Mitterer, H., Arvanitidou, M., Balzan, J., Borg, D., Demanuele, R., Engels, L.-V. A., Grixti, N., Koczyk, P., & Thorsteinsdottir, A. (2026). Rhotics in general, but not trills in particular, are associated with roughness: Experimental evidence from Maltese. *Glossa Psycholinguistics*, 5(1): 4, pp. 1–19. DOI: https://doi.org/10.5070/G6011.47317

Rhotics in general, but not trills in particular, are associated with roughness: Experimental evidence from Maltese

Holger Mitterer, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta; Hanyang Institute for Phonetics and Cognitive Sciences of Language, Hanyang University Seoul, South Korea, holger.mitterer@um.edu.mt

Martha Arvanitidou, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, martha.arvanitidou.24@um.edu.mt

Julia Balzan, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, julia.m.balzan.19@um.edu.mt

Daniela Borg, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, daniela.borg.20@um.edu.mt

Ritienne Demanuele, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, ritienne.demanuele.20@um.edu.mt

Laura-Viviane Anna Engels, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, laura.engels.19@um.edu.mt

Nadia Grixti, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, nadia.grixti.16@um.edu.mt

Patrycjusz Koczyk, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, patrycjusz.koczyk.21@um.edu.mt

Anna Thorsteinsdottir, Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta, anna.thorsteinsdottir.24@um.edu.mt

Recent work on sound symbolism has established a relation between roughness and the alveolar trill (e.g., Ćwiek et al., 2024). The available evidence is, however, ambiguous as to whether the phonetic properties of an alveolar trill or the phonological properties of rhotics are associated with roughness. This was tested in Maltese, which uses both alveolar approximants and alveolar trills for /r/. Participants were presented with a product that had either a rough or a smooth surface and were asked to choose a potential product name from two options, one containing an /r/ and the other an /l/. The /r/ was produced either as a trill or an approximant. The /r/-bearing word was chosen more often when the product was rough, but this was independent of the phonetic implementation of the /r/. This indicates that the phonological properties of rhotics, and not the phonetic properties of the trill, support the association of /r/ with roughness.

1. Introduction

Recent work in linguistics has challenged the very core assumptions of what the design principles of language are, including recursion (Everett, 2005; Karlsson, 2007), language universals (Evans & Levinson, 2009) and arbitrariness (Dingemanse, 2012; Hinton et al., 1994). In the latter line of research, Winter et al. (2022) suggested that alveolar trills (i.e., the way /r/ is often produced, e.g., in Russian or Spanish) are associated with roughness, and that this association is based on an iconic relation that links auditory amplitude modulation with the haptic experience of exploring rough surfaces. They surveyed the vocabulary of different languages and found that the likelihood of a word containing an /r/ is much higher if that word has a "rough" meaning (e.g., *abrasive*). Importantly, this pattern was not observed for languages that used another version of an /r/ sound – called *rhotics* in phonology – such as a uvular fricative [ʁ] or an alveolar approximant [ɹ].

Anselme et al. (2025) questioned Winter et al.'s (2022) coding methodology for rhotic variants (which is a more general issue in illustrations of the IPA; see Anselme et al., 2023). Their analysis found an association of any rhotic, not just trilled /r/, with roughness in lexicons of various languages. We see two explanations for this finding. The first explanation relies on sound change. Rhotic sound change often moves away from the alveolar trill (Lancien et al., 2023; Mitterer et al., 2025; Sebregts, 2015; Wiese, 2001), and the alveolar trill is often considered the ur-rhotic (Chabot, 2019; Wiese, 2003). Therefore, it is conceivable that these associations might arise while a language uses an alveolar trill for the rhotic, and then the association between the rhotic and roughness remains in the language even when the phonetic substance of the rhotic changes. This is, in fact, how Winter et al. (2022) explain the relation between roughness and rhotics in English, which mostly uses approximants for /r/. Another possibility is that the phonological properties of the rhotic phoneme /r/, rather than the phonetic properties of the alveolar trill [r], are associated with roughness (Anselme et al., 2025). In the current study, we aim to provide relevant experimental evidence bearing on this issue.

The distinction between the phonetic and phonological properties of a phoneme is particularly important for rhotics, which form an unusual class. Dutch, for instance, is said to have 21 different allophones for /r/ (Sebregts et al., 2023), some of which have little in common, both acoustically and articulatorily (such as a uvular fricative and an alveolar approximant). Ladefoged and Maddieson (1996, p. 217) see no other basis for these sounds forming a class other than orthography, that is, being written with the letter <r> or the Greek letter *rho*.

More recent work has suggested that rhotics are held together either by a sort of family resemblance, which can be understood in terms of diachronic changes (Sebregts, 2015), or by phonological principles (Chabot, 2019). Regarding the former, alveolar and uvular trills are obviously acoustically related; both have amplitude modulation in the signal. Within each place of articulation, lenition processes can turn trills into fricatives or approximants (Iskarous & Kavitskaya, 2018). In this way, a link between sounds as apparently different as alveolar approximants and uvular trills can be made. A different approach to explain the relatedness of rhotics is phonological in nature and argues that rhotics are sounds that easily cluster with other consonants, independent of their phonetic substance. Evidence for this claim comes, for instance, from loanword adaptations, where Moroccan Arabic uses an alveolar tap for French loans containing an /r/, even though Moroccan Arabic contains back fricatives that are much closer to the uvular fricative that French uses for /r/. This indicates that rhotics are understood in terms of their phonological properties, independent of their phonetic substance.

