

Published in final edited form as:
Proc Meet Acoust. ; 14: .

Temporal regularity in speech perception: Is regularity beneficial or deleterious?

Eveline Geiser* and Stefanie Shattuck-Hufnagel

Abstract

Speech rhythm has been proposed to be of crucial importance for correct speech perception and language learning. This study investigated the influence of speech rhythm in second language processing. German pseudo-sentences were presented to participants in two conditions: 'naturally regular speech rhythm' and an 'emphasized regular rhythm'. Nine expert English speakers with 3.5 ± 1.6 years of German training repeated each sentence after hearing it once over headphones. Responses were transcribed using the International Phonetic Alphabet and analyzed for the number of correct, false and missing consonants as well as for consonant additions. The over-all number of correct reproductions of consonants did not differ between the two experimental conditions. However, speech rhythmicization significantly affected the serial position curve of correctly reproduced syllables. The results of this pilot study are consistent with the view that speech rhythm is important for speech perception.

Introduction

Speech rhythm is of crucial relevance for speech acquisition and speech perception. For example, it facilitates speech segmentation (Cutler, 1996; Dilley & McAuley, 2008) and has been suggested to facilitate language acquisition, as prosodic aspects of speech are among the first speech characteristics infants learn to distinguish (Nazzi & Ramus, 2003).

Relatedly, it has been proposed that individual languages can be classified according to their preferred rhythmic, that is, temporal pattern. Researchers have searched for acoustic and phonological measures capturing an underlying temporal periodicity (isochrony) in speech (Abercrombie, 1967). Newer approaches refrain from this notion and suggest that speech does not rely on isochrony (Patel, 2008). It is suggested instead that the regular alternation between syllables, feet and phrases of various prominences, rather than the exact timing of these elements in the speech signal, is what induces a percept of regularity (Arvaniti, 2009; Dauer, 1983). Indeed, listeners anticipate prominence on the basis of the rhythmic pattern, which influences not only lexical and syntax processing but also word recognition (Dilley & McAuley, 2008; Snedeker & Casserly, 2010).

There is evidence that emphasizing the prominence pattern could be helpful to speakers who face speech challenges. For example, synchronous speaking is achieved by exaggerating the prominence pattern of a sentence. Moreover, speech impairment for example due to traumatic brain injuries is treated in a similar manner. Patients with traumatic brain injuries to the language areas of the left hemisphere are reported to benefit from treatment with melodic intonation therapy (MIT) in which stressed syllables are sung on the higher of two pitches (Schlaug, Norton, Marchina, Zipse, & Wan, 2011; Stahl, Kotz, Henseler, Turner, &

*Corresponding author's address: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Cambridge, MA 02139, egeiser@mit.edu.

Geyer, 2011). Thus, since syllables with higher pitch are often associated with phrase-level prominence, MIT seemingly comprises the exaggeration of accented syllable prominence.

Interestingly, both of the situations above involve not only an emphasized prominence pattern but also increased temporal periodicity. In fact the two closely interact. E.g. although MIT focuses on the melodic intonation of sentences, a periodic rhythmic pace is an integral part of the method as well. Namely, the patient's left hand is regularly tapped 1× per syllable (Norton, Zipse, Marchina, & Schlaug, 2009). From non-linguistic auditory processing we know that temporal regularity facilitates perception e.g. by reducing auditory thresholds (Ellis & Jones, 2010; Jones, Moynihan, MacKenzie, & Puente, 2002). This effect is associated with reduced processing load in the auditory cortex (Geiser, Notter, & Gabrieli, 2012) and explained by temporal expectation that is derived from the temporally regular structure of a sound sequence (Barnes & Jones, 2000; Chapin et al., 2010; London, 2004). Thus, it seems plausible that temporal periodicity and the associated exaggeration of syllable prominence in speech might facilitate speech perception and/or production in general.

