001a)	–	–	Exp 4	–	Exp 1, 2, 3	–
Weber (2001b)	–	–	Exp 4	–	Exp 1, 2, 3, 5	–
Weber (2002)	–	–	–	–	Exp 1	–
Lindsey (2013)	–	–	–	–	Exp 1, 2	–
Present study	–	–	Exp 1	–	–	Exp 1, 2

Adapted from Scott (2019a, p. 286, Table 5.1). C represents an obstruent consonant. For nasal assimilation, N represents the nasal (application environment) and V represents the preceding vowel (conditioning environment). For dorsal fricative assimilation, F represents the fricative (application environment) and X represents the preceding consonant or vowel (conditioning environment). Application environments are in italics. Conditioning environments are in bold. Listening targets, whether the object of interest *x* or the subsequent phone *y*, are underlined. Weber (2001a) reports Weber (2001b) experiments 1–4; Weber (2002) reports Weber (2001b) experiment 5.

analogous to English [k]–[c] in *cougar–keener*. They are typically described as standing in complementary distribution (e.g., Hall, 1989, 2022, p. 680; Wiese, 1996). The word-internal environment preceding any dorsal fricative is argued to determine whether it surfaces as [x] or [ç], conditioned by a preceding back or front vowel (e.g., *Buch* [bu:x], “book, SG.” vs. *Bücher* [ˈby:.çɐ] “book, PL.”, *kochen* [ˈkɔ:.xn] “cook, INF” vs. *weich* [vaɪç] “soft, weak”) or coronal consonant (e.g., *Milch* [mɪlç] “milk,” *Dolch* [dɔlç] “dagger,” *Mönch* [mœnç] “monk”), with a few morphologically derived exceptions (e.g., *Kuh* [ku:] “cow” vs. *Kuhchen* [ˈku:.çən] “cow, DIM”). In dialects that lack DFA or palatal [ç] (e.g., in Switzerland), [x] may surface even after front sonorants (e.g., *echt* [ɛçt] ~ [ɛxt] “real(ly), actual(ly)”), and [k] or [ʃ] may substitute for [ç] in loan words (e.g., *China* [ˈki:.na]/[ˈʃi:.na] “China,” *Chemie* [ke.ˈmi:]/[ʃe.ˈmi:] “chemistry,” Hall, 2022, p. 767–772; cf. invariant <ch> in *Chemnitz* [ˈkɛm.nɪts] “city of Chemnitz, Saxony”; see Hall, 2014, 2022, for more on dialectal variation). As DFA applies only to the dorsal subset of German fricatives (not /f v s z ʃ ʒ h/), feature spreading must be at a lower tier. See, for example, Figure 1C, which adds [CORONAL] to the [DORSAL]-specified Place node (cf. Hall, 1997; Iverson and Salmons, 1992). There has been a lack of clear consensus in the German phonological literature about precisely which feature triggers DFA for nearly a century (Hall, 2022, p. 1–6); oft-cited approaches employ [CORONAL] (e.g., Robinson, 2001), [+back] (e.g., Hall, 1989, 1992), and [front] (e.g., Wiese, 1996). Most recently, Hall (2022) presents a comprehensive review of the problem as reflected in the literature and weaves together historical, dialectal, and synchronic data for a coherent analysis of DFA (so-called because it is not assimilatory in all varieties). Hall characterizes DFA as a specific case of a more general velar fronting process (recall English *cougar–keener*), versions of which arise in many German dialects as well as national standard varieties.<sup>3</sup> For thorough reviews of phonological accounts of DFA, see Steinberg (2014, p. 27–35) and Hall (2022).

<sup>3</sup> Hall further argues that [ç] and [x] have different phonological statuses—phonemic, quasi-phonemic, or allophonic variants—in different regional dialects of German.

### 3.4.2 Phoneme detection investigations of German DFA

Weber (2001a,b, 2002) conducted a series of phoneme detection experiments to investigate the distribution of German dorsal fricatives [x] and [ç] and the psychological reality of DFA as an obligatory assimilation for L1 German speakers, compared to L1 Dutch speakers (see Table 1). Taken together, in the subjective experience of Weber’s participant groups, listening targets varied between experiments in terms of whether they were orthographic or phonemic and whether they indicated phonemic, subphonemic, or non-native sounds regarding the L1 phonology. Weber’s DFA experiments focused listener attention directly on a specific allophonic variant of the dorsal fricative (*x*, the object of interest) rather than a subsequent phone (*y*) as the Persean approach does.

I also note that Weber (2001a) collected German data in Regensburg, Bavaria, a mainly velar-fronting dialect area with some enclaves that lack DFA and the fronted [ç] (Hall, 2022, p. 104, 427), while Weber (2002) collected data in Hannover, which lies squarely in a velar-fronting dialect area (Hall, 2022, p. 137–143). Given the potential for regional dialect variation in German DFA, dialect background, including regional exposure profiles of L1 German listeners, should be regarded as an important factor in future studies. For the present study, a time-limited opportunity for *in situ* L1 group data collection arose in Stuttgart, Baden-Württemberg. For this reason, the dialect background for the present study’s L1 German group more closely resembles Lipski’s (2006) Stuttgart sample than Weber’s.

Based on the RT data elicited from the non-word DFA experiments, Weber (2001a) detected a small facilitation effect for L1 German listeners when presented with a front–back mismatch sequence (e.g., \*[ɛx]) while listening for [x] in mono- and disyllables, whether the non-words were described as German or Dutch. In contrast, Weber (2002) found no RT effect—neither facilitation nor inhibition—when L1 Germans listened for [ç] in disyllables, regardless of whether it occurred in a licit front–front sequence (e.g., [i:ç]) or illicit back–front sequence (e.g., \*[a:ç]). Weber (2001a,b, 2002) argues that this result, which departs from the inhibition reported for RNA by Otake et al. (1996), arises due to the

interaction between the direction of the assimilation (progressive) and a reaction to novelty. In short, regressive assimilation such as RNA creates a strong expectation for which few phones may follow the first. For instance, once velar [ŋ] is perceived, internalized knowledge of obligatory RNA narrows down the possibilities significantly. In German, either [g] or [k] must follow; violating this strong expectation hinders the recovery of the underlying phonolexical representations (Darcy et al., 2007; Key, 2014) and causes processing inhibition (e.g., Otake et al., 1996; cf. visual attention experiments; Posner et al., 1978). In contrast, the earlier phone in DFA creates a much less restrictive expectation for the following phone. For example, a front vowel [i:] gives rise to the weak expectation that the next phone will not violate DFA (i.e., merely not \*[x]); whether it conforms to DFA ([ç]) or is irrelevant to DFA (e.g., [p b m t d n g k ŋ]). Only an illegal sequence (e.g., \*[i:x]) violates this weak expectation. If the sequence is nonetheless attested in the language (e.g., [u:ç] in *Kuhchen* /ku:-çən/ → [ˈku:.çən] “cow, DIM”), then no RT effect arises. But if the sequence is both in violation of DFA and novel (e.g., \*[bɪxt], \*[blɪn.xən]), then *novel popout* may yield a small facilitation effect, as Weber (2001a) argues (Christie and Klein, 1996; Johnston and Schwarting, 1996, 1997; Weber, 2001b, p. 40–41, 53).

Lindsey’s (2013) replication and expansion of Weber’s (2001a, 2002) study included both L1 (Southern) German speakers and advanced L1 American English L2+ German learners. For this group, German dorsal fricatives were novel TL phones (Plag et al., 2009, p. 5–6). Contra Weber, Lindsey found processing inhibition in both groups for all DFA violations, not only front–back sequences. Additionally, Lindsey’s study included both front and back listening target conditions (i.e., <ch> = [ç] in *Eiche* [aiçə] “oak” and *euch* [əʏç] “you, 2P.PL.” vs. <ch> = [x] in *Bach* [bax] “brook” and *Bauch* [baʊx] “stomach”) for all participants. These were presented as two separate blocks, completed in alternating orders, whereas Weber’s participants were tested on just one listening target, depending on the location of data collection (Regensburg vs. Hannover). Weber’s and Lindsey’s studies offer independent evidence for the psychological reality of DFA to L1 German speakers, albeit with differing response patterns that suggest variable representation of DFA among different groups of German speakers (e.g., regional dialect differences).

This review of Weber’s and Lindsey’s studies highlights four important insights. Weber posits a potential explanation for processing facilitation effects with violations of progressive assimilation to highlight nuanced differences in strong or weak expectations that may arise due to the direction of assimilation.<sup>4</sup> Lindsey’s study demonstrates that advanced L2+ learners (may) acquire sensitivity to assimilation violations in a TL. Like Otake et al.’s (1996) experiments 1, 2, and 4 with nasals, Weber’s and Lindsey’s experiments use the application environment—namely, the subphonemic variants of the dorsal fricative—as the listening target (Table 1). Finally, it is important to recall the implications of using the <ch> label for listening targets in these studies. For L1 German speakers, <ch> represents an L1 phoneme with multiple

context-dependent surface forms, [x] and [ç]. Thus, Weber’s and Lindsey’s tasks required L1 German listeners to focus their attention on subphonemic variants with optimal (or highly overlearned) phonological and orthographic representations (Bordag et al., 2021; Strange, 2011). In contrast, Lindsey’s L2 learners focused their attention on German <ch>, a *foreign* phone without L1 analog, which has variable surface forms in the TL and context conditions that must be acquired. Depending on learners’ exposure to the target language and current IL representations, <ch> may not have optimal phonological or orthographic representations, and L1 representations (e.g., *cheese* [tʃi:z]) may be activated and interfere. Thus, the label <ch> likely differs in meaning between groups in the phonological and orthographic domains, as well as the mappings between them. In this light, the same behavioral task could be a *phoneme* or *allophone* detection task for an L1 group, but a *phone* or “fuzzy” *phoneme* detection task for cross-language listeners or L2+ learners. To ensure task parity between groups, these crucial factors have been addressed in the experiments of this study by stipulating a Persean listening target—that is, an adjacent obstruent (y) that is uncontroversially a phoneme of both L1 and TL for all participants. The results of this task will tell us something about the representation of the neighboring fricatives.

## 4 Beyond place assimilation: syllable phonotactic constraint of /h/

In addition to adjacency effects (i.e., place assimilation and associated transition cues), I aimed to test phoneme detection on a fundamentally different type of phonological knowledge: a prosodic/phonotactic constraint governing a segment’s distribution in syllabic structure. The voiceless glottal fricative /h/ is uncontroversially phonemic in both English and German. A phonotactic well-formedness condition (Selkirk, 1984, p. 114) bans /h/ from syllable codas in both languages (e.g., *ahead* [ə.ˈhɛd], *heat* [hi:t] and German *Ahorn* [ˈa.hɔɐ̯n] “maple,” *Hut* [hu:t] “hat” vs. \*[ti:h]). This is henceforth referred to as the \*Coda-/h/ constraint (Scott, 2019a, p. 338–339). This restricted distribution is language-specific rather than universal (e.g., Turkish *tahta* [ˈtah.ta] “(wooden) board,” Persian *šāh* [ʃɔ:h] “king”). See Davis and Cho (2003), Jessen (1998, p. 152–153), and Scott (2019a, p. 83–100) for thorough discussion of /h/ and [h], including in English and German. The \*Coda-/h/ constraint informs the design of experiment 2, which undertakes the first ever phoneme detection investigation of a non-linear prosodic structure constraint on segment distribution.

## 5 Detargeting the target: design motivations for adjacent listening targets

### 5.1 Where to direct attention in phoneme detection tasks: giving Perseus a shield

The selection of labels to focus listener attention for phoneme detection experiments must navigate three non-exclusive non-equivalence types in L1–L2+ scenarios. Each of these non-equivalence scenarios bears on GPC activation and impacts

<sup>4</sup> Expectations about upcoming phones in the speech stream may be considered a phonological/phonotactic prediction during speech perception with implications for processing and reaction times (McMurray and Jongman, 2011; see Kaan and Grüter, 2021, for examples in other linguistic domains).



potential interpretations of different groups' responses in phoneme detection tasks. (a) The underlying phonological categories may not be equivalent. The chosen label may point to non-equivalent categories between languages. For example, the letter <N> does not represent the same phonological information to L1 Japanese speakers as it does to L1 Dutch speakers; similarly, <CH> predominantly indicates a different phonological category in English than it does in German (Lindsey, 2013; Otake et al., 1996). (b) A label may represent a phonological category of one language that does not exist in the other. For example, the letter <G> represents /g/ in German, which is a foreign phoneme to L1 Dutch speakers (Weber, 2001a,b), and German <CH> points to a phoneme without a correspondent in American English (Lindsey, 2013; Plag et al., 2009, p. 5–6). (c) A third type of phoneme detection task non-equivalence arises from which phone in a sequence is designated the listening target. In the case of assimilation, three types of listening targets are possible. Otake et al.'s (1996) experiments 1, 2, and 4 with nasals; Weber's (2001a,b, 2002) dorsal fricative experiments; and Lindsey's (2013) replication experiment all direct listener attention to the *application* environment of RNA or DFA, respectively (Figure 1)—that is, the object of interest ( $x$ ). In contrast, Otake et al.'s experiments 5 and 6 and Weber's nasal experiment target the *conditioning* environment of RNA. This latter focus has the benefit of equivalence of the [p] and [k] listening targets as phonemic /p/ and /k/ between languages, at least in Weber's case of L1 German and L1 Dutch groups. For them, these phones are represented discretely on the segmental level in L1 orthography, whereas L1 Japanese learn these segmental representations through their secondary *rōmaji* script. In my study, a third approach (the Persean approach) is introduced, in which the listener is directed to focus on an adjacent phone that is neither an application environment nor a conditioning environment. Table 1 places the present study in methodological context by summarizing these listening target alignments.

## 5.2 What does Persean phoneme detection gain us?

The aim of the Persean approach to target selection for the phoneme detection task is to employ the phoneme detection paradigm while preventing the task and procedure from constituting different cognitive tasks to different groups due to how the listening target is stipulated. By targeting a subsequent adjacent phone, the Persean approach affords L2+ researchers four key advantages for intergroup task parity. First, it allows researchers to study objects of interest  $x$  that do not make good listening targets themselves because the languages do not spell the sound congruently, such as the difference between what orthographic <u> represents in French vs. Japanese *rōmaji*, or consistently, such as the numerous sounds that orthographic <ch> may represent in English. Instead, researchers can choose to focus listener attention on a target with a transparent one-to-one GPC common to multiple orthographies. This broadens the pool of potential stimulus designs substantially. Second, researchers can choose listening targets that are phonemically equivalent (or much closer) between language background groups. This is important for studies involving mixed L1 or previous L2 backgrounds (e.g., Canadian adults with extensive

French and English exposure) and crucial for comparisons between L1–L2+ pairings in phonological acquisition research. A third advantage may benefit studies that investigate multiple levels of TL proficiency using cross-sectional or longitudinal designs. Group (or time-point) differences may include different degrees of orthographic or phonological optimization of fuzzy TL categories (Ontogenesis Model; Bordag et al., 2021), which we may associate with learners' gains in overlearning and automaticity of selective perception routines while processing TL input (Automatic Selective Perception Model; Strange, 2011). Fourth, the Persean phoneme detection variant draws listeners' explicit attention away from the object of interest  $x$  (i.e., detargets the target of research interest), allowing investigation of implicit phonological knowledge. For investigations of implicit phonological knowledge, it is cleanest to direct listener attention away from both the application environment and conditioning environment, both of which are crucial to the phonological phenomenon of interest. The [t]-detection conditions used in this study benefit from these advantages.

