The impact of rhotic allophony on spoken-word recognition during sound change

Holger Mitterer

Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta

Hanyang Institute for Phonetics and Cognitive Science, Hanyang University, Seoul, South Korea

Eva Reinisch

Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria

Maltese, a Semitic language spoken on the Mediterranean island of Malta, is experiencing a rhotic sound change from an alveolar trill to an alveolar approximant. In word onset position, approximants and trills appear with a similar frequency, which enables the investigation of allophonic effects in the absence of a frequency bias. Given claims that the trill is the most prototypical exemplar of the rhotic category (e.g., Ladefoged & Maddieson, 1996, "Sounds of the world's languages"), we hypothesized that the trills lead to faster word recognition. This was tested in two visual-world eye-tracking experiments. The first used images and focused on word recognition in reasonably predictable sentences. The second experiment used printed words to also investigate competition effects, with the hypothesis that the phonetic similarity between the rhotic allophone and other segments modulates the amount of activation of similar words. Over the two experiments, an advantage for the trill variant was only found in the second experiment. Both studies also tested the prediction from gestural theories of speech perception that listeners should more easily recognize words that match their own productions. This was not observed.

Variation in the speech signal that spoken-word recognition has to deal with stems from various factors such as speech rate (Mitterer, 2018; Reinisch et al., 2011), prosody (Cho et al., 2017), casual versus formal speech styles (Ernestus, 2014), anatomic differences (Whiteside, 2001), regional dialects (Sumner & Samuel, 2009) and sociophonetic factors (Drager, 2011), to name but a few. Effects from regional dialects and sociophonetic factors are especially salient for rhotics (Lawson et al., 2011; Wiese, 2001), a class of phones that generally displays unusual amounts of variation (Ladefoged & Maddieson, 1996), often with a high number of allophones even within the same language (Sebregts, 2014). While some of this variation may be phonetically driven (e.g., devoiced trills in clusters with unvoiced obstruents), other types of variation (uvular trill versus alveolar approximant as a rhotic in Dutch) are extremely unlikely to arise as a consequence of coarticulation and hence must be intended by the respective speaker. Such differences are often related to a sound change in a given language.

In this paper, we investigate how allophonic variation in the production of rhotics impacts spoken-word recognition in Maltese, a language, in which, due to an ongoing sound change, there is (currently) no overall frequency bias towards one of the allophones. This allows us to investigate whether the trill allophone has an intrinsic advantage, a

Holger Mitterer, https://orcid.org/0000-0003-4318-0032

Eva Reinisch, https://orcid.org/0000-0002-1400-5473

We have no known conflict of interest to disclose. This work was supported by a Research Excellence Grant (REP-2023-010) to the first author by The Malta Council of Science and Technology.

Correspondence concerning this article should be addressed to Holger Mitterer, holger.mitterer@um.edu.mt.

hypothesis based on linguistic and experimental evidence (see below). The ongoing nature of the sound change also allows us to investigate whether there is a link between perception and production, so that participants perceive the variant better that they themselves prefer.

Rhotics frequently play a role in sound change, presenting an intriguing paradox: most documented shifts tend to move away from the alveolar trill, for instance in German (Schiller, 1998), Dutch (Velde & Hout, 1999), and Standard French (Haden, 1955) with similar changes ongoing in Canadian French (Sankoff et al., 2001). It is often assumed that the move away from the trill is caused by functional factors such as the difficulty in acquiring the trill (Schiller, 1998). Indeed, in an survey of language acquisition in 27 languages, the alveolar trill was one of the latest mastered segments (McLeod & Crowe, 2018), while the uvular trill seemed to be mastered much earlier (70 vs. 36 month, though the database included only one language, Brazilian Portuguese, that used a uvular variant). Interestingly, rhotic variation is one of the few issues in which there even is awareness of phonetic detail in collective consciousness; for instance with Russian having a pejorative term for children that do not manage to produce an alveolar trill and use a uvular instead (картавый [kɐrtavj]). This is especially noteworthy because listeners are usually explicitly unaware of the phonetic details of spoken language (Gessinger et al., 2021; Kemps et al., 2004), even if they might use them implicitly to identify the accent of the speaker.

Given these observations of recent or ongoing sound changes away from the alveolar trill to other rhotic phones in multiple languages, one might expect that the alveolar trill is a rare segment in the languages of the world. However, it is often assumed that the alveolar trill is the most common of the rhotics in the languages of the world (Chabot, 2019; Ladefoged & Maddieson, 1996). However, the claim that the alveolar trill is the most common may have to be taken with a grain of salt, since it remains an open question whether this claim stands up to scrutiny, or whether it is due to an inconsistency of using the symbol [r] whenever a language uses some form of rhotic segment. In an overview of rhotics in Illustrations of the International Phonetic Alphabet (IPA), Anselme et al. (2023, p. 1027) argue that: “We find that although the phoneme /r/ is not rare (105 illustrations do have an /r/), [r] seems to become less and less frequent as one approaches the phonetic reality (here, as captured to various degrees by the phonetic transcriptions present in some of the illustrations).”

While this may temper the enthusiasm over [r] being the canonical implementation of /r/, the question still remains why many languages still use the alveolar trill as the implementation of the rhotic category, despite the well-documented challenges in acquisition. If there is a disadvantage in production, this must be outweighed by other factors. The example of the pejorative term for speakers not achieving a trill in Russian suggests that there might be socio-phonetic reasons for this. However, another, maybe more likely possibility is that the difficulties in production are outweighed by an advantage in perception. We will pursue this explanation here, since there are multiple hints from disparate strands of research that suggest such an advantage. First of all, the auditory system is mostly sensitive to change (Kluender et al., 2003), and the amplitude modulations of a trill provide exactly that. In line with that, a study using the selective-adaptation paradigm suggested that trills have a special status in perception, with non-trill rhotics being more strongly adapted by trill adaptors than adaptors using the *same* allophone (Mitterer & Reinisch, 2023). This would mean that words should be recognized easier if produced with an alveolar trill compared to other rhotic phones. This is the first prediction to be tested in the present study.

The question how rhotic allophonic variation impacts spoken word recognition was already investigated by Llompart et al. (2021) for German. They measured the efficiency of word recognition in a visual-world paradigm (VWP) with slightly predictive sentences (e.g. *Der Kellner kommt mit der Rechnung*, Engl. ‘the waiter brings the bill’, see Figure 1 for an example of the current study). While hearing this sentence, participants saw three images on the screen representing a bill

("Rechnung", the target), a bowl of soup (a competitor, as another possible continuation of the sentence) and a blanket as a distractor. Participants were instructed to click on the image corresponding to the word mentioned in the sentence, and the main manipulation was whether the /r/-initial target word was produced with an alveolar trill (a variant common in Southern Germany where the study took place) or a uvular fricative (the most common realization of /r/ in German and considered the standard). The results showed that German native speakers recognized words faster when produced with a uvular fricative. Llompart et al. (2021) also tested Italian and French learners, with Italian learners (used to a trilled [r] in their native language) finding it easier to recognize words when produced with a trill, while French learners did not have a preference either way. The latter result was somewhat surprising if considered more closely. The French learners were likely used to a uvular fricative in their native French, just like the native German participants and (probably) were even unable to produce an alveolar trill. Nevertheless, they had no problem identifying /r/-initial words when produced with the alveolar trill, despite their presumably very limited experience with this allophone in German. This may indicate that the trill has an intrinsic advantage in perception that becomes apparent when testing second language learners who are less influenced by the L1 usage statistics of rhotic allophones than native listeners.

The results of Llompart et al. (2021) hence indicate a major obstacle in investigating the potentially privileged status of the alveolar trill as rhotic allophone: frequency biases. Frequency bias is a major determining factor in spoken-word recognition (Connine, 2004; Magnuson et al., 1999). A potential remedy to this problem is provided by a recent study on rhotics in Maltese (Mitterer et al., 2025), since it was found that there is roughly a 50/50 split of alveolar trills versus approximants in word onset position. Moreover, this variation is not solely attributable to differences between speakers; even within individual speakers, there is some inconsistency in the production of word-initial rhotics, with the same speaker potentially using both, trills and approximants. Nevertheless, speakers differ reliably in their production patterns. Older speakers are more likely to use trills, which is indicative of ongoing language change. Additionally, speakers who use English more frequently tend to produce more approximants, suggesting cross-talk between the phonetic systems. That is, despite some variability even within speakers, speakers still have reliable differences that can be linked to variables such as age and usage.

This provides an ideal testing ground for examining whether the alveolar trill holds a privileged status in perception. Maltese listeners are familiar with both the alveolar trill and approximant but they should not be subject to a frequency bias as the listeners in Llompart et al. (2021). In the overall language community, there is a roughly equal split between these variants and listeners will hence encounter both frequently. Moreover, there is a surprising amount of entropy in the production data of Mitterer et al. (2025). While syllable position has a major influence on rhotics—approximants dominate in the coda position, in the onset position, some speakers alternate between trills and approximants. That is, listeners are used to speakers that use different types of allophones for the onset position and “know” that the type of allophone is not predictable.

Given that the frequency of an allophone is a clear influence on how well listeners recognize it (Llompart et al., 2021), one may ask whether it is even useful to investigate an advantage for the alveolar trill in a speech community that is familiar with it. It might hence be more useful to investigate the potential advantage of the trill over other allophones in listeners unfamiliar with any of these allophones. Given the absence of such allophones in the mental lexicon of these listeners, this would naturally have to be a perceptual task, such as a discrimination or detection task. However, for an advantage to really count for a community, it would have to manifest in the very task that matters for communication: recognizing existing words in the mental lexicon. That is, if we want to explain why the alveolar trill is still relatively

popular in the languages of the world,[1] we need to test whether it provides an advantage in spoken-word recognition in a given speech community. Only this would provide a clear advantage for the use and maintenance of the alveolar trill within a speech community.

The distribution of rhotic allophones in Maltese allows to test whether and how allophonic detail influences lexical activation. In a similar study on sound change and spoken-word recognition, Soo and Babel (2025) found a perceptual advantage for one allophone. They investigated Cantonese, in which syllable-initial /n/ is produced more and more commonly as [l]. They find that speakers retain the sensitivity to distinguish these two sounds, but do not differentiate them for lexical processing; there is only a small effect of identity versus variant priming in immediate priming and no effect whatsoever in long-term priming. However, somewhat surprisingly, participants often fail to recognize words when produced with the "older" form, starting with [n]. This is even more surprising given that, in a corpus study, 18% of /n/ initial words were found to be produced with [n].

This raises the question how allophonic detail is represented in the spoken-word recognition system. Soo and Babel (2025) argue for a model where allophonic variation is filtered pre-lexically, so that both allophones map onto a phonemic representation before lexical access is achieved, aligning with the model of Bowers et al. (2016) and contrasting with the model of Mitterer et al. (2018), who argued that allophonic variation is represented on a lexical level[2].

It also important to note that the type of sound change investigated here differs from the case of Soo and Babel (2025). They look at a sound merger, in which a phonological contrast is lost. In our case, the change is an allophonic shift, that maintains all contrasts of the phonological system. As we will examine in the General Discussion, these different types of sound changes (merger vs shift) may lead to different results, especially with regard to the relation between perception and production.

The first experiment of the current study examined the question how rhotic variation, which is influenced by sound change, influences lexical access in Maltese. That is, we asked whether Maltese listeners would recognize /r/-initial target words faster if they are produced with an alveolar trill compared to an alveolar approximant. We focus on this variation—rather than on the difference between a trill and a tap—since producing an approximant rather than a trill is clearly "planned" articulatory variation, rather than a reduction phenomenon as reducing a trill to a tap, similar to reduction of stops to fricatives (Mitterer & Ernestus, 2006) or stops to flaps (Warner et al., 2009).

Following earlier examples (Eger et al., 2019; Llompart et al., 2021), we used a visual-world paradigm (VWP) with eye-tracking to investigate this issue. In this paradigm, participants listen to linguistic material that can vary in length, from brief phrases like "click on the beaker" (Allopenna et al., 1998) to extended discourses consisting of multiple sentences (Dahan et al., 2002; Tanenhaus et al., 1995). A simultaneously shown visual display contains an array of potential referents, one of which usually corresponds to the linguistic material that is presented. The referents can be pictures or written words (McQueen & Viebahn, 2007), and, in tasks focussing on spoken-word recognition, participants are usually asked to click on the matching referent. How fast fixations converge onto that target is taken as a measure of how easy or difficult word recognition was (Allopenna et al., 1998; Magnuson et al., 1999; for a review, see

[1] It is unsolved conundrum why the alveolar trill is so common in the languages of the world even though language change pre-dominantly moves away from the alveolar trill. Even stronger, we are not aware of a single sound change towards the alveolar trill. It is an open question how to reconcile these two contradictory observations.

[2] The data of Soo and Babel (2025) would be in line with both models, nevertheless. The situation in Cantonese would allow a new test of the model of Mitterer et al. (2018), which would predict that learning that a speaker does not produce a clear [n] would not generalize to [l] and that no selective adaptation can be observed between [n] and [l] over and above acoustic similarity and saliency effects (see also Mitterer & Reinisch, 2023).

Reinisch & Mitterer, 2022). The prediction for the first experiment of the present study hence is that participants' fixations should converge on /r/-initial target referents faster when the /r/ is produced as an alveolar trill than when the /r/ is produced as an alveolar approximant.

In the second experiment of the current study, the focus was not only on target recognition but also patterns of lexical competition depending on the properties of the input, that is, the rhotic allophone heard. We focused on /r/ in word-initial position, but—over the word boundary—in intervocalic position. In this intervocalic context, we can expect the /r/ to be produced as either an approximant or with a tapped variant of the trill, with a single alveolar contact and no stable open phase (Mitterer et al., 2025). Speakers, who tend to produce a word-initial trill, with the preceding word ending on a consonant, tend to produce a tap in intervocalic position, a pattern that is also attested in Spanish (Lewis, 2004). If an /r/ is produced as a tap, it resembles a /d/ (in fact, the tap has the same IPA symbol as a flap, which is an allophone of /d/ in American English). If the /r/ is produced as an approximant, it rather resembles an /l/. The question then is whether these difference in similarity between allophones of /r/ and other segments influence processing at the lexical level. If so, we might find activation of /d/-initial words when hearing an /r/-initial word produced with a tap, but if the /r/ is produced as an approximant, participants might co-activate /l/-initial words. This question pertains to the debate between cascaded processing and fixed-stage processing in cognitive science. While this debate seems to be more or less resolved towards the view of cascaded processing (Weber & Scharenborg, 2012), a recent finding suggests that phonetic information may sometimes not be used for word recognition immediately. Galle et al (2019) found that frication noise may not be used immediately in spoken-word recognition. Therefore, the question arises whether we can see effects of allophone similarity reflected at the lexical level. We are not aware of many papers that successfully show that phonetic similarity between segments influences lexical activation (see Mitterer, 2011; for an example). Experiment 2 of the current study examines this issue.