One study (Ćwiek et al., 2024) provided further evidence for a link between the alveolar trill and roughness. The study used a sequential matching procedure between auditory stimuli that are best described as long, syllabic versions of an alveolar trill or a lateral /l/ and smooth and jagged lines. This is similar to studies that investigated an association between high vowels and voiceless stops with jagged objects, as well as an association between round vowels and voiced consonants with round objects (often labelled the "kiki-bouba" effect; see Bremner et al., 2013; Köhler, 1947; Ramachandran & Hubbard, 2001). Participants saw the two lines and, on a given trial, heard a syllabic version of either [r̩] or [l̩] and were then asked to choose which line better matched the auditory stimulus, with the instruction to focus on how the lines feel to the touch. Over a thousand participants from different languages participated in this two-trial study. The results show a very large effect size in matching (close to 90%) that was slightly lowered when the participants' native language used an alveolar trill as primary allophone. However, the design of the study confounds the phonological and phonetic properties of the alveolar trill. That is, participants had a choice between an /l/-bearing stimulus and an [r]-bearing stimulus. However, the [r]-bearing syllable was also /r/-bearing.[1] Therefore, the results do not allow us to distinguish between the claims that the alveolar trill, or rhotic segments in general, are associated with roughness.

We investigated this issue in Maltese, a Semitic language spoken in Malta. A recent corpus study on Maltese (Mitterer et al., 2025) indicated that alveolar trills and alveolar approximants occur as free allophones and are roughly equiprobable in the onset position. While there is a general trend that speakers who are more English-dominant use more approximants, there is a large amount of intra-speaker variation, so that most speakers will use both approximants and trills for /r/, even within the same syllabic position. We even observed, in a phrase such as *rajt rota* (Engl. 'I saw a bike'), that one /r/ is produced as a trill and the other as an approximant.

---

[1] This requires an additional assumption here that listeners unfamiliar with the trill would recognize it as /r/. Phonological theories (Chabot, 2019) argue that the alveolar trill is the ur-rhotic and may be therefore, recognized as such, while some data on L2 perception may suggest that it might be perceived as non-speech if dissimilar to native speech sounds (Best, 1995). As such, both assumptions ([r] is recognized or not as /r/) would require further testing.

This provides an ideal testing ground for investigating the sound-symbolic properties of the alveolar trill. First, the two variants are encountered frequently by Maltese listeners, that is, in linguistic terms, they are both unmarked. Secondly, the two allophones differ acoustically, so that only the trill is obviously related to roughness, given the fast amplitude modulation. The approximant, on the other hand, is a rather sonorant consonant, with full voicing that usually reaches an amplitude that is not strongly different from surrounding vowels. These properties are similar to /l/.[2] That is, phonetically, there is no reason to associate approximants with roughness. Note how that would be different for a uvular fricative, which, as an unvoiced sound, can be conceived as naturally rougher.

Since it is challenging to get hundreds of Maltese speakers for online studies (e.g., at the time of writing, hardly any can be found on the Prolific platform), we adapted the task to make it a multi-trial study. **Figure 1** shows the basic task setup, in which participants choose one of two different nonwords as a name for a given product, with the additional instruction that the name should reflect how it feels to touch the object. This follows on a long tradition of product naming in sound-symbolism research (Klink, 2000; Lowrey & Shrum, 2007; for a recent overview, see Arora et al., 2022).

The products were typically perceived as either rough or smooth to touch, and one of the nonwords contained /l/, while the other contained /r/. It varied over trials whether the /r/ was produced as [ɹ] or [r]. The experimental design, hence, uses the two factors Product Surface (smooth vs. rough) and Allophone of /r/ (approximant vs. trill). The dependent variable is whether participants choose the /r/-bearing or /l/-bearing (non-)word for the product. If rhotics in general are associated with rough objects, a main effect of the *product-surface* factor is expected, with more /r/-word choices for rough-surface products. If particularly alveolar trills are associated with roughness, there should be an interaction between the two factors, with a stronger difference in /r/-choices between smooth and rough products if the /r/ is a trill. As **Figure 1** already indicates, we did not use minimal pairs, because the use of minimal pairs generates a demand characteristic that the experiment is about /l/ vs. /r/ (see 2.2 for details)

Demand characteristics means that the experiment suggests to the participant what they should react to. An interesting example is supplied in slope perception, in which initial reports suggested that wearing a heavy backpack makes a slope look steeper (Bhalla & Proffitt, 1999). In these experiments, participants were supplied with the backpack with no obvious reason for it. Durgin et al. (2009) argued that participants could, hence, expect that the backpack was supposed to influence their reactions. Indeed, when the backpack was justified as measuring equipment, its influence on slope perception disappeared.

[2] The two sounds are mostly distinguished by the location of the third formant. There is no a-priori reason to assume that a low F3 is related to roughness.