The findings mentioned above led us to explore, in this pilot study, the effect of emphasized regular temporal patterns in speech on the perception and reproduction of difficult phonological sequences in healthy adults. Fitting speech into a more or less periodic temporal grid not only alters the rhythmic pattern of the syllables, but is also likely to result in an emphasized degree of prominence, that is, duration and intensity of accented syllables (Kochanski, Grabe, Coleman, & Rosner, 2005). Expert English speakers with some knowledge of German were presented with German pseudosentences, and were asked to repeat them immediately. English and German share rhythmic characteristics resulting in a categorization of both languages as 'stress-timed' (Abercrombie, 1967). At least for some speakers the two languages show similar values on statistical measures of speech rhythm (Grabe & Low, 2002). This rhythmic similarity seemed advantageous considering the difficulty the task represented for our participants. The pseudosentences were presented in two rhythmic conditions: spoken with a) a 'naturally regular' or 2) an 'emphasized regular' prosody. To create the emphasized regular speech stimuli, the speaker fitted prominent syllables into a more rigid temporal grid with the help of a metronome. We hypothesized that this emphasized regular speech rhythm would facilitate the number of correctly heard, remembered, and reproduced speech elements.

Methods

Participants

Nine volunteers (6 females; mean age of 20 ± 1.7 years) participated in the experiment. All participants were expert English speakers and had an average of 3.5 ± 1.6 years of German training, resulting in a self-attributed CEFR level of A2–B1 which ranges from elementary to threshold and intermediate. All participants gave written informed consent in accordance with the requirements of the MIT Institutional Review Board and were compensated for their participation.

Stimuli & Procedure

The stimulus material comprised a total of 30 German pseudo-sentences that were syntactically and phonotactically well formed but semantically meaningless.

Example: „Das Fristelon graft den tospen Blieger.“

Spoken renditions of the stimulus sentences were elicited from a trained, native speaker of German (female) in two conditions: 15 such sentences were produced with a naturally regular speech rhythm, and 15 different sentences were produced with emphasized

regularity. For this emphasized rhythmicity condition, the tempo of the speech was established by a metronome, resulting in the perceptually prominent occurrence of accented syllables (Geiser, Zaehle, Jancke, & Meyer, 2008), presumably with greater than usual temporal regularity. Sentences produced in the two rhythmic conditions were matched in syntax, CV-structure and number of syllables, and had an average of 10 syllables (range 8–11 syllables) per sentence. The utterances in the two experimental conditions did not differ with respect to mean duration ($t_{29} = 1.388$, $p = 0.176$). Furthermore, the utterances were matched for mean intensity on a root-means-square based measure. Consequently the presented stimuli did not differ on mean intensity ($t_{29} = -1.551$, $p = 0.132$) or peak intensity ($t_{29} = -1.645$, $p = 0.111$). However, there was a qualitative difference in the intensity envelope associated with our stimulus manipulation. That is, the naturally regular utterances compared to the emphasized regular sentences display a steeper decrease in the intensity of the prominent syllables (Figure 1).

The experiment adopted an immediate recall paradigm. Participants listened to the stimuli over headphones and were instructed to repeat each sentence into a microphone. The experiment was carried out in two blocks, with one block per experimental condition, and the serial order of the conditions was randomized over participants. The experiment lasted approximately 30 minutes.

Data analysis

The data analyses quantified the degree of CV similarity between each stimulus sentence and participant response. Participants' responses were transcribed by two trained phoneticians using the International Phonetic Alphabet (IPA). All of the consonants produced by the speaker were encoded and missing consonants were indicated by a placeholder. Vowels were coded as present or not present. Three measures of successful reproduction were analyzed in separate ANOVAs: the performance rate (average percent of correct consonant, syllable, and word reproductions per sentence), the error rate (average percent of missing, added, and wrong consonants per sentence), and the serial position curve (average percent of correct syllables per syllable position in the sentence). Inter-rater reliability was estimated using Pearson Product Moment correlation over both conditions. Descriptive statistics are given in mean percent of correct answers and standard errors per sentence and condition.

