ARE ALVEOLAR TRILLS PERCEIVED AS "STRONG" RHOTICS?

Eva Reinisch[1], Holger Mitterer[2,3]

[1]Austrian Academy of Sciences, [2]University of Malta, [3]Hanyang University Seoul, South Korea
Eva.Reinisch@oeaw.ac.at, holger.mitterer@um.edu.mt

ABSTRACT

Rhotics are a highly variable sound class. In German, realizations as different as alveolar trills [r] and uvular fricatives [ʁ] co-occur as allophones of a single rhotic phoneme. Previous evidence indicated that listeners perceive [r] as the strongest/most salient version of /r/. The present experiments tested whether this leads to a form of speaker adaptation, in which weak forms of /r/ are deemed less acceptable, if the speaker also produces the strong from (i.e., [r]). In a 2AFC task, listeners categorized a continuum from [ʁ] to [h] while also hearing the same speaker use [r] or not. In two experiments, we found that participants in the [r]-exposure group accepted fewer stimuli from the [ʁ]-[h] continuum as /r/ than those in the [ʁ]-only exposure group. A third experiment tested whether this effect is speaker specific by exposing participants to [r] from a different speaker, but those results were inconclusive.

Keywords: speech perception, rhotics, German, perceptual saliency, speaker specificity

1. INTRODUCTION

Rhotics form a highly variable sound class [1]. They can be produced at different places and manners of articulation with variants as different as alveolar trills and uvular fricatives. Cross-linguistically, the most frequent variant is the alveolar trill, despite its articulatory complexity [1]. Trills are also considered the perceptually "strongest" or "most salient" variants, which is due to their "clearly modulated signal, distinct from other speech segments" ([2], see also [3],[4]). In the present study, we asked whether different rhotic allophones are interpreted relative to each other such that the presence of a strong [r] makes listeners less willing to accept other, "weaker" rhotic allophones as an instance of /r/. Note that such a finding may partly explain the popularity of [r] in the world's languages despite its articulatory complexity.

Specifically, we assessed the perception of alveolar trills compared to the articulatorily and acoustically maximally different uvular fricatives. German was used as the example language, since in German both, alveolar trills and uvular fricatives (among others) co-occur as allophones of a single rhotic phoneme [5]. While in syllable onset [ʁ] or [ʀ] are considered the standard, in the south of Germany, many speakers use [r]. In general, speakers are relatively consistent in their use of rhotic allophones in prevocalic position. With regard to the status of [r] as the perceptually strongest rhotic allophone, one study on German already found some support using a selective adaptation paradigm [6].

Selective adaptation is the effect that if a given stimulus is heard repeatedly, a following ambiguous stimulus is less likely to be perceived as an instance of the previously heard adaptor. In the case of rhotics, if listeners are exposed to /r/-initial words produced with [ʁ], then ambiguous stimuli from a test continuum between [ʁ] and [h] should be less likely perceived as the rhotic. In [6], this full match condition (among others) was compared to a condition in which the adapter stimuli were produced with [r], an allophone with little acoustic overlap to [ʁ] in the test continuum. A stronger effect of selective adaptation was *expected* for the full match condition but *found* for the acoustically dissimilar [r]-adaptors. This greater effectiveness of the alveolar trill as adaptor was interpreted to result from this allophone's greater perceptual saliency compared to other German rhotic allophones.

The present study set out to further assess the role of trilled rhotic allophones relative to other variants in speech perception. Using a different experimental setup, we asked whether native German listeners who have heard /r/-initial words produced with [r], would be less willing to accept [ʁ] as an instance of a rhotic than when no alveolar trills have been heard.

In three experiments, native German listeners performed blocks of forced-choice minimal word pair identification tasks categorizing [ʁ]-[h] and [r]-[l] continua. Different continua for the different rhotic allophones were necessary to minimize the acoustic distance between the rhotics and the other option. In Experiment 1 the rhotic allophones were produced on different words but spoken by the same speaker. Experiment 2 then went on to test categorization of the different allophones on the same word in two different minimal pairs, again spoken by the same speaker. Experiment 3 used different word pairs and the [r]-[l] continuum was spoken by a different speaker. Responses to the [ʁ]-[h] continuum were compared between a group of listeners who categorized both continua and one who only ever heard the [ʁ]-[h] continuum. We hypothesized that if

[r] is perceived as a strong/salient variant of the rhotic, then hearing a speaker produce this allophone should reduce listeners' willingness to categorize sounds along the [ʁ]-[h] continuum as the rhotic.

2. EXPERIMENTS 1 AND 2

2.1. Methods

2.1.1. Participants

Thirty and 34 participants for Experiments 1 and 2 respectively were recruited via the online platform prolific.co [7]. Participants were between 20 and 42 years of age and declared themselves as native speakers of German. They performed the experiments via the internet and were paid for their services.