The allophony of /r/ also allows to investigate the relation between speech perception and speech production. It can be argued that gestural theories of speech perception, especially when they assume the involvement of the motor system in speech perception, make the prediction that participants should find it easier to perceive words when they matched their own production patterns. To test this, participants in both experiments first performed a short production task. In contrast to other observations on /r/ allophony (Sankoff et al., 2001; Sebregts, 2014), Maltese speakers tend to be variable in using approximants and trills, even within the same position (Mitterer et al., 2025). The production study in Mitterer et al. (2025) used a sentence memory task to avoid read-aloud speech, and for the present study we used a subset of items from this task. We recorded not only items with /r/ in word-initial position (matching those used in the perception experiments), but also /r/ in initial consonant clusters (the position most likely to elicit trills in Mitterer et al., 2025) and coda clusters (the position least likely to elicit trills). The inclusion of these items allowed us to differentiate between speakers at the floor and ceiling of trill likelihood to generate a predictor variable that should differentiate well how likely a given speaker is to use a trill overall. This overall likelihood of trill use was then related to the perception experiments to test whether listeners' own production preference influenced their lexical processing.

To summarize, we investigate three theoretical issues. First, whether the popularity of the alveolar trill in the phonological inventories of the languages of the world may be ascribed to a perceptual benefit of the trill over other rhotic sounds. Secondly, we test whether allophonic details influence lexical activation. Experiment 2 focuses on this question. Finally, we ask whether there is a relation between perception and production so that participants' production patterns influence how efficiently they can recognize words produced with one of the other allophone.

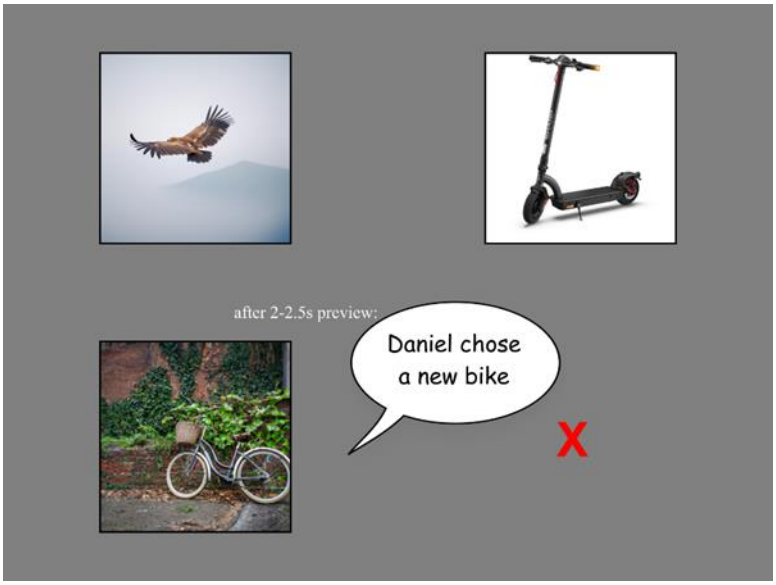

Figure 1. A schematic example of a trial with the translation of the Maltese stimulus sentence (*Daniel għażel rota ġdida*) with the bicycle as target, the scooter as competitor, and the eagle as distractor. Preview duration depended on the target onset in the stimulus sentence (see Methods for details).

Experiment 1

Experiment 1 set out to testing how allophonic variation impacts spoken word recognition. It investigates whether trills facilitate word recognition (our first question) and whether this potential advantage is moderated by the listeners own production patterns (our third question).

We measure efficiency of word recognition in a VWP. To focus on the perceptual advantage that a trill might convey for the recognition of /r/-initial words, it is most useful to use semantic competitors to the target in a VWP instead of a phonological competitor. Phonological cohort competitors (as 'beaker' vs. 'beetle' in Allopenna et al., 1998) necessarily have to be /r/-initial for /r/-initial targets, and if trills are perceptually beneficial, this benefit would help both the /r/-initial target and its cohort competitor. When using a competitor that carries some overlap (as the rhyme competitor 'speaker' to the target 'beaker', see Allopenna et al., 1998), the activation of the competitor may depend on the allophonic variation as well (as tested in Experiment 2). However, without a competitor, any effects of allophonic variation may be difficult to see in a VWP, because the target fixations quickly rise to ceiling. Adding a competitor may make effects more likely to be measurable in the fixation patterns. Therefore, we use a version of the VWP in which the target is somewhat predictable, but the visual display contains a semantic competitor that would also fit the target sentence. Figure 1 provides an example; the first words of the sentence "Daniel chose a new bicycle" make the picture of the bicycle and the scooter a more likely target than the picture of the eagle. This is a similar design to that used by Eger et al. (2019) and Llompart et al. (2021).

We aimed for 50 participants since this provides an acceptable power (>0.8) for finding a correlation of 0.4 between perception and production patterns, which we consider the smallest effect size of interest. Using a correlation here is appropriate, because production pattern is better conceptualized as a continuous variable rather than a categorical one, based on the results of the previous production study (Mitterer et al., 2025).

Method

Participants

This study was performed in accordance with the procedures of the Faculty Research Ethics Committee (FREC) of the Faculty of Media and Knowledge Sciences at the University of Malta (MAKS-2023-00131) and all participants provided informed consent before participation. Fifty-three participants

from the University of Malta student population took part in the experiment. The data of 51 participants could be fully used. For one participant, calibration of the eyetracker failed due to reflection from glasses and for one participant, the *gazeR* package (see below) found too much trackloss. The data of this participant was used for the analysis of the behavioral responses but excluded from the analysis of the eye-tracking data. The final sample size with eye-tracking data of 51 slightly exceeded the planned sample size, as additional slots were allocated to account for no-shows.

The final sample consisted of 53 participants aged 18–23 (mean = 19.26), with 36 identifying as females and 17 as males. Self-assessed language proficiency on a 1–4 scale was high in both English and Maltese, with a slight mean preference for English in the written modality (Spoken: 3.7 in English, 3.6 in Maltese; Written: 3.8 in English, 3.4 in Maltese), likely due to school and university assessments being primarily conducted in English[3].

Participants reported predominantly using Maltese in infancy (0–4 years; mean = 60%, range: 20–90%), with usage decreasing during childhood (mean = 54%, range: 10–90%) and stabilizing during adolescence (mean = 48%, range: 10–80%) and adulthood (mean = 52%, range: 10–90%).

Materials

For the production task, which was implemented in the SR Research Tool *ExperimentBuilder*, we used 20 items in total taken from the 128 items in the production study by Mitterer et al. (2025). Five items had /r/ in the coda position, five items in onset clusters and ten items had /r/ as a simple onset. Items were selected, first, according to their likelihood to provide valid data (that is, participant did not tend to forget the sentence) and then by providing a good range of trill likelihood within that category.

For each production-task trial, participants saw a picture with a sentence underneath, with the picture relating to the sentence. For example, for the sentence *Ganni was a rebel when he was young*, the image depicted an adolescent boy in a leather jacket with a defiant stance. Participants were asked to remember the sentence. The written sentence then disappeared while the picture remained visible, and at this point participants were asked to repeat the sentence. This procedure was chosen to ensure that the speech was not merely read aloud.

For the VWP task, 42 Maltese /r/-initial words were selected such that they were depictable. The number of items was limited by the availability of /r/-initial words in the Maltese lexicon that were deemed depictable and recognizable to most participants. For each word, a sentence was generated that made the target word plausible while allowing for other continuations (see Table A1 for a list of the sentences). Another plausible alternative continuation served as a competitor picture, and a third, unrelated picture was included as a distractor. Note that our main question—is an /r/-initial word easier to recognize when produced with a trill—is asked "within" items. Variation to the extent that one target item is more plausible in the sentence frame than another does not influence the effect of target allophone.

For example, for the word *rutella* (Engl. 'tape measure'), the sentence was *Il-ħaddiem nesa li kellu rutella żejda fil-van* (Engl. 'The worker forgot that he had a spare tape measure in the van'), with *spanner* as the competitor picture and *cherry* as the distractor (see Table A1 in the Appendix for the full list of items, including a gloss and a translation of the Maltese sentences). We also created 60 additional filler trials with a similar structure, where target words started with other phonemes than /r/.

The target words were recorded by a speaker capable of producing both trills and approximants (similar to the speaker in Llompart et al., 2021). The speaker was recorded using the Speechrecorder software (Draxler & Jänsch, 2004) using a Scarlett Focusrite external sound card and microphone. The speaker was instructed to speak casually as if speaking to a friend in a

[3] This might be surprising given that the lack of orthographic transparency in English leads to a slower reading acquisition (Aro & Wimmer, 2003). In fact, the Maltese orthography is relatively transparent, with the exception that vowel length is not always coded clearly.

pub without strong prosodic variation to enable the cross-splicing (see below). Despite this semi-professional set-up (often used for blogs), there was a 50Hz noise from the electrical grid. This is a common problem in Malta. Therefore, we used denoising in Praat (Boersma & Weenink, 2024), based on the initial silence in the recordings.

The filler sentences were recorded once, while the target sentences were recorded four times—twice with the instruction to produce a trill and twice with the instruction to produce an approximant. To avoid confounds in comprehensibility, the experimental items were cross-spliced. Since the speaker naturally tended to produce trills, there was a possibility that the surrounding words would be pronounced more carefully (due to increased focus on pronunciation) or less carefully (due to focus on the target word) when producing an approximant. To eliminate such potential confounds, we cross-spliced a trilled [r] and an approximant [ɹ] as well as their neighbouring segments into a single token of the carrier sentence (see Table 1).

The speaker was instructed to speak the sentences fluently without any breaks and with a broad focus (i.e., in reply to a question such as "What happened"). This aligns with the earlier production study, in which participants generally produced the sentences as one intonational phrase—unless they had trouble recalling a word—and with a broad focus. This means that there are no obvious cues that would allow the listener to predict whether the speaker would use a trill or an approximant.

From the two productions per target word with each the approximant or trill the following selection criteria were applied: For the approximants, the token with the longer stable consonant phase was selected. For the trills, the token that showed the stronger amplitude modulation was selected. This was based, first, on the number of tongue contacts (= amplitude drops in the signal) and then on the strength of these drops. If the critical word-initial /r/ was preceded by a consonant, which was the case for 29 stimuli, the speaker produced a trill with mostly two contacts (three cases contained three contacts). If the /r/ was preceded by a vowel, the speaker typically produced a tap (10 of 13 cases), but for all of these a full closure was achieved with usually a period of near silence of around 20ms. One of the other two sentences was then used as the carrier sentence, with the constraint that, in the whole set, half of the carrier phrases were taken from recordings with the instruction to use a trill and the other half from recordings with the instruction to use an approximant. Splicing points were identified that allowed for relatively smooth cross-splicing and contained the segments neighbouring the /r/, since those are typically coloured by coarticulation with the /r/ (see Table 1 for an example). The speaker generally maintained a stable f0, facilitating cross-splicing without audible artifacts. Cross-splicing was performed using Praat's "Concatenate with overlap" function (Boersma, 2001), with the overlap time set to 10ms if one end was unvoiced or to the duration of one glottal period (i.e., 1s/f0) at the splicing point when splicing during a voiced interval.

For the pictures, prompts were generated with Dall-E using batch API requests (the code for this method is available at: https://osf.io/apcyb/). If pictures were not fitting well, the prompt was changed to indicate angle or general context. Using DALL-E ensured a similar "look-and-feel" of all pictures.

Table 1. *Cross-splicing of the experimental items. The splicing points were the midpoint of the [l:] geminate and the closure for the /t/ in rutella.*

Soundfile	Content
[r] source	*Il-ħaddiem nesa li kellu rutella żejda*
[ɹ] source	Il-ħaddiem nesa li kellu rutella żejda
Carrier sentence	Il-ħaddiem nesa li kellu rutella żejda
cross-spliced [r] item	Il-ħaddiem nesa li kel*lu ru*tella żejda
cross-spliced [ɹ] item	Il-ħaddiem nesa li kellu rutella żejda

Procedure and Apparatus

Participants first signed a consent form and then completed the production task with 20 items, which took approximately five minutes. After the production task, the Scarlett Focusrite microphone used for recording was removed and a chin rest was attached to the table where the participants were seated. Participants were then presented with instructions about the VWP on the experiment monitor. The instructions explained that three pictures would appear on the screen, with the lower right quadrant left empty. Participants were instructed to click on one of the pictures if it matched a word in the sentence or to click on the empty lower right quadrant if none of the pictures matched. Using a fixed "empty" quadrant simplified the randomization procedure, as there were now only three rather than four possible target positions. With 21 trials per participant in each condition (i.e., 42 experimental items and two conditions), the target could be presented seven times in each target position for each combination of participant and condition. After the instruction, the eye tracker (SR Research EyeLink 1000, desktop mount) was calibrated using a 9-point calibration, and the procedure was prepared using SR Experiment Builder software.

After the calibration, the VWP experiment started. Trials orders were randomized differently for each participant. On each trial, before the sentence started, participants could preview the pictures for 2s or 2.5s, with the longer preview if the target occurred within 1s of sentence onset. This was done to ensure that participants had time to scan the display before hearing the target word. Each order started with 5 practice trials. For each order, 42 fillers were randomly selected to be used after one of the 42 experimental items. Each participant was presented with 21 experimental items with a trill and 21 with an approximant, and items were presented equally often in both conditions over participants. The randomization procedure also made sure that the target appeared 7 times each in each of the three possible screen locations.

Preprocessing and analysis strategy

All data processing was done in R (version 4.3.3; R Core Team, 2024) and all code from raw data to final analysis is available on OSF (https://osf.io/wvns3/). The data were pre-processed with the package *gazeR* (Geller et al., 2020) with the modifications that some non-critical columns in the data files were deleted during processing, since the long data format of *gazeR* otherwise required too much memory.