Results

Performance rate

There was a high correlation between the two raters with respect to correct consonant reproduction ($r = 0.904$, $n = 18$, $p < 0.001$), syllable reproduction ($r = 0.924$, $n = 18$, $p < 0.001$), and word reproduction ($r = 0.814$, $n = 18$, $p < 0.001$). The performance rate is thus reported as the average rating of the two raters. The number of correct reproductions of consonants (natural: $62.4 \pm 3.5\%$, emphasized: $63.5 \pm 3.2\%$), syllables (natural: $49.2 \pm 3.0\%$, emphasized: $49.0 \pm 4.0\%$) and words (natural: $38.0 \pm 3.4\%$, emphasized: $38.4 \pm 3.2\%$) did not differ between the two experimental conditions.

Serial position curve

There was a significant interaction between serial position and experimental condition observed ($F_{10,70} = 3.079$, $p < 0.05$, Figure 3). This interaction took the form of a higher correct response rate in the emphasized regularity condition toward the end of each utterance and a lower correct response rate in that condition in the beginning of each utterance. There was a main effect of position over the eleven syllable positions ($F_{10,70} = 12.769$, $p < 0.001$). No main effect of condition was observed.

Error analysis

There was high correlation between rater 1 and rater 2 on consonant omission ($r = 0.767$, $n = 18$, $p < 0.001$) and wrong consonants ($r = 0.937$, $n = 18$, $p < 0.001$). However, the correlation between rater 1 and rater 2 on the consonant intrusions ($r = 0.014$, $n = 18$, $p < 0.96$) was not significant. Rater 1 reported significantly more consonant intrusions in the emphasized regular speech ($5.9 \pm 0.6\%$) compared to the naturally regular speech ($4.3 \pm 0.8\%$, $F_{1,8} = 10.737$, $p < 0.05$, Bonferroni corrected). This effect was not present in the data of rater 2 (emphasized regular speech: $5.0 \pm 0.7\%$, naturally regular speech: $4.3 \pm 0.7\%$, $F_{1,8} = 0.763$, $p = 0.41$). Consonant omissions and wrong consonants did not differ between experimental conditions (Figure 4).

Discussion

This study is an initial attempt to quantify how emphasized regular speech affects the reproduction of phonological sequences that are difficult for the participants because they are in the participants' incompletely-mastered second language (L2). Participants' responses were analyzed on the basis of the number of correctly reproduced consonants per sentence and per serial position in the sentence, as well as of the number of wrong consonants, consonant intrusions and consonant omissions. The measures for correctly reproduced consonants showed high inter-rater agreement, indicating that the results are reliable.

Contrary to our prediction, emphasized regular compared to naturally regular speech rhythm did not improve the number of correctly reproduced speech elements. That is, listeners did not benefit from emphasized regularity in the presented renditions. Since both the naturally regular and the emphasized regular rhythm resulted in the same performance rate, we must assume that neither the perception, the remembering, nor the reproduction was facilitated by the combination of prominence exaggeration and regular timing in the emphasized-regularity condition. The participants correctly reproduced an average of 6 out of 10 consonant combinations, which is slightly below the average serial recall rate for unrelated items in young adults (Cowan, Saults, Elliott, & Moreno, 2002). We originally hypothesized that emphasized regularity increases sensory perception and subsequent reproduction. However, since no increase was observed, at least in phonological reproduction, we must assume that participants gained no additional benefit from emphasized regular speech, at least in this difficult task. The beneficial effect of emphasized speech rhythm might thus apply to circumscribed situations only, such as to patients suffering from injury related speech difficulties, but not to healthy young adults carrying out a second language reproduction task. Although one might argue that they were not sensitive to the stimulus differences, this seems unlikely considering that the serial position curve differed between experimental conditions as discussed in the next sections.