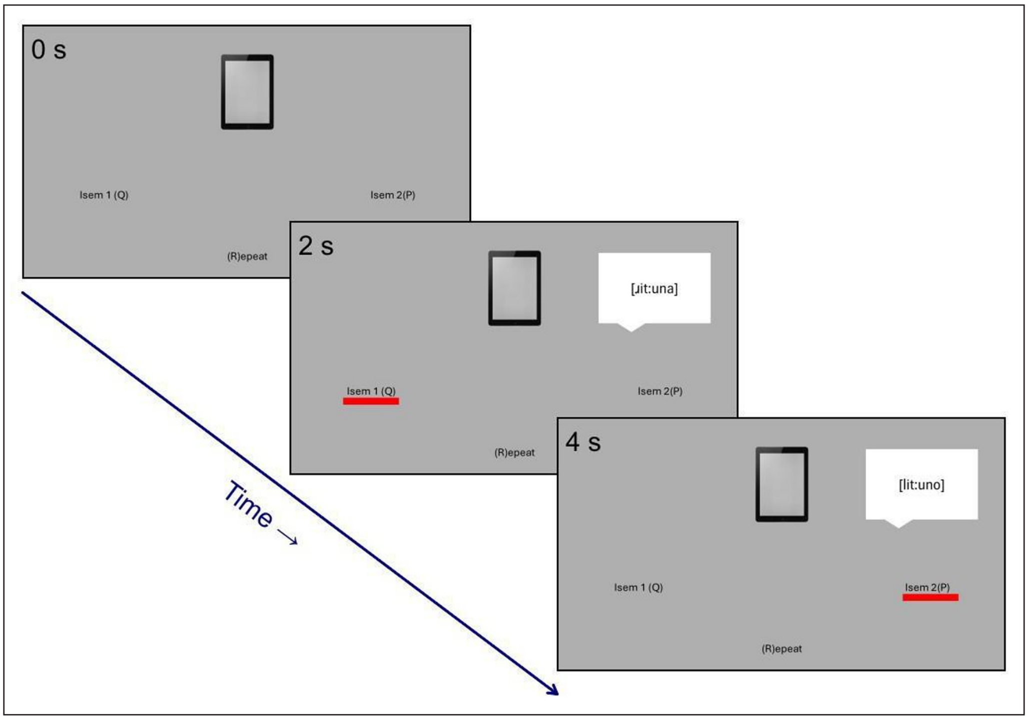

Figure 1: Schematic representation of a trial with screenshots at trial onset and after 2 s and 4 s. (The time indications here were not visible to participants.) Participants first saw a product (here, a tablet) and then heard two potential names for it (*isem* is Maltese for 'name') as indicated inside the speech bubbles, which illustrate the auditory stimuli and were not shown visually during the experiment, and then pressed either <p> or <q> to choose one of the names or <r> to hear the names again.

In a similar way, Ćwiek et al. (2024) presented syllabic versions of the consonants only and asked participants to select either a jagged or a smooth line. Participants are, hence, invited to elaborate on the relation between sound and touch, and have to focus on the one speech sound presented during the trial. This creates a similar demand characteristic, as in experiments on slope perception with and without a backpack (for further discussions of overly confirming research strategies in sound-symbolism research, see Dingemanse et al., 2016). To provide an analogy, if we ask whether Maltese listeners use epenthetic glottal stops as signs for prosodic boundaries, we could simply present the same phrase with and without glottal stop. This creates the demand characteristic that glottal stop should influence the response. To prevent that, Mitterer et al. (2021) presented stimuli with final lengthening as an additional cue, so that listeners could do the task without using the information from the glottal stop. For the experiment in Cwiek et al. (2024), a similar strategy would have been to use stimuli such as /V1rV1/ vs. /V2lV2/, with the identity of the vowels counterbalanced over participants. In this case, participants would not have been forced to consider [r] crucial to the task. Along these lines, the current experiment uses stimuli that differ in an aspect other than just between /l/ and /r/. That is, participants are not forced to consider that /r/ and /l/ are crucial to the task.

2. Method

2.1 Participants

56 participants were recruited from the student body of the University of Malta. The participants were aged between 18 and 40 (mean: 21.7, median: 20), and 24 identified as male and 32 as female. **Table 1** shows the usage characteristics of Maltese and indicates that participants, in general, used Maltese more when younger, but usage stabilized in adulthood. While four participants were clearly English dominant, using Maltese only for 15–25% of their language use since secondary school, all participants were at least familiar enough with Maltese to be familiar with both approximants and trills as allophones of /r/.

Table 1: Usage of Maltese in the participant sample during different periods of life.

	Period			
	Pre-school	**Primary school**	**Secondary School**	**Adulthood**
Mean	59.0%	56.4%	50.0%	51.7%
Median	72.5%	60%	50%	50%
Range	0–100%	0–95%	10–90%	15–90%

2.2 Materials

50 products were used that are prototypically smooth (tablet) or rough (doormat). Images were generated with Dzine (https://www.dzine.ai/) as well as ChatGPT and then rescaled to fit a 500 × 500 pixel frame on a transparent background. Two stimuli were used for practice trials and paired with nonwords that contained neither /r/ nor /l/. The practice trials were intended to familiarize participants with the procedure.

Rough and smooth products were chosen based on face validity. This was based on the assumption that testing the association of roughness with the alveolar trill would indicate whether the face-validity judgements were indeed valid. However, since two reviewers independently suggested a norming study, we added such a study. To do this as a web-experiment, we used Italian participants, since Maltese participants are not easy to reach online (i.e., there is no established payment structure). Italy is a neighboring country with similar cultural associations. This is reflected, first, in linguistic and cultural similarity, with more than half of Maltese words having an Italian origin (Aquilina, 2006) and Malta and Italy showing an average Euclidean Distance on the Hofstede cultural dimensions (Hofstede et al., 2010) of about 21, which is smaller than the average distance between other EU countries (which is about 28). 24 native-speakers of Italian residing in Italy were recruited via prolific.com and participated in the study. They were presented with the two practice items as warm-up trials and then with one of 24 pre-compiled random orders of the 48 product images. Below the image there was a visual-analogue scale whose ends were indicated as *molto liscio* and (Engl., 'very smooth) and *molto ruvido* (Engl., 'very rough'). Participants could click anywhere on the scale, and the experiment only moved on to the next trial when they made a choice. Ratings were coded such that a click on the "smooth" end of the scale results in a zero and a click on the "rough" end results in a value of one hundred. Higher values, hence, indicate roughness. Results are presented in Section 3.