All analyses made use of (generalized) linear-mixed effects models with participants and items as random effects using the lmerTest package (Kuznetsova et al., 2015) with the non-default *bobyqa* optimizer. Modelling started with a maximal-random effects structure. In case of convergence issues, first, correlations between random effects were removed and then the random effect with the smallest standard deviation was removed iteratively until convergence was achieved. Two fixed factors were used: The *Stimulus /r/* predictor was contrast coded with trill mapped onto 0.5 and the approximant mapped onto -0.5. As a mnemonic, this reflects the assumption that the trill is the "better" allophone of the rhotic category. The predictor for the production pattern of the participants was operationalized as the (scaled) likelihood of a trill produced by the listener and will therefore be called *trill preference*. A facilitatory effect of a match between production and perception should hence lead to a positive regression weight between the predictors Stimulus /r/ and Trill preference in production.

Given the range of use of Maltese in our sample, we also generated a scaled variable for the use of Maltese in infancy and childhood (since this period is critical for phonetic development, see Flege et al., 2006). This was used in the analysis whether trills have a perceptual advantage and whether the advantage is moderated by participants' language profiles.

Holger Mitterer, https://orcid.org/0000-0003-4318-0032

Eva Reinisch, https://orcid.org/0000-0002-1400-5473

We have no known conflict of interest to disclose. This work was supported by a Research Excellence Grant (REP-2023-010) to the first author by The Malta Council of Science and Technology.

Correspondence concerning this article should be addressed to Holger Mitterer, holger.mitterer@um.edu.mt.

For the eye-tracking data, the dependent variable was a target-preference measure based on a time window starting at 200ms (when the signal starts to influence the eye-tracking record) and ended when the fixation functions reached an asymptote. In this time-window, we determined the difference (in empirical logits) between target and competitor fixation proportions.

Transparency and Openness

Citation of Data, Materials, and Code: All data, materials, and analysis code associated with this study are appropriately cited in the manuscript and are available at https://osf.io/wvns3/ for the current study and https://osf.io/apcyb/ for the code to make batch API requests for image generation.

Data, Materials, and Code Availability: The data, experimental stimuli, and analysis scripts used in this study are publicly available at https://osf.io/wvns3/. Access is unrestricted under CC BY 4.0.

Reporting Standards: We followed the Journal Article Reporting Standards (JARS) for quantitative data (see https://apastyle.apa.org/jars/quant-table-1.pdf)

Pre-registration: The study was not pre-registered. We consider the hypothesis to be self-evident based on the existing literature.

Results

Production data

The production data were first forced-aligned using the WEBMaus algorithm provided by the LMU Munich (Strunk et al., 2014). Then, the critical rhotic was categorized binarily as being either containing an amplitude minimum (i.e., a trill or tap; here summarized as "trill") or as another instantiation. Note that in the onset, other variants than trill/tap and approximant are rare. The results replicate the general finding from the previous production study (Mitterer et al., 2025) with most trills in onset clusters (84.6%), fewer trills in the simple onset (50.3%) and highly unlikely in coda clusters (8.3%). From these data, we calculated a trill likelihood for each participant that capitalized on the trill likelihood in the simple onsets (participant_trill = ½ *%trillOnsets + ¼ * (% trillOnsetClusters + %trillCodaClusters). This variable had roughly a normal distribution with a mean of 48.3%, a median of 50% and a range from 0 to 85%. It was used as a predictor in the analysis of perception data, using the z-transformation of the original data, so that the intercept reflects the grand mean. While the analyses will make use of the continuous data, we will use median splits to illustrate the impact of this variable.

The preference for trills in production correlated moderately with the use of Maltese during infancy and childhood, though not significantly ($r(51) = 0.24$, $p = 0.083$). This is line with findings from our earlier production study (Mitterer et al., 2025). However, given the relatively small correlation, there is not issue with collinearity when using both variables in the same regression analysis.

Click Responses

Participants clicked on the intended target in about 90% of the cases. Most of the remaining clicks when to the "target absent" object in the lower right quadrant (6.8%) or the competitor (3.7%). Both participants above and below the median split of trill usage in production gave slightly more correct responses when they heard a target in which the /r/ of the stimulus word was produced as an approximant (trill-leaning participants: 90.1% vs. 88.0%, approximant-leaning participants: 90.4% vs. 87.9%). A generalized linear mixed-effects model, however, revealed no significant differences caused either by the Stimulus /r/ (i.e., which /r/ allophone was heard, b = -0.29 (0.19), z = -0.976, p = .333), the trill-preference variable (b = 0.01 (0.13), z = 0.04, p = .965), or the interaction between these two variables (b = -0.19 (0.18), z = -1.03, p = .303). Similarly, neither was there a main effect of early Maltese usage (b = -0.18 (0.28), z = 0.66, p = .511) nor its interaction with Stimulus /r/ (b = -0.53 (0.39), z = -1.37, p = .170).

For reaction times, we see a similar pattern, with slower responses to trills than to approximants for both participants with a trill-preference in production (2114ms for heard trills vs. 2075ms for approximants) and an approximant-preference in production (2160ms vs. 2073ms). However, as for the accuracy measures, the linear mixed-effects model revealed no significant differences caused either by the Stimulus /r/ (b = -16 (30),

t(49) = -0.55, p = .588), the trill preference (b = 40 (54), t(49) = 0.75, p = .458) or the interaction between these variables (b = 32 (30), t(46) = 1.051, p = .299). As for accuracy, there was no significant main effect of early Maltese usage (b = -86 (107), t(49) = -0.80, p = .428) nor a significant interaction with Stimulus /r/ (b = -104 (64), t(48) = -1.63 p = .110).

Eye-tracking data

Figure 2 shows the fixation proportions from just before target onset until 1500ms after target onset. The data indicate that participants were using the sentence context to predict the target picture, since there are more looks to the target and (semantically matching) competitor picture than to the distractor already around target onset. The data also show that the target fixations start reaching their asymptote around 800ms after target onset. We therefore decided on using target preference in the time window 200-800ms as the dependent variable.

Figure 2 shows no clear effect of Stimulus /r/ but a small preference in each participant group for their preferred allophone, that is, trill-leaning participants looked more at the target when the stimulus contained a trill while approximant-leaning participants looked at the target slightly more often when the stimulus contained an approximant. These preferences are small and not consistently observed in the 200-800ms window as is evident from the overlapping lines in that time window in Figure 2. It is also reflected in the statistical analysis, which reveals no significant effects of any predictor (Stimulus /r/: b = -0.05 (0.2), t(2010) = -0.287, p = .774; trill-preference: b = 0.04 (0.13), t(36) = 0.32, p = .747; interaction: b = 0.24 (0.22), t(837) = 1.11, p = .267). Similarly, neither the main effect of early Maltese usage (b = -0.12 (0.24), t(47) = -0.489, p = .627) nor its interaction with Stimulus /r/ (b = 0.35 (0.44), t(2011) = 0.80, p = .427) were significant.

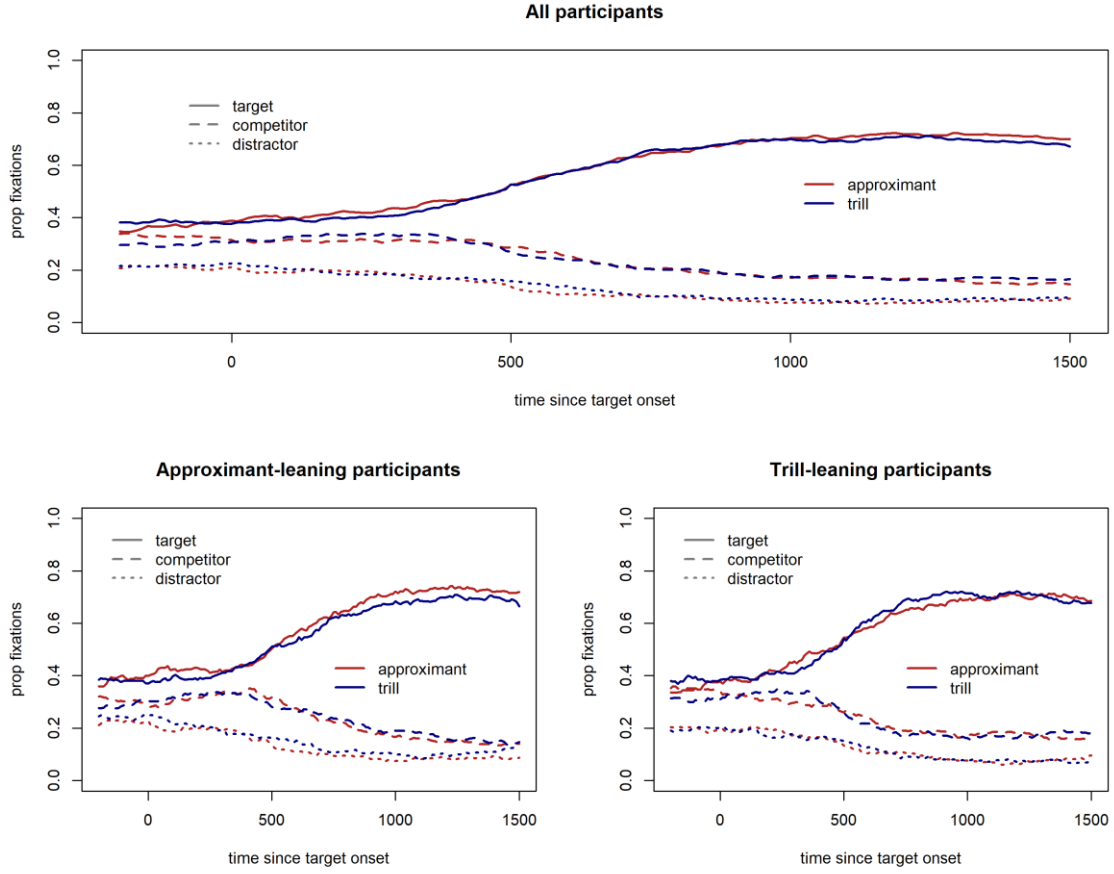

Figure 2. Fixation proportions depending on the allophone in the stimulus (line type) as well as the production pattern of the participants (different panels).

Discussion

The experiment examined whether rhotic allophonic variation influences spoken word recognition, with two hypotheses. The first hypothesis was that trilled (or tapped) versions of the rhotic with strong amplitude modulation may lead to more efficient word recognition. The second hypothesis was that participants may be faster if a stimulus aligns with their own production patterns. The results failed to support either of these hypotheses. The failure to find an effect of Stimulus /r/ on word recognition is in line with the results of Llompart et al. (2021), who reported an effect of allophone frequency on recognition efficiency in a visual-world paradigm. In the current case, the data show overall similar production frequencies for these two allophones in the onset position in Maltese, so that no difference is expected to arise for perception in Maltese on the basis of frequency alone.

This obviously leads to the issue that null-effects are difficult to interpret. However, with 42 items and 50 participants, our data set is relatively large for an eye-tracking study. This is reflected in Figure 2, in which the upper panel showing the data for all participants (independent of their production pattern) shows a near complete overlap of the fixation curves for approximants and trills as stimuli, indicating that there is little noise left in the aggregate data. Moreover, if phonological effects arise in a VWP paradigm, they are often around or larger than one logit unit (Mitterer et al., 2019, one logit unit is around a 20% difference around 50% and a 5% difference around 90%), and the standard error in the current experiment for the eye-tracking analysis was around 0.2 logit units. That is, effects of around 0.4 logit units would have been significant. This indicates that, if an effect of allophone exists, it may be smaller than the typical phonological effects found in VWP paradigms. As such, the current experiment seems to indicate that there is no huge perceptual advantage for the alveolar trill in spoken word recognition.

It may be surprising that the usage of Maltese during infancy and childhood did not influence the results. This is less surprising given the Maltese linguistic landscape, in which Maltese is basically unavoidable in certain situations (church, political gatherings, the football ground). Apparently, this small amount of exposure is sufficient to get "fluent" in perception and accept the trill as a possible variant of the rhotic in Maltese.

The situation is less clear for the relation between production and perception. Here, the interaction is in the expected direction, with trill-leaning speakers preferring trilled variants in perception and vice versa for approximant-leaning speakers. However, the interaction was not significant. Therefore, the current experiment does not allow us to draw strong conclusions.

Experiment 2

As previewed in the introduction, one of our questions is whether allophonic detail affects lexical activation. This is the main question investigated in this experiment. If an /r/ is produced with an amplitude dip as in a tap, it resembles a /d/, so we might find activation of /d/-initial words when hearing an /r/-initial word produced with a tap. Note that the tap is arguable more similar to a /d/ than a full trill with multiple contacts. Therefore, we used a phrase in which the /r/-initial words were preceded by a word that ended on a vowel (*kelma* Engl, 'word'), which is likely to give rise to a tapped version of /r/ (see the Materials section of Experiment 1). If the /r/ is, however, produced as an approximant, /l/ is phonetically closer, and participants might co-activate /l/-initial words, while the lack of an amplitude dip would rule out /d/-initial candidate words.

Clearly, the similarity between speech sounds is influenced by the phonological status in a given language (Cutler, 2012). The phonological relations here are, however, the same across conditions. Independent of whether the /r/ is produced as approximant or tap, both /l/ and /d/ are always phonologically contrasting with the phone in the target utterance. Note that we are not investigating whether tap and approximant are perceptually more similar to each other due their allophonic status.

The results of Experiment 1 could be read as meaning that allophonic variation does not matter at the lexical stage (but see Llompart et al., 2021). Competition patterns provide

another way to test whether allophonic differences are first resolved on a pre-lexical level (as potentially for unvoiced fricatives, see Galle et al., 2019) or whether there is cascaded processing from the phonetic to the lexical level.

To maximize the visibility of competition effects, we make use of a target-absent version of the VWP. That is, on some trials, participants hear an /r/-initial word that is not represented on the screen, which maximises the likelihood of observing competition effects as fixations to other referents that do not fully match the spoken target (Huettig et al., 2011). We will call these items that have the closest match to the spoken word of all three referents on the screen pseudo-targets. The instructions explicitly mentioned that there would be target-absent trials. Participants were informed that the screen will contain three printed words and a large red X in the lower right quadrant of the screen, which should be clicked on if none of the three printed words on the screen matches the spoken target.