2.1.2. Materials

A female native speaker of German recorded the minimal word pairs *Reis-heiß* (rice-hot), *Reiher-Leiher* (heron-lyre), *Rose-Hose* (rose-pants) and *Rose-Lose* (rose-lottery ticket). The rhotics were produced as [ʁ] in the minimal pairs contrasting with /h/ and as [r] when contrasting with /l/, matching the places of articulation. For [r], a token with a clear amplitude modulation was chosen; for [ʁ] one with a steady-state fricative noise. The words with /h/ and /l/ were selected to match the /r/-initial words in duration and f0 to facilitate the creation of continua, for which we used audio morphing with STRAIGHT [8]. A set of pre-tests with participants who did not take part in the main experiments was run to find 6 continuum steps (from originally 21) that spanned the range from good tokens of /r/ to the other phoneme. The continua *Reis-heiß* and *Reiher-Leier* were used in Experiment 1, *Rose-Hose* and *Rose-Lose* in Experiment 2.

2.1.3. Procedure

Except for the use of different minimal word pairs, the procedure was the same in Experiments 1 and 2. Participants performed a 2-alternative forced choice task (2AFC). On each trial they were presented with an audio stimulus (i.e., one token from the continua) and saw two pictures on the computer screen that represented the words from the respective minimal pair. They had to respond by button press (left vs. right arrow key) whether the left or right picture better matched the audio stimulus.

In each experiment, participants were randomly assigned (by the JavaScript program based on [9]) to the "same" condition or the "mixed" condition. Participants in the same condition only ever heard stimuli from the [ʁ]-[h] continuum. Participants in the mixed condition received alternating blocks from the continuum in which the rhotic was produced as [ʁ] or [r]. Each participant completed 30 blocks, where one block consisted of all 6 stimuli from a given continuum. Table 1 illustrates the design.

Block	Same condition	Mixed condition
1	uvular fricative	alveolar trill
2	uvular fricative	uvular fricative
3	uvular fricative	alveolar trill
...		
30	uvular fricative	uvular fricative

Table 1: Design of Experiments 1 and 2.

2.2. Analyses and Results

The two experiments were analysed separately but in identical manner. Analyses were conducted in R [10] using generalized linear mixed-effects models with a logistic linking function using the lme4 [11] and lmerTest [12] packages. In all models, the dependent variable was whether participants heard the word as /r/-initial or not. Fixed factors were Continuum Step, Block (both variables rescaled to range from -0.5 to 0.5), and Condition ("mixed" coded as 0.5, "same" as -0.5). Condition was allowed to interact with the other two factors. The models used the maximal random-effects structure that converged.

Two types of analyses were conducted. First, we compared responses to only the even numbered blocks, that is blocks, in which participants in both conditions responded to the [ʁ]-[h] continuum. Secondly, we compared the first 15 blocks of the same condition to the even numbered blocks in the mixed condition, again only comparing responses to the [ʁ]-[h] continuum. Here conditions matched on experience with the [ʁ]-[h] continuum.

Tables 2 and 3 report the statistics for the analyses of even numbered blocks in Experiments 1 and 2 respectively, and Fig. 1 and 2 illustrate the results. Analyses for both experiments show a significant effect of Condition, suggesting that fewer /r/-responses were given by participants in the mixed conditions. That is, if listeners are presented with rhotics produced as [r], then [ʁ] was less likely to be accepted as a rhotic than when only [ʁ] was heard. This was the case when different /r/-initial words were used in Experiment 1 as well as when the same word was heard in both conditions in Experiment 2.

The second analyses that took into account the amount of experience with [ʁ] showed similar results. The effect of Condition in Experiment 1 just reached significance ($b_{(\text{Condition})}$=-0.82, z=-1.98, p=0.048) as did the effect of Condition in Experiment 2 ($b_{(\text{Condition})}$=-1.15, z=-2.25, p=0.024). All data and analyses are made available on OSF [13].

Factor	b	z	p
Intercept	-0.20	-1.03	0.303
Step	3.66	13.72	< 0.001
Condition	-0.99	-2.52	0.012
Block	0.15	0.92	0.359
Step:Condition	0.61	1.25	0.260
Condition:Block	-0.42	-1.32	0.186

Table 2: Statistics for Experiment 1; comparison of even numbered blocks between conditions.

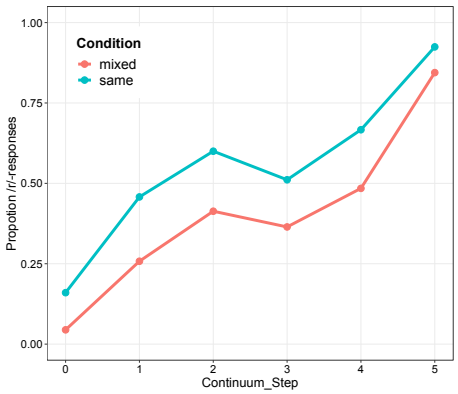

Figure 1: Results of Experiment 1: /r/-responses along the [ʁ]-[h] continuum depending on whether the speaker was also heard producing [r] on a different /r/-initial word.