For each product, product names were based on two different nonword frames with an empty slot (the empty slot was then filled with an alveolar trill, an alveolar approximant, or an /l/; as shown in **Table 2**) that had the same syllable structure and differed in one or two phonemes. The frames were kept as similar as possible without leading to existing words (when the empty slot was filled with either /r/ or /l/). The slot for /r/ or /l/ was word-initial for half of the stimuli, and word-medial, but syllable-initial, for the other half. This leads to six auditory stimuli for each product (see **Table 2**). On a given trial, two nonwords were chosen, with either the first or second frame containing an /l/, and the other containing either an [r] or an [ɹ]. This leads to four different possible trials for each product, by crossing the choice of which frame bears the /r/ with the two allophones for implementing /r/. Using a Latin-Square design, the different types of trials were presented to participants so that each participant heard 24 trials with an approximant [ɹ] and 24 with a trilled [r], and each product was presented once in each of the four versions over a set of four participants. Through counterbalancing, we remove any confounds that might arise from the other phonological differences between the two frames on a given trial. The /r/-bearing word was randomly allocated to the first or second sound played on a given trial.

Table 2: Two example stimuli for a rough (doormat) and a smooth product (tablet).

	Segment filling the empty slot					
	[ɹ]		**[r]**		**[l]**	
	Frame					
Product	Frame 1	Frame 2	Frame 1	Frame 2	Frame 1	Frame 2
doormat	[ɹomenu]	[ɹimenu]	[romenu]	[rimenu]	[lomenu]	[limenu]
mirror	[keʃɹitɑ]	[kɑʃɹitɑ]	[keʃritɑ]	[kɑʃritɑ]	[keʃlitɑ]	[kɑʃlitɑ]

The nonwords were produced by a female native speaker of Maltese. In some cases, trills were not achieved (a common phonetic pattern for trills; see Iskarous & Kavitskaya, 2018). Items were then re-recorded. In the final recordings, 45 items had one tongue contact, 46 items had two tongue contacts, and 5 items had three tongue contacts, which fits the typical properties of singleton trills (Ladefoged & Maddieson, 1996). Since prosodic differences also contribute to sound symbolism (Dingemanse et al., 2016), an additional step equalized all six stimuli for one product using Praat (Boersma, 2001). To achieve that, all recordings were first annotated for syllable boundaries, with boundaries aligned to zero-crossings of the speech signal. For each syllable, the maximum amplitude was determined and then adjusted to the mean of this syllable (first, second, third, etc.) within a set of six stimuli. Then, the average duration of the six recordings as well as the median f0 value for each syllable were measured, from which the mean duration and the mean f0 contour over the syllables of the set were calculated. Using the PSOLA algorithm, the overall duration of the stimuli within one set was equalized. The original pitch contour was removed and replaced by a pitch contour that provided one pitch point at each syllable midpoint, which was the mean of the median values of the six stimuli within one set. The resulting stimuli had a generally falling f0 contour, which sounded somewhat monotone and/or uninterested, but nevertheless still sounded natural.

2.3 Procedure

Participants first signed a consent form and then filled out a questionnaire about their use of English and Maltese (the two official languages of Malta) during different periods of their lives. After that, they performed the main experimental task, which lasted around 5 minutes. Participants were paid 3 euro for their participation. Different randomizations were generated offline, implementing the Latin Square design that distributed the four different trial types for one product over participants. After this allocation, the order of trials for a given participant was randomized without further constraints.

The experiment was run on a standard Windows PC, and stimuli were presented to participants via a 21-inch monitor. The experiment was controlled using PsychoPy (Peirce et al., 2019), and at the start of the experiment, the experimenter put in a participant number, which was used to pick one of the experimental orders. Auditory stimuli were presented through Focusrite Scarlett headphones.

Figure 1 shows the timeline for a single trial. After every 20 trials, participants were informed about how many of the 50 trials they had already completed. They could take a break here, but most participants preferred to quickly proceed.

3. Results

3.1 Rating Task

The rating task showed a complete separation of the two categories of products that were a-priori categorized as smooth and rough. Mean ratings for the items categorized a-priori as smooth ranged from 5.3 to 39.3 (mean = 18.8, SD = 8.91), and mean ratings for items categorized a-priori as rough ranged from 60.7 to 97.1 (mean = 81.9, SD = 9.25). The “roughest” smooth item (a car key) was still two standard deviations away from the smoothest “rough” item (artificial grass). This indicates that our judgements by face validity can be replicated in a norming task.

3.2 Product-naming Task

Through experimenter error, the first participant performed more than 48 experimental trials, and for one participant, the last two trials were lost. We retained the first 48 trials for the first participant and the 46 available for the second participant. Participants were fairly decisive, with 30 participants never hearing product names more than once. Only on 6 trials (out of 2687) did participants choose to hear the object names three times, while hearing the options twice was more frequent (136 out of 2687 trials). Of the 26 participants who did repeat trials, the median number of repeats was 4 (out of 48 trials per participant), with a range from 1 to 18.

Figure 2 shows how often participants chose the /r/-bearing word, depending on the surface of the product (rough vs. smooth) and the allophone of /r/. Note that participants always had a choice between an /r/-bearing word and an /l/-bearing word. However, /l/ is not one of the independent variables. It figures implicitly in **Figure 2**, as the proportions of /l/-choices are the inverse of the (visualized) proportion of /r/-choices.