Since it was not possible to find sufficient items in which there was both an /l/- and /d/-initial word was similar to an /r/-initial word, we selected separate /r/-/l/ and /r/-/d/ word pairs (see below for details). That is, the experiment had a two-by-two design with the first factor being the initial segment (either /d/ or /l/) of pseudo-targets similar to the presented /r/-initial word and the second factor being whether the /r/-initial word was heard with an approximant or a tap. For these trials, the screen hence contained three words, and only one of them, the /d/- or /l/-initial pseudo-target bore any similarity to the spoken target. Since the spoken /r/-initial target word and the /d/- or /l/-initials written pseudo-targets were not obviously semantically related, we also changed the type of the carrier phrase, since it was impossible to use predictive sentences that would match both the actual spoken words and pseudo-targets. We therefore decided to use the fixed carrier phrase *Fittex il-kelma* TARGET (Engl. “Look for the word TARGET”), similar to other studies that looked at phonological competition (e.g., Allopenna et al., 1998; McQueen & Viebahn, 2007; Mitterer & McQueen, 2009; Reinisch & Sjerps, 2013).

Finaly, because the main experimental trials always involve an /r/-initial spoken target that is not present on the screen, participants might learn to associate these targets with target-absent trials. Therefore, we also used trials in which an /r/-initial word was the target; that is, the written counterpart of the target was on the screen to be clicked on. This provides us with the possibility for another test of the other two research questions, whether or not trills/taps have an advantage over approximants in target recognition and whether the listeners’ own production pattern influences which allophone more efficiently activates words in perception.

Method

Participants

The experiment was performed in accordance with the procedures of the Faculty Research Ethics Committee (FREC) of the Faculty of Media and Knowledge Sciences at the University of Malta (MAKS-2023-00131) and all participants provided informed consent before participation. Thirty-six participants from the University of Malta student population took part in the experiment, none of which had taken part in Experiment 1. The sample consisted of 36 participants aged 18–24 (mean = 19.19, 22 female/14 male). Self-assessed language proficiency on a 1–4 scale was high in both English and Maltese, with a slight preference for English in the written modality (Spoken: 3.5 in English, 3.7 in Maltese; Written: 3.9 in English, 3.5 in Maltese), again likely due to school assessments being primarily conducted in English.

Participants reported predominantly using Maltese in infancy (0–4 years; mean = 59%, range: 20–90%), with usage decreasing during childhood (mean = 57%, range: 10–90%) and stabilizing during adolescence (mean = 44%, range: 10–80%) and adulthood (mean = 55%, range: 10–90%). The sample size was based on prior findings that phonological effects reliably arise in eye-tracking when more than 30 participants are tested with more than 48 items (see below) in two conditions (see the discussion of Exp. 3 in Mitterer et al., 2019).

Materials

For the main experimental trials, 24 word-pairs were selected which overlapped for the first few segments after an initial difference between the initial phonemes of the auditory target and visual pseudo-targets (i.e., /r/-/d/ and /r/-/l/; where 24 pairs were selected for each type of pairs, see Tables A2 and A3). For instance, a pair with /r/ and /d/ was *riżultat* (Engl. 'result') and *diżunur* (Engl. 'disgrace'), with a three-phoneme overlap [(d/r)ɪzʊ]); a pair with /r/ and /l/ was *raqqad* (Engl. 'to make someone sleep') and *laqqam* (Engl. 'to vaccinate'), also with a three-phoneme overlap [(r/l)laʔ:a]). Across the 24 pairs, the stress pattern was the same in 71% of the /d/-/r/ pairs and 75% of the /l/-/r/ pairs. Both sets contained four minimal pairs, and the average phoneme overlap was similar: 2.7 phonemes for /d/-/r/ pairs and 2.9 for /l/-/r/ pairs. For each word pair, we identified two additional words that were not obviously phonologically or semantically related to the auditory target or written pseudo-target word. Note that slight imbalances in overlap between the two types of pairs are not critical for the main prediction that the activation of similar words depends on the form of the allophone in the input, since that hypothesis was tested within items.

Additionally, 48 trials were generated in half of which /l/- and /d/-initial words were the target and, as described above, 38 trials in which an /r/-initial word was the target (see Table A4). There were an additional 60 target-present filler trials in which the neither the target (i.e., spoken word and printed word) nor the two additional printed words on the screen started with /d/, /r/, or /l/. This use of /d/-, /l/-, and /r/-initial trials with a matching printed-word targets ensured that neither the presence of /l/- or /d/-initial printed words on the screen nor the use of an /r/-initial spoken word indicated a target-absent trial. These filler trials contained no phonologically similar competitors. This is similar to the experimental trials, in which the participants heard an /r/-initial word, which is not among the three words on the screen. Instead, there was the /d/- or /l/-initial pseudo-target with some phonological overlap with the /r/-initial stimulus word. During the filler trials, these /d/- or /l/-initial targets matching the spoken word appeared without phonological competitors on the screen, ensuring that the filler trials with /d/- and /l/-initial targets were not distinguishable from target-absent trials with /d/- or /l/-initial pseudo-targets coupled with an /r/-initial spoken word.

The target phrases (*fittex il-kelma…*, Engl. 'Look for the word ….', with the empty slot taken up by the target words) were recorded by the same speaker who recorded the items for Experiment 1. Each /r/-initial target was produced four times, two times with a tap and two times with an approximant. The preceding context 'il-kelma' ensured that the word-initial /r/ was intervocalic, which usually triggers a tap-variant that maximised the similarity to /d/. If participants would have heard a trill with two contacts, the presence of two amplitude minima would mismatch the amplitude envelope associated with a /d/, which only contains one amplitude minima for the closure. Figure 3 presents the critical part of the junction "il-kelma CV_{Target}" for the targets *rata* Engl. 'installment' with an approximant (Panel A), a tap (Panel B), the stimulus *lazz* (Panel C) and the stimulus *data* (Panel D). In terms of amplitude envelope, there is an obvious similarity between the approximant /r/ (Panel A) and the /l/ (Panel C) and the tap /r/ (Panel C) and the /d/ (Panel D). Note how the use of a trill with two amplitude minima would have increased the difference between the /d/ and the /r/ stimulus, which motivated the choice for a carrier phrase which made the word-initial /r/ intervocalic over the word boundary.

All /r/-initial words were recorded twice and then the recording with the clearer amplitude dip (for the taps) and the smaller amplitude dip (for the approximants) was used. Given that the sentence frame was constant, we used the unedited recordings after noise reduction (using denoising in PRAAT based on the initial silence in each recording) and annotated the onset of the /r/ in these recordings in order to normalize the fixation pattern to the onset of the target words. Figure 3 also shows how these onsets were determined. For /d/ and tapped /r/, the start of the amplitude dip was used, and for /l/ and approximant /r/, the onset of the stable phase of the consonant was used.

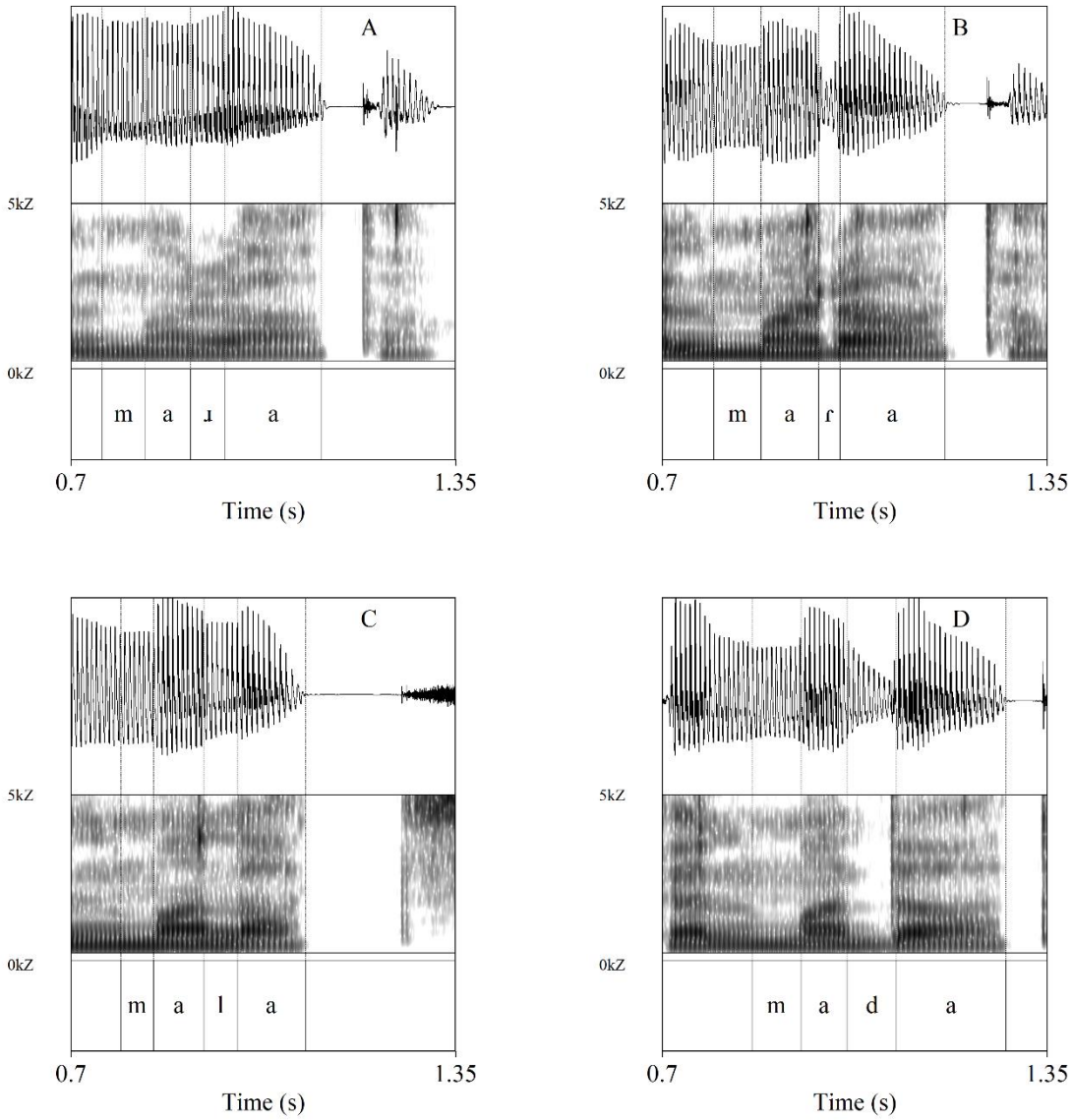

Figure 3. The critical part of the carrier phrase (the second syllable of the word kelma) and the onset of the spoken target for an initial approximant /r/ (i.e.,[ɹ], see Panel A), an initial tapped /r/ (i.e., [ɾ], see Panel B), an initial /l/ (Panel C) and an initial /d/ (Panel D).

Procedure, apparatus, pre-processing and analysis strategy

Procedure, apparatus, pre-processing and analysis strategy were similar as in Experiment 1. Participants filled in an informed consent and a language background questionnaire, finished the short production task that was also used in Experiment 1 and then proceeded to the VWP task. The instructions of the eye-tracking task mentioned that the display will contain a large red X in the lower-right quadrant of the screen, which should be clicked on if the spoken word did not match any of the written words.

For the VWP task, trial orders were randomized individually for each participant. The randomization balanced how often /d/- and /l/-initial pseudo-targets on the target-absent trials were presented with an approximant or tap across participants. For participants, the randomization also paired half of the 24 /d/- and /l/-initial pseudo-targets with an approximant /r/ and half with a tapped /r/, so that each participant completed twelve trials each for each condition arising from crossing the predictors Stimulus /r/ (tap vs. approximant) and Initial Segment of the pseudo-target (/d/ vs. /l/). On each trial,

participants saw the targets for 1.6s before the stimulus sentence started.

From the target-present trials, five were used as initial practice trials. Each experimental target-absent trial was followed by at least two target-present trials. These target-present trials were "set apart" at the start of the randomization routine. The randomization procedure first randomized the order of the target-absent trials and then added two of the reserved target-present trials. After this order was generated, the remaining target-present trials were inserted at random positions within the existing order. For each participant, within each condition, the target appeared four times in each of the three available positions (upper-left, upper-right, and lower-left quadrants) with the lower-right quadrant being reserved for the red X to be clicked when there was no match.

For the analysis of the eye-tracking data, the dependent variable—given that there was no competitor—is the empirical-logit transform of the proportion of target (or pseudo-target in target-absent trials) fixations in a time window that started at 200ms after target onset and ended when the fixation curves reach an asymptote. This was done separately for target-present and target-absent trials. The categorical predictors were contrast coded. For the initial segment of the pseudo-target, /d/ was mapped on 0.5 and /l/ on -0.5. Stimulus /r/ was coded as in Experiment 1 (tap = .5, approximant = -0.5). With this coding, the hypothesis of a modulation of lexical competition by allophonic properties predicts a positive regression weight for the interaction (more looks to /d/-initial pseudo-targets when the Stimulus /r/ is a tap and less if it is an approximant, and vice versa for for /l/-initial pseudo-targets).

Transparency and Openness

Citation of Data, Materials, and Code: All data, materials, and analysis code associated with this study are appropriately cited in the manuscript and are available at https://osf.io/wvns3/ for the current study and https://osf.io/apcyb/ for the code to make batch API requests for image generation.

Data, Materials, and Code Availability: The data, experimental stimuli, and analysis scripts used in this study are publicly available at https://osf.io/wvns3/. Access is unrestricted under CC BY 4.0.

Reporting Standards: We followed the Journal Article Reporting Standards (JARS) for quantitative data (see https://apastyle.apa.org/jars/quant-table-1.pdf)

Pre-registration: The study was not pre-registered. We consider the hypothesis to be self-evident based on the existing literature.

Results

Production data

The production data were processed as in Experiment 1 (forced alignment followed by manual classification of allophones). The results were similar as in Experiment 1, with trills or taps being rare in coda clusters (10.0%), common in onset clusters (75.0%), and accounting for nearly half of the productions in singleton onsets (43.6%). As in Experiment 1, we generated a measure for how likely each participant was to produce trills or taps over all positions; this measure ranged from 10.0% to 85.0%, with a mean of 42.5% and a median of 43.0%.

Competition effects

For this analysis, we begin by including all trials in which the participant indicated that none of the printed words matched the spoken word by clicking on the red X, as well as trials in which they clicked on the printed word most similar to the spoken word (i.e., the pseudo-target). Only for the reaction time analysis, trials were removed in which the participants did not correctly reject the pseudo-target word on the screen, since correct rejection and a false alarm are categorically different responses.