## 6 Research questions

The central motivation for this study is to expand on Otake et al.'s (1996) crucial insight that phoneme detection aimed at an adjacent segment for the listening target equips us to investigate questions of phonological representation in a language-neutral manner. First, I use Otake's method to investigate RNA in a new language (L1 German) to both calibrate the method with relatively clear predictions for an assimilation process and explore the potential to investigate the nuances of phonological representation using phoneme detection (experiment 1). Second, I investigate the potential of this Persean variant of the phoneme detection method to explore broader questions in L2+ phonological acquisition. I apply this design variant for the first time to investigate German DFA in both L1 and L2+ German groups, using a novel [t]-detection condition. Finally, I use the Persean [t]-detection task in the first phoneme detection investigation of a non-assimilatory phonological constraint (\*Coda-[h]) with L1 and L2+ German groups in parallel (Experiment 2).

### 6.1 Research questions for RNA

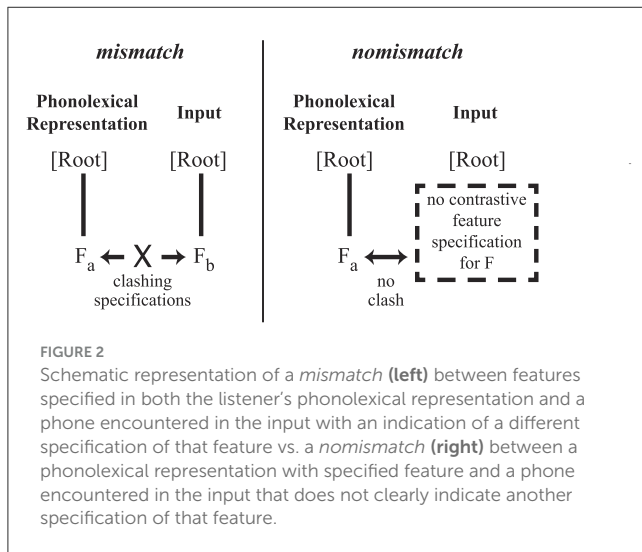
Experiment 1 investigates the first two research questions.

RQ1: Does violation of RNA in an obligatory context in German yield RT patterns of consistent inhibition, variable inhibition, or no inhibition?

It is hypothesized that RNA violation will inhibit RT (i.e., slower reactions) in L1 German speakers. Experiment 1 undertakes replication of the findings for RNA with L1 German speakers by Weber (2001a,b) with [k]- and [p]-detection. A secondary motivation of experiment 1 is theoretical. The [p]- and [k]-detection blocks of experiment 1 investigate the impact of feature specification on mismatched Place violations of RNA on processing.

RQ2: What manner of feature specifications do RT data from violation of German RNA support?

An unlikely null effect would suggest tolerance of RNA violation, undermining its obligatoriness. A single distinction between match vs. mismatch would support the obligatoriness of RNA but not



assumptions of a psychologically real distinction between the underapplication of RNA in cases of underspecification (e.g., ?[np nk]) and true place-clash sequences, such as \*[mk] or \*[ɲp] that involve fully specified Place features.

The third possibility of a ternary distinction between place-match [mp ɲk], underspecified mismatch ?[np nk], and place-specified mismatch \*[mk] or \*[ɲp], even with non-word stimuli, might lend support to underspecification theories. One example is the ternary logic of the Featureally Underspecified Lexicon (FUL; Lahiri and Reetz, 2010, p. 50–51). This model posits a processing distinction between input that clashes with a phonological representation, for example, an explicit *mismatch* of specified Place features that cannot be interpreted as the output of an overlearned rule for which an underlying form is recoverable (Darcy et al., 2007; Strange, 2011) vs. input that merely fails to match a phonological representation, for example, *nomismatch* of Place features, due to underapplication of a rule to specify an underlyingly underspecified Place. These scenarios are schematized in Figure 2.

## 6.2 Research questions for DFA

RQ3: Does the [t]-detection task show processing facilitation or processing inhibition in response to violation of DFA in an obligatory context in German?

Previous phoneme detection studies report contradictory RT results for violation of DFA (Lindsey, 2013; Weber, 2001a,b, 2002), which limits the basis for predictions. Weber found processing facilitation (i.e., faster RT) for violation of DFA with L1 German listeners but only under certain conditions or with certain regional populations, whereas Lindsey found processing inhibition (i.e., slower RT) in all conditions for both L1 German and L2+ learner groups. This study aims at replication with similar groups, but without directing the attention of either group to attend to <CH> directly, to ensure task parity between language background groups and add to our empirical knowledge of phonological processing of DFA violations. In the novel [t]-detection condition of experiment

1 and in the similar DFA condition of experiment 2, the listening target is (a) familiar to both language groups as an L1 phoneme with similar acoustic realizations, and (b) irrelevant to the progressive assimilation that precedes it. The novel [t]-detection approach in experiments 1 and 2 pursues independent replication of findings with DFA in L1 German speakers or L2+ German learners, but with listener focus on a different target.

RQ4a: Do L1 American English L2+ German learners exhibit sensitivity to violations of German DFA?

RQ4b: Do L1 German speakers exhibit sensitivity to violations of German DFA?

RQ4c: Is the adjacent [t]-detection task able to detect sensitivity to violations of progressive assimilation that precede the listening target?

Experiment 2 undertakes replication of Lindsey (2013) and extends the adjacent target technique in two ways. First, it includes an L1 English L2+ German group as a further replication of Lindsey (2013), to investigate the fourth group of research questions. A demonstrated sensitivity to DFA violation in either experiment could independently confirm either Weber's or Lindsey's findings, at least with the population sampled here. It would also demonstrate that the instrument is sufficiently sensitive to detect RT effects for violation of progressive assimilation. Failure to find any significant RT effects for DFA violation could indicate that the instrument is susceptible to Type II error for this type of assimilation. Experiment 2 also adds a novel non-assimilation condition with illicit /h/ in syllable codas.

## 6.3 Research questions for ban of [h] in syllable codas

The final research questions expand on previous research by undertaking the first phoneme detection investigation of the prosodic constraint banning /h/ in codas. This ban is exceptionless in both languages, despite cross-language perceptual assimilation patterns for dorsal fricatives (see Section 4; Scott and Darcy, 2023), so violations should yield strong RT inhibition (cf. RNA).

RQ5a: Do L1 American English L2+ German learners exhibit sensitivity to violations of \*Coda-/h/ ?

RQ5b: Do L1 German speakers exhibit sensitivity to violations of \*Coda-/h/ ?

RQ5c: Is the adjacent [t]-detection task suitable for investigating types of phonotactic knowledge other than assimilation processes?

Experiment 2 also includes a novel investigation of the phonotactic/prosodic constraint \*Coda-/h/ with both language background groups, to explore the task's utility with other types of phonological constraints. Demonstrated sensitivity to violation of \*Coda-/h/ by either population would demonstrate that the instrument is sufficiently sensitive to detect RT effects for violation of prosodic/phonotactic well-formedness constraints that create strong expectations for upcoming phones. Failure to find any significant RT effects for \*Coda-/h/ violation could indicate that the instrument is susceptible to Type II error for this type of phonotactic/prosodic constraint.

## 7 Experiment 1: detection of conditioning environment or Persean target in L1

Experiment 1 focuses on nasals and fricatives in L1 German. The phoneme detection task in this experiment includes two types of listening targets, represented by the following obstruent: (a) focus on the conditioning environment of RNA ([p] or [k]) and (b) focus on an adjacent phone unrelated to DFA ([t]). Two experiment blocks, [p]- and [k]-detection tasks, focus on obligatory RNA in German, with the aim to replicate previous findings of RT inhibition when RNA is violated in monosyllable codas and listener focus is directed to the obstruent that conditions assimilation of the preceding nasal. Another block employs Persean [t]-detection to investigate German DFA. It diverges from previous phoneme detection studies of DFA by directing listener attention to a following obstruent [t], which plays no role in the rule. It also has the advantage of avoiding the use of orthographic <CH> as the listening target for multiple allophonic variants (cf. Lindsey, 2013; Weber, 2001a,b, 2002). The aim is to test whether violation of DFA in an obligatory monosyllabic context yields RT inhibition (cf. Lindsey) or facilitation (cf. Weber) for L1 German speakers. On the assumption that RT effects for place mismatches carry over to influence listening targets that immediately follow the phones involved in assimilation, experiment 1 should yield shifts in RT in response to violation of RNA (RQ1 and RQ2) and DFA (RQ3), if these are psychologically real to L1 German speakers.<sup>5</sup>

### 7.1 Method

#### 7.1.1 Participants

Seventeen L1 German speakers (13 female; ages 18–35;  $M = 25.2$ ,  $SD = 4.764$ ) received €5 for completing experiment 1. Eleven completed the task in Stuttgart, Baden-Württemberg, Germany, seven of whom had previously completed experiment 2 because experiment 1 was offered as an optional additional experiment in a session. Six Germans attending universities in the Midwestern United States were recruited for supplementary data collection. Additional participant details are provided in [Supplementary material B](#).

#### 7.1.2 Stimuli

A tabular summary of experiment 1 trial types by condition is provided in [Supplementary Table C1](#). Non-word stimuli ( $N = 304$ ) were prepared for three assimilation types (Nasal-[p]-Detection vs. Nasal-[k]-Detection vs. Fricative-[t]-Detection). The nasal condition was balanced for three conditions of match type ( $n = 18$ ; 3 each for Match [mp], [ɪk] vs. Underspecified Mismatch ?[np], ?[nk] vs. Specified Mismatch \*[mk], \*[ɪp]). The Fricative condition was balanced for four conditions of Match type ( $n = 20$ ; 5 each for Front Match [eç] vs. Back Match [ax] vs. Back-Front Mismatch \*[aç] vs. Front-Back Mismatch \*[ex]). All target stimuli

included the listening target as the final obstruent of a monosyllable coda (e.g., [p] in [zɪmp], [k] in [zɔŋk], [t] in [glɛçt], [glaxt]). The 38 critical trials were balanced by 114 distractors with the listening target in non-final positions (27 with [p], 27 with [k], 60 with [t]) and 152 fillers without the listening target (36 in [p]- and [k]-detection blocks, 80 in [t]-detection) so that [p]- and [k]-detection blocks totaled 72 trials and [t]-detection totaled 160, yielding a 1:3 ratio of critical trials to distractors and 1:1 ratio of trials with the listening target (critical trials + distractors) to fillers without it (Keating and Jegerski, 2015; p. 16). [Supplementary material F](#) provides a complete list of stimuli.

At least three tokens of each item were digitally recorded in a sound-attenuated booth (sampling rate 44,100 Hz) by a phonetically trained L1 German female talker from Saxony who spoke and taught Standard German professionally in the United States. The researcher selected one token of each item for recording quality. Six training non-words were also recorded: *Tiesel*, *gamisch*, *frettig*, *Skirm*, *Prasen*, and *Schloft*. Training trials for the [t]-detection block and all distractors and fillers were drawn from experiment 2 stimuli. Files were manually cut and normalized for volume by a Praat (Boersma and Weenink, 2014; Version 5.4) script; the task was presented with OpenSesame (Mathôt et al., 2012; Version 2.9).

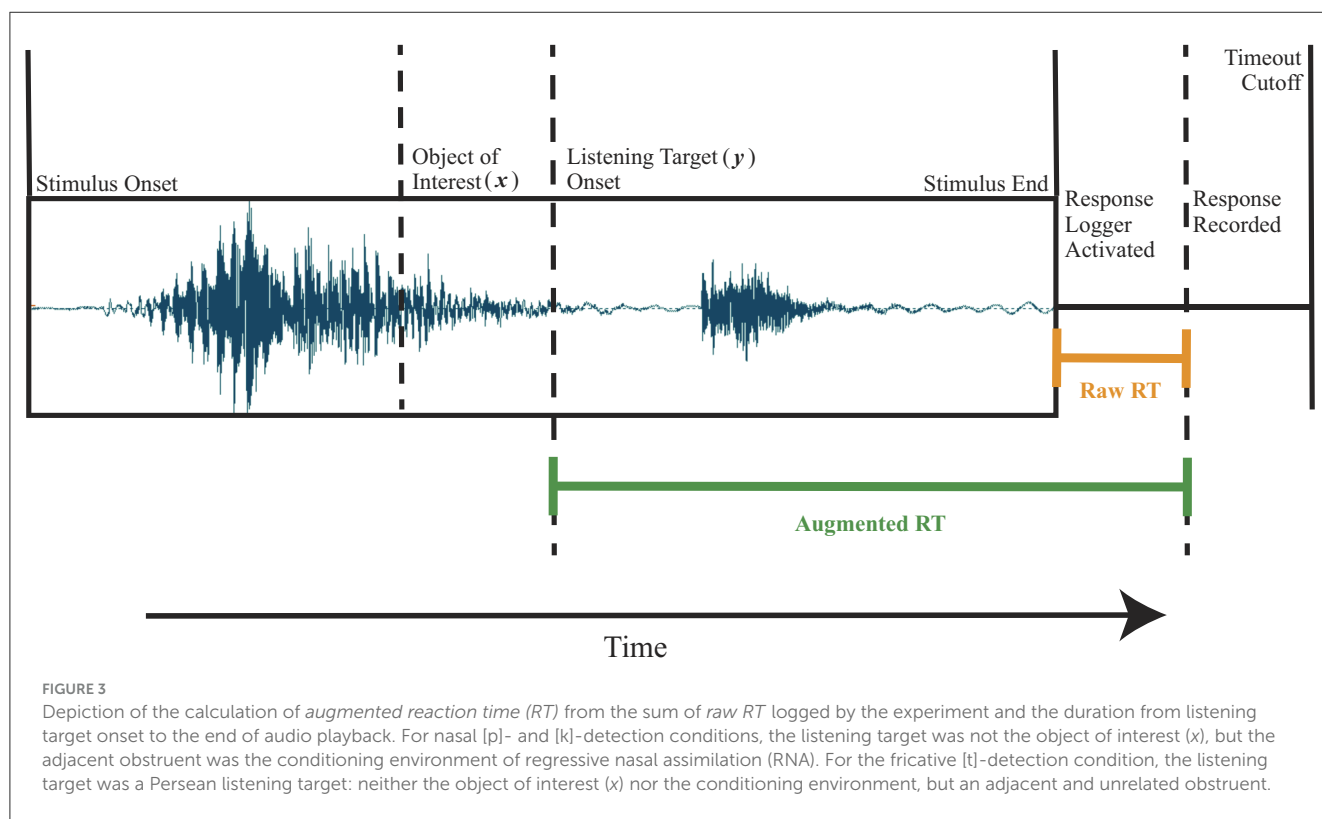
Listening targets in the generalized phoneme monitoring procedure (Frauenfelder and Seguí, 1989) may occur in different parts of the stimuli (distractors vs. critical trials); furthermore, each individual token of a listening target varies in duration. To compensate for varying duration, it was more accurate to combine RT measurements collected by the software (see Section 7.2.2) with the duration from listening target onset to the end of the audio file for each trial to derive an *augmented RT* measurement that reflects participants' processing time. The calculation of augmented RT for each trial is depicted in [Figure 3](#). Segment boundaries were marked in Praat (Boersma and Weenink, 2014; Version 6.0.19), and the onset and duration of each listening target and the phone preceding it were extracted. [Supplementary Tables C2, C3](#) describe these durations by condition in aggregate; see [Supplementary material F](#) for the extracted durations of each stimulus and Scott (2019a, p. 299–302) for additional analyses of the stimuli. Because the listening target for critical condition trials always appears at the end of the stimulus, the sum of the listening target duration and the automatically recorded RT yielded the *augmented RT*, which serves as the dependent variable for analysis.