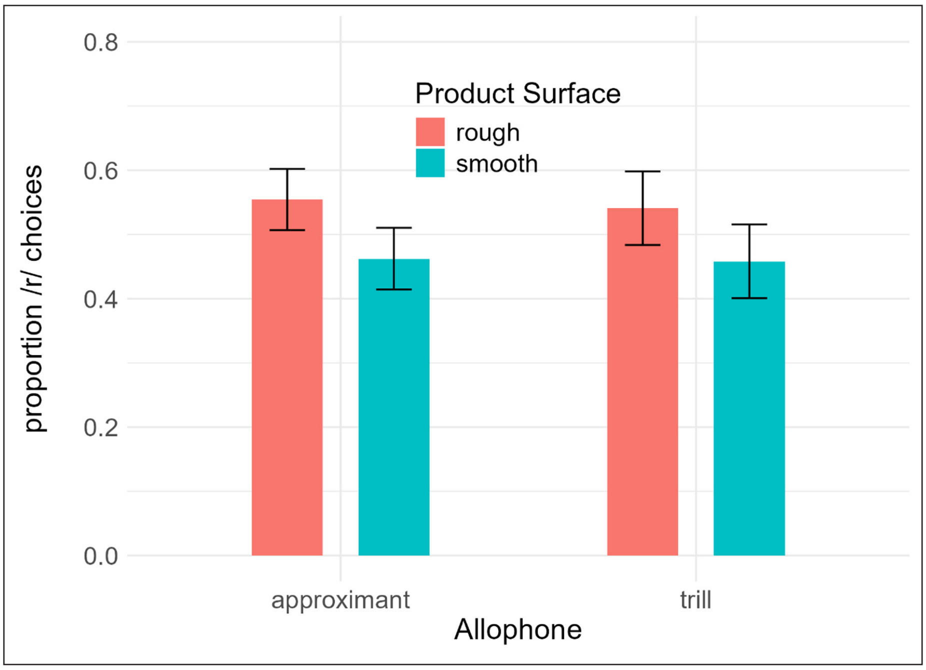

Figure 2: Mean proportions of /r/-choices in the four conditions of the experiment. Error bars are based on the 95% credible intervals derived from the statistical model.

The figure shows that participants chose the /r/-bearing word over the /l/-bearing word more often when the product was rough than when it was smooth, independent of the allophone used for /r/. The critical issue is whether there is an interaction between /r/ allophone and Product Surface. To assess this, a Bayesian logistic mixed-effects regression model was constructed using the *brms* package (Bürkner et al., 2024) in R (v4.3.4). The model used a maximal random-effects structure.

We performed the analysis with two different priors. The first prior was a weakly informative prior centred on zero, using a Cauchy distribution with a scale of 2.5, which has been suggested as a default prior for logit models (Gelman et al., 2008). This prior prioritizes values between –5 and 5 logit units, which relates to a change from, for example, 7% to 93%. A second prior also took into account the fact that we expect positive regression weights, given how the contrast coding was set up. For this, we used a normal distribution with a mean and standard deviation of two. This was based on the effect size in Ćwiek et al. (2024), who found matching responses at 90%. This means 90% /r/ choices for the rough stimulus and 10% /r/ choices for the smooth stimulus, which corresponds roughly to a logit difference of four. This distribution, then, expects effect sizes between zero and four, using the effect size in Ćwiek and colleagues as an upper limit.

The two predictors were contrast-coded. For /r/ allophone, the trill was mapped on 0.5 and the approximant on –0.5 (as mnemonic, the trill is considered the “better” rhotic than the approximant) and for the product-surface factor, *rough* was mapped on 0.5 and *smooth* on –0.5. With this coding, the predicted effects (main effect of product type and the interaction) should have positive regression weights. With contrast coding and the experimental design, this also means that there is no issue of collinearity between the predictors.

The analysis indicated that the effect of Product Surface is small, at 0.35 for the first analysis and 0.36 for the second analysis, but likely to be reliable, since the credible interval excludes zero (first analysis: [0.11, 0.60], second analysis: [0.13, 0.60]). The effect of Allophone – which is irrelevant for the theoretical alternatives – is close to zero, and the credible interval includes zero (–0.04 [–0.22, 0.15], identical in both analyses). The critical interaction term also is close to zero, with a credible interval that includes zero (first analysis: –0.04 [–0.39, 0.32], second analysis: –0.02 [–0.38, 0.34]). We used the Savage-Dickey ratio to approximate a Bayes Factor for the interaction, which is the ratio of the heights of the density functions for the prior and the posterior at zero. In both cases, the Bayes Factor indicates that the null hypothesis is more than ten times as likely as the alternative hypothesis (BF = 17.0 for the first analysis, BF = 18.2 for the second analysis).

Another exploratory analysis for the influence of phonetic properties on sound symbolism can be achieved by exploiting the natural variation within our stimuli set, which contained trills with one, two, or three contacts, though there were very few items with three contacts. Potentially, with more tongue contacts, the /r/-bearing word becomes the more attractive choice for rough products, but less attractive for smooth products, relative to an /l/-bearing nonword. This prediction, which suggests an interaction of product type and the number of contacts in the trill, was tested by focusing on trials with trilled /r/s only. We again used a *brm* model with a logit-link function and estimated a Bayes Factor for the interaction using the Savage-Dickey ratio, using a weakly informative prior for the first analysis and one informed by the previous data for the second analysis. The resulting Bayes Factors are 8.9 and 9.8 in favour of the null. However, our recordings were not controlled strongly for phonetic detail, so this should be considered an additional exploratory analysis.

Two additional analyses investigated whether properties of either items or participants influence the product-naming task. First, we asked whether the roughness ratings influenced the product-naming choices. For this, we extracted the random slopes of the product category from the linear mixed effects model. We inverted the values for smooth objects, because a positive random slope means a higher rate of /r/-choices for rough products, but fewer /r/-choices for smooth products. By inverting the latter, the values indicate more /r/-choices for higher values. These were then correlated with the raw roughness ratings. For the roughness ratings, we calculated the deviation from the category mean, that is, to what extent the object is considered rough (or less smooth) within its category. We then correlated these measures, using the function correlation BF from the package BayesFactor (Morey et al., 2024), and calculated the credible interval from the posterior distribution with 10000 iterations. The correlation was estimated at –0.18 (CrI: –0.43 –0.08) and the BayesFactor with a correlation of 0.33 as H1 was 0.80, hence not indicating strong evidence either way.