The distribution of correctly reproduced consonants as each utterance unfolds over time was different between the experimental conditions (Figure 3). That is, naturally regular speech compared to emphasized regular speech significantly increased the percent of correct consonant reproductions at the beginning of the sentence. This increase in performance was achieved at the cost of lower performance in the end of the sentence. This finding partially parallels the qualitative difference in the intensity envelope between the two experimental conditions, in the sense that the higher intensity of the first syllable in the naturally regular speech condition could explain the increased performance at the beginning of this utterance. However, the performance in the middle and at the end of the sentences does not parallel the intensity envelope. Thus, the difference in performance per syllable position seems not directly related to the syllable prominence as measured by the intensity envelope.

An alternative interpretation focuses the timing differences between the two conditions. Participants might have had a propensity to process the individual syllables as less related to each other in the emphasized regular speech than in the naturally regular speech. Indeed, the performance pattern in the emphasized regular condition parallels a serial position curve for unrelated items, such as the reproduction of single consonants or numbers (Cowan et al., 2002). This serial position curve shows better performance in the beginning and in the end compared to the middle of the sequence, an effect commonly understood as the combination of a 'primacy effect' and a 'recency effect'. It seems as if the emphasized regular speech rhythm condition altered the performance rate so that the serial reproduction curve approximated the response curve expected for unrelated item reproductions, at least at the ends of the utterances. It is important to note that all pseudo-sentences presented in this experiment had a grammatically well-formed syntactic structure and, consequently, listeners were able to identify the subject noun, the object noun, the adjective, and the verb in all sentences. We therefore interpret this finding as possible evidence that sentences with naturally regular speech rhythm trigger the processing of the relationship between the individual syllables and words of the sentences. This consequently affects the perceptual grouping which in turn will modulate the serial position curve in the reproduction task (Cowan et al., 2002). However, we must consider the possibility that the speech rhythm affected not only the perception and reproduction of consonants by the listener, but also the production by the speaker who originally produced the stimuli, potentially affecting the intelligibility of the consonants. This alternative interpretation will need careful consideration in future experiments. For now, we interpret the observed difference in serial reproduction as an effect of perceptual grouping. This interpretation focuses on the importance of rhythm in speech perception.

While the phonological performance rate across the whole sentence did not differ between the two experimental conditions, emphasized regular speech stimuli resulted in significantly more consonant intrusions than did naturally regular stimuli (Figure 4). This effect was only present in the coding of one rater and the coding of the two raters did not correlate on this measure either. We assume that the introduced consonants were not as clearly spoken as the correctly reproduced ones, so that the two raters did not agree on their intelligibility; the low number of intrusions may have increased the statistical effect. Although this finding could suggest that emphasized regular speech encourages participants to insert additional consonants, – possibly due to an induced rhythmic impetus which reduces carefulness or self-judgment –, future investigations will be needed to further corroborate this finding. For now, this finding primarily highlights the importance of multiple rater comparisons for speech production experiments.

In sum, the reproduction of difficult phonological sequences of a foreign language in healthy adults does not benefit from emphasized regularity and the related increase in prominence of accented syllables in the stimulus renditions. Future studies should focus on the temporal context of the reproduction as a potential influence on performance. However, the manipulation of syllable prominence clearly influences which elements of the sentence are remembered. This indicates that speech rhythm influences speech processing.

Acknowledgments

This research was supported by the Swiss National Science Foundation (SNF; PBZHP1-123304) and by the National Institutes of Health (R01DC8780). We thank Dana Bullister and Christine Park for the coding of the participants' responses. We thank the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig for assistance in stimulus recording.