#### 7.1.3 Procedure

Stuttgart data were collected in a quiet computer lab of six identically configured computers running Windows 7 (Professional 9 Service Pack 1, 64-bit) with a 3.2 GHz processor, 4 GB RAM, and 1680 × 1050-pixel screen resolution. Mobile data collection in the United States used a single Dell XPS 12 two-in-one laptop (laptop and tablet modes) running Windows 8 or Windows 10 Pro (64-bit) with an Intel i7-4510U 2.6 GHz processor, 8 GB RAM, and 1080 × 1920-pixel screen resolution. Stimuli were presented through high-quality circumaural headphones.

Participants completed a language background questionnaire (see the Open Science Framework resources). The researcher briefly explained the phoneme detection task in German, and participants

<sup>5</sup> This section presents a reanalysis of a data set originally collected in 2015 and reported by Scott (2019a, Chapter 5).



**TABLE 2** Signal detection rates for experiment 1 after participant exclusion (N = 12).

Condition	Signal detection			
	Hit	False alarm	Correct rejection	Miss
<b>&lt;P&gt;/&lt;K&gt;</b>				
Match	0.972	-	-	0.028
Mismatch (underspecified) <sup>a</sup>	0.958	-	-	0.042
Mismatch (specified)	0.958	-	-	0.042
<b>&lt;T&gt;</b>				
Match	0.942	-	-	0.058
Mismatch	0.892	-	-	0.117
Distractors <sup>b</sup>	0.908	-	-	0.092
Fillers	-	0.016	0.984	-

Dashes indicate fields for which no rate is possible with the “go”/“no-go” format.

<sup>a</sup>Nasal Specified and Underspecified Mismatch conditions exhibited similar accuracy in aggregate, with individual variation in different subconditions.

<sup>b</sup>Distractor Miss rate may include responses entered before the experiment response logger initiated.

read instructions on the screen that explained that they would hear invented words in three blocks. In each block, they were to listen for “T,” “K,” or “P” somewhere in the word and indicate when they heard the listening target by pressing the space bar as quickly as possible. If the target was absent, participants waited for the next trial without responding. Text examples of the listening

targets present in various positions (or absent) were displayed with explanations of appropriate responses, and then participants completed a six-trial training phase in blocks of two for [t]- (*Tiesel, gamisch*), [k]- (*frettig, Skirm*), and [p]-detection (*Prasen, Schloft*). The practice trials alternated with training instructions explaining the block-specific listening targets and the potential for the target to occur anywhere in the word. The order of the three blocks was random, and the trials were randomized within blocks. Every 16 trials, participants had the option to pause and resume when ready. Each trial began with a fixation point on the screen, followed by audio playback. The experiment recorded responses and RTs. The OpenSesame response logger was located immediately after playback so that a recorded RT of 0 ms corresponded to the end of the stimulus just after the release of the syllable-final listening target in critical trials (see [Figure 3](#)). Experiment 1 lasted ~25 min.

## 7.2 Results

### 7.2.1 Exclusion criterion

This study employed a “go”/“no-go” phoneme detection response format (i.e., was the phoneme detected? Affirmative responses only). In terms of signal detection theory as used in perception research, this format only records hits (e.g., accurate detection of the target) and false alarms (e.g., spurious indication of target presence when it is absent). Correct rejections (e.g., correct indication of target absence) and misses (i.e., failure to indicate target presence when it is present) are not recorded. In this format, some correct rejections could additionally result from non-response bias, while misses could result from lack of sensitivity, non-response

TABLE 3 Descriptive statistics and tests of normality for experiment 1 by condition.

Condition	<i>n</i>	$\mu$ (ms)	$\sigma$	Skewness		Kurtosis		Shapiro–Wilk	
					<i>SE</i>		<i>SE</i>	<i>W</i>	<i>p</i>
Nasal	207	628	122.4	0.315	0.169	0.250	0.337	0.989	0.132
Match	70	564	115.3	0.272	0.287	0.121	0.566	0.987	0.681
Underspecified	69	632	96.1	0.247	0.289	−0.177	0.570	0.989	0.807
Specified	68	689	122.1	0.529	0.291	−0.181	0.574	0.968	0.079
Fricative	220	559	105.6	0.648	0.164	0.187	0.327	0.964	<0.001
Match	113	565	99.0	0.464	0.227	−0.022	0.451	0.980	0.095
Mismatch	107	554	112.3	0.820	0.234	0.393	0.463	0.940	<0.001

RT values rounded to the nearest millisecond. Observed positive skewness is not an artifact of data trimming, which excluded just one non-filler fast response (< 100 ms), a \* [ɲp] Nasal Mismatch.

bias, or both.<sup>6</sup> A minimum threshold of five hit responses was set for each Match Type condition according to Assimilation Type to ensure that no single response could influence the mean of any condition too much. This excluded five participants with fewer than five hit responses in any of the five conditions, Nasal Match, Nasal Underspecified Mismatch, Nasal Specified Mismatch, Fricative Match, or Fricative Mismatch, while retaining 428 responses from the remaining 12 participants (i.e., 183 Match and 245 Mismatch trials; 208 Nasal and 220 Fricative trials). Table 2 displays the remaining 12 participants' signal detection rates.

## 7.2.2 Data trimming and preparation

Data trimming removed the upper and lower ends as follows. The response logger timed out at 700 ms, so trials with raw RT of 700 ms were excluded. Extremely short RTs might indicate responses prior to presentation of the listening target in playback, so one non-filler \* [ɲp] trial with raw RT below 100 ms was excluded. The remaining non-filler trials were included for analysis of their augmented RT. Fillers were excluded.

Scott (2019a) examined these data regarding assumptions of normality, ultimately abstaining from log-transformation. It has been common practice to log-transform data to satisfy assumptions of normality that raw behavioral RT data often violate. However, log-transformation of data has recently come under much criticism (e.g., O'Hara and Kotze, 2010). For example, as data may not approximate a log-normal distribution, there is no guarantee of reduced skewness and, indeed, some risk of increased skewness; furthermore, log-transformation can often increase variability (Feng et al., 2014, p. 106). Some type of transformation is always available to increase or reduce the variability of original data, making their value questionable (p. 107). Finally, hypothesis testing on the log-transformed data may not address the hypothesis for the original data (p. 108). For these reasons, I apply no data transformation. Instead, descriptive statistics transparently include skewness, kurtosis, and Shapiro–Wilk tests of normality.

<sup>6</sup> The Gardner Lab at Stanford University offers a concise introduction to the logic and terminology of signal detection theory for human perception research online at: <https://gru.stanford.edu/doku.php/tutorials/sdt>. For more detailed introductions in a psychology framework, see Heeger's (1997) handout or online summary at: <https://www.cns.nyu.edu/~david/handouts/sdt/sdt.html>.

## 7.2.3 Combined analysis

Table 3 displays central tendency statistics and normality tests for experiment 1 by condition. Equivalent mean RT of Match conditions across Assimilation Types (Nasal:  $M = 564$  ms; Fricative:  $M = 565$  ms) establishes the baseline performance for experiment 1 for both conditions against which to compare other conditions. Both Match conditions satisfy the assumption of normality, and an independent samples *t*-test detects no significant difference between the two Match conditions,  $t_{(181)} = -0.067$ ,  $p = 0.946$ . Mean RT between Nasal conditions shows an ordinal pattern: Underspecified Mismatch ( $M = 632$  ms) RT is inhibited with respect to Match, and Specified Mismatch is slower still ( $M = 689$  ms). This pattern of processing inhibition does not hold for the Dorsal Fricative conditions, where Mismatch appears to be slightly faster ( $M = 554$ ) than Match ( $M = 565$ ).

Scott (2019a) coded both types of Nasal Mismatch together for a  $2 \times 2$  model (Match condition vs. Assimilation type) but noted a regular difference in RT between the Specified Mismatch and Underspecified Mismatch conditions (p. 319). To investigate this, the present analysis distinguishes all three levels. This yields three levels of condition for nasals (Nasal Match, Nasal Underspecified Mismatch, Nasal Specified Mismatch) and two levels for fricatives (Fricative Match, Fricative Mismatch), which precludes a meaningful comparison within a single model. Separate analyses for nasals and fricatives follow.

## 7.2.4 Nasals analysis

See Table 3 for descriptive statistics and normality tests for nasal conditions. A linear mixed-effects model was run on the nasal RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (3 levels: Nasal Match, Nasal Underspecified Mismatch, Nasal Specified Mismatch) and Target (2 levels: K, P) were declared as fixed effects. To construct a maximal initial model (Barr et al., 2013), participants and items were declared as random effects grouping factors with Condition, Target, and Condition\*Target as random effects. The data set could not support the maximal random effects structure, so the model was incrementally simplified to what the data can support (Matuschek et al., 2017). The final model included Condition, Target, and Condition\*Target as fixed effects with random intercept. Table 4 reports the estimated marginal means and parameter estimates. Figure 4 depicts the distribution of

**TABLE 4** Estimated marginal means (ms, hits only) of experiment 1 (nasals), SE, and 95% confidence interval (top), with parameter estimate, variability, SE, df, t-value, and p-value (bottom).

Condition	Mean RT	SE	95% confidence interval		
			Lower bound	Upper bound	
<b>Match condition</b>					
Match					
K-detection	551	23.9	504	598	
P-detection	578	23.9	531	625	
<b>Mismatch</b>					
Underspecified					
K-detection	614	23.9	567	661	
P-detection	655	24.1	608	702	
<b>Specified</b>					
K-detection	634	23.8	587	681	
P-detection	752	24.5	704	800	
<b>Target condition</b>					
<b>K-detection</b>					
Match	551	23.9	504	598	
<b>Mismatch</b>					
Underspecified	614	23.9	567	661	
Specified	634	23.8	587	681	
<b>P-detection</b>					
Match	578	23.9	531	625	
<b>Mismatch</b>					
Underspecified	655	24.1	608	702	
Specified	752	24.5	704	800	
Fixed effects	Estimate	SE	df	t	p
Intercept	631	16.04	11.97	39.33	<0.001
[Condition (1) = Underspecified Mismatch]	-66	11.28	12.16	-5.85	<0.001
[Condition (2) = Specified Mismatch]	4	11	12.28	0.33	0.748
[Target (1) = P]	-31	8	12.28	-3.87	0.002
[Condition = Underspecified Mismatch]*[Target = P]	17	11	12.17	1.55	0.148
[Condition = Specified Mismatch]*[Target = P]	10	11	12.29	0.93	0.372
Parameter	Variance	$\sigma$			
Residual	8,958	95			
Participant	2,318	48			
Item	369	19			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Target reference level is K. Significance calculation includes Holm-Bonferroni correction for multiple comparisons.

responses by condition, showing an apparent ordinal trend overall toward slower RT from Match to Underspecified Mismatch to Specified Mismatch (Figure 4A), although this trend appears to be less robust in [k]-detection conditions (Figure 4B, left).

With the type III tests of mixed effects, the F-tests show a main effect of match condition,  $F_{(2,12.27)} = 21.474, p < 0.001$ , driven by the observed difference in mean RT between match and underspecified mismatch conditions. Although the mean RT for the specified mismatch is relatively slower, its contribution to this model does not achieve significance. There is also a main effect of the listening target,  $F_{(1,12.27)} = 14.940, p = 0.002$ , confirming that [p]-detection trials have consistently slower RT than [k]-detection. The interaction of these factors is marginal, not significant,  $F_{(2,12.28)} = 3.089, p = 0.082$ . An independent samples t-test comparing underspecified and specified mismatch conditions reveals a significant difference,  $t_{(135)} = 3.022, p = 0.003$ , a medium effect,  $d = 0.516, SE = 0.177$ . Rerunning the model with only underspecified and specified mismatch trials shows the same main effects, but again there is not a significant interaction effect. In summary, experiment 1 replicates a robust inhibition effect for violation of RNA (cf. Otake et al., 1996; Weber, 2001a,b) and shows a significant difference between [k]-detection and the relatively slower [p]-detection conditions. There is a significant difference between underspecified mismatch [nk np] trials and slower specified \*[ɲp mk] trials, and the differences of estimated marginal mean RTs between [p]-detection and [k]-detection are relatively wider for specified than for underspecified trials (118 vs. 41). Nonetheless, this interaction is marginal, so it does not conclusively indicate a greater inhibition effect for \*[ɲp] than \*[mk].

### 7.2.5 Fricatives analysis

Descriptive statistics and normality tests for fricative conditions are displayed in Table 5. A linear mixed-effects model was run on the fricative RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (Match vs. Mismatch) and Vowel Context ([a] vs. [ɛ]) were declared as fixed effects. Participants and items were declared random effects grouping factors. As before, the maximal model was incrementally simplified, resulting in a final model that includes Condition, Context, and Condition\*Context as fixed effects with random intercept. Table 6 reports the estimated marginal means and parameter estimates. Figure 5 depicts the distribution of responses by condition: licit [eç], illicit \*[ɛx], licit [ax], illicit \*[aç].

With the type III tests of mixed effects, the F-tests show no main effect of condition,  $F_{(1,15.69)} = 0.682, p = 0.421$ , nor is there a main effect of vowel context,  $F_{(1,15.66)} = 1.531, p = 0.234$ . The interaction is also not significant,  $F_{(1,15.74)} = 1.728, p = 0.207$ . In summary, experiment 1 replicates neither a facilitation effect (cf. Weber, 2001a,b) nor an inhibition effect (cf. Lindsey, 2013) for L1 German speakers. Subsequent analyses of these fricative assimilation data by Scott (2019a, p. 322–326) revealed a high degree of variation between participants, including individuals with strong inhibition or strong facilitation but mostly neither.

Recall from Section 3.4.2 Weber's (2001a; 2001b, p. 40–41, 53) claim that facilitation may be a novel popout reaction to DFA-violating sequences that are truly novel in the language, such as \*[ix], and not those that are merely rare (e.g., [u:ç] in *Kuhchen*

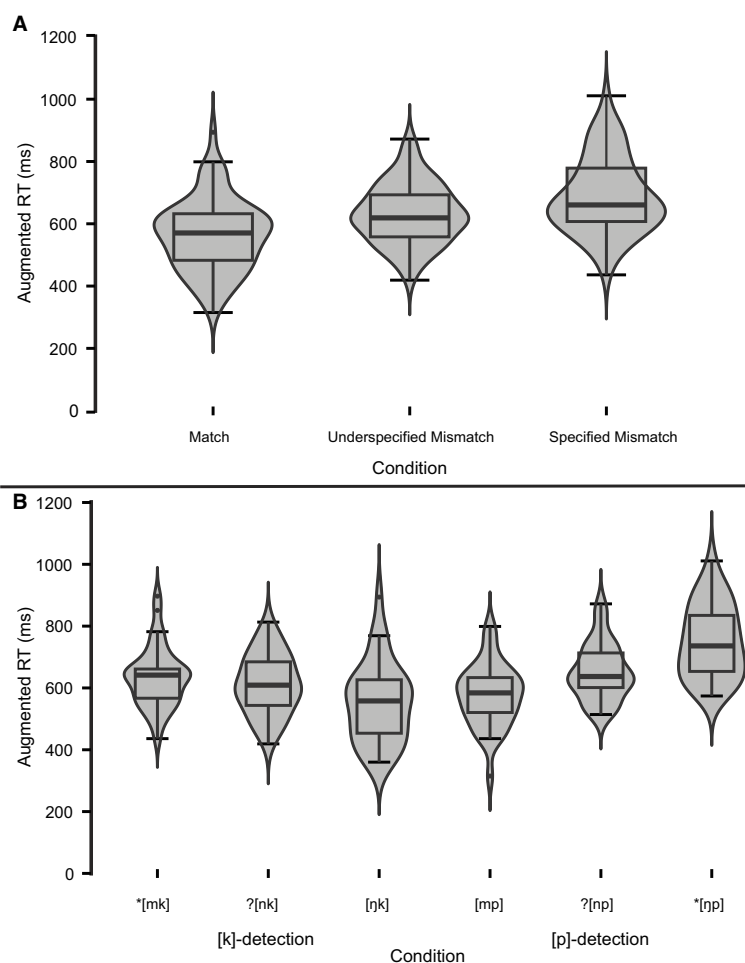


FIGURE 4 Experiment 1 (nasals): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data (A) by condition (match, underspecified mismatch, specified mismatch) and (B) additionally by target (K vs. P).