Looking at where participants clicked on the target-absent trials, participants mostly correctly rejected a match between the spoken word and the printed referents. If errors occurred, they were due to clicks on /l/-initial pseudo-targets (in 36 out of 853 trials, i.e., 4.2%) with only one click on an /d/-initial pseudo-target. Since one of the cells of the design (approximant stimulus for a /d/-target) had a zero value, we randomly recoded one trial to be an error to prevent a perfect separation in the generalized mixed-effects

logistic regression model. With one case recoded, the model converged but indicated no significant effects (Initial Segment of pseudo-target, /d/ vs /l/: b = 2.43 (1.8), z = 1.35, p = .18, Stimulus /r/: b = -0.05 (1.77), z = -0.03, p = .977, or an interaction between these two factors (b = -0.63 (2.12), z = -0.3, p = .77).

Looking at the reaction times for correct rejections, responses were faster when a /d/-initial pseudo-target had to be rejected (1581ms when the stimulus contained an approximant /r/ and 1582 when the stimulus contained a tap) than when an /l/-initial pseudo-target had to be rejected (1619ms when the stimulus contained an approximant /r/ and 1627ms when the stimulus contained a tap). However, these differences were not significant in a linear mixed-effects model (Initial Segment of pseudo-target: b = -39.67(40.14), t(46) = -0.99, p = .328, Stimulus /r/: b = -0.05 (1.77), z = -0.03, p = .977, interaction (b = -0.63 (2.12), z = -0.3, p = .77).

Figure 4 shows the eye-tracking data. There were more looks to the pseudo-targets than the distractors, with a slight modulation of the competition effects. The /d/-initial pseudo-targets received more looks when the stimulus contained a tap and the /l/-initial pseudo-targets received more looks when the stimulus contained an approximant, as predicted based on the assumption that phonetic detail influences lexical activations. For the statistical analysis, we used the mean fixation proportions to the pseudo-targets in a time window from 200ms to 1200ms post-stimulus, given that the looks to the empty target area reach an asymptote around 1200ms[4].

This dependent variable was used in a linear mixed-effects model with the predictors Stimulus /r/, Initial Segment of pseudo-target and their interaction. While the main effects of Stimulus /r/ (b = 0.06(0.07), t(35) = 0.83, p = .415) and Initial Segment of the pseudo-target (b = -0.14(0.19), t(56) = -0.76, p = .453) were not significant, there was a significant interaction between these two factors in the predicted positive direction (b = 0.32(0.14), t(1554) = 2.26, p = .024).

We then tested the competition effects separately by running an analysis with the factor allophone for both /l/ and /d/ targets. Although these simple effects tests at each level of target did not reach conventional significance levels (/l/-targets: b = -0.10 (0.10), t(793) = -0.99, p = .321; /d/-targets: b = 0.22 (0.12), t(34) = 1.82, p = .077), the crossover pattern in the descriptive data aligned with our hypothesis: /d/ targets would receive more looks with taps in the input and /l/-targets more looks with approximants in the input. The significant interaction indicates that the effect of competition differed reliably across levels of target type.

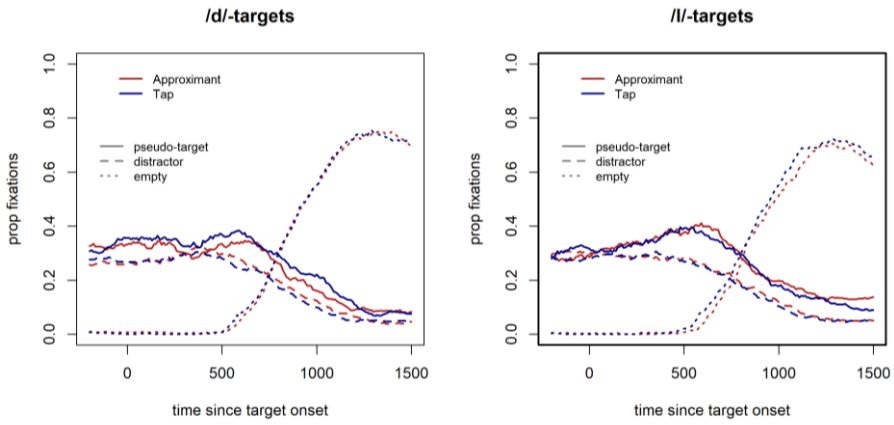

Figure 4. Looks to the /d/-initial and /l/-initial pseudo-targets, distractors, and the empty field in the experimental trials of Experiment when /r/-initial words are presented as taps or approximants.

[4] Note that this is a different time window than in Experiment 1; however, it follows from our strategy to use a time window from 200ms till the target region reaches an asymptote.

This indicates that the competition effects were modulated by the allophonic properties of the stimulus /r/, with stronger activation of /d/-initial words if the stimulus was a tap and stronger activation of /l/-initial words if the stimulus was an approximant.

Allophonic effects on word recognition

The trials with /r/-initial targets (i.e., printed /r/-initial words matching the spoken word) were also analysed, testing for an influence of the allophonic differences and the production pattern of the listeners, similar to Experiment 1. Since there were only 5 errors in click responses among the 1358 trials, error rate was deemed too low to allow for an analysis of error proportions.

Reaction times were slightly longer when the stimulus was an approximant, and this was the case for both approximant-leaning participants (approximant: 1348ms vs. tap: 1260ms) and trill-leaning participants (approximant: 1183ms vs. tap 1142ms). A linear-mixed effects model found this to be a significant effect of Stimulus /r/ (b = -50.33 (17.73), t(1307) = -2.83, p = .004), while the effect of participants' production pattern (b = -35.01 (41.33), t(33) = --0.84, p = .404) and the interaction of the two factors (b = 14.72 (18.48), t(1313) = 0.796, p = .425) were not significant. Similarly, neither the main effect of Maltese usage (b = -93.56, (83.55), t(34) = -1.12, p = .271) nor its interaction with Allophone (b = 22.21 (37.12), t(1313) = 0.589) were significant.

Figure 5 shows the eye-tracking data for the same two variables, Stimulus /r/ and production pattern. The results show more fixations to the target when the Stimulus /r/ was a tap than when it was an approximant. As for the reaction-time analysis, this effect is, if anything, somewhat larger for participants who prefer to use an approximant themselves. The figure also shows that the fixation curves asymptote around 700ms. Therefore, the empirical-logit transform of target fixations in a time-window from 200-700ms after target onset was used as the dependent variable in the eye-tracking analysis.

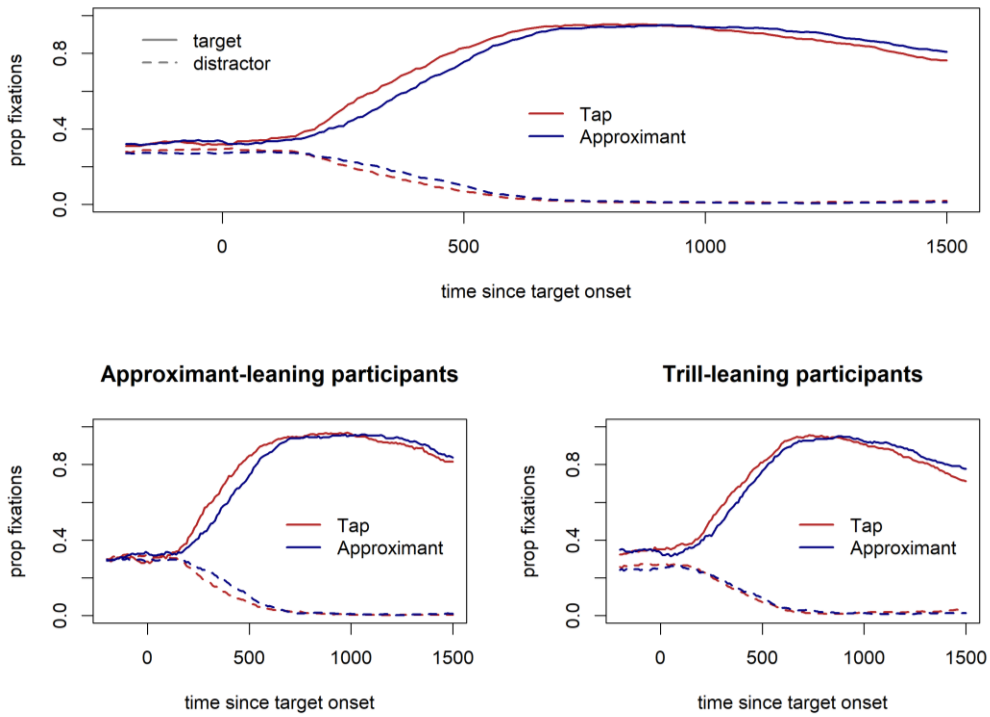

Figure 5. Proportion of fixations to the target and unrelated distractors on trials with an /r/-initial target, depending on the Stimulus /r/ as well as the production pattern of the listener.

As for the reaction-time analysis, this effect is, if anything, somewhat larger for participants who prefer to use an approximant themselves. The figure also shows that the fixation curves asymptote around 700ms. Therefore, the empirical-logit transform of target fixations in a time-window from 200-700ms after target onset was used as the dependent variable in the eye-tracking analysis.

A linear mixed-effects model showed that the target preference for words produced with a tap was significant in the fixation proportions measure, with a significant effect for Stimulus /r/ (b = 0.44 (0.13), t(33) = 3.25, p = .003) while neither the effect of participants' production pattern (b = -0.09 (0.11), t(34) = -0.85, p = .403) nor the interaction of the two factors (b = -0.09 (0.13), t(33) = -0.67, p = .510) were significant. Similarly, neither the main effect of Maltese usage (b = 0.11 (0.23), t(39) = 0.471, p = .640) nor its interaction with Allophone (b = 0.006 (0.282), t(33) = 0.24, p= 0.981) were significant.

The perception-production relation re-analysed

In both experiments, we examined whether participants' production patterns influence the efficiency with which different rhotic allophones activate lexical candidates. In both studies, we found a null effect. In order to synthesize an overall verdict of our data, we use a Bayesian analysis which allows us to combine Bayes Factors (BFs) by multiplication. While BFs cannot always be multiplied, in our case, the prior is based on the assumption that there is a link between perception and production, and that theoretical conviction is not changed by observation (McQueen et al., 2023). In this situation, we can multiply BFs.

To obtain a BF for each study, we re-analyzed both data sets with a Bayesian linear mixed-effects model using the brms package (Bürkner et al., 2025). We specified a full model with all possible random slopes. The priors were generally weakly informative (following Gelman et al., 2008). However, for the critical interaction between perception and production allophone, we generated a prior based on the correlations observed in Cheng et al. (2022), which were around 0.5. We generated expected regression weights from this using the formula:

$$b = r * (SD_{outcome} / SD_{Predictor})$$

The prior for the critical interaction was then modelled as a normal distribution with a mean on the expected value and a standard deviation of half that, so that values below zero are deemed unlikely (following Dienes, 2014). The models were estimated using 4 chains with 4,000 iterations per chain, including 2,000 warm-up iterations, resulting in 8,000 post-warm-up samples. Convergence was indicated by R-hat values of 1.00 for all parameters and no divergent transitions. Then, a Savage-Dickey approximation was used to indicate how the data shift the likelihood that there is a positive relation between perception and production: For both the prior and the posterior distribution, the density at zero was calculated. The ratio of these odds indicates the BF for the hypothesis that there is a positive relation between perception and production.

For Experiment 1, this BF_{01} was 1.42, indicating that the data do not substantially shift the balance between the alternative and null hypotheses. For Experiment 2, where the descriptive data show a slight difference in the opposite direction, the BF_{01} was 10.39, indicating strong evidence for the null hypothesis over the hypothesis of a positive relation. When combined with the BF from Experiment 1, the overall effect was BF_{01} = 14.74. This constitutes strong evidence supporting the assumption that there is no relation between perception and production in our data.

Discussion

The results of Experiment 2 reveal two clear patterns. First, the phonetic detail caused by allophonic variation influences lexical competition, since the consideration of /l/- and /d/-initial words when hearing an /r/-initial word is modified by the allophonic variant of the /r/-initial word. This finding complicates psycholinguistic work, since competition effects in models of spoken-word recognition are usually based on phonemic information. For instance, Shortlist B (Norris & McQueen, 2008) uses a phoneme-confusion matrix to estimate the likelihood of a word being present in the input, but the current data

indicate that phoneme-confusion data may be dependent on the /r/ allophone. The current data also indicate that whether an /l/ is mistaken for an /r/ or not may depend on whether that /r/ is produced as an approximant or as a trill/tap, and these differences cascade to lexical processing. Other models of spoken-word recognition that assume that lexical activation can be computed from abstract phonological features in the input (Lahiri & Reetz, 2002) are also challenged by the current data, since the activation of phonological features would not (necessarily) reflect allophonic differences in the input. That is, models of spoken-word recognition need to consider allophonic variation.

Secondly, the current experiment shows that spoken-word recognition seems to benefit from a tapped variant of /r/ over an approximant variant. Note that taps and trills are rather similar, and in most languages, a tap is typical for intervocalic /r/ even when the /r/ otherwise is a trill. Note that this is also the case for our speaker. In Experiment 1, most /r/-initial words were preceded by a consonant, so that the /r/ was not intervocalic and mostly was produced as a trill. That is, tap and trills can be considered rather similar variants, and it is unlikely that this difference caused the difference in results, with a preference for taps in this experiment but no such effect for the mix of trills and taps in Experiment 1. In fact, if anything, the tap is considered a "reduced" or weaker variant of a trill, so that if a tap leads to a perceptual advantage (as in this experiment), we also should expect a trill to produce an advantage (in Experiment 1). However, no advantage for the trilled variants over approximants was observed in Experiment 1. A better candidate to explain the difference in results is the difference in carrier sentences that contained the /r/-initial word. In Experiment 2, participants had nothing to go on but the signal, as we used a fixed carrier phrase that allowed for any word to occur in the slot of the target word. In Experiment 1, by contrast, the /r/-initial words were slightly predictable from the sentence context. The absence of any contextual cues may hence lead to a stronger effect of bottom-up cues.

The data from the /r/-initial trials also indicate that the preference for the tap variant was not moderated by the production patterns of the participants. In Experiment 1, the descriptive data indicated such an effect but failed to reach significance; in Experiment 2, the effect again failed to reach significance, and the descriptive data were in fact going in the opposite direction. A combined Bayes-Factor analysis indicated that these patterns of results provide strong evidence for an overall null-effect, so that the allophonic production patterns of listeners appear not to influence their spoken-word recognition processes.

General Discussion

Two experiments investigated how rhotic allophonic variation impacts spoken-word recognition in Maltese. We investigated three questions, each with their own theoretical significance. First, is there a perceptual advantage for the trill in spoken-word recognition? Such an advantage would explain the popularity of the trill in the languages of the world despite its disadvantages in production. Second, does allophonic detail influence lexical activation? Third, is there a relation between production and perception, so that speakers that more often use trills also recognize words more efficiently when produced with a trill (and vice versa for speakers that tend to use approximants).