References

- Abercrombie, D. Elements of general phonetics. Edinburgh: Edinburgh University Press; 1967.
- Arvaniti A. Rhythm, timing and the timing of rhythm. *Phonetica*. 2009; 66:46–63. [PubMed: 19390230]
- Barnes R, Jones MR. Expectancy, attention, and time. *Cognitive Psychology*. 2000; 41(3):254–311. [PubMed: 11032658]
- Chapin H, Zanto TP, Jantzen KJ, kelso JAS, Steinberg F, Large EW. Neural responses to complex auditory rhythms: The role of attending. *Frontiers in Auditory Cognitive Neuroscience*. 2010; 1:224.
- Cowan N, Saults JS, Elliott EM, Moreno MV. Deconfounding serial recall. *Journal of Memory and Language*. 2002; 46:153–177.
- Cutler, A. Prosody and the word boundary problem. In: Morgan, JL.; Demuth, K., editors. Signal to syntax: Bootstrapping from speech to grammar in early acquisition. 1996. p. 87-100.
- Dauer RM. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*. 1983; 11:5–62.
- Dilley LC, McAuley D. Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*. 2008; 59:294–311.
- Ellis RJ, Jones MR. Rhythmic context modulates foreperiod effects. *Attention Perception & Psychophysics*. 2010; 72(8):2274–2288.
- Geiser E, Zaehle T, Jancke L, Meyer M. The neural correlate of speech rhythm as evidenced by metrical speech processing. *Journal of Cognitive Neuroscience*. 2008; 20(3):541–552. [PubMed: 18004944]
- Geiser E, Notter M, Gabrieli JDE. A cortico-striatal neural system enhances auditory perception through temporal context processing. *Journal of Neuroscience*. 2012; 32(18):6177–6182. [PubMed: 22553024]
- Grabe, E.; Low, EL. Acoustic correlates of rhythm class. In: Gussenhoven, C.; Warner, N., editors. *Laboratory phonology*. Vol. 7. Berlin/New York: Mouton de Gruyter; 2002. p. 515-546.
- Jones MR, Moynihan H, MacKenzie N, Puente J. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*. 2002; 13(4):313–319. [PubMed: 12137133]
- Kochanski G, Grabe E, Coleman J, Rosner B. Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America*. 2005; 118(2):1038–1054. [PubMed: 16158659]
- London, JM. *Hearing in time: Psychological aspects of musical meter*. New York: Oxford University Press; 2004.
- Miller GA. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*. 1956; 63:81–97. [PubMed: 13310704]
- Nazzi T, Ramus F. Perception and acquisition of linguistic rhythm by infants. *Speech Communication*. 2003; 41(1):233–243.
- Norton A, Zipse L, Marchina S, Schlaug G. Melodic intonation therapy: Shared insights on how it is done and why it might help. *Annals of the New York Academy of Sciences*. 2009; 1169:431–436. [PubMed: 19673819]
- Patel, AD. *Music, language and the brain*. New York: Oxford University Press, Inc.; 2008.
- Schlaug G, Norton A, Marchina S, Zipse L, Wan CY. From singing to speaking: Facilitating recovery from nonfluent aphasia. *Future Neurology*. 2011; 5(5):657–665. [PubMed: 21088709]
- Snedeker J, Casserly E. Is it all relative? effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*. 2010; 25(7–9):1234–1264.
- Stahl B, Kotz SA, Henseler I, Turner R, Geyer S. Rhythm in disguise: Why singing may not hold the key to recovery from aphasia. *Brain*. 2011; 134:3083–3093. [PubMed: 21948939]