TABLE 5 Descriptive statistics and tests of normality for experiment 1 fricatives by condition.

Condition	n	μ (ms)	σ	Skewness		Kurtosis		Shapiro–Wilk	
					SE		SE	W	p
Match	113	565	99.0	0.464	0.227	-0.022	0.451	0.980	0.095
[ax] + t	56	581	105.6	0.343	0.319	-0.173	0.628	0.982	0.556
[εç] + t	57	549	90.2	0.508	0.316	0.154	0.623	0.973	0.226
Mismatch	107	554	112.3	0.820	0.234	0.393	0.463	0.940	<0.001
*[aç] + t	54	552	117.0	1.117	0.325	1.040	0.639	0.897	<0.001
*[εx] + t	53	556	108.4	0.469	0.327	-0.298	0.644	0.963	0.102

Reaction time values rounded to the nearest millisecond. Observed positive skewness is not an artifact of data trimming, which excluded no non-filler fast responses (<100 ms) in [t]-detection conditions.

/ku:çəŋ/ → [ˈku:çəŋ] “cow, DIM.”). To explore this claim, an independent samples *t*-test between Back–Front Mismatch (\*[aç]; *M* = 552, *SD* = 117) and Front–Back Mismatch (\*[εx]; *M* = 556, *SD* = 108) conditions was run. No significant difference was revealed, *t*(105) = 0.163, *p* = 0.871. In contrast, a comparison of the corresponding Match conditions showed a marginal difference,

*t*(111) = 1.721, *p* = 0.088, suggesting a trend toward a slower baseline RT for matching back [ax] sequences (*M* = 581, *SD* = 106) than matching front [εç] (*M* = 549, *SD* = 90). Thus, despite the slightly faster mean (by 11 ms) observed for the Mismatch condition, the statistical model of experiment 1 results does not conclusively support Weber’s novel popout argument.

TABLE 6 Estimated marginal means (ms, hits only) of experiment 1 (fricatives), SE, and 95% confidence interval (top), with parameter estimate, variability, SE, df, t-value, and p-value (bottom).

Condition	Mean RT	SE	95% confidence interval			
			Lower bound	Upper bound		
<b>Match condition</b>						
Match						
[ax] + t	584	22.0	541	627		
[eç] + t	552	21.9	509	595		
<b>Mismatch</b>						
*[aç] + t	557	22.1	514	600		
*[ex] + t	558	22.2	515	601		
<b>Vowel Context</b>						
Back [a]						
[ax] + t	584	22.0	541	627		
*[aç] + t	557	22.1	514	600		
Front [e]						
[eç] + t	552	21.9	509	595		
*[ex] + t	558	22.2	515	601		
<b>Fixed effects</b>		<b>Estimate</b>	<b>SE</b>	<b>df</b>	<b>t</b>	<b>p</b>
Intercept		563	19.04	11.25	29.55	<0.001
[Condition (1) = Mismatch]		5	6.39	15.69	0.83	0.421
[Context (1) = e]		8	6.38	15.66	1.24	0.234
[Condition = Mismatch]* [Context = e]		8	6.39	15.74	1.32	0.207
<b>Parameter</b>		<b>Variance</b>	<b>σ</b>			
Residual		7,490	87			
Participant		3,852	62			
Item		132	11			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Vowel Context reference level is back [a]. Significance calculation includes Holm–Bonferroni correction for multiple comparisons.

### 7.2.6 Discussion of experiment 1

The combined analysis (Section 7.2.3) establishes baseline performance on the task for Match conditions across assimilation types. The nasals analysis (Section 7.2.4) addresses the first two research questions. As expected, violation of RNA results in consistent, pronounced RT inhibition (RQ1). Regarding the second research question, the results show that the type of mismatch, whether underspecified due to underapplication of RNA ([n]) or specified with a clash of non-coronal place features (\*[mk ɲp]), makes a significant difference in RT by degrees. This distinction is primarily driven by strong processing inhibition for illicit \*[ɲp] sequences, which never occur in German, whereas [mk] sequences, rare yet possible (e.g., *Imker*), manifest less delay. This may be due to the uncontroversial phonemic status of /m/ in German, which does not require derivation via RNA. In

contrast, [ɲ] derives exclusively from RNA in pre-velar context. This demonstrates processing differences between Place-assimilated, Place-mismatched underspecified, and Place-mismatched specified nasals, constituting psycholinguistic evidence for the incremental ungrammaticality of phonotactic violations (RQ2). This ordinal differentiation suggests that there may be a continuum of processing inhibition for assimilation of this type according to the intensity of the RT effect:

\* (phonotactic constraint) > ? (phonological underapplication) > lexical rarity<sup>7</sup>

The possibility remains that the phonemic status of /m/ may also play a role, as argued by Otake et al. (1996). Note the difference in Table 4 between the quick baseline RT for homorganic labial [mp] and the much slower RT for heterorganic \*[mk], which violates the Coda condition (Itô, 1989; p. 224). Because [mk] sequences occur rarely in German, labeling them questionable ?[mk] may be more appropriate. Their mean RT here also aligns better with underspecified ?[nk np]. Inhibition with ?[nk np] seems to reflect that they do not occur in German. However, following Darcy et al. (2007), unassimilated sequences ought to be recoverable as cases of the underspecified nasal due to highly overlearned (automatic) perception routines (Strange, 2011). This may play a role in the reduced degree of processing inhibition observed. In contrast, Specified Mismatch \*[ɲp], which lacks any phonological or lexical motivation, exhibits more severe processing inhibition than all other conditions. This partly aligns with the ternary logic of the FUL model (Lahiri and Reetz, 2010), which describes an algorithm for comparing features extracted from the acoustic signal with features encoded in the mental lexicon to recover the speaker’s intended lexical meaning. By the FUL algorithm, we might expect a *mismatch* for \*[ɲp], which lacks motivation via RNA or lexical precedent, to inhibit processing more than a *nomismatch* for unassimilated /Nk/ or /Np/ surfacing with coronal [n] (see Figure 2). However, the present nasal model’s marginal factor interaction does not conclusively support a difference between [k]-detection mismatches and relatively slower [p]-detection mismatches, so experiment 2 cannot address how incorrect application of RNA might interact differently with the labial feature of phonemic /m/ vs. the RNA-acquired velar feature of [ɲ]. Alternatively, this pattern may partly owe to relatively weaker invariant cues for [p] Ohala (1996, p. 1720; Weber, 2001a, p. 111). Future research with a larger data set designed to investigate match, underspecified mismatch, and specified mismatch with labials and velars, in words and non-words, is needed to address this conclusively.

Finally, the fricative analysis (Section 7.2.5) yielded no significant results for violation of DFA. It is possible that subtle RT effects may be cloaked here by the slower baseline for trials with [ax] (Table 5). Alternatively, the [t]-detection task may not be sensitive enough to detect small facilitation effects in the preceding context, or greater individual variation in sensitivity to DFA violations may

<sup>7</sup> A fourth category, *lexical unprecedentedness*, could logically arise as the systematic result of a phonotactic constraint or as the accidental result of a lexical gap, where novel potential words may yet comply with well-formedness principles. Whether a complete lexical gap would affect processing similarly to a phonotactic ban is an open empirical question.



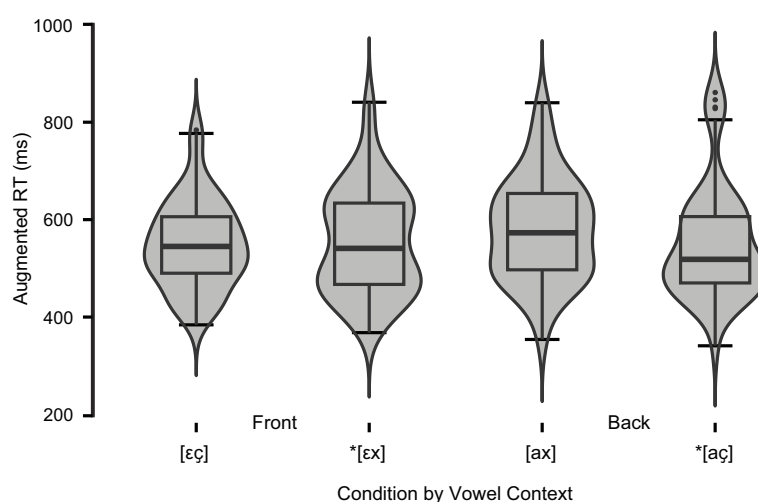


FIGURE 5

Experiment 1 (fricatives): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data by condition (match vs. mismatch).

necessitate more statistical power (Type II error). Finally, DFA may not exert much influence on perception for this L1 German sample. Thus, experiment 2 takes up the third research question again.

## 8 Experiment 2: detection of Persean targets in L1 and L2+

Experiment 1 validated the phoneme detection task generally by replication of an RT inhibition effect in response to violation of RNA with an L1 German group. However, the regressive (left-to-right) directionality of the RNA condition only shifts the listening target one step away from the object of interest  $x$  onto the following [p] and [k], both of which still play a role in the RNA process as the triggers of place assimilation. To take the next step away from  $x$  itself and stipulate a true Persean listening target  $y$ , the [t]-detection task investigated nearby DFA, a place assimilation in which the [t] plays no role. This condition was inconclusive in experiment 1 with the L1 German group, so in experiment 2, this is attempted again and with an additional L2+ learner group. Experiment 2 also novelly tests this Persean listening target on the prosodic/phonotactic constraint governing /h/ in codas.<sup>8</sup>

Experiment 1 also established a baseline performance pattern between phoneme detection tasks that use an uncontroversial phonemic listening target, whether that target is part of the conditioning environment (i.e., [k] and [p] in RNA context; Otake et al., 1996, experiments 5 and 6; Weber, 2001a,b, experiment 4) or it is an adjacent Persean target, unrelated to the assimilation of interest in DFA context. The aim of experiment 2 is to investigate sensitivity to phonotactic violations in both L1 speakers and adult L2+ learners, without drawing participant attention to segments that are directly involved in the phonological principles of interest, because such segments are likely attached to

differing mental representations between groups (e.g., application or conditioning environments). To achieve this, experiment 2 uses [t], which is uncontroversially phonemic /t/ in both German and English. Furthermore, the [t]-detection task uses only released stops, avoiding language-specific alternatives such as optionally unreleased final stops in English. Thus, the task employs phonetic realizations that unambiguously instantiate the phoneme /t/ in both languages.

Experiment 2 draws additional motivation from research on the perception of German dorsal fricatives by L1 American English speakers. Scott (2019a) and Scott and Darcy (2023) report that prosodic and phonotactic contexts modulate perceptual assimilation (Best, 1995; Best and Tyler, 2007) of dorsal fricatives [x] and [ç] and confusability with [k], [ʃ], and [h]. This includes perceptual assimilation to [h] in coda positions, despite the phonotactic/prosodic constraint \*Coda-/h/ (§4). For L1 English speakers' perception of German, such a perceptual assimilation mapping in cross-language and IL perception violates syllable well-formedness principles of both the L1 and the TL. This indicates attention to phonetic detail over phonological patterning during early exposure. To complement the explicit attention to phonetic detail inherent to perceptual assimilation or phoneme detection focused on subphonemic variants (Lindsey, 2013; Weber, 2001a,b, 2002), experiment 2 investigates implicit processing reactions while listener attention is directed elsewhere as a proxy for *automatic* or *optimal* phonological knowledge (Bordag et al., 2021; Hui and Jia, 2024; Strange, 2011).

Following the [t]-detection block of experiment 1, experiment 2 aims to replicate the findings on German DFA reported by Weber (2001a,b, 2002) or Lindsey (2013). Acknowledging the negative DFA result of experiment 1, it also remains to be shown whether this task design is sensitive enough to detect RT effects on processing caused by violations of a preceding progressive assimilation. I then undertake the first investigation of the prosodic/phonotactic constraint \*Coda-/h/. Experiment 2 should yield shifts in RT in response to violation of DFA and

<sup>8</sup> This section presents a reanalysis of a data set originally collected in 2015 and reported by Scott (2019a, Chapter 6; Scott, 2019b).

TABLE 7 Experiment 2 response tallies by group after exclusion criteria.

Group	Dorsal fricative		Glottal fricative		Totals
	Licit	Illicit	Licit	Illicit	
	[axt εçt]	[açt εxt]	<sub>σ</sub> [hVCt]	[...Vht] <sub>σ</sub>	
L1 (n = 9)	135	123	60	56	374
L2+ (n = 14)	194	195	100	98	587
Totals	329	318	160	154	N = 961

Fillers are excluded from the table and analysis. All critical trials were monosyllables with the listening target [t] in final position. In licit glottal fricative trials, the penultimate consonant was always licit in the position in both English and German.

\*Coda-/h/ in each participant group for whom these phonological patterns are mentally represented.

## 8.1 Method

### 8.1.1 Participants

Two participant groups were recruited for experiment 2. Data collection for L1 German speakers (12 females, 2 males; ages 20–29,  $M = 22.9$ ,  $SD = 2.492$ ), took place in Stuttgart, Baden-Württemberg, Germany during a data collection period of ~3 weeks. Seven of these also completed experiment 1 more than 30 min later after intervening experiments to avoid back-to-back presentation of two phoneme detection experiments. **Supplementary material D** presents additional participant details for this group. A one-semester data collection period with L1 American English adult learners of German was conducted at a Midwestern university. Participants were recruited via advertisement to all students enrolled in the university’s second-semester German course during that term (150+ students) and voluntarily scheduled with the researcher to attend a data collection session at a campus computer lab during two data collection periods of ~2 weeks. Ten learners completed the task at midterm and 19 at finals. Low enrollment in the study made a cross-sectional analysis unfeasible, so both time points were collapsed for analysis, selecting the second time point for those who completed both ( $n = 6$ ). This yielded 22 unique L2+ participants (10 females, 12 males; ages 18–23,  $M = 19.6$ ,  $SD = 1.170$ ). One L2+ participant at the first time point reported simultaneous bilingualism in English and Latvian. Another reported Spanish exposure since birth, first use at age 3. Two reported initial exposure to German at age 3 or “very young” but no use until late teens. One reported birth and residence in Australia until age 4.