Maltese provides an interesting testbed for these questions, because there are two variants of the rhotic, alveolar trills (or taps, depending on the phonological context) and alveolar approximants, that occur with comparable frequencies in the overall population but both variants are used by individual speakers to different degrees. This was evident in an earlier production study (Mitterer et al., 2025) as well as in the current data, in which the likelihood of trills was overall around 50% in both experiments. This makes it possible to investigate allophonic effects in the absence of an overall frequency bias.

The first question, whether spoken-word recognition is faster when trilled variants are used, was investigated in both Experiments. The hypothesis that was motivated by claims within linguistics that the alveolar trill is the

most prototypical rhotic (Chabot, 2019; Ladefoged & Maddieson, 1996) and the observation that the acoustics of trills are likely facilitating perception by its amplitude modulation, which is known to be a strong driver of neural activity in the auditory system (Kluender et al., 2003). Experiment 1 used a VWP with images as targets and sentences that were slightly predictive of the target and another competitor image, following earlier, similar studies (Eger et al., 2019; Llompart et al., 2021). However, no advantage was found for /r/-initial words produced with a trilled or tapped variant over words produced with an approximant. In contrast, an advantage for the tapped variant was found in Experiment 2, which used written words as targets and a neutral (and repetitive) sentence frame. This indicates that there is a perceptual preference for trilled variants, but that preference is relatively small and easily overridden by contextual constraints. Since most words are at least somewhat predictable in a normal discourse, this would indicate that the perceptual benefit that trills have is probably quite small and too small to outweigh the very clear disadvantages that trills have in language acquisition, as has been discussed in the introduction.

One caveat to the claim that the popularity of the alveolar trill is unlikely to be caused by a perceptual advantage is that we used what might be considered "lab speech". We cannot rule out that an advantage might arise in more conversational speech. This raises the question why we did not use conversational speech in our experiment. Indeed, some studies have managed to use conversational speech in eye-tracking experiments (Brouwer et al., 2012, 2013), making use of an existing corpus in the target language. These experiments were done in Dutch, relying on the large Corpus Gesproken Nederlands (Corpus of Spoken Dutch, Oostdijk, 2000). A comparable corpus does not exist for Maltese. Therefore, using lab speech was the only way forward with Maltese as the target language. However, we instructed the speaker to speak fluently and not overly clearly.

While it is hence possible that an advantage might arise in spontaneous speech, this is not particularly likely. First of all, there is no theory that would directly predict a radical change in spontaneous speech. Secondly, we presented whole sentences produced fluently (rather than single-word utterances), and conversational speech contains stretches with minimal phonetic reduction similar to such registers (see Figure 2 in Warner, 2023). Others also have argued that lab speech can closely resemble much of spontaneous speech (Wagner et al., 2015; Xu, 2010). While the proof of the pudding remains in the eating, it would hence be surprising if our results would not hold for conversational speech.

The relative popularity of trills in the languages of the world is hence unlikely to be due to the small perceptual advantages for the trilled variants, as observed in Experiment 2. In search for the reason for the trill's popularity in the languages of the world, one might have to search elsewhere. Potentially, the trill has an advantage for social reasons. It might be considered a sort of shibboleth or a sort of a phonetic "peacock feather". As already mentioned in the introduction, some languages that use a trill (e.g., Italian and Russian) have a pejorative term for speakers that do not trill properly. Similarly, German theatre schools enforced the alveolar trill as the correct pronunciation on stage into the later parts of the 20th century, even though standard usage had switched to uvular variants well before that time (Wiese, 2001). The failure to find a clear perceptual benefit for trilled /r/ hence suggests that it might be worthwhile to investigate how trilled /r/ influences person perception, since the observations above would lead to the prediction that a trilled /r/ may lead to a positive person evaluation.

The question whether allophonic detail influences lexical activation was investigated in Experiment 2. It asked the question whether phonological competition in word recognition is influenced by allophonic variation. It has recently been reported that processing in spoken-word recognition is not always cascading (Galle et al., 2019). If this is the case for rhotics, we should see that phonetic detail of the rhotics should not influence lexical competition, simply because phonetic information only influences lexical activation once the rhotic is clearly recognizable as such. To test this, we selected pairs of words

that started with either /r/ and /d/ or /r/ and /l/ that had some form overlap after the initial consonants (e.g., *raqqad* and *laqqam*). We then used a target-absent version of the visual-world paradigm to increase the amount of looks to these similar words (i.e., when participants heard the word *raqqad*, the screen showed the words *laqqam* and two unrelated words). The results showed that competition was modulated by allophonic detail, so that the amount of looks to /d/-initial and /l/-initial words was modulated by how the /r/-initial word was produced. This indicates that phonetic information can cascade to the lexicon, in contrast to the findings of Galle et al. (2019). There are different possible conclusions to be drawn from this. One possibility is that the results of Galle et al. (2019) should be questioned. This, however, is unlikely since these results have been conceptually replicated (Kim et al., 2025). Another option is that the difference arises between segments, with the strongest buffering effects found for fricatives (Muegge et al., 2025). This discrepancy hence raises the question which information is immediately used in speech perception, and which is buffered. One possible conjecture is that amplitude modulations are immediately fed forward, given their primary status in auditory perception.

The final question examined in this study was whether a listener's own production patten influences how efficiently words are recognized, so that a match between production and perception leads to a facilitation. While the results of Experiment 1 were inconclusive, the results of Experiment 2 strongly suggest the absence of such an effect. This is reminiscent of other studies in which a mismatch between a listeners' own productions and the input did not lead to processing costs. Sumner and Samuel (2009) investigated rhoticity in New York and General American English and argued for the concept of the fluent listener, that is, a listener who speaks with a General American accent but is fluent in comprehending a New York accent. Similarly, Witteman et al. (2013) showed that Dutch listeners can learn to fluently process German-accented Dutch when exposed to it regularly.[5] With regard to rhotic allophony, Mitterer and Ernestus (2008) showed that Dutch listeners can shadow /r/-initial words produced with an alveolar or uvular place of articulation with the same speed, independent of their own production. Since shadowing first requires recognition, this also indicates that a mismatch between a participant's preferred /r/ and the heard /r/ does not slow down perception. This hence indicates that perception can adapt to a multitude of inputs without affecting production. In line with this assumption, Kraljic et al. (2008) showed that perceptual learning in speech can occur without affecting production.

These findings challenge theories that assume a strong relation between perception and production, such as direct realism (Fowler & Smith, 1986) and Motor theory (Galantucci et al., 2006). While the two theories differ in whether they posit an innate, specialized module for perceiving speech as articulatory gestures (see Diehl et al., 2004), both assume that there is a tight link between perception and production. This predicts that hearing an allophone that one would be unlikely to use oneself should lead to processing costs, such as slower recognition (Fowler et al., 2003). Another type of model that assumes a tight link between perception and production are the prediction models (e.g., Pickering & Gambi, 2018). In these models, the production system is used to predict upcoming linguistic material. However, if such predictions operate primarily at abstract symbolic levels (phonemes, words) rather than detailed phonetic specifications, these models would not necessarily predict the processing costs that Motor Theory and Direct Realism would. Our results are thus compatible with prediction models that operate at symbolic rather than phonetic levels.

The lack of an effect of production patterns on perception is somewhat surprising, because it contradicts findings from late

[5] An auxiliary assumption here is that their Dutch does not become slightly German accented, which has some face validity.

bilinguals suggest a connection between the two domains (Cheung & Babel, 2022; Eger & Reinisch, 2019; Mitterer et al., 2020). However, even researchers who question a tight relation between speech production and perception agree that some connection must exist between the two processes; otherwise we would not be able to learn to speak (Ohala, 1996; Scott et al., 2009). In particular, Rauschecker and Scott (2009) argue for two streams in speech perception, one for lexical access during conversation and one for learning to speak. Late L2 learners are typically still developing new articulatory routines and may therefore rely more heavily on the latter, learning-related stream. Nevertheless, when their task is to recognize words in fluent speech, a dissociation has been observed: a distinction that is at least partly mastered in production is not consistently used in perception (Eger et al., 2019).

The lack of an effect of production patterns on perception is also somewhat surprising given that most studies on sound change find moderate correlations between perception and production at least at the group level (Cheng et al., 2022; Coetzee et al., 2018). However, these results are not necessarily contradictory, because they differ in both the perceptual measures used and the type of sound change investigated. Most of the examples above involve a neutralizing sound change that collapses a phonological contrast. In contrast, in our study, the change from a trill/tap to an approximant variant of /r/ maintains all phonological contrasts of Maltese. In this sense, it is not surprising that we see a correlation between perception and production when contrasts are reduced. Speaker on opposite “sides” of the sound change then use different phonological systems. This is a situation not unlike L2 speakers, and here we also see a clear link between perception and production. Dutch and German speakers of English tend to have trouble in contrasting between /æ/ and /ɛ/ in both perception and production (Eger & Reinisch, 2019). Similarly, younger speakers learning the phonological system in a neutralizing sound-change situation simply learn a phonological system in which a contrast is marginalized, hence they have less of a need to learn a contrast. This is not the case in Maltese where the typical allophone is changing but the phoneme stays the same.

This also brings us to a second difference between the current study and those cited above. The other studies measured perception as the ability to discriminate a contrast in a categorization or discrimination task. In the current study, however, perception was assessed through the efficiency of word recognition. Consider the case of Cantonese syllable-initial /n/ and /l/, which are almost fully merged (Cheng et al., 2022; Soo & Babel, 2025). Cheng et al. (2022) find that speakers who tend to produce historically /n/- and /l/-initial words consistently as [l]-initial also tend to a reduced pattern of categorical perception of that contrast. However, it is far from clear that necessarily implies that the efficiency of lexical access is reduced when such speakers are confronted with [n]-initial words. In fact, if a speaker performs poorly on the categorization of an [n]-[l] continuum, we might predict that they have no problems to recognize words that are produced with [n], even if they themselves would pronounce them with [l]. This is similar to what is observed in L2 perception. Dutch speakers listening to English, who have trouble with the /æ/-/ɛ/ contrast easily activate the word *lamp* when hearing [lɛmp] instead of the correct [læmp].

Interestingly, in L2 research (Díaz et al., 2012; Eger et al., 2019; Llompart et al., 2025), a dissociation is generally observed between the ability to categorize speech sounds consistently (similar to how most sound-change papers evaluate perception) and to use the contrast for lexical access. Notably in the present case for Maltese, we tested lexical access for different allophones of the same phoneme, hence phonemic or lexical contrast were not affected. Rather, lexical access appeared to be somewhat modulated by words that were phonetically similar to one or the other allophone (i.e., /d/ vs. /l/-initial competitors in Experiment 2).

All in all, our findings suggest that listeners are able to have acoustic targets for speech sounds that are common in their speech community, and those can be used for word recognition (Holt et al., 2005; Ohala, 1996), even if they are not able or unlikely to produce

these sounds themselves. This is in line with earlier findings that, even when participants are asked to shadow (i.e., repeat back as fast as possible), a mismatch between the heard and produced speech gesture does not lead to a latency cost (Mitterer & Ernestus, 2008; Mitterer & Müsseler, 2013).

To summarize, the current study shows that trilled variants of /r/ are slightly easier to recognize than approximant variants. This advantage is, however, easily overturned by some amount of predictability of the /r/-bearing word. As such, it remains an open question why the alveolar trill, a phone that is relatively difficult to acquire, is relatively frequent in the languages of the world. This opens the possibility to look for socio-linguistic reasons why the trill may be popular. We also showed that rhotic allophony impacts lexical processing, since the activation of similar words depends on the allophone used for the rhotic. This fits in the picture of spoken-word recognition being finely attuned to phonetic detail in the input (Davis et al., 2002; Gow, 2003, p. 200; Salverda et al., 2003, 2014). Finally, our results indicated that speech perception seems relatively independent of speech production, since there was no benefit if listeners heard /r/-initial words in the same allophone that they tend to prefer in production.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439. https://doi.org/10.1006/jmla.1997.2558

Anselme, R., Pellegrino, F., & Dediu, D. (2023). What's in the r? A review of the usage of the r symbol in the Illustrations of the IPA. *Journal of the International Phonetic Association*, *53*(3), 1003–1032. https://doi.org/10.1017/S0025100322000238

Aro, M., & Wimmer, H. (2003). Learning to read: English in comparison to six more regular orthographies. *Applied Psycholinguistics*, *24*, 621–635.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*, 341–345.

Boersma, P., & Weenink, D. (2024). *Praat* (Version 6.4.12) [Computer software]. University of Amsterdam.

Bowers, J. S., Kazanina, N., & Andermane, N. (2016). Spoken word identification involves accessing position invariant phoneme representations. *Journal of Memory and Language*, *87*, 71–83. https://doi.org/10.1016/j.jml.2015.11.002

Brouwer, S., Mitterer, H., & Huettig, F. (2012). Speech reductions change the dynamics of competition during spoken word recognition. *Language and Cognitive Processes*, *27*(4), 539–571. https://doi.org/10.1080/01690965.2011.555268

Brouwer, S., Mitterer, H., & Huettig, F. (2013). Discourse context and the recognition of reduced and canonical spoken words. *Applied Psycholinguistics*, *34*, 519–539. https://doi.org/10.1017/s0142716411000853

Bürkner, P.-C., Gabry, J., Weber, S., Johnson, A., Modrak, M., Badr, H. S., Weber, F., Vehtari, A., Ben-Shachar, M. S., Rabel, H., Mills, S. C., Wild, S., Popov, V., Kosmidis, I., Schneider, B., Kallioinen, N., Silva, W. J., Carvalho, L., & Pekin, S. (2025). *brms: Bayesian Regression Models using 'Stan'* (Version 2.23.0) [Computer software]. https://cran.r-project.org/web/packages/brms/index.html

Chabot, A. (2019). What's wrong with being a rhotic? *Glossa: A Journal of General Linguistics*, *4*(1), Article 1. https://doi.org/10.5334/gjgl.618

Cheng, L. S. P., Babel, M., & Yao, Y. (2022). Production and perception across three Hong Kong Cantonese consonant mergers: Community- and individual-level perspectives. *Laboratory Phonology*, *13*(1). https://doi.org/10.16995/labphon.6461

Cheung, S., & Babel, M. (2022). The own-voice benefit for word recognition in early bilinguals. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.901326

Cho, T., Kim, D., & Kim, S. (2017). Prosodically-conditioned fine-tuning of coarticulatory vowel nasalization in English. *Journal of Phonetics, Mechanisms of Regulation in Speech*, *64*, 71–89. https://doi.org/10.1016/j.wocn.2016.12.003

Coetzee, A. W., Beddor, P. S., Shedden, K., Styler, W., & Wissing, D. (2018). Plosive voicing in Afrikaans: Differential cue weighting and tonogenesis. *Journal of Phonetics*, *66*, 185–

216. https://doi.org/10.1016/j.wocn.2017.09.009

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, *11*, 1084–1089.