average intensity envelope

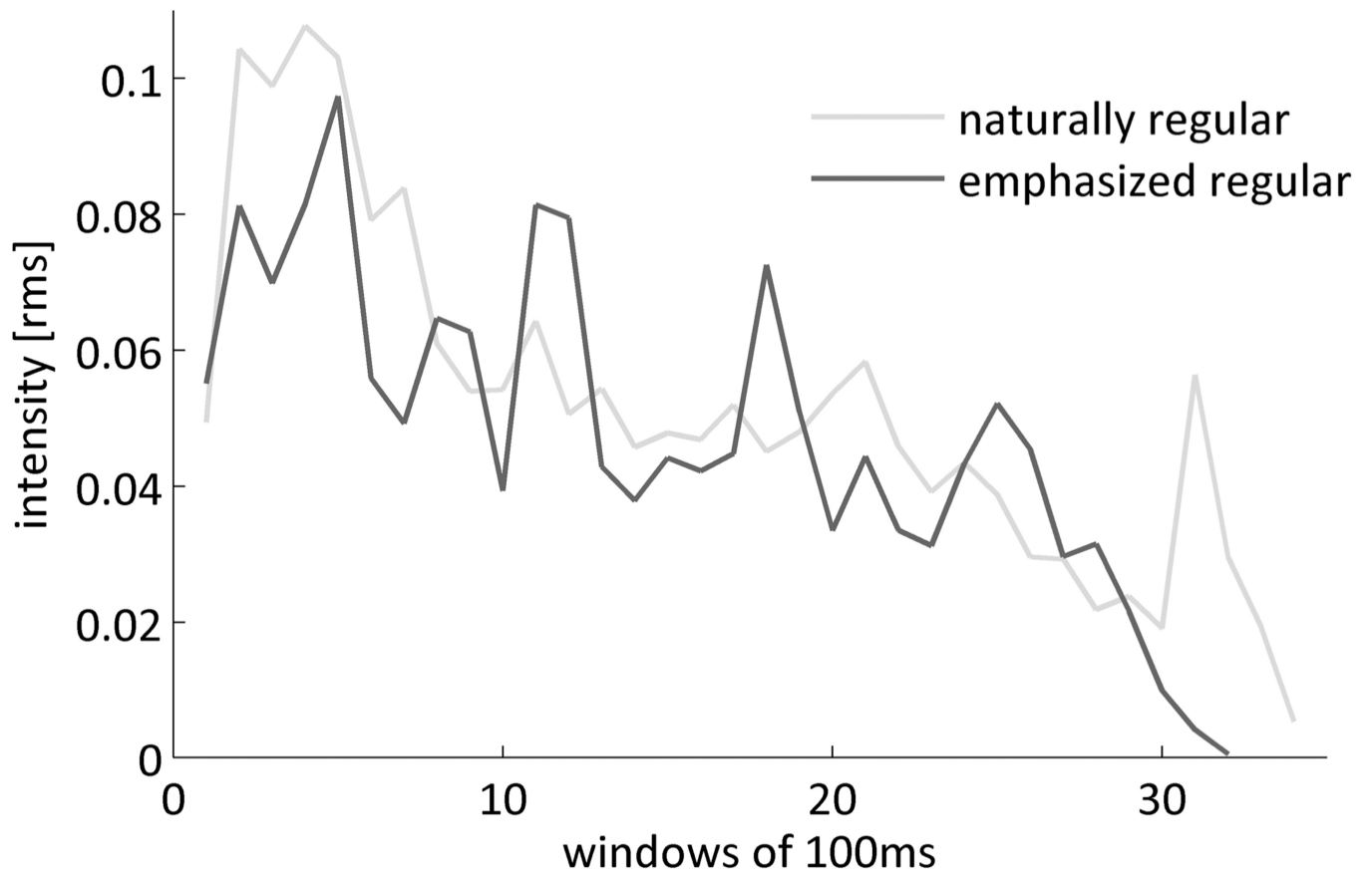


Fig 1. The figure shows the average intensity envelope of naturally regular and emphasized regular experimental stimuli.