### 8.1.2 Stimuli

Experiment 2 investigates phonotactic awareness of both dorsal fricatives and [h], which is phonemic /h/ in English and German. **Supplementary Table E1** summarizes conditions and trial types. The single block of items included 48 critical trials balanced for the Licit and Illicit contexts in six conditions ( $n = 8$  trials per condition). As long as a dorsal fricative matches the place of the preceding vowel (DFA), German phonotactics permits both [ç] and [x] in coda clusters followed by [t]. In contrast, /h/ is never allowed in syllable codas, simple or complex. Licit conditions were Front Match [εç]

TABLE 8 Signal detection rates for experiment 2 after participant exclusion ( $n_{L1} = 9$ ,  $n_{L2+} = 14$ ).

Group × condition	Signal detection			
	Hit	False alarm	Correct rejection	Miss
<b>L1</b>				
Dorsal licit [ax εç]	0.938	–	–	0.063
Dorsal illicit *[aç εx]	0.854	–	–	0.146
Glottal licit Onset-[h]	0.833	–	–	0.167
Glottal illicit *Coda-[h]	0.778	–	–	0.222
Distractors <sup>a</sup>	0.803	–	–	0.197
Fillers	–	0.012	0.988	–
<b>L2+</b>				
Dorsal licit [ax εç]	0.866	–	–	0.134
Dorsal illicit *[aç εx]	0.871	–	–	0.129
Glottal licit Onset-[h]	0.893	–	–	0.107
Glottal illicit *Coda-[h]	0.875	–	–	0.125
Distractors <sup>a</sup>	0.772	–	–	0.228
Fillers	–	0.021	0.979	–

Dashes indicate fields for which no rate is possible with the “go”/“no-go” format.

<sup>a</sup>Distractor Miss rate may include responses entered before the experiment response logger initiated.

(e.g., [glεçt]), Back Match [ax] (e.g., [glaxt]), and Onset-[h] (e.g., [hamt]). Illicit counterparts were Back-Front Mismatch \*[aç] (e.g., \*[glaçt]), Front-Back Mismatch \*[εx] (e.g., \*[glεxt]), and \*Coda-[h] (e.g., \*[gaht]). Similar to experiment 1, these were counterbalanced by 144 distractors with [t] in other positions and 192 fillers without [t] for a total of 384 trials with 1:3 critical:distractor ratio and 50% fillers. **Supplementary material F** provides a complete list of stimuli. Stimuli were recorded as in experiment 1.

Segment boundary marking and extraction of stimulus onset and duration were conducted as in experiment 1. **Supplementary Table E2** describes these by condition in aggregate; see **Supplementary material F** for the extracted durations of each stimulus and **Scott (2019a)**, p. 351–354 for additional analyses of the stimuli. Augmented RT for analysis was derived as in experiment 1.

### 8.1.3 Procedure

The same computer lab in Stuttgart was used for experiment 2, which was always administered before experiment 1 for those who completed both. Sessions lasted 100–120 min and included a language background questionnaire and two additional experiments of a different type. This group received €15 payment.

U.S. data collection took place in a university language laboratory, starting with a language background questionnaire and then testing on desktop computers running Windows 7 Service Pack 1 (64-bit). Additional specifications varied by computer: 2.6, 3.4, or 3.6 GHz processor; 4, 8, or 16 GB RAM; screen resolutions of 1024 × 768, 1440 × 900, or 1680 × 1050. Stimuli were presented with high-quality Sanako (Tandberg Educational) SLH-07 circumaural headphones; participants could adjust the volume themselves.

TABLE 9 Descriptive statistics and tests of normality for experiment 2 by condition (L2+).

Condition	n	μ (ms)	σ	Skewness		Kurtosis		Shapiro-Wilk	
					SE		SE	W	p
Licit	291	543	128.2	0.816	0.143	0.081	0.285	0.941	<0.001
DFA match	192	543	130.5	0.877	0.175	0.075	0.349	0.926	<0.001
Onset-[h]	99	543	124.3	0.689	0.243	0.140	0.481	0.961	0.005
Illicit	291	555	126.8	0.476	0.143	-0.258	0.285	0.977	<0.001
DFA mismatch	194	519	116.0	0.614	0.175	-0.164	0.347	0.960	<0.001
*Coda-[h]	97	628	116.1	0.469	0.245	-0.191	0.485	0.972	0.035

Reaction time values rounded to the nearest millisecond. Observed positive skewness is most likely not an artifact of data trimming, which excluded only three DFA trials (2 × [εç] + t Match, 1 \* [εx] + t Mismatch) and one Glottal trial (Onset-[he]) as non-filler fast responses (<100 ms).

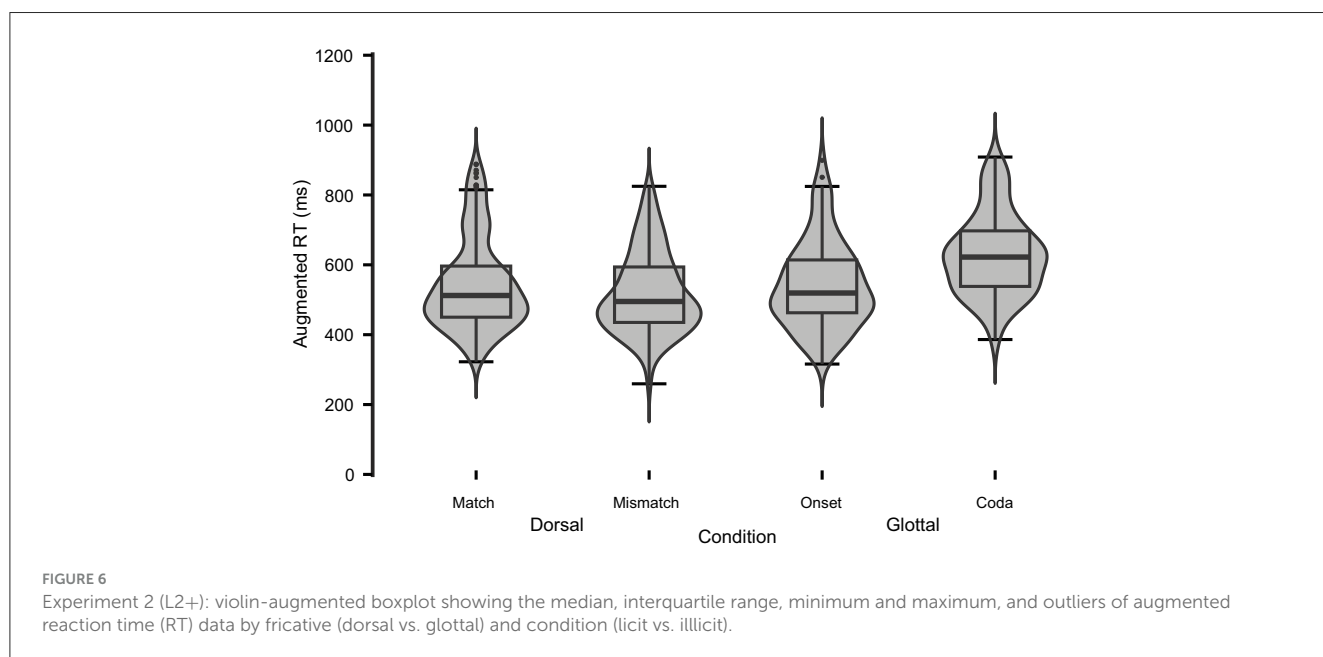


FIGURE 6 Experiment 2 (L2+): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data by fricative (dorsal vs. glottal) and condition (licit vs. illicit).

Task presentation, training phase, within-block randomization, and trial structure were the same as for the [t]-detection portion of experiment 1 but with OpenSesame (Mathôt et al., 2012; Version 2.8) and opportunity for self-paced breaks every 32 trials. With breaks in between, the experiments lasted ~90 min. Participants in midterm data collection received US\$10; those at the finals session received a 1% bonus German course credit. Returning participants were entered in drawings for one US\$50 cash prize per 10 returning participants.

## 8.2 Results

### 8.2.1 Exclusion criterion

The same minimum threshold of five or more hit responses was applied to the four conditions of experiment 2: Dorsal Licit [ax εç], Dorsal Illicit \*[aç εx], Onset-[h] Licit, and \*Coda-[h] Illicit. This criterion retained 9 of the L1 German group and 14 of the L1 English group. I acknowledge that these are small numbers. This is due to difficulty in recruiting participants to voluntarily schedule a long laboratory session during limited data collection

periods at each location (e.g., <20% of the course enrollment of L2+ learners during the semester) as well as a deliberate effort to avoid recruiting linguistics majors in Stuttgart (the most available and willing group), which would have skewed the L1 German data through their atypical metalinguistic awareness. Nonetheless, the statistical techniques employed in this study should be able to handle the small data set and provide useful insights that may be tested with replication. Table 7 describes the remaining data set, and Table 8 displays signal detection rates by group and condition. The L2+ group exhibits little variation between critical conditions, whereas the L1 German group shows consistently lower hit rates for illicit contexts than for licit contexts within each fricative type condition. They also exhibit lower hit rates for the glottal [h] fricative type conditions than for the dorsal fricative type.

### 8.2.2 Data trimming and preparation

The data trimming procedure was as in experiment 1. The fast trial cutoff (<100 ms) excluded four trials from the L1 German data set (2 × \*[aç], 1 × [εç], 1 × \*[ah]) and four trials from the L2+

TABLE 10 Estimated marginal means (ms, hits only) of experiment 2 (L2+), SE, and 95% confidence interval (top), with parameter estimate, variability, SE, df, t-value, and p-value (bottom).

Condition	Mean RT	SE	95% confidence interval			
			Lower bound	Upper bound		
<b>Licitness condition</b>						
Licit						
Dorsal [ax eç] + t	550	19.3	512	588		
Glottal Onset-[h]	547	19.3	509	585		
Illicit						
Dorsal *[aç ex] + t	522	19.3	484	560		
Glottal *Coda-[h]	623	19.4	590	666		
<b>Fricative</b>						
Dorsal						
Licit [ax eç] + t	550	19.3	512	588		
Illicit *[aç ex] + t	522	19.3	484	560		
Glottal						
Licit Onset-[h]	547	19.3	509	585		
Illicit *Coda-[h]	623	19.4	590	666		
<b>Fixed effects</b>		<b>Estimate</b>	<b>SE</b>	<b>df</b>	<b>t</b>	<b>p</b>
Intercept		562	15.27	15.64	36.79	<0.001
[Condition (1) = Mismatch]		-13	6.63	44.17	-2.02	0.050
[Fricative (1) = Glottal]		-26	7.31	24.18	-3.51	0.002
[Condition = Mismatch]* [Fricative = Glottal]		27	6.62	44.09	4.11	<0.001
Parameter	Variance		$\sigma$			
Residual	11,366		107			
Participant	2,648		51			
Fricative	133		12			
Item	931		31			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Fricative reference level is Dorsal. Significance calculation includes Holm–Bonferroni correction for multiple comparisons.

German data set (2 × [eç], 1 × \*[ex], 1 × [he]). As in experiment 1, augmented RT data are not log-transformed. Skewness, kurtosis, and Shapiro–Wilk tests of normality are reported. See Scott (2019a) for further discussion of normality conformity and violation in these data.

### 8.2.3 Intergroup comparison

Supplementary material G summarizes the descriptive statistics and tests of normality for L2+ (Supplementary Table G1) and L1 (Supplementary Table G2) groups. Note that, except for two vowel–consonant pairing conditions (licit [eç] and illicit \*[ah]), the L1 German group exhibits a slower mean RT across the board.

Furthermore, while the L1 group data satisfy the test of normality for licit trials and all licit subconditions, the L2+ learner group data do not. With RT experiments, such intergroup differences may reflect real behavioral differences rooted in the fundamental difference between L1 speakers and early L2+ learners. Alternatively, as the groups were tested in different laboratory settings, this may reflect hardware or software latency differences or a combination of these factors. The present research questions do not entail intergroup comparisons, and retaining Group as an additional fixed effect might hinder statistical modeling with this small data set. Therefore, separate models for L2+ and L1 groups are reported here.

### 8.2.4 L2+ German analysis

Descriptive statistics and normality tests for the L2+ group in DFA and [h] conditions (licit and illicit for each) are displayed in Table 9. Figure 6 depicts the distribution of responses by condition, with generally longer RT for illicit coda-[h] than all other conditions. Supplementary Table G1 shows that, for the L1 English L2+ German group, mean RT was similar overall between Licit conditions, most of which violate the assumption of normality. This establishes baseline task performance for the L2+ group across fricative conditions. A linear mixed-effects model was run on the L2+ RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (Licit vs. Illicit) and Fricative (Dorsal vs. Glottal) were declared as fixed effects. Participants and items were declared as random effects grouping factors. The maximal model was incrementally simplified, resulting in a final model that included Condition, Fricative, and Condition\*Fricative as fixed effects with random intercept and Fricative as a random effect under participant. Table 10 reports the estimated marginal means and parameter estimates. The parity of licit conditions between fricative types is evident in Table 9 and Figure 6.

With the type III tests of mixed effects, the F-tests show the main effects for Condition,  $F_{(1,44.17)} = 4.061, p = 0.050$ , and for Fricative,  $F_{(1,24.18)} = 12.288, p = 0.002$ . The interaction of these is also significant,  $F_{(1,44.09)} = 16.861, p < 0.001$ . In summary, the model finds the RT for illicit trials is significantly different from the licit trials, driven largely by pronounced processing inhibition in response to violation of \*Coda-/h/. To check whether the relatively fast RT trend with illicit \*[aç ex] trials indicates a facilitation effect arising from sensitivity to violation of DFA, dorsal fricative match and mismatch RTs were compared directly. An independent samples t-test reveals a marginal trend,  $t_{(384)} = 1.936, p = 0.054$  (small effect size,  $d = 0.197, SE = 0.102$ ). The model confirms that the reliable inhibition effect observed for the glottal fricative in the illicit \*Coda-[h] context vs. the licit Onset-[h] context is not shared by the dorsal fricative type, which instead may show a slight facilitation for illicit \*[aç ex] contexts as compared to licit [ax eç]. In short, robust inhibition for violations of \*Coda-[h] drives the significant effects of the mixed-effects model, although this group also shows a marginal trend of facilitation for violations of DFA (cf. Weber, 2001a,b).

### 8.2.5 L1 German analysis

Descriptive statistics and normality tests for the L1 group in DFA and [h] conditions (licit and illicit for each) are displayed in Table 11.

TABLE 11 Descriptive statistics and tests of normality for experiment 2 by condition (L1).

Condition	n	μ (ms)	σ	Skewness		Kurtosis		Shapiro–Wilk	
					SE		SE	W	p
Licit	194	573	122.4	0.209	0.175	−0.260	0.347	0.990	0.203
DFA match	134	554	118.4	0.436	0.209	0.123	0.416	0.980	0.049
Onset-[h]	60	615	121.3	−0.290	0.309	−0.062	0.608	0.986	0.716
Illicit	176	588	135.6	0.472	0.183	−0.418	0.364	0.971	0.001
DFA mismatch	121	552	118.5	0.607	0.220	0.236	0.437	0.974	0.018
*Coda-[h]	55	666	135.8	0.039	0.322	−1.017	0.634	0.962	0.083

Reaction time values rounded to the nearest millisecond. Observed positive skewness is most likely not an artifact of data trimming, which excluded only three DFA trials [2 × \* [aç] + t Mismatch, 1 × [εç] + t Match] and one Glottal trial [\*Coda-[ah]] as non-filler fast responses (<100 ms).