Factor	b	z	p
Intercept	-0.93	-3.96	< 0.001
Step	2.95	7.44	< 0.001
Condition	-1.06	-2.28	0.023
Block	0.16	1.03	0.306
Step:Condition	0.42	0.54	0.587
Condition:Block	1.07	3.36	< 0.001

Table 3: Statistics for Experiment 2; comparison of even numbered blocks between conditions.

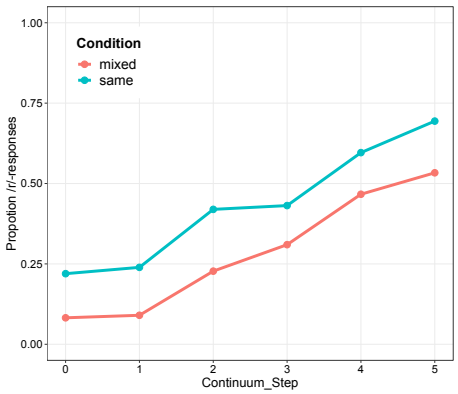

Figure 2: Results of Experiment 2: /r/-responses along the [ʁ]-[h] continuum depending on whether the speaker was also heard producing [r] on the same /r/-initial word.

2.3. Discussion

Experiments 1 and 2 confirmed our hypothesis that German listeners would be less willing to accept [ʁ] as an instance of the rhotic if they heard the speaker produce [r] as well. This effect is clearly visible in the descriptive data with a shift of roughly 15% fewer /r/-responses if an alveolar trill was heard. However, the effect is not strong statistically, just passing the p.<05 criterion. This is due to large interindividual differences in how often listeners perceive the rhotic on the [ʁ]-[h] continuum. However, our manipulation has to be between participants. Nevertheless, the effects were found in both types of analyses and replicated when the two rhotic allophones were heard on different words as well as on the same words.

3. EXPERIMENT 3

Experiment 3 addressed the question to what extent the decrease in /r/-responses to [ʁ] when also [r] was heard might be driven by listeners' expectation that a given speaker produces only one rhotic allophone. It therefore set out to replicate the mixed condition of Experiment 1, but with a male speaker producing the [r]-[l] continuum. The [ʁ]-[h] continuum was the same as in Experiment 1. We hypothesized that the effect of fewer /r/-responses in this new mixed rhotics/ different speaker condition might disappear relative to the mixed rhotics/ same speaker condition.

3.1. Methods

3.1.1. Participants

A new set of 40 participants was recruited via the online platform prolific.co [9]. Additionally, the data from 15 participants from Experiment 1 was re-used - those who had been assigned to the mixed condition.

2.1.2. Materials

The same [ʁ]-[h] continuum was used as in Experiment 1. In addition, a male native speaker of German recorded the minimal pair *Reiher-Leier* with the rhotic produced as [r]. According to the same criteria as reported above, tokens were selected and morphed into a 21-step continuum. Again, six steps were selected in a pre-test.

2.1.3. Design and Procedure

The "same speaker" condition here was identical to the mixed condition in Experiment 1. The female speaker was heard producing both, the [r]-[l] and [ʁ]-[h] continua, again in 30 alternating blocks, where one block refers to all 6 steps of a given continuum.

The "mixed speaker" condition was identical in design and procedure, with the only difference that the [r]-[l] continuum was produced by the male speaker. The randomization procedure assigned participants to the same- or mixed-speaker conditions with a probability of 1/3 vs. 2/3, which led to 12 versus 28 of the new participants assigned to the same versus mixed condition, respectively (15 were added to the same-speaker condition from Experiment 1).

3.2. Analyses and Results

We analyzed responses to the female speaker's [ʁ]-[h] continuum, which was identical across conditions. Condition now refers to whether the [r]-[l] continuum was spoken by the female or male speaker. The statistical model was identical to that in the first two experiments, again using the maximally converging random-effects structure (see OSF [13]). Table 4 reports the statistics and Fig. 3 illustrates the results.

Factor	b	z	p
Intercept	-0.43	-2.19	0.082
Step	3.99	15.96	< 0.001
Condition	-0.51	-1.32	0.187
Block	-0.31	-1.25	0.211
Step:Condition	-0.17	-0.35	0.725
Condition:Block	0.64	1.31	0.192

Table 4: Statistics for Experiment 3. Comparison of responses to the [ʁ]-[h] continuum.