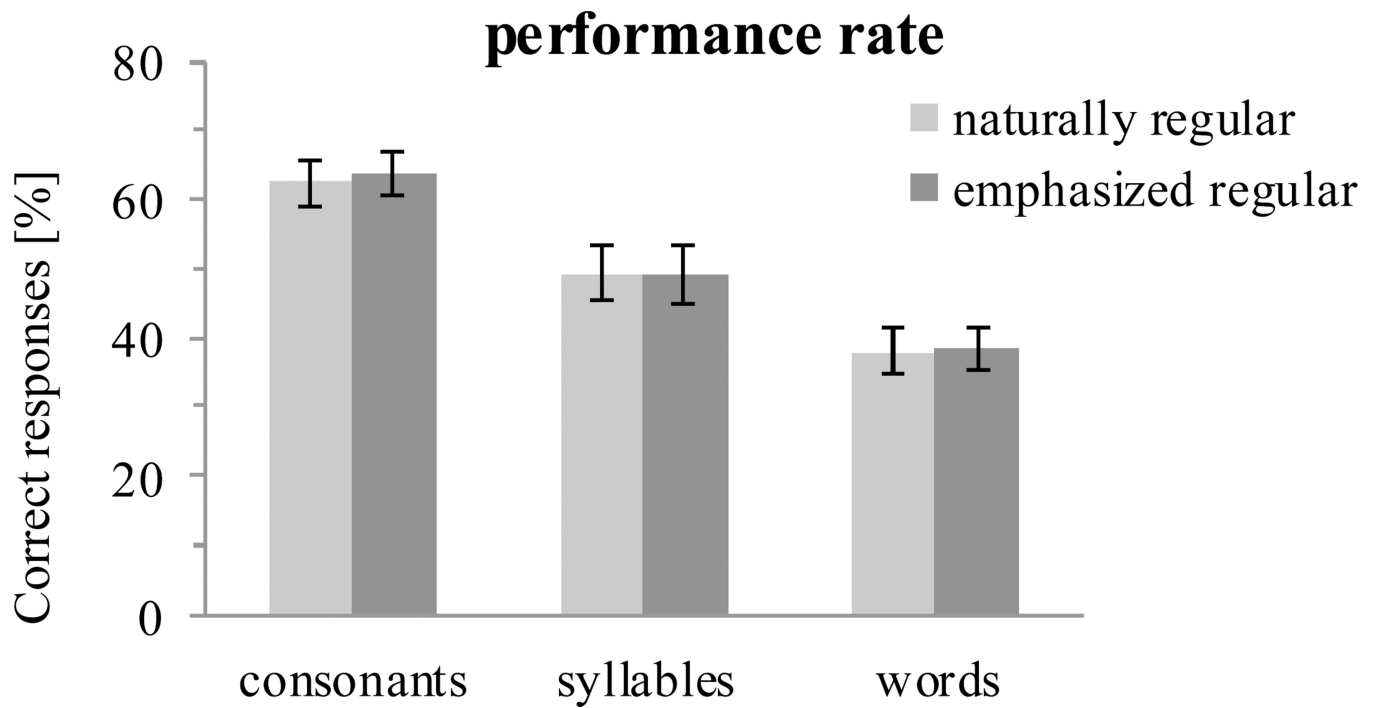


Fig 2.
The figure shows the % of correct responses as average of the two raters for consonants, syllables, and words, separately for the two experimental conditions. The two rhythmic conditions did not differ on any of these measures.