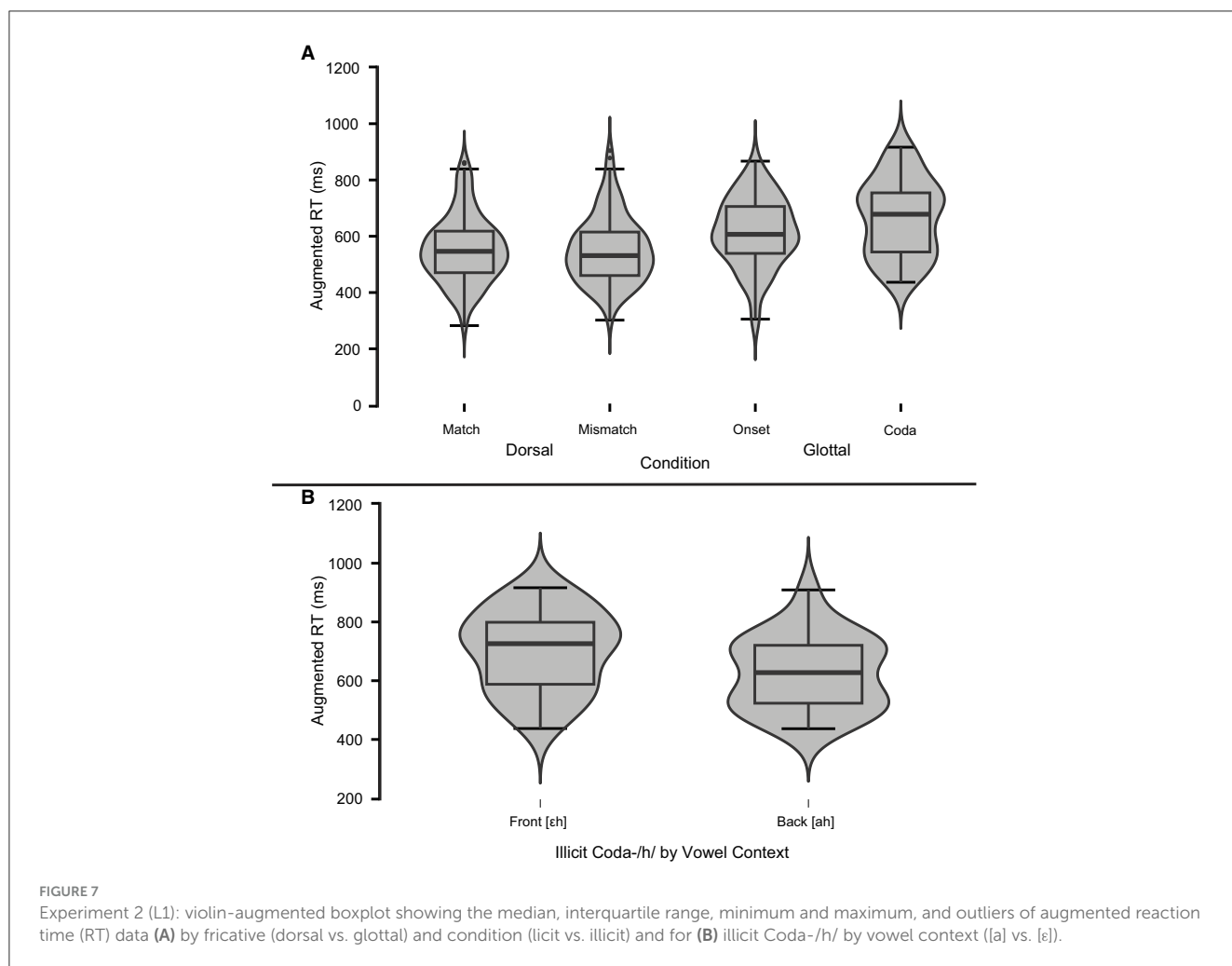


Figure 7A depicts the distribution of responses by condition, showing longer RT for both [h] conditions, especially illicit coda-[h], as compared to dorsal fricatives. Supplementary Table G2 shows that all L1 German licit condition data satisfy the assumption of normality; however, this group’s glottal fricative conditions tend to have slower RT than the dorsal conditions. In addition, the L1 group’s RT means are reliably slower in conditions containing velar [x] than in palatal [ç] conditions, a pattern not shared by the L2+ group.

A linear mixed-effects model was run for the L1 RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (Licit vs. Illicit) and Fricative (Dorsal vs. Glottal) were declared as fixed effects. Participants and items were declared as random effects grouping factors. The maximal model was incrementally simplified, resulting in a final model that includes Condition, Fricative, and Condition\*Fricative as fixed effects with random intercept and Fricative as a random effect under participant (as with the L2+ model). Table 12 reports the estimated marginal means and

TABLE 12 Estimated marginal means (ms, hits only) of experiment 2 (L1), SE, and 95% confidence interval (top), with parameter estimate, variability, SE, df, t-value, and p-value (bottom).

Condition	Mean RT	SE	95% confidence interval		
			Lower bound	Upper bound	
<b>Licitness condition</b>					
Licit					
Dorsal [ax eç] + t	554	26.6	502	606	
Glottal Onset-[h]	619	29.9	560	677	
Illicit					
Dorsal *[aç ex] + t	556	26.7	503	609	
Glottal *Coda-[h]	672	30.2	613	731	
<b>Fricative</b>					
Dorsal					
Licit [ax eç] + t	554	26.6	502	606	
Illicit *[aç ex] + t	556	26.7	503	609	
Glottal					
Licit Onset-[h]	619	29.9	560	677	
Illicit *Coda-[h]	672	30.2	613	731	
<b>Fixed effects</b>					
	Estimate	SE	df	t	p
Intercept	600	24.9	8.63	24.07	<0.001
[Condition (1) = Mismatch]	-14	7.16	46.987	-1.92	0.060
[Fricative (1) = Glottal]	-45	9.01	12.31	-5.00	<0.001
[Condition=Mismatch]*[Fricative = Glottal]	13	7.16	46.96	1.79	0.080
Parameter	Variance	$\sigma$			
Residual	9,124	96			
Participant	5,138	72			
Fricative	268	16			
Item	933	31			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Fricative reference level is Dorsal. Significance calculation includes Holm–Bonferroni correction for multiple comparisons.

parameter estimates. In contrast to the L2+ data, a clear difference in licit conditions between fricative types is evident in Table 11 and Figure 7A.

With the type III tests of fixed effects, the F-tests show a significant main effect for Fricative,  $F_{(1,12.31)} = 24.986, p < 0.001$ . There is a marginal effect of Condition,  $F_{(1,46.99)} = 3.701, p = 0.060$ , and the interaction of these is also marginal,  $F_{(1,46.96)} = 3.210, p = 0.080$ . In summary, the model confirms only that the relatively slower RT for the glottal fricative trials is significant. Unlike the L2+ group, the L1 group’s licit and illicit dorsal fricative trials have nearly identical distributions, and the observed difference between licit and illicit glottal fricative trials is not sufficient to drive an

unambiguous main effect for Condition or an interaction effect, given that glottal trials represent only one third of the data set. However, an independent samples t-test conducted without the dorsal fricative data shows that the difference between licit Onset-/h/ and illicit \*Coda-/h/ trials is significant,  $t_{(113)} = 2.101, p = 0.038$ , with a small-to-medium effect size,  $d = 0.392, SE = 0.190$ . Thus, an RT inhibition effect for violating \*Coda-/h/ appears to single-handedly drive the marginal trends observed in the model. Like experiment 1, this model provides no clear evidence of sensitivity to violation of DFA in the L1 German group.

It is apparent in Supplementary Table G2 and Figure 7B that the glottal fricative illicit \*[eh] subcondition is markedly slower than the illicit \*[ah] subcondition, although both equally violate \*Coda-/h/. To investigate whether the inhibition observed for \*Coda-/h/ trials might be specific to one vowel context, these subconditions were compared directly. An independent samples t-test shows that this difference is significant,  $t_{(53)} = 2.084, p = 0.042$ , with a medium effect size,  $d = 0.564, SE = 0.280$ . Thus, the illicit \*[eh] subcondition specifically contributes most strongly to the marginal trends observed in the model. In contrast, the illicit \*[ah] subcondition has a distribution more comparable to the licit Onset-/h/ condition and a mean RT relatively closer to the licit [ax] condition. This suggests that there may be a principled difference in the L1 German group’s processing of /h/ in syllable codas based on the preceding vowel, such that \*[ah] does not reliably trigger (as much) inhibition. I return to this in Section 8.3.2.

### 8.3 Discussion of experiment 2

#### 8.3.1 Facilitation trend with DFA violations in L2+ learners not found in Swabian Germans

Experiment 2 uses the [t]-detection innovation, which stipulates a Persean listening target that has equivalent phonemic status in English and German, and with stimuli that exhibit a release burst realization appropriate to the coda position in both languages. The dorsal fricative condition undertook to replicate either Weber’s (2001a,b, 2002) processing facilitation findings or Lindsey’s (2013) processing inhibition findings. Either of these RT effects for either participant group would answer the third research question (RQ3) in the affirmative for each of the L2+ (RQ4a) and L1 German (RQ4b) groups and the methodological question of whether the Persean task design is sensitive enough for investigating progressive assimilation processes (RQ4c). The L2+ group shows only a marginal trend, with a small effect size, suggesting that the early L2+ learners may have a weak tendency toward facilitation (faster RT) in response to violation of DFA (RQ4a; cf. Weber, 2001a,b). The L1 German group shows no facilitative or inhibitory effect (RQ4b; contra Lindsey, 2013, and Weber). Thus, no previously reported RT effect is replicated here (RQ3), yet the adjacent [t]-detection task’s ability to detect listener sensitivity to violations of progressive assimilation may be provisionally affirmed (RQ4c), pending more conclusive replication.

The negative result for DFA with L1 German listeners may be due, in part, to the primary location of data collection (Stuttgart). Lipski’s (2006) magnetoencephalography study, which found no effect of DFA violation, also collected data in Stuttgart and nearby

Tübingen, where the regional dialect is Swabian. According to Hall (2022, p. 82–85), [x] and [ç] both occur only in a post-sonorant position in Swabian, and they are in an allophonic relationship through DFA (a case of velar fronting). But southern Baden-Württemberg, including Swabia, abuts High Alemannic and transition dialect areas within 80–120 km, where only velar [x] occurs. Both university cities, and Stuttgart as regional capital, afford ample opportunity for exposure to speakers from a broader dialect area and thus for more variable local input in terms of DFA than in Regensburg or Hannover, where Weber sampled. In short, the failure of experiments 1 and 2 to replicate previous DFA findings may arise from the fact that, for many L1 German speakers in Stuttgart, the Front-Back Mismatch [ɛx] sequence, although illicit and unprecedented in Standard German, is not unusual to encounter in natural speech and thus less likely to trigger any reliable RT effects for groups in aggregate—neither facilitation nor inhibition of processing. Scott's (2019a, p. 292, 317–324) additional analysis of individuals' differences of means reveals that this variability is not limited to participants tested in Stuttgart or who reported a Swabian dialect background. This highlights that sensitivity to DFA likely depends greatly on the dialect exposure of individual speakers. The present experiments may not have detected sensitivity to DFA violations simply because that sensitivity is not a robust feature of this L1 German sample group.

Experiment 2 found only a marginal facilitation trend for the L2+ group when presented with violations of DFA. This response pattern, which was less internally variable (more uniform) than for the L1 German group, further suggests that dialect variation among L1 Germanophones plays an important role in the perception of dorsal fricatives. It may also be the case that the choice of listening target, a phone occurring *after* a progressive assimilation, exchanges some degree of task sensitivity to small effects as the methodological price of avoiding explicit attentional focus on conditioning or application environments in phoneme detection: a trade-off for cross-group task equivalence. Additional research is required to address these questions.

### 8.3.2 Inhibition with Coda-/h/ : broad in L2+ learners, context-dependent for L1 Germans

I included the glottal fricative condition to test the Persean variant of the phoneme detection task's ability to investigate sensitivity to violation of a phonological constraint that is fundamentally different from place assimilation. This is the first phoneme detection study to investigate sensitivity to violation of a constraint on syllable well-formedness: the phonotactic/prosodic constraint \*Coda-/h/.<sup>9</sup> The L2+ group's responses to violations of this constraint exhibit unambiguous and pronounced RT inhibition

<sup>9</sup> In postvocalic coda position, both German and (rarely) English orthography use <h> not to indicate a consonant, but rather as a diacritic of vowel duration (in English, perhaps also vowel quality), for example, German *Stahl* [ʃta:l], "steel," *Mehl* [me:l], "flour," and English *yeah* [jæ:], "yes," *nah* [næ:], "no," or *ah* [a:] and *meh* [me:] (interjections). These contextual uses of <h> may predispose both groups in this study to process any perceived frication noise in the stimuli as cue evidence *against* the perception of /h/ in the signal. This may contribute to the activation of other fricatives as more viable competitor candidates. As with the dorsal fricatives, stipulating /t/ as the listening target discourages explicit attention to the [h] itself.

(RQ5a). The L1 German group's response pattern is more selective: Only the \*[ɛh] subcondition shows inhibition comparable to the L2+ group (RQ5b). In contrast, RT in the \*[ah] subcondition is more comparable to licit conditions such as Onset-/h/ or the [ax] sequence. Together, these results demonstrate that the adjacent [t]-detection task unambiguously detects sensitivity to violations of this phonotactic/prosodic constraint governing syllable well-formedness (RQ5c).

Regarding the \*Coda-/h/ constraint, the intergroup results align and differ in important ways. The L2+ group exhibits robust inhibition, but the L1 group's specific sensitivity to \*[ɛh] in codas complicates the scenario. Crucially, processing inhibition is mitigated for its counterpart \*[ah], not differing markedly from the licit [ax] subcondition. This is a likely candidate for *phonotactic assimilation*, described by Seguí et al. (2001, p. 198), of which they outline three types:

- (1) the listener "ignores" individual phonemes (or stress patterns: Dupoux et al., 1997) that are present in the signal;
- (2) the listener "perceives" illusory phonemes that have no acoustic correlate in the signal;
- (3) the listener "transforms" one phoneme in (sic!) another.

The lack of difference in perception between illicit \*[ah] and licit [ax] in syllable codas likely falls under the third type—that is, the L1 German group appears to perceptually "transform" the contextually illicit [h] into an instance of a spectrally similar fricative (Supplementary material A) that is licensed to follow back [a]—namely, into [x]—and then reacts accordingly, without RT inhibition. This crucial difference between \*[ah] and its corresponding \*[ɛh], which is much less easily repaired in an analogous way due to spectral differences between [h] and [ç], detracts from the overall sensitivity to \*Coda-/h/ violations for this group. Indeed, how RT experiment methodology interacts with prelexical phonotactic assimilation is an open empirical question relevant to broader questions about prediction during language processing more generally (e.g., Kaan and Grüter, 2021; Key, 2014). Alternatively, as the \*[ɛh]-condition is not phonotactically assimilated in this way, this difference may serve as indirect support for the argument that L1 German speakers do, in fact, maintain a phonological distinction between which vowel + dorsal fricative pairs are permitted (RQ4b). This may depend on the interaction of the frontness/backness of each segment, despite the negative result for violations of DFA in the present study.