In a similar way, we correlated the participants' mean usage of Maltese (logit-transformed) with the participant random slope of the product-category effect. The estimated correlation was 0.01 (CrI: –0.25 –0.25), and the BayesFactor with a correlation of 0.33 as H1 was 0.30, hence providing moderate evidence for the null hypothesis.

4. Discussion

Based on the question of whether alveolar trills in particular or rhotics in general are related to roughness, we investigated this relationship, using a product-naming task with rough and smooth products and alveolar trills and approximants in Maltese, where both allophones are common and, therefore, unmarked. This creates a strong foundation for examining potential bottom-up effects of trills without the confounding influence of a frequency bias for one of the allophones. The results indicated a role for rhotics in general and not particularly for trills.

First of all, the current results confirm a sound-symbolic relation between /r/ and roughness in a paradigm in which the relation between dependent and independent variables was less obvious to participants, thereby lowering the demand characteristics of the experiment. At this juncture, it might be useful to unpack the term *sound symbolism*. On the one hand, it is a useful umbrella term, but on the other hand, it is somewhat imprecise. It conflates iconic relationships that are based on cross-modal associations, as assumed for the kiki-bouba phenomenon, and systematic relationships that are based on a co-variation of phonetic properties of words and their semantic or even syntactic properties that are language-specific (Monaghan et al., 2007).

This distinction may also help to explain the current finding. As reviewed above, Winter et al. (2022) argued that the relation between roughness and /r/ is based on English having used the alveolar trill in the past, with the relation now remaining in the language. This account explains only the relation in the English lexicon, but leaves open the question whether the relation between /r/ and roughness is relevant for current speakers of English. The results of a study by Bergen (2004) indicate that this might be the case. He showed that English listeners learn the phonaestheme *gl-* in the absence of an obvious iconic relationship. Relatedly, Farmer et al. (2006) present evidence that listeners are sensitive to (language-specific) phonological cues to word class. This suggests that speakers of English may also learn (implicitly) the relation between /r/ and roughness, independent of the allophone used for /r/. Since all of our participants were fluent in English, they hence may have learned the relationship. That is, exposure to a lexicon that contains the "/r/ is rough" relation may be sufficient for listeners to learn that link. If that is the case, it is, however, still noteworthy that the putative iconic relationship does not add anything to the systematic one for our Maltese listeners. That is, when there is a systematic relationship between /r/ and roughness, the potential iconic relation between [r] and roughness does not seem to contribute a lot anymore in language processing.

It may be argued that the lack of clear correlation between participant and item characteristics and the choice of /r/ for rough items may be informative of the difference between an iconic or a systematic relationship. For instance, it may be argued that exposure to Maltese should increase the likelihood of choosing /r/ for rough items. However, we do not agree with this prediction, because English clearly also has a relation between /r/ and roughness (see Winter et al., 2022). Since Maltese is also likely to have a systematic relationship between /r/ and roughness (being fed by Italian and Arabic words), relative exposure to both languages is unlikely to influence the association of /r/ with roughness in the mental lexicon of our Maltese participants. On the other hand, it may be argued that the lack of correlation between item roughness within category and choices for /r/ indicates a categorical and not an iconic relationship. Independent of the outcome, which showed no clear association between the two variables, we would argue that even a positive correlation would simply show that the product images activate the concepts "rough" and "smooth" to a different degree, which then modulates the fit of a "rough" name. The correlation does, in our view, not speak to the reason why /r/ is considered rough by participants.

Considering the field of L2 speech perception, it is not surprising if we consider that perception of speech sounds is strongly influenced by learning, which is especially visible in the countless demonstrations of failures in L2 speech perception (Cutler, 2012). The finding is also reminiscent of an article on cross-modal associations by Mitterer and Jesse (2010). They tested the association between a sustained piano sound and a "systematic" co-variate (i.e., seeing the key being pressed) and the "iconic" co-variate (seeing the hammer hitting the key). They found a stronger cross-modal association for the learned pattern with the key than for the hammer. Importantly, the key press is, in fact, an "anti-iconic" correlate of a sustained piano sound, since, usually, keeping in contact with a surface leads to a muffled sound (as it does on a vibraphone). Our current results dovetail well with this finding that cross-modal associations are dominated by learning and, once learning is established, iconic relations seem to play a minor role.

How does this relate to the findings of Ćwiek et al. (2024)? How do speakers of all the languages tested in Ćwiek and colleagues associate a trill with roughness? We see at least two explanations for this finding. Research on cross-modal associations (Guzman-Martinez et al., 2012) argues for a relation between visual-spatial and auditory frequencies that is mediated by tactile roughness. The apparent problem with linking visual-spatial and auditory frequency is that their denominators differ: time in the auditory domain and space in the visual domain. Guzman-Martinez et al. (2012) argued that this link is achieved by assuming that humans explore surfaces with their hands at roughly similar speeds, which gives the spatial component a temporal dimension. However, the question remains how to conceptualize these associations. Is it the case that cross-modal associations influence early perceptual processing (van Wassenhove et al., 2005), or are the two streams processed independently and then merged into a percept (Massaro, 1998)? This is related to the question whether knowledge influences perception in a top-down fashion, which also questions the modularity of perceptual processes. And, indeed, many of the pitfalls identified by Firestone and Scholl (2016) for top-down modulation of perception also apply to cross-modal interactions. Early interactions often turn out to be based on relatively general-purpose attentional mechanisms (Vroomen & Stekelenburg, 2010).