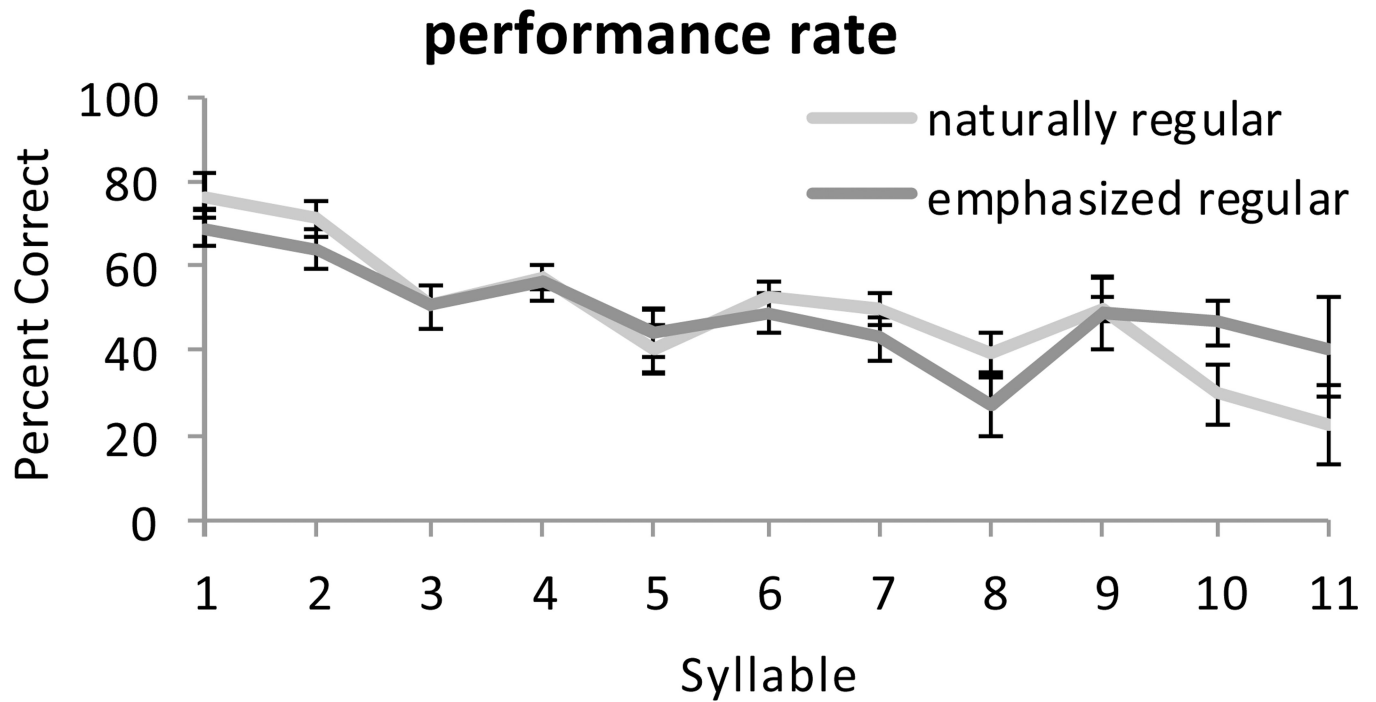
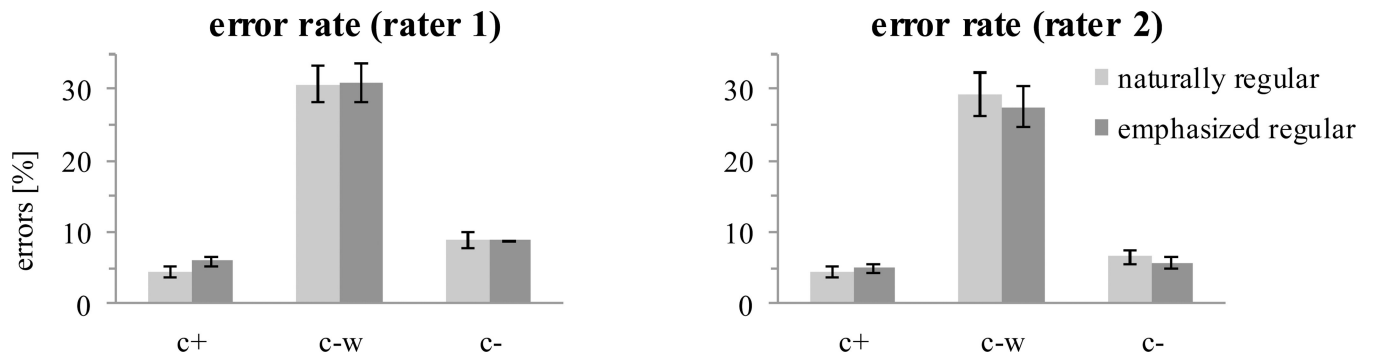


Fig 3. The figure shows the % correct consonants per syllable position as average of the two raters. There was a significant interaction between the experimental conditions and syllable position.

**Fig 4.**

The figure shows the % of consonant intrusions (c+), consonant omissions (c-) and wrong consonants (c-w) separately for the two experimental conditions. The left plot displays the estimates of rater 1 and the right plot displays the estimates of rater 2. There was a significant difference between the experimental conditions for consonant intrusions as estimated by rater 2. No other effects were observed.