## 9 General discussion and conclusion

### 9.1 Implications and context of results

The central aim of this study was to develop a version of the phoneme detection task with task parity for participant groups with different language backgrounds and to test the method on various phonological phenomena. The first experiment with L1 German speakers unambiguously replicated processing inhibition (slower RT) findings in response to violations of RNA (RQ1; Otake et al., 1996; Weber, 2001a,b). In addition, reanalysis of Scott's (2019a) nasal data with a model that differentiates underspecified

mismatch and specified mismatch conditions suggests that there may be a continuum of processing inhibition for assimilation of this type according to the intensity of the RT effect such that phonotactic constraints may trigger more severe inhibition than mere underapplication of place assimilation, while rare but precedented sequences may trigger a lesser effect than both. An attempt to analyze the observed difference between specified mismatch \*[ŋp] vs. \*[mk] was inconclusive, leaving the phonemic status of [m] vs. the context-dependent allophone [ŋ] an open question for future research with a larger data set (RQ2).

Both experiments attempted to replicate any sort of consistent RT effect in response to violation of DFA with an L1 German group (RQ3), but the samples in this study exhibited neither processing facilitation (Weber, 2001a,b) nor processing inhibition (Lindsey, 2013). Experiment 2 revealed a weak facilitation (faster RT) effect in L2+ learners in reaction to violation of DFA (RQ4a). This result is similar to Weber's (2001a,b) L1 German result, and contra Lindsey's L2+ learner result, but does support the conclusion that phoneme detection with the Persean listening target [t] is sensitive enough to investigate progressive assimilation (RQ4c). An unexpected asymmetry in the results between the illicit Coda-[h] conditions by vowel context suggests that the L1 German group may have some phonotactic preference for front-back agreement between a nuclear vowel [a] or [ɛ] and a following dorsal fricative despite the inconclusive result in the DFA condition (RQ4b).

Finally, to test this method with more than place assimilation, experiment 2 undertook the first phoneme detection investigation of sensitivity to the prosodic/phonotactic constraint against /h/ in syllable codas in both German and English. Both groups exhibited unambiguous processing inhibition in response to illicit Coda-/h/ (RQ5a, RQ5b), confirming that the Persean listening target technique can be fruitfully employed to investigate a non-assimilation type of phonological knowledge (RQ5c). Additionally, the observed asymmetry in the L1 German group's responses—inhibition specifically for the \*[ɛh] subcondition but not for the \*[ah] subcondition—suggests that this group's prelexical perception may reanalyze the illicit [h] as a licit [x] following the back vowel [a] as a sort of phonotactic assimilation (Seguí et al., 2001).

This study provisionally supports argumentation (e.g., Weber, 2001a,b) that violation of strong phonological expectations (e.g., regressive assimilation, phonotactic/prosodic constraints) yields profound inhibition in phoneme detection, whereas violation of weak phonological expectations (e.g., progressive assimilation) yields a smaller facilitation effect. Future research is needed to establish further systematic predictions about which phoneme detection task and listening target designs elicit which RT effects (facilitation or inhibition) and intensity (small or weak effects) with various populations (L1, L2+), whose representations may diverge in precision and robustness (Bordag et al., 2021; Cook et al., 2016; Darcy et al., 2007, 2013).

The second experiment undertook the first phoneme detection investigation of /h/ ungrammatically located in syllable codas, revealing different patterns of processing inhibition between L1 and L2+ groups. This result has interesting implications for methodological approaches to investigating questions in L2+ phonological theory. The robust processing inhibition effect exhibited in the L2+ learner group clearly validates the

utility of the Persean listening target technique for investigating prosodic/phonotactic constraints other than the segmental adjacency effects of place assimilation. The L1 German group's inhibition pattern generally confirms this. But the lack of this effect specifically in the \*[ah]+[t] condition in the L1 German group suggests an additional, subtler context-dependent sensitivity. Recalling the three types of phonotactic assimilation outlined by Seguí et al. (2001), finding reports of listeners phonotactically assimilating input in illicit contexts is not difficult. Listeners may fail to detect individual phonemes in the signal (Type 1, e.g., L1 Mandarin failure to discriminate syllables with English coda laterals from open CV syllables; Wang, 2023) or illusorily perceive phonemes that have no acoustic basis in the input (Type 2, e.g., L1 Korean epenthesis of a vowel within word-initial consonant clusters in English; Darcy and Thomas, 2019). In contrast with the L1 German group's significant processing inhibition in response to Coda-[h] following the front vowel [ɛ], the same group's lack of RT effect in response to Coda-[h] following the back vowel [a] suggests that they did not perceive the [h] as /h/ at all, but instead as [x], an allophone of the dorsal fricative that would be licit following /a/ (Type 3, transformation of one phoneme to another). Interestingly, the fact that this prelexical perceptual repair strategy does not appear to be viable in the front vowel [ɛ] context suggests that this group has a preference for agreement of place between a vowel and a following dorsal fricative, despite the negative result for the DFA conditions. Nonetheless, phonotactic assimilation is constrained to some degree by an acoustic similarity between the signal (here, [h]) and the potential percept (here, [x] in the back context and [ç] in the front environment). This interaction of two types of phonological knowledge suggests that further research may be able to intentionally leverage such subtleties to investigate one phenomenon of interest (e.g., place assimilation) through phoneme detection tasks aimed at obliquely related phenomena (e.g., prosodic/phonotactic constraints on placement of a phone with similar acoustic cues).

## 9.2 Design phoneme detection experiments for statistical power *a priori*

Particular care is warranted in task design to maintain statistical viability when investigating smaller effects such as processing facilitation in response to violation of weak phonological expectations (e.g., progressive assimilation). The present reanalyses draw finer distinctions within factors than Scott (2019a) initially designed for (i.e., two subtypes of nasal mismatch and sorting coda glottal fricative trials by vowel). This subdivides relatively small data sets to explore additional questions *ad hoc*. In practice, the bar of maximal models set by Barr et al. (2013) is too high for many data sets to meet (Matuschek et al., 2017). The present study simplifies model dimensions so that the results derive sufficient support from the available data. These results may lack enough statistical power to conclusively address specific empirical questions about each of the speech perception phenomena investigated, particularly for small effects. For example, the overall results signal that violation of RNA and \*Coda-/h/ induces RT inhibition generally, but smaller RT effects, such as facilitation, may be marginal, risking Type II



error (false negative). To model several fixed and random effects simultaneously, this study employs linear mixed-effects models. When these data were collected in 2015, there was no standard for *a priori* power analysis of such models with multilevel factors (Kumle et al., 2021, p. 2,528–2,529). Statistical practice also indicates against post-experimental power calculations (so-called *observed power*; e.g., Hoening and Heisey, 2012; Lydersen, 2019). Thus, no power analysis is provided here. Future studies using phoneme detection to investigate phonological and phonotactic sensitivity should include simulation-based *a priori* power analysis at the design stage to determine the appropriate number of trials and sample sizes (see Kumle et al., 2021). Additionally, more statistical research is needed to establish appropriate field standards for how to statistically model RT task data sets, which often violate assumptions of normality.

### 9.3 Design phoneme detection L2+ experiments for task parity

Most phoneme detection studies have focused on phonological interactions between adjacent segments (e.g., place assimilation) and the effects of phonetic transition cues with L1 speakers (e.g., Foss, 1969; Frauenfelder and Seguí, 1989). Although some include non-learner comparison groups from other language backgrounds (cross-language listeners; e.g., Otake et al., 1996; Weber, 2001a,b), phoneme detection investigations of L2+ learner groups are less common (e.g., advanced L2+ learners, Lindsey, 2013). Our collective awareness and comfort with RT psycholinguistic methods generally in L2+ acquisition research is still developing with necessary caution (Hui and Jia, 2024).

When building RT experiments for L2+, thorough methodological consideration is necessary to reduce noise from unintended sources of RT differences between groups, such as from hardware and software latency, stimuli that could be phonetically ambiguous for one language group, and the like. Phoneme detection tasks may require presenting the listening target within a very narrow temporal window (e.g., adjacent phones) for small RT effects to remain detectable (i.e., to avoid Type II error).

One underexamined challenge is achieving task parity between language background and target language proficiency groups in terms of phonological and orthographic representation as discussed in this study. Phoneme detection studies historically vary by the relationship between the listening target and the phenomenon of interest. Listening targets may be subject to a phonological principle (the object of interest *x*; e.g., nasals in RNA, fricatives in DFA and \*Coda-/h/ contexts), they may trigger application of a phonological principle (e.g., obstruents in RNA, vowels in DFA), or they may be merely adjacent without being implicated in the phonological principle of interest (a Persean listening target *y*, e.g., [t] following fricatives in DFA and \*Coda-/h/ contexts). There are countless fundamental representational differences between the phonology, orthography, and the GPCs of any given listening target for different L1 and L2+ populations. As a result, using phoneme detection—or any task design that requires explicit labels for phones or phonemes—in L2+ acquisition research requires careful consideration of the equivalence of the relationship of the listening target and its label to potential mental representations for each language group under investigation. When learner and L1

language groups have equivalent relationships with the experimental task, the experimenter can be more confident that RT results show evidence that the L2+ mental representation is different from the L1 mental representation for the phenomenon of interest, not the trivial result that different cognitive tasks have different performance speeds. As we improve our understanding of the representational challenges that come with listening target labels in phoneme detection, we gain a powerful methodological tool for investigating phonological knowledge in a wide variety of cross-language and L2+ learning scenarios.

### 9.4 Design phoneme detection experiments for L1 varieties and L2+ learner trajectories

This study suggests that the phoneme detection paradigm can be leveraged to investigate underlying mental representations, such as theories of feature underspecification. We can ask a variety of interesting theoretical questions: (How) do L2+ learners' representations of phonological (un)grammaticality change from L1-based to L2-based in the course of IL development? What do the earliest steps of IL development look like? What does ultimate attainment look like in L2+ phonological perception, and does it ever become target-like/optimal? By stipulating an adjacent, phonologically uninvolved, and acoustically consistent Persean listening target, phoneme detection can be a cognitively equivalent task for different groups when investigating important phonological questions with a variety of L1 dialect and regiolect groups and L2+ learners at different stages of IL development. Phoneme detection provides an instrument that enables investigation of implicit, prelexical processing, even with participants who lack any TL experience. Future research should also consider a wider variety of scenarios. This instrument is readily adaptable for uninstructed L2+ learners or non-reading immigrant groups in a TL environment or for comparing groups whose L1 literacies use different non-alphabetic scripts.

### 9.5 Design phoneme detection experiments with Persean listening targets

This study addresses the labeling problem of phonology, orthography, and GPCs for perception research in both L1 and L2+ scenarios with a methodological solution. The experiments presented here establish that directing listener focus away from application and conditioning environments to a Persean listening target (*y*) can meet the challenge of stipulating representationally equivalent listening targets across groups while also more clearly tapping into implicit knowledge of the objects of interest (*x*) that are not attended to explicitly. This is achieved while retaining sufficient task sensitivity to investigate implicit or automatized explicit linguistic knowledge in speech processing for a variety of assimilation phenomena, cue weighting and fusion strategies, and prosodic/phonotactic constraints.<sup>10</sup> Different types of phenomena

<sup>10</sup> For similar work in other areas of L2+ grammar, see Rebuschat (2013) and Suzuki (2017).

may need to be investigated separately, as different RT effects (processing facilitation or inhibition) may not lend themselves to combined models.

Directing attention to a specific phone, such as in identification and perceptual assimilation tasks (Best, 1995; Best and Tyler, 2007), probes perception at a highly explicit level rather than implicit phonological knowledge (cf. *optimal*, Bordag et al., 2021; *automatic*, Strange, 2011). When we are more interested in implicit knowledge including abstract phonological representations such as phonemes or phonotactics, stipulating the object of investigation  $x$  as the listening target can place unintended emphasis on subphonemic phonetic detail in  $x$ . Moreover, for investigation of L2+ *acquisition* of phonological knowledge, stipulating the object of investigation  $x$  as the listening target can confound both intergroup congruence (of phonological and orthographic domains and GPC mappings) and assessment of pre-learner and IL developmental stages in L2+ scenarios. Avoiding the target of investigation as the target of listener attention and redirecting focus to a reliable and congruent Persean listening target  $y$  affords access to crucial questions in L2+ phonology, just as the shield's reflection enabled Perseus to strike true.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation. The datasets analyzed for this study, including language background questionnaires, stimuli, experiment code, and anonymized data set, can be found in the Open Science Framework at <https://osf.io/prfyw/>.

## Ethics statement

The studies involving humans were approved by Indiana University, Bloomington and Universität Stuttgart. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because collection of signed consent forms would have created additional identifying records of participants for otherwise de-identified data, increasing risk of loss of confidentiality.

## Author contributions

JS: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

## References

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by a short-term research grant from the Universität Stuttgart (Sonderforschungsbereich 732), a 2015 Language Learning Dissertation Grant, and a 2016 Graduate Travel Award from the College of Arts and Sciences at Indiana University. Partial funding for open access provided by the University of Maryland School of Languages, Literatures, and Cultures and by UMD Libraries' Open Access Publishing Fund.

## Acknowledgments

I would like to thank Isabelle Darcy, Ken de Jong, Phil Lesourd, and Susanne Even for their guidance and feedback on the dissertation that led to this article. Thanks also to Christiane Kaden for patient hours in a recording booth, Bronson Hui for guidance with statistical analysis in JASP, Janet Scott for reading drafts and graphics assistance, the reviewers for their insightful comments, and the editors of this special collection. This article is based in part on parts of the author's dissertation and a conference proceedings article by the author, cited herein as Scott (2019a,b).

## Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/flang.2024.1254956/full#supplementary-material>

Bassetti, B., Escudero, P., and Hayes-Harb, R. (eds.) (2015a). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguist.* 36.