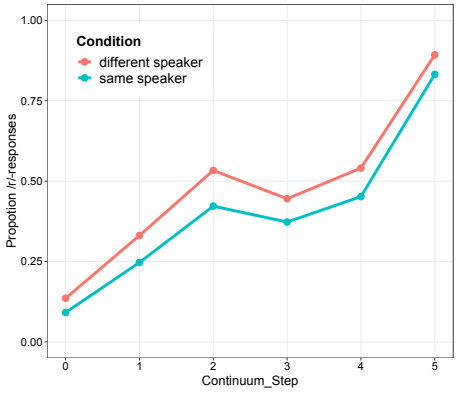

Figure 3: Results of Experiment 3: /r/-responses along the [ʁ]-[h] continuum spoken by the female speaker depending on whether the [r]-[l] continuum was produced by the same or a different speaker.

3.3. Discussion

Experiment 3 showed a numerical difference in how often [ʁ] was accepted as an instance of the rhotic between the conditions in which [r] was heard from the same versus a different speaker. This difference was, however, not statistically significant, due to the large interindividual differences in how often /r/ was perceived on the [ʁ]-[h] continuum. Therefore, we interpret the results of Experiment 3 as inconclusive. We cannot say whether the tendency to reject [ʁ] as an instance of the rhotic depends on hearing them from the same speaker.

4. GENERAL DISCUSSION

The present study built on the common assumption that trilled rhotics are perceptually "strong" or "salient" variants of rhotic phonemes [2,4,5]. It asked whether this strength of trills would affect the perception of other rhotic allophones, specifically the uvular fricative [ʁ]. We used German as the test language since a variety of rhotic allophones is found in native speakers, including [r] and [ʁ]. Moreover, as discussed in the introduction, a recent study [6] has shown that [r] triggers stronger selective adaptation than other rhotic allophones. The present study adds to this finding by showing that listeners are less willing to accept [ʁ] as a rhotic if also [r] is heard. This may be one of the reasons that [r] is able to remain the standard rhotic in so many languages despite its articulatory complexity.

That is, the data support hypothesis that the alveolar trill can be interpreted as a perceptually strong/ salient from of the rhotic. However, this finding might be somewhat surprising in the light of the language that we used to show the effect. While the saliency of the alveolar trill has been used as an argument to explain the prevalence of this allophone across languages despite its articulatory complexity [1], German does not entirely fit the pattern. Specifically, German has undergone a recent historical shift from [r] to [ʁ] as the standard rhotic allophone (for discussions see [4] and [5]).

We speculate that the fact that [ʁ] is considered the German standard, in addition to the occurrence of a large variety of rhotic allophones in German local varieties and dialects, might have contributed to the large interindividual differences in the present study. Note that a previous study on spoken-word recognition in German [14] showed that native speakers of German were faster to recognize /r/-initial words if they were produced with [ʁ] rather than [r]. Further research will therefore have to assess to what extent acoustic properties interact with language or speaker-specific characteristics of allophones during speech perception.

5. ACKNOWLEDGMENTS

This work was supported by a grant to (a.o.) HM by the Ministry of Education of the Republic of Korea

and the National Research Foundation of Korea (NRF-2021S1A5C2A02086884).

6. REFERENCES

[1] Ladefoged, P., Maddieson, I. (1996). *The Sounds of the World's Languages.* Oxford: Blackwell.
[2] Solé, M.-J. (2002). Aerodynamic characteristics of trills and phonological patterning. *J. Phon.* 30, 655–688.
[3] Sebregts, K. (2015). *The sociophonetics and phonology of Dutch r.* Poland: LOT Publications.
[4] Schiller, N. O. 1999. The phonetic variation of German /r/. In: Butt, M., Fuhrhop, N. (eds), *Variation und Stabilität in der Wortstruktur: Untersuchungen zu Entwicklung, Erwerb und Varietäten des Deutschen und anderer Sprachen*. Hildesheim, 261-287.
[5] Wiese, R. 2003. The unity and variation of German /r/. *Zeitschrift für Dialektologie Linguistik.* 70, 25–43.
[6] Mitterer, H., Reinisch, E. 2023. Selective adaptation of German /r/: A role for perceptual saliency. *Attention, Percept., Psychophys.*, 85, 222-233.
[7] Prolific, https://www.prolific.co/.
[8] Kawahara, H., Masuda-Katsuse, I., & de Cheveigné, A. 1999. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Comm.* 27, 187–207
[9] de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47, 1–12.
[10] R Core Team. 2021. R: A language and environment for statistical computing. (version 4.1.2) R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.R
[11] Bates, D. M., & Sarkar, D. 2007. lme4: Linear mixed-effects models using S4 classes (Version 0.999375–27) [Computer software]. http://www.r-project.org/: R Foundation for Statistical Computing.
[12] Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *J. Statist. Softw.* 82:13.
[13] OSF, supplementary materials. https://osf.io/2jpba/.
[14] 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.