Cutler, A. (2012). *Native Listening: Language Experience and the Recognition of Spoken Words*. MIT Press.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292–314. https://doi.org/10.1016/S0749-596X(02)00001-3

Davis, Matthew. H., Marslen- Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 218–244. https://doi.org/10.1037//0096-1523.28.1.218

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Diehl, R., Lotto, A. J., & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology*, *55*, 149–179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00781

Drager, K. K. (2011). Sociophonetic variation and the lemma. *Journal of Phonetics, Cross-Language Speech Perception and Variations in Linguistics Experience*, *39*(4), 694–707. https://doi.org/10.1016/j.wocn.2011.08.005

Draxler, C., & Jänsch, K. (2004). Speechrecorder—A universal platform independent multi-channel audio recording software. *Proceedings of LREC*, 559–562.

Eger, N. A., Mitterer, H., & Reinisch, E. (2019). Learning a new sound pair in a second language: Italian learners and German glottal consonants. *Journal of Phonetics*, *77*, 100917. https://doi.org/10.1016/j.wocn.2019.100917

Eger, N. A., & Reinisch, E. (2019). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *45*, 552–571. https://doi.org/10.1037/xlm0000599

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27–41. https://doi.org/10.1016/j.lingua.2012.12.006

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). Degree of foreign accent in English sentences produced by Korean children and adults. *Journal of Phonetics*, *34*(2), 153–175. https://doi.org/10.1016/j.wocn.2005.05.001

Fowler, C. A., Brown, J. M., Sabadini, L., & Welhing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396–413. (ISI:000185269200006). https://doi.org/10.1016/s0749-596x(03)00072-x

Fowler, C. A., & Smith, M. (1986). Speech perception as 'vector analysis': An approach to the problems of segmentation and invariance. In J. Perkell & D. Klatt (Eds), *Invariance and variability of speech processes* (pp. 123–136). Lawrence Earlbaum Associates.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin and Review*, *13*(3), 361–377. https://doi.org/10.3758/BF03193857

Galle, M. E., Klein-Packard, J., Schreiber, K., & McMurray, B. (2019). What Are You Waiting For? Real-Time Integration of Cues for Fricatives Suggests Encapsulated Auditory Memory. *Cognitive Science*, *43*(1), e12700. https://doi.org/10.1111/cogs.12700

Geller, J., Winn, M. B., Mahr, T., & Mirman, D. (2020). GazeR: A Package for Processing Gaze Position and Pupil Size Data. *Behavior Research Methods*, *52*(5), 2232–2255. https://doi.org/10.3758/s13428-020-01374-8

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*(4), 1360–1383. https://doi.org/10.1214/08-AOAS191

Gessinger, I., Raveh, E., Steiner, I., & Möbius, B. (2021). Phonetic accommodation to natural and synthetic voices: Behavior of groups and individuals in speech shadowing. *Speech Communication*, *127*, 43–63. https://doi.org/10.1016/j.specom.2020.12.004

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*, 575–590. https://doi.org/10.3758/BF03194584

Haden, E. F. (1955). The Uvular r in French. *Language*, *31*(4), 504–510. https://doi.org/10.2307/411363

Holt, L. L., Stephens, J. D. W., & Lotto, A. J. (2005). A critical evaluation of visually-moderated phonetic context effects. *Perception & Psychophysics*, *67*, 1101–1112.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. R. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, *90*, 17–127. https://doi.org/10.1016/S0093-934X(03)00425-5

Kim, H., Muegge, J. B., & McMurray, B. (2025). Extending the buffer: Real-time cue integration requires encapsulated auditory memory, even at absurdly long durations. *The Journal of the Acoustical Society of America*, *157*(4_Supplement), A170–A171. https://doi.org/10.1121/10.0037795

Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication*, *41*, 59–69.

Kraljic, T., Brennan, S., & Samuel, A. G. (2008). Accommodating variation: Dialects, idiolects, and speech processing. *Cognition*, *107*, 51–81. https://doi.org/10.1016/j.cognition.2007.07.013

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R Package Version*, *2*(0).

Ladefoged, P., & Maddieson, I. (1996). *Sounds of the world's languages*. Blackwell Publishers.

Lahiri, A., & Reetz, H. (2002). Underspecified recognition. In C. Gussenhoven & N. Warner (Eds), *Laboratory Phonology 7* (pp. 637–676). Mouton de Gruyter.

Lawson, E., Scobbie, J. M., & Stuart-Smith, J. (2011). The social stratification of tongue shape for postvocalic /r/ in Scottish English1. *Journal of Sociolinguistics*, *15*(2), 256–268. https://doi.org/10.1111/j.1467-9841.2011.00464.x

Lewis, A. M. (2004). Coarticulatory effects on Spanish trill production. *Proceedings of the 2003 Texas Linguistics Society Conference*, *116*, 127. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4a9df10150ac8c8ad3eb5e570a42ee4d29bbae81

Llompart, M., Eger, N. A., & Reinisch, E. (2021). Free Allophonic Variation in Native and Second Language Spoken Word Recognition: The Case of the German Rhotic. *Frontiers in Psychology*, *12*, 5289. https://doi.org/10.3389/fpsyg.2021.711230

Llompart, M., Gorba, C., & Prieto, P. (2025). Revisiting the link between second-language sound identification and word recognition with an eye on methodological similarity. *Studies in Second Language Acquisition*, 1–13. https://doi.org/10.1017/S0272263125101113

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (1999). Spoken Word Recognition in the Visual World Paradigm Reflects the Structure of the Entire Lexicon. In *Proceedings of the Twenty-first Annual Conference of the Cognitive Science Society*. Psychology Press.

McLeod, S., & Crowe, K. (2018). Children's Consonant Acquisition in 27 Languages: A Cross-Linguistic Review. *American Journal of Speech-Language Pathology*, *27*(4), 1546–1571. https://doi.org/10.1044/2018_AJSLP-17-0100

McQueen, J. M., Jesse, A., & Mitterer, H. (2023). Lexically Mediated Compensation for Coarticulation Still as Elusive as a White Christmash. *Cognitive Science*, *47*(9), e13342. https://doi.org/10.1111/cogs.13342

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *Quarterly Journal of Experimental Psychology*, *60*, 661–671. https://doi.org/10.1121/1.419865

Mitterer, H. (2011). The mental lexicon is fully specified: Evidence from eye-tracking. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 496–513. https://doi.org/10.1037/a0020989

Mitterer, H. (2018). The singleton-geminate distinction can be rate dependent: Evidence from Maltese. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *9*(1), 6. https://doi.org/10.5334/labphon.66

Mitterer, H., Eger, N. A., & Reinisch, E. (2020). My English sounds better than yours: Second-language learners perceive their own accent as better than that of their peers. *PLOS ONE*, *15*(2), e0227643. https://doi.org/10.1371/journal.pone.0227643

Mitterer, H., & Ernestus, M. (2006). Listeners recover /t/s that speakers lenite: Evidence from /t/-lenition in Dutch. *Journal of Phonetics*, *34*, 73–103. https://doi.org/10.1016/j.wocn.2005.03.003

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition*, *109*(1), 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of Maltese. *Journal of Memory and Language*, *108*, 104034. https://doi.org/10.1016/j.jml.2019.104034

Mitterer, H., & McQueen, J. M. (2009). Processing reduced word-forms in speech perception using probabilistic knowledge about speech production. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(1), 244–263. (ISI:000262838300021). https://doi.org/10.1037/a0012730

Mitterer, H., & Müsseler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception & Psychophysics*. https://doi.org/10.3758/s13414-012-0407-8

Mitterer, H., Pantelmann, Julia, Cassar, Rachel, & Reinisch, E. (2025). *Can a second language trigger sound change in the first? An apparent time study on rhotic production in Maltese*. https://osf.io/9fuqg

Mitterer, H., & Reinisch, E. (2023). Selective adaptation of German /r/: A role for perceptual saliency. *Attention, Perception & Psychophysics*, *85*(1), 222–233. https://doi.org/10.3758/s13414-022-02603-2

Mitterer, H., Reinisch, E., & McQueen, J. M. (2018). Allophones, not phonemes in spoken-word recognition. *Journal of Memory and Language*, *98*(Supplement C), 77–92. https://doi.org/10.1016/j.jml.2017.09.005

Muegge, J. B., Kim, H., & McMurray, B. (2025). Decoupling speech processing from time. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2025.05.017

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian Model of Continuous Speech Recognition. *Psychological Review*, *115*, 357–395. https://doi.org/10.1037/0033-295X.115.2.357

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America*, *9*, 1718–1725. https://doi.org/10.1121/1.414696

Oostdijk, N. (2000). Het corpus gesproken nederlands. *Nederlandse Taalkunde*, *5*(3), 280–284.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002–1044. https://doi.org/10.1037/bul0000158

R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (Version 4.3.3) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience*, *12*(6), 718–724. https://doi.org/10.1038/nn.2331

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 978–996. https://doi.org/10.1037/a0021923

Reinisch, E., & Mitterer, H. (2022). Phonetics and eye-tracking. In J. Setter & R. A. Knight (Eds), *The Cambridge Handbook of Phonetics* (pp. 457–479). Cambridge University Press.

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89. https://doi.org/10.1016/S0010-0277(03)00139-2

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Sankoff, G., Blondeau, H., & Charity, A. H. (2001). *Individual roles in a real-time change: Montreal (r->R) 1947-1995*. https://www.semanticscholar.org/paper/Individual-roles-in-a-real-time-change%3A-Montreal-Sankoff-Blondeau/bfde934c988b50a8b5a241b6b07aeb4e1333c3c9

Schiller, N. (1998). The phonetic variation of German /r/. In M. Butt & N. Fuhrhop (Eds), *Variation und Stabilität in der Wortstruktur. Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen* (pp. 261–287). Olms.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action: Candidate roles for the motor cortex in speech perception. *Nature Reviews Neuroscience*, *10*, 295–302. https://doi.org/10.1038/nrn2603

Sebregts, K. (2014). *The sociophonetics and phonology of Dutch r*. LOT, Netherlands Graduate School of Linguistics.

Soo, R., & Babel, M. (2025). Processing pronunciation variation with independently mappable allophones. *Journal of Phonetics*, *110*, 101402. https://doi.org/10.1016/j.wocn.2025.101402

Strunk, J., Schiel, F., & Seifart, F. (2014). Untrained Forced Alignment of Transcriptions and Audio for Language Documentation Corpora using WebMAUS. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk, & S. Piperidis (Eds), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014* (pp. 3940–3947). European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2014/summaries/1176.html

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*, 487–501.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. https://doi.org/10.1126/science.7777863

Velde, H. van de, & Hout, R. van. (1999). The Pronunciation of (r) in Standard Dutch. *Linguistics in the Netherlands*, *16*(1), 177–188. https://doi.org/10.1075/avt.16.16van

Wagner, P., Trouvain, J., & Zimmerer, F. (2015). In defense of stylistic diversity in speech research. *Journal of Phonetics, The Impact of Stylistic Diversity on Phonetic and Phonological Evidence and Modeling*, *48*, 1–12. https://doi.org/10.1016/j.wocn.2014.11.001

Warner, N. (2023). Advancements of phonetics in the 21st century: Theoretical and empirical issues of spoken word recognition in phonetic research. *Journal of Phonetics*, *101*, 101275. https://doi.org/10.1016/j.wocn.2023.101275

Warner, N., Fountain, A., & Tucker, B. V. (2009). Cues to perception of reduced flaps. *Journal of the Acoustical Society of America*, *125*, 3317–3327.

Weber, A., & Scharenborg, O. (2012). Models of spoken-word recognition. *WIREs Cognitive Science*, *3*(3), 387–401. https://doi.org/10.1002/wcs.1178

Whiteside, S. P. (2001). Sex-specific fundamental and formant frequency patterns in a cross-sectional study. *The Journal of the Acoustical Society of America*, *110*(1), 464–478. https://doi.org/10.1121/1.1379087

Wiese, R. (2001). The unity and variation of (German) /r/. *The Unity and Variation of (German) /r/*, (4), 11–26.

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics*, *75*(3), 537–556. https://doi.org/10.3758/s13414-012-0404-y

Xu, Y. (2010). In defense of lab speech. *Journal of Phonetics*, *38*(3), 329–336. https://doi.org/10.1016/j.wocn.2010.04.003

Appendix

Table A1. Experimental items used in Experiment 1.