- Bassetti, B., Escudero, P., and Hayes-Harb, R. (2015b). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguist.* 36:393. doi: 10.1017/S0142716414000393
- Bertelson, P. (ed.). (1986a). The onset of literacy. *Cognition* 24.
- Bertelson, P. (1986b). The onset of literacy: liminal remarks. *Cognition* 24, 1–30. doi: 10.1016/0010-0277(86)90002-8
- Best, C. T. (1995). “A direct realist view of cross-language speech perception,” in *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*, ed. W. Strange (Timonium, MD: York Press), 171–204.
- Best, C. T., and Tyler, M. (2007). “Nonnative and second-language speech perception. Commonalities and complementarities,” in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds. O. S. Bohn and M. J. Munro (Philadelphia: Benjamins), 13–34.
- Boersma, P., and Weenink, D. (2014). *Praat: Doing Phonetics by Computer* (Version 5.3.66) [Computer Program]. Available at: <http://www.praat.org/> (accessed September 27, 2024).
- Bordag, D., Gor, K., and Opitz, A. (2021). Ontogenesis model of the L2 lexical representation. *Bilingualism* 25, 185–201. doi: 10.1017/S1366728921000250
- Castro-Caldas, A. (2004). Targeting regions of interest for the study of the illiterate brain. *Int. J. Psychol.* 39, 5–17. doi: 10.1080/00207590344000240
- Christie, J., and Klein, R. (1996). Assessing the evidence for novel popout. *J. Exp. Psychol. Gen.* 125, 201–207. doi: 10.1037/0096-3445.125.2.201
- Connine, C. M. (1994). “Vertical and horizontal similarity in spoken word recognition,” in *Perspectives on Sentence Processing*, eds. C. Clifton, Jr., L. Frazier, and K. Rayner (Hillsdale, NJ: Lawrence Erlbaum Associates), 107–120.
- Connine, C. M., and Titone, D. (1996). Phoneme monitoring. *Lang. Cogn. Process.* 11, 635–646. doi: 10.1080/016909696387042
- Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345
- Cutler, A., Butterfield, S., and Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *J. Mem. Lang.* 26, 406–418. doi: 10.1016/0749-596X(87)90099-4
- Cutler, A., and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *J. Mem. Lang.* 33, 824–844. doi: 10.1006/jmla.1994.1039
- Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Mental Lexicon* 8, 372–420. doi: 10.1075/ml.8.3.06dar
- Darcy, I., Peperkamp, S., and Dupoux, E. (2007). “Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners,” in *Laboratory Phonology 9*, eds. J. Cole and J. I. Hualde (Berlin: De Gruyter Mouton), 411–442.
- Darcy, I., and Thomas, T. (2019). When *blue* is a disyllabic word: perceptual epenthesis in the mental lexicon of second language learners. *Biling. Lang. Cogn.* 22, 1141–1159. doi: 10.1017/S1366728918001050
- Davis, S., and Cho, M.-H. (2003). The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications. *Linguistics* 41, 607–652. doi: 10.1515/ling.2003.020
- Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci.* 12, 635–647. doi: 10.1162/089892900562390
- Dollenmayer, D. B., Crocker, E. W., and Hansen, T. S. (2014). *Neue Horizonte: Student Activities Manual, 8th Edn.* Boston, MA: Heinle, Cengage Learning.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *J. Exp. Psychol. Hum. Percept. Perform.* 25, 1568–1578. doi: 10.1037//0096-1523.25.6.1568
- Dupoux, E., Pallier, C., Sebastian, N., and Mehler, J. (1997). A destressing deafness in French? *J. Mem. Lang.* 36, 406–421. doi: 10.1006/jmla.1996.2500
- Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., et al. (2014). Log-transformation and its implications for data analysis. *Shanghai Archiv. Psychiat.* 26, 105–109. doi: 10.3969/j.issn.1002-0829.2014.02.009
- Foss, D. J. (1969). Decision processes during sentence comprehension: effects of lexical item difficulty and position upon decision times. *J. Verb. Learn. Verb. Behav.* 8, 457–462. doi: 10.1016/S0022-5371(69)80089-7
- Frauenfelder, U. H., and Seguí, J. (1989). Phoneme monitoring and lexical processing: evidence for associative context effects. *Mem. Cogn.* 17, 134–140. doi: 10.3758/BF03197063
- Glover, J. (2014). *Liquid Vocalizations and Underspecification in German Dialects* (Dissertation). Indiana University, Bloomington, IN, United States.
- Hall, T. A. (1989). Lexical phonology and the distribution of German [ç] and [x]. *Phonology* 6, 1–17. doi: 10.1017/S0952675700000920
- Hall, T. A. (1992). *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Max Niemeyer.
- Hall, T. A. (1995). “Remarks on coronal underspecification,” in *Leiden in Last. HIL Phonology Papers I*, eds. H. van der Hulst and J. van de Weijer (The Hague: Holland Academic), 187–203.
- Hall, T. A. (1997). *The Phonology of Coronals*. Amsterdam: John Benjamins.
- Hall, T. A. (2010). Nasal place assimilation in Emsland German and its theoretical implications. *Zeitschrift für Dialektologie und Linguistik* 77, 129–144. doi: 10.25162/zdl-2010-0005
- Hall, T. A. (2014). Alveopalatalization in Central German as markedness reduction. *Trans. Philol. Soc.* 112, 143–166. doi: 10.1111/1467-968X.12002
- Hall, T. A. (2022). *Velar Fronting in German Dialects: A Study in Synchronic and Diachronic Phonology* (Open Germanic Linguistics 3). Berlin: Language Science Press. doi: 10.5281/zenodo.7185567
- Heeger, D. (1997). *Signal Detection Theory*. New York, NY: New York University. Available at: <http://www.cns.nyu.edu/%7Edavid/handouts/sdt-advanced.pdf> (accessed September 27, 2025).
- Hoening, J. M., and Heisey, D. M. (2012). The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am. Statist.* 55, 19–24. doi: 10.1198/000313001300339897
- Hui, B., and Jia, R. (2024). Reflecting on the use of response times to index linguistic knowledge in SLA. *Ann. Rev. Appl. Linguist.* 2024, 1–11. doi: 10.1017/S0267190524000047
- Hura, S. L., Lindblom, B., and Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Lang. Speech* 35, 59–72. doi: 10.1177/002383099203500206
- Itô, J. (1989). A prosodic theory of epenthesis. *Nat. Lang. Linguist. Theor.* 7:2172259. doi: 10.1007/BF00138077
- Iverson, G. K., and Salmons, J. (1992). The place of structure preservation in German diminutive formation. *Phonology* 9, 137–143. Available at: <https://www.jstor.org/stable/4420048>
- JASP Team (2023). *JASP* (Version 0.18.1) [Computer Software]. Available at: <https://jasp-stats.org/download/> (accessed September 27, 2024).
- Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German*. Amsterdam: John Benjamins.
- Johnston, W. A., and Schwarting, I. S. (1996). Reassessing the evidence for novel popout. *J. Exp. Psychol. Gen.* 125, 208–212. doi: 10.1037/0096-3445.125.2.208
- Johnston, W. A., and Schwarting, I. S. (1997). Novel popout: an enigma for conventional theories of attention. *J. Exp. Psychol. Hum. Percept. Perform.* 23, 622–631. doi: 10.1037//0096-1523.23.3.622
- Kaan, E., and Grüter, T. (2021). *Prediction in Second Language Processing and Learning*. Amsterdam: John Benjamins. doi: 10.1075/bpa.12
- Kalmar, T. M. (2015). *Illegal Alphabets and Adult Bilinguality: Latino Migrants Crossing the Linguistic Border, 2nd Edn.* New York, NY: Routledge.
- Keating, G. D., and Jegerski, J. (2015). Experimental designs in sentence processing research. *Stud. Sec. Lang. Acquisit.* 37, 1–32. doi: 10.1017/S0272263114000187
- Key, M. (2014). Positive expectation in the processing of allophones. *J. Acoust. Soc. Am.* 135, EL350–EL356. doi: 10.1121/1.4879669
- Koda, K. (1998). The role of phonemic awareness in second language reading. *Sec. Lang. Res.* 14, 194–215. doi: 10.1191/026765898676398460
- Kumle, L., Vö, M. L.-H., and Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R. *Behav. Res. Methods* 53, 2528–2543. doi: 10.3758/s13428-021-01546-0
- Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonet.* 38, 44–59. doi: 10.1016/j.wocn.2010.01.002
- Lindsey, J. (2013). *L2 Acquisition of the Context-Dependent [ç]-[x] Alternation in American English Learners of German* (Unpublished undergraduate thesis). Indiana University, Bloomington, IN, United States.
- Lipski, S. C. (2006). *Neural Correlates of Fricative Contrasts Across Language Boundaries* (Dissertation). University of Stuttgart, Stuttgart, Germany. Available at: <http://elib.uni-stuttgart.de/>
- Lydersen, S. (2019). Statistisk styrke - før, men ikke etter! [Statistical power: Before, but not after!] *Tidsskrift for Den norske legeförening* 139:847. doi: 10.4045/tidsskr.18.0847
- Mann, V. A. (1986). Phonological awareness: the role of reading experience. *Cognition* 24, 65–92. doi: 10.1016/0010-0277(86)90005-3
- Marslen-Wilson, W., and Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychol. Rev.* 101, 653–675. doi: 10.1037/0033-295X.101.4.653
- Martin, J. G., and Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *J. Acoust. Soc. Am.* 69, 559–567. doi: 10.1121/1.385484

- Martin, J. G., and Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 473–488. doi: 10.1037//0096-1523.8.3.473
- Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences (Version 2.8.0). *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001
- McMurray, B., and Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychol. Rev.* 118, 219–246. doi: 10.1037/a0022325
- Mills, C. B. (1980). Effects of context on reaction time to phonemes. *J. Verb. Learn. Verb. Behav.* 19, 75–83. doi: 10.1016/S0022-5371(80)90536-8
- Morais, J., Bertelson, P., Cary, L., and Alegria, J. (1986). Literacy training and speech segmentation. *Cognition* 24, 45–64. doi: 10.1016/0010-0277(86)90004-1
- Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *J. Acoust. Soc. Am.* 99, 1718–1725. doi: 10.1121/1.414696
- O'Hara, R. B., and Kotze, J. (2010). Do not log-transform count data. *Methods Ecol. Evol.* 1, 118–122. doi: 10.1111/j.2041-210X.2010.00021.x
- Otake, T., Yoneyama, K., Cutler, A., and Van der Lugt, A. (1996). The representation of Japanese moraic nasals. *J. Acoust. Soc. Am.* 100, 3831–3842. doi: 10.1121/1.417239
- Plag, I., Braun, M., Lappe, S., and Schramm, M. (2009). *Introduction to English Linguistics, 2nd Edn.* Berlin: Mouton de Gruyter.
- Posner, M., Nissen, M., and Ogden, W. (1978). "Attended and unattended processing modes: the role of set for spatial location," in *Modes of Perceiving and Processing Information*, eds. H. L. Pick, Jr., and E. Saltzman (Hillsdale, NJ: Lawrence Erlbaum Associates), 137–157.
- Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Lang. Learn.* 63, 595–626. doi: 10.1111/lang.12010
- Robinson, O. W. (2001). *Whose German? The Ach/Ich Alternation and Related Phenomena in "Standard" and "Colloquial" [ProQuest Ebook Central Version]*. doi: 10.1075/cilt.208
- Scott, J. H. G. (2019a). *Phonemic and Phonotactic Inference in Early Interlanguage: Americans Learning German Fricatives in L2 Acquisition* (Dissertation). Indiana University, Bloomington, IN, United States. Available at: <https://www.proquest.com/dissertations-theses/phonemic-phonotactic-inference-early/docview/2235974900/se-2> (accessed October 4, 2024).
- Scott, J. H. G. (2019b). "Who follows the rules? Differential robustness of phonological principles," in *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference, ISSN 2380-9566*, eds. J. Levis, C. Nagle, and E. Todey (Ames, IA: Iowa State University), 213–225. Available at: <https://apling.engl.iastate.edu/conferences/pronunciation-in-second-language-learning-and-teaching-conference/psl1t-archive/> (access October 4, 2024).
- Scott, J. H. G., and Darcy, I. (2023). Prosodic location modulates listeners' perception of novel German sounds. *Lab. Phonol.* 14, 1–42. doi: 10.16995/labphon.6428
- Scott, J. H. G., Lim, R. Z. J., and Russell, C. B. (2022). "Sound-spelling correspondences in FL instruction: same script, different rules," in *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference, held June 2021 virtually at Brock University, St. Catharines, ON*, eds. J. Levis and A. Guskaroska. St. Catharines, ON. doi: 10.31274/psl1t.13361
- Seguí, J., Frauenfelder, U., and Hallé, P. (2001). "Phonotactic constraints shape speech perception: implications for sublexical and lexical processing," in *Language, Brain, and Cognitive Development: Essays in honor of Jacques Mehler*, ed. E. Dupoux (Cambridge, MA: MIT Press), 195–208.
- Seguí, J., and Frauenfelder, U. H. (1986). "The effect of lexical constraints on speech perception," in *Human Memory and Cognitive Abilities: Symposium in Memoriam Hermann Ebbinghaus*, eds. F. Klix and H. Hagendorf (Amsterdam, North-Holland), 795–808.
- Selkirk, E. (1984). "On the major class features and syllable theory," in *Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students*, eds. M. Aronoff, F. Kelley, and B. K. Stephens (Cambridge, MA: MIT Press), 107–136.
- Speeter Beddor, P., and Evans-Romaine, D. (1995). Acoustic-perceptual factors in phonological assimilations: a study of syllable-final nasals. *Rivista di Linguistica* 7, 145–174. Available at: <http://linguistica.sns.it/RdL/1995.html> (accessed October 4, 2024).
- Steinberg, J. (2014). *Phonotaktisches Wissen: Zur prä-attentativen Verarbeitung phonotaktischer Illegalität. [Phonotactic knowledge: On the pre-attentive processing of phonotactic illegality]*. Berlin: De Gruyter.
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2010a). Preattentive phonotactic processing as indexed by the Mismatch Negativity. *J. Cogn. Neurosci.* 22, 2174–2185. doi: 10.1162/jocn.2009.21408
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2010b). Activation and application of an obligatory phonotactic constraint in German during automatic speech processing is revealed by human event-related potentials. *Int. J. Psychophysiol.* 77, 13–20. doi: 10.1016/j.ijpsycho.2010.03.011
- Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2011). Phonotactic constraint violations in German grammar are detected in auditory speech processing: a human event-related potentials study. *Psychophysiology* 48, 1208–1216. doi: 10.1111/j.1469-8986.2011.01200.x
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *J. Phonet.* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001
- Streeter, L. A., and Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *J. Acoust. Soc. Am.* 65, 1533–1541. doi: 10.1121/1.382917
- Suzuki, Y. (2017). Validity of new measures of implicit knowledge: distinguishing implicit knowledge from automatized explicit knowledge. *Appl. Psycholinguist.* 38, 1229–1261. doi: 10.1017/S014271641700011X
- Swinney, D. A., and Prather, P. (1980). Phonemic identification in a phoneme monitoring experiment: the variable role of uncertainty about vowel contexts. *Percept. Psychophys.* 27, 104–110. doi: 10.3758/BF03204296
- Thomson, R. I. (2011). Computer assisted pronunciation training: targeting second language vowel perception improves pronunciation. *CALICO J.* 28, 744–765. doi: 10.11139/cj.28.3.744-765
- Thomson, R. I. (2018). *English Accent Coach [Computer Program]. Version 4.0.* Available at: <https://www.englishaccentcoach.com/> (accessed September 27, 2024).
- Valaczkai, L. (1998). *Atlas deutscher Sprechlaute: Instrumentalphonetische Untersuchung der Realisierung deutscher Phoneme als Sprechlaute [Atlas of German speech sounds: Instrumental phonetic examination of the realization of German phonemes as speech sounds]*. Vienna: Edition Praesens.
- Vance, T. J. (1987). *An Introduction to Japanese Phonology*. Albany, NY: SUNY.
- Wang, Y. (2023). Processing of English coda laterals in L2 listeners: an eye-tracking study. *Lang. Speech* 2023, 1–15. doi: 10.1177/00238309231203899
- Weber, A. (2001a). Help or hindrance: how violation of different assimilation rules affects spoken language processing. *Lang. Speech* 44, 95–118. doi: 10.1177/00238309010440010401
- Weber, A. (2001b). *Language-Specific Listening: The Case of Phonetic Sequences*. Wageningen: Ponsen & Looijen.
- Weber, A. (2002). Assimilation violation and spoken-language processing: a supplementary report. *Lang. Speech* 45, 37–46. doi: 10.1177/00238309020450010201
- Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Percept. Psychophys.* 35, 49–64. doi: 10.3758/BF03205924
- Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Percept. Psychophys.* 50, 351–360. doi: 10.3758/BF03212227
- Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon Press.
- Wiese, R. (2011). *Phonetik und Phonologie [Phonetics and Phonology]*. Paderborn: Wilhelm Fink.
- Winters, S. J. (2003). *Empirical Investigations Into the Perceptual and Articulatory Origins of Cross-Linguistic Asymmetries in Place Assimilation* (Dissertation). The Ohio State University, Columbus, OH, United States. Available at: <https://linguistics.osu.edu/sites/linguistics.osu.edu/files/dissertations/winters2003.pdf> (accessed October 4, 2024).