The results in Ćwiek et al. (2024) do not force the conclusion of early cross-modal interactions, as participants were explicitly invited to search for an analogy. This issue is aggravated by the fact that the stimulus for the /r/-bearing syllable was not a typical example of an alveolar trill as it occurs in normal, conversational speech. Ladefoged and Maddieson (1996, p. 218) report that alveolar trills usually contain two to three contacts with geminates, maximally having six to eight contacts. Inspection of **Figure 1** in Ćwiek et al. (2024) indicates that their stimulus had about 17 contacts and is, therefore, not an ecologically valid example of an [r]. In addition to this unusual rendition of the trill, participants were required to focus on the properties of one speech sound given the nature of the syllabic consonant stimuli. This may have created demand characteristics; in other words, Ćwiek et al. (2024) employed what Firestone and Scholl (2016) termed an overly confirmatory research strategy.

Another explanation for the results of Ćwiek et al. (2024) may be found in a learning account. The association between rhotics and roughness might be caused by the learning of speech sounds. Learning seems to be a more potent force in speech perception than the bottom-up characteristics of speech sounds. Countless demonstrations of failures of L2 speech perception (see Cutler, 2012, for a review) show that speech sound differences that are obvious to L1 listeners, and, hence, perceivable bottom-up, are intractable for L2 learners. The Maltese listeners in the current study may have learned to equate the alveolar trill and the alveolar approximant, and this learning may be potent enough to override any allophonic effect of the alveolar trill. Such a learning-based effect would also explain why Anselme et al. (2025) observed the association of roughness with rhotics overall. Given that the most observed sound changes for rhotics are usually away from the alveolar trill, it is conceivable that most languages –or at least one of their ancestral stages– once made use of an alveolar trill. This historic relation may underlie the association between rhotics and roughness, originating from the phonetic properties of the trill. This, then, led to a learning bias that rhotics as a phonological class should be associated with roughness, even when the phonetic substance of that rhotic changed. Due to an initial association of the alveolar trill with roughness, the lexicon provides a source for learning. Consequently, /r/ then works as a learned cue for roughness, independent of the phonetic implementation.

How can we investigate to what extent any potential iconic relationship adds to (or exists above) the systematic relationship? One way to do so is to see whether sub-phonemic detail enhances the effect of a cross-modal match. Importantly, though, this should be achieved in a design in which the sub-phonemic detail remains subtle –rather subtle manipulations have been shown to influence lexical access (Salverda et al., 2003)– and the task should allow participants to succeed without relying on the sub-phonemic manipulation (i.e., other cues support the decision).

Another possible way to distinguish symbolic and iconic relations would be to investigate whether learning of new words for objects can be based on consistent association between object properties and word properties, even if they are "anti-iconic". For instance, if participants learn that, in an artificial language, voiceless stops are associated with round objects, would this create a bias that makes it easier for them to learn new words following this pattern, even though it conflicts with the well-known kiki-bouba effect? Cross-modal associations can be learned, and learned associations even seem stronger than natural ones (Mitterer & Jesse, 2010). Overall, this indicates that sound symbolism may rely much more on what Hinton et al. (1994) called "conventional sound symbolism", somewhat counter to the idea of bottom-up determined synaesthetic relationships (Ramachandran & Hubbard, 2001), even if there is a small natural cross-modal association.

Data accessibility statement

All materials, raw data, data-preprocessing script, processed data, and analysis code are available here: https://osf.io/ahner/.

Ethics and consent

The study was run in line with the ethical guidelines of the University Research Ethics Committee of the University of Malta (application number: MAKS-2023-00131). All participants signed an informed consent form before commencing the study.

Acknowledgements

This study has been supported by a grant from The Malta Council for Science and Technology (REP-2023-010) to Holger Mitterer. The authors thank Shravan Vasishth for his help in implementing the Bayes-Factor calculations

Competing interests

The authors have no competing interests to declare.

Authors contributions

This project originated as part of the course work for the Master in Cognitive Science at the University of Malta, supervised by Holger Mitterer.

Holger Mitterer: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Supervision.

Martha Arvanitidou: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Julia Balzan: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Daniela Borg: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Ritienne Demanuele: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Laura-Viviane Anna Engels: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Nadia Grixti: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Patrycjusz Koczyk: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

Anna Thorsteinsdottir: Methodology, Resources (stimuli generation), Investigation, Data curation, Writing – review & editing

ORCiD IDs

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

Martha Arvanitidou: https://orcid.org/0009-0006-7447-477X

Julia Balzan: https://orcid.org/0009-0007-8825-7490

Daniela Borg: https://orcid.org/0009-0004-7599-3817

Ritienne Demanuele: https://orcid.org/0000-0003-1289-9845

Laura-Viviane Anna Engels: https://orcid.org/0009-0004-7796-6257

Nadia Grixti: https://orcid.org/0009-0003-0086-3519

Patrycjusz Koczyk: https://orcid.org/0009-0005-8569-8894

Anna Thorsteinsdottir: https://orcid.org/0009-0001-8105-1218

References

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

Anselme, R., Pellegrino, F., & Dediu, D. (2025). Not just the alveolar trill, but all "r-like" sounds are associated with roughness across languages, pointing to a more general link between sound and touch. *Scientific Reports*, *15*(1), 12930. https://doi.org/10.1038/s41598-025-94850-0

Aquilina, J. (2006). *Concise Maltese English, English Maltese Dictionary*. Midsea Books.

Arora, S. V., Kalro, A. D., & Sharma, D. (2022). Positioning products through names: Effect of sound symbolism on perception of products with blended brand names. *Journal of Product & Brand Management*, *32*(3), 361–378. https://doi.org/10.1108/JPBM-12-2021-3794

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, *80*, 290–311.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. Pp. 167–200. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 167–200). York Press.

Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(4), 1076–1096. https://doi.org/10.1037//0096-1523.25.4.1076

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

Bremner, A. J., Caparos, S., Davidoff, J., de Fockert, J., Linnell, K. J., & Spence, C. (2013). “Bouba” and “Kiki” in Namibia? A remote culture make similar shape–sound matches, but different shape–taste matches to Westerners. *Cognition*, *126*(2), 165–172. https://doi.org/10.1016/j.cognition.2012.09.007

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. (2024). *brms: Bayesian regression models using “Stan”* (Version 2.22.0) [Computer software]. https://cran.r-project.org/web/packages/brms/

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

Cutler, A. (2012). *Native listening: language experience and the recognition of spoken words*. MIT Press. https://doi.org/10.7551/mitpress/9012.001.0001

Ćwiek, A., Anselme, R., Dediu, D., Fuchs, S., Kawahara, S., Oh, G. E., Paul, J., Perlman, M., Petrone, C., Reiter, S., Ridouane, R., Zeller, J., & Winter, B. (2024). The alveolar trill is perceived as jagged/rough by speakers of different languages. *The Journal of the Acoustical Society of America*, *156*(5), 3468–3479. https://doi.org/10.1121/10.0034416

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics Compass*, *6*, 654–672. https://doi.org/10.1002/lnc3.361

Dingemanse, M., Schuerman, W., Reinisch, E., Tufvesson, S., & Mitterer, H. (2016). What sound symbolism can and cannot do: Testing the iconicity of ideophones from five languages. *Language*, *92*(2), e117–e133. https://doi.org/10.1353/lan.2016.0034

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaughnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review*, *16*(5), 964–969. https://doi.org/10.3758/PBR.16.5.964

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448. https://doi.org/10.1017/S0140525X0999094X

Everett, D. L. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology*, *46*(4), 621–646. https://doi.org/10.1086/431525

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203–12208. https://doi.org/10.1073/pnas.0602173103

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, *39*, e229. https://doi.org/10.1017/S0140525X15000965

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

Guzman-Martinez, E., Ortega, L., Grabowecky, M., Mossbridge, J., & Suzuki, S. (2012). Interactive coding of visual spatial frequency and auditory amplitude-modulation rate. *Current Biology*, *22*(5), 383–388. https://doi.org/10.1016/j.cub.2012.01.004

Hinton, L., Nichols, J., & Ohala, J. J. (Eds.). (1994). Introduction: Sound-symbolic processes. In *Sound symbolism* (pp. 1–12). Cambridge University Press. https://doi.org/10.1017/CBO9780511751806.001

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind, Third Edition*. McGraw Hill Professional.

Iskarous, K., & Kavitskaya, D. (2018). Sound change and the structure of synchronic variability: Phonetic and phonological factors in Slavic palatalisation. *Language*, *94*(1), 43–83. https://doi.org/10.1353/lan.2018.0001

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*, 365–392. https://doi.org/10.1017/S0022226707004616

Klink, R. R. (2000). Creating brand names with meaning: the use of sound symbolism. *Marketing Letters*, *11*(1), 5–20. https://doi.org/10.1023/A:1008184423824

Köhler, W. (1947). *Gestalt psychology; An introduction to new concepts in modern psychology* (Rev. Ed.). Liveright.

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

Lancien, M., Hutin, M., Stuart-Smith, J., Adda-Decker, M., & Vasilescu, I. (2023, August). /r/ lenition in Quebec French: Evidence from the distribution of 9 allophones in large corpora. *20th International Congress of Phonetic Sciences*. https://hal.science/hal-04312598

Lowrey, T. M., & Shrum, L. J. (2007). Phonetic symbolism and brand name preference. *Journal of Consumer Research*, *34*(3), 406–414. https://doi.org/10.1086/518530

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. MIT Press.

Mitterer, H., & Jesse, A. (2010). Correlation versus causation in multisensory perception. *Psychonomic Bulletin & Review*, *17*(3), 329–334. https://doi.org/10.3758/pbr.17.3.329

Mitterer, H., Kim, S., & Cho, T. (2021). The role of segmental information in syntactic processing through the syntax–prosody interface. *Language and Speech*, *64*(4), 962–979. https://doi.org/10.1177/0023830920974401

Mitterer, H., Pantelmann, J., Cassar, R., & 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

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*(4), 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Morey, R. D., Rouder, J. N., Jamil, T., Urbanek, S., Forner, K., & Ly, A. (2024). *BayesFactor: Computation of Bayes Factors for common designs* (Version 0.9.12-4.7) [Computer software]. https://lib.stat.cmu.edu/R/CRAN/web/packages/BayesFactor/index.html

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Ramachandran, V. S., & Hubbard, E. M. (2001). Synaesthesia –a window into perception, thought and language. *Journal of Consciousness Studies*, *8*(12), 3–34.

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

Sebregts, K. (2015). *The sociophonetics and phonology of Dutch r* [Doctoral dissertation], Universiteit Utrecht. https://www.lotpublications.nl/the-sociophonetics-and-phonology-of-dutch-r

Sebregts, K., Hout, R. van, & Velde, H. V. de. (2023). Sociophonetics and rhotics. In C. Strelluf (Ed.), *The Routledge handbook of sociophonetics* (1st ed., pp. 195–213). Routledge. https://doi.org/10.4324/9781003034636-10

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181–1186. https://doi.org/10.1073/pnas.0408949102

Vroomen, J., & Stekelenburg, J. J. (2010). Visual anticipatory information modulates multisensory interactions of artificial audiovisual stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Wiese, R. (2003). The Unity and Variation of (german) /r/. *Zeitschrift Für Dialektologie Und Linguistik*, *70*(1), 25–43. http://www.jstor.org/stable/40504887

Winter, B., Sóskuthy, M., Perlman, M., & Dingemanse, M. (2022). Trilled /r/ is associated with roughness, linking sound and touch across spoken languages. *Scientific Reports*, *12*(1), 1035. https://doi.org/10.1038/s41598-021-04311-7