Target Translation	Sentence in Maltese Gloss Translation	Competitor / Distractor
rivolver revolver	Il-ħallelin kellhom rivolver f'idejhom. DEF-thieves have.PAST.3PL revolver in hands.PL-their The thieves had a revolver in their hands.	crowbar/ tweezers
remote remote	Għandha bzonn remote għat-televison. have.3SG need remote for-DEF-television We need a universal remote for the television	speaker/ saw
rigg rig	Iz-ziju jaħdem fuq rigg fil-Libya. DEF-uncle work.3SG.M on rig in-DEF-Libya My uncle works on a rig in Libya	boat/ chair
rota bicycle	Daniel għażel rota ġdida. Daniel choose.PAST.3SG.M bicycle new.F Daniel chose a new bike	scooter/ eagle
riga ruler	Issellift riga mingħand sieħbi. borrow.PAST.1SG ruler from friend.my I borrowed a ruler from my friend	Tennis racket/ buffalo
rixa feather	Il-ħamiema tilfet rixa fil-glieda. DEF-pigeon lose.PAST.3SG.F feather in-DEF-fight The pigeon lost a feather during the fight	eye/ sausage
ras head	Waqajt u ħbat rasi mal-bankina. fall.PAST.1SG and bump.PAST.3SG.M head.my with-DEF-pavement I fell and bumped my head on the pavement	shoulder/ tv
rigal present	L-istudenti tawni rigal fi tmiem is-sena skolastika. DEF-students give.PAST.3PL-me present in end DEF-year scholastic The students gave me a present at the end of the scholastic year	chocolates/ sofa

re king	L-Ingliżi għandhom re ġdid. DEF-English have.3PL king new.M The English have a new king	coin/ beach
ross rice	Għamilt platt ross bit-tiġieġ. make.PAST.1SG plate rice with-DEF-chicken I made a plate of rice with chicken	pasta/ paprika
rakun raccoon	Ġieli nittawwal u nsib rakun fil-ġnien ifittex l-ikel. sometimes look.1SG and find.1SG raccoon in-DEF-garden search.PROG DEF-food There are times when I peep and find a racoon in the garden looking for food	rat/ stapler
rettili reptile	Hemm ħanut il-Belt li jbiegħ rettili ta' kull tip. there.is shop DEF-city REL sell.3SG.M reptiles of every kind There's a shop in Valletta that sells reptiles of all kinds	mobiles/ brain
ramel sand	Meta ninżel sal-bajja nispiċċa mimlija ramel. when go_down.1SG to-DEF-beach end_up.1SG full.F sand When I go to the beach I end up covered in sand	salt/ paint
rulett roulette	Dejjem tilgħab rulett meta tkun il-kasino. always play.3SG.F roulette when be.3SG.F DEF-casino She always plays roulette when she's at the casino	dice/ hand mirror
raba fields	Joey għandu raba' ħad-Dingli b'veduti kostali Joey have.3SG.M fields at-Dingli with-views coastal Joey has a field in Dingli with coastal views	house/ wasp
radjatur radiator	Fix-xitwa nħobb nixgħel radjatur fis-salott. in-DEF-winter love.1SG turn_on.1SG radiator in-DEF-living_room In winter I like to turn on the radiator in the living room	fireplace/ pool
radju radio	Fil-kċina għandi radju antik. in-DEF-kitchen have.1SG radio antique In the kitchen I have an old radio	burglar/ bathtub
ragħaj shepherd	In-nagħaġ għandhom bżonn ragħaj biex jiggwidahom. DEF-sheep have.3PL need shepherd to guide.3SG.M-them Sheep need a shepherd to guide them	sheepdog/ carrot

radar radar	Ix-xjentisti jużaw radar biex jistudjaw it-temp u l-klima. DEF-scientists use.3PL radar to study.3PL DEF-weather and DEF-climate Scientists use a radar to study the weather and climate	satellite/ corn
raxx rash	X'hin messet il-pjanta qabadha raxx. when touch.PAST.3SG.F DEF-plant catch.PAST.3SG.M-her rash When she touched the plant she got a rash	blisters/ fridge
razzett farm	Ħdejn l-iskola hemm razzett antik b'bieb tal-injam. next_to DEF-school there.is farm antique with-door of-DEF-wood Next to the school there is an old farm with a wooden door	hotel/ zebra
rebbieħ winner	F'logħba bejn tnejn, dejjem ikun hemm rebbieħ. in-game between two always be.3SG.M there.is winner In a game of two, there is always a winner.	loser/ ear
rassa crowd	Weħilna f'nofs rassa qabel ma dħalna. get_stuck.PAST.1PL in-middle crowd before that enter.PAST.1PL We got stuck in the middle of a crowd before we got in.	Traffic jam/ suitcase
ristorant restaurant	Hemm ristorant tajjeb ħafna fil-pjazza. there.is restaurant good very in-DEF-square There is a very good restaurant in the square	bar/ petrol station
rekorder recorder	Jekk ma nużax rekorder fil-klassi, ninsa kollox. if NEG use.1SG-NEG recorder in-DEF-class forget.1SG everything If I don't use a recorder in class, I forget everything.	camera/ toy
riġel leg	Gidmitni xi ħaġa f'riġlejja. bite.PAST.3SG.F-me some thing in-legs.my Something bit me on my leg	hand/ curtain
risiver telephone receiver	M'għadux jintuża risiver fid-djar. NEG-anymore-NEG use.PASS.3SG.M receiver in-DEF-houses A telephone receiver is no longer used in houses.	typewriter/ horseshoe

rokit rocket	Kuginti taħdem fuq rokits fl-Istati Uniti. cousin.my work.3SG.F on rockets in-DEF-States United My cousin works on rockets in the US	submarine/ fire
rummiena pomegranate	Fil-klassi pinġejt xena sempliċi ta' rummiena u lumija in-DEF-class paint.PAST.1SG scene simple of pomegranate and lemon In class I painted a simple scene of a pomegranate and a lemon	apple/ detergent
russett heron	Ġie russett fuq il-lag waqt li konna qed nistadu fl-Iskozja come.PAST.3SG.M heron on DEF-lake while that be.PAST.1PL PROG fish.1PL in-DEF-Scotland A heron came by on the lake while we were fishing in Scotland	otter/ floor lamp
rutella tape_measure	Il-ħaddiem nesa li kellu rutella żejda fil-van. DEF-worker forget.PAST.3SG.M that have.PAST.3SG.M tape_measure spare in-DEF-van The worker forgot that he had a spare tape measure in the van	spanner/ cherry
rużarju rosary	Ommi kisbet rużarju mill-Vatikan waqt vaganza tal-familja mother.my get.PAST.3SG.F rosary from-DEF-Vatican during holiday of-DEF-family My mother got a rosary from the Vatican during a family holiday	keychain/ bear
rixtellu gate	Quddiem il-bieb hemm rixtellu tal-hadid. in_front DEF-door there.is gate of-DEF-iron In front of the door there is a metal gate	letterbox/ armchair
ravjul ravioli	Għandi aptit platt ravjul ghall-ikel illejla. have.1SG appetite plate ravioli for-DEF-food tonight I feel like a plate of ravioli for dinner tonight.	spaghetti/ nails
romblu pastry_roll	Oħti għamlet romblu ta l-għaġina bil haxix u l-ġobon. sister.my make.PAST.3SG.F pastry_roll of DEF-dough with-DEF vegetables and DEF-cheese My sister made a pastry roll with vegetables and cheese.	pie/ tin

rejżer razor	Insejt nippakkja rejżer fil-bagalja. forget.PAST.1SG pack.1SG razor in-DEF-luggage I forgot to pack a razor in my luggage.	toothbrush/ vacuum
referì referee	Kellhom referi tajjeb għal-logħba. have.PAST.3PL referee good for-DEF-game They had a good referee during that game.	goalkeeper/ kiwi
respiratur respirator_mask	Meta kont ma niflaħx ħsibt li ħa jkolli bżonn respiratur. when be.PAST.1SG NEG well.1SG-NEG think.PAST.1SG that will have.1SG need respirator When I was unwell I thought I would need a respirator	thermometer/ dinosaur
rukola rocket_salad	Ħaga waħda li ma togħbnix fuq il-pizza hija rukola. thing one REL NEG like.3SG.F-me-NEG on DEF-pizza be.3SG.F rocket One thing I don't like on pizza is rocket.	pineapple/ wardrobe
reġina queen	Liema pajjiżi huma mmexxija minn reġina? which countries be.3PL led.PASS from queen Which countries are run by a queen?	government/ giraffe
ram brass	Hija moda li tuża ram f'disinn tal-kċina. be.3SG.F fashion that use.3SG.M brass in-design of-DEF-kitchen It is fashionable to use brass in kitchen design	marble/ sock
rokna corner	It-tifla staħbiet minn sħabha f'rokna fid-dlam. DEF-girl hide.PAST.3SG.F from friends.her in-corner in-DEF-darkness The girl hid from her friends in a corner in the dark.	cupboard/ cup

Note that the items were presented as pictures, we hence only present the English picture names for the competitor and the distractor.

Table A2. Experimental items with d-initial target used in Experiment 2. In each cell, the upper word is the Maltese word with an English translation below.

Spoken Target	/d/-initial Word	overlap	Competitor.1	Competitor.2
risposta	disponibbli		attent	tgħannaq
response	available	4	attentive	embrace
residenti	destinazzjoni		ewlieni	spiss
residents	destination	2	chief	often
raqqaq	daqqa		ġid	qalb
to slim sth.	blow	3	great	heart
riżorsi	diżorganizzat		kbir	sellem
resources	disorganized	3	good	greeting
rata	data		tbissima	fenek
rate	date	3	smile	rabbit
riżultat	diżunur		koppja	merkant
result	dishonor	3	couple	merchant
riċerka	diċenti		papra	speċjali
research	decent	3	duck	special
raħal	daħal		żomm	tafda
village	entered	4	keep	trust
relazjoni	delizzjuz		kiesaħ	tmiem
relation	delicious	2	cold	end
rispett	dispett		mera	studju
respect	disrespect	5	mirror	study
riskju	diskors		tul	żmien
risk	speech	2	long	time
roża	doża		papoċċ	sabiħ
pink	dose	3	slipper	beautiful
rettur	dettalji		kor	gwerra
rector	details	2	chorus	war
relatat	delikat		ħasla	nemel
related	delicate	2	washing	ants
rabbejt	dabbar		ġardinar	avkuata
I brought up	to acquire	2	gardener	lawyer
reċita	deċiżjoni		vulkan	xalla
recite	decision	3	volcano	scarf
redikolu	dedikat		toqba	spiżjar
ridiculous	dedicated	3	hole	pharmacist
ruħ	duħħan		tiġbed	impjegat
soul	smoke	2	draw	employee
reklam	deklinat		poeżija	bajja
advertisement	declined	3	poem	beach
rizza	dizzjunarju		mħabba	triq
heap	dictionary	2	love	road
rutella	dutrina		matul	għalqet
tape measure	doctrine	2	during	closed
rixtellu	dixxiplina		għasel	imxi
gate	discipline	2	honey	walk
romblu	domanda		injorant	ħolma
roller	demand	2	ignorant	dream
rinunzja	dinosawru		aħwa	madum
renunciation	dinosaur	2	brothers	tiles

rilevant	dilettant		ċivili	xibka
relevant	amateur	3	civilian	net
riżenja	diżerta		ġelat	stenna
resignation	desert	2	ice cream	wait

Table A3

Experimental items with d-initial target used in Experiment 2. In each cell, the upper word is the Maltese word with an English translation below.

Spoken Target	/l/-initial Word	overlap	Distractor.1	Distractor.2
rabat	labar		perit	għasfur
to tie	needles	3	architect	bird
rebaħ	lebaniż		ieqaf	mewġ
to win	lebanese	3	stop	waves
rotta	lottu		naddaf	għalija
route	lotto	3	cleaning	for me
riċenti	liċenzja		tapit	fwieħa
recent	license	4	carpet	perfume
reġina	leġġenda		waħx	diska
queen	legend	2	horrible	record
razza	lazz		xita	skola
race	lace	4	rain	school
ribell	libell		lejl	kif
rebell	libellous	3	night	how
ramel	lamenta		sellun	ftakar
sand	lament	2	stairs	remember
rabta	labra		oratorju	iltqajt
bond	needle	2	oratory	I met
russett	lussuż		moħħ	bram
heron	luxurious	4	mind	jellyfish
raqad	laqat		avża	bard
sleep	hit	4	warn	cold
rampa	lampa		ġmiel	flus
ramp	lamp	3	beautiful	money
randan	lanqas		inżul	prietka
lent	not even	3	landing	sermon
raħħas	laħħaq		sfortunat	ġemma
cheap	reached	3	unfortunate	collected
raqqad	laqqam		tqażżeż	muntanja
sleep	call	4	slip	mountain
rokit	lokit		żelaq	skopra
rocket	lock	2	slide	discovered
ravjul	lavanda		quċċija	tlift
ravioli	lavender	2	peak	I lost
ragħaj	lagħab		stieden	taġen
gragher	played	2	inviting	pan
resqa	lesti		veduta	lezzjoni
approach	ready	4	view	lesson
ranġament	lanġas		skuża	jitlef
arrangement	pear	2	sorry	he loses
remissa	lemaħ		qatgħa	tar
remission	seen	3	cut	fly
ringiela	lingwa		inka	siket
row	language	2	inca	silence
reġistru	leġittmu		prinċep	faddal
registry	legitimate	3	prince	save

rema	lemin		qares	iknes
remove	right	2	sour	sweeps
rinforz	linfa		tarbija	ġeru
reinforcement	lymph	3	baby	dog
rimarka	limitat		bankina	inkwatru
remark	limited	3	sidewalk	framed

Table A4

/r/-initial targets and their distractors in Experiment 2.

/r/ targets	Distractor.1	Distractorr.2
ross	merluzz	ċkejken
rice	cod	small
rkoppa	toppu	parroċċa
knob	toppu	parish
rabja	wiżgħa	tazza
anger	fear	cup
risq	pazjent	fjakkoli
risq	patient	torches
ritmika	taraġ	xagħar
rhythmic	stairs	hair
ritwal	foloz	imnieħer
ritual	fake	nose
romanz	bizzilla	ftit
romance	lace	little
raġun	ċint	demm
reason	fence	blood
ras	quddiem	għadira
head	front	pond
rapport	galletta	fjakkajt
report	biscuit	I fainted
raġel	stqarr	mument
man	confessed	moment
raġġ	ordnajt	soppa
ray	I ordered	soup
radju	tieghu	xewqa
radio	his	wish
rattab	festa	morna
softened	party	we went
rustiku	żebbuġ	tweġiba
rustic	olives	answer
rotond	widna	ħlief
round	ear	except
ritorn	entrata	nokkli
return	entrance	noodles
rispettat	ġwienaħ	mara
respected	wing	woman
riproduzzjoni	għadma	kelma
reproduction	bone	word
riserva	eċċellenti	frawli
reserve	excellent	strawberries
rikorrenza	tagħma	xiħa
recurrence	blindness	old
regola	problema	eżerċizju
rule	problem	exercise
rakkont	kamra	famuż
Erzählung	room	famous

reċipjent	isfel	kok
recipient	downstairs	cook
redentur	ħobża	sajjied
redeemer	loaf	fisher
refgħa	bużżieqa	dejqa
lift	bubble	narrow
re	neputi	ħbieb
king	nephew	friends
replika	ċeramika	fjamma
replicate	ceramics	flame
rivelat	nstab	boċċi
revealed	found	balls
rvinat	ħass	kwiekeb
rowned	lettuce	stars
rank	basla	arka
rank	onion	ark
rigal	arblu	bniedem
gift	pole	man
reliġjuz	pulizija	karru
religious	police	carriage
radd	sinjal	saru
return	sign	became
ra	waħħal	ibqaw
see	attach	stay
rotazzjoni	żejt	nifs
rotation	oil	breath
rikostruzzjoni	qiegħ	stampa
reconstruction	bottom	picture
ripetut	illum	xirja
repeated	today	run