Trochaic foot benefit in speech comprehension

Mathias Scharinger[1,3,4]
Ulrike Domahs[2,3,4]

[1]Research Group »Phonetics«, Institute of German Linguistics, Philipps-Universität Marburg, Germany
[2]Research Group »Neurolinguistics«, Institute of German Linguistics, Philipps-Universität Marburg, Germany
[3]Research Center »Deutscher Sprachatlas«, Philipps-Universität Marburg, Germany
[4]Center for Mind, Brain & Behaviour »CMBB«, Universities of Gießen & Marburg, Germany

emails: mathias.scharinger@uni-marburg.de; ulrike.domahs@uni-marburg.de

ABSTRACT

There is evidence that speech comprehension benefits from prosodic regularities, possibly through segmentation of continuous acoustic signals into foot structure. We probed this assumption in a modified Sternberg memory task, using 8-digit-numbers with either trochaic or iambic grouping obtained by prosodic manipulations. A visual 4-digit probe number was used to test recall performance. The probe was either congruent or incongruent with the foot structure of the 8-digit-numbers. Recall performance was measured by accuracy and reaction times. Accuracy was significantly better for probes congruent with the foot structure, evidencing that feet are indeed used during the encoding of speech. Reaction times for correct recall were faster for probes with congruent foot structure and in the trochaic condition. Together, these findings provide evidence for the role of feet during speech comprehension in a memory task during speech comprehension, and for a trochaic bias in German.

Keywords: speech comprehension, prosody, foot structure, trochee

1. INTRODUCTION

There is mounting evidence that speech comprehension involves a segmentation mechanism describable by oscillator models [1-3]. The main assumption is that the continuous speech signal is segmented into temporally discrete units that are optimal for working-memory processing [1, 4]. The sizes of these units roughly correspond to the syllable (150-300 ms) and to short phrases (500-2000 ms). Correspondingly, the neural oscillations underlying the segmentation process during speech perception are most prominently in the theta- (3-7 Hz) and in the delta-band (0.5-2 Hz) [5, 6]. In this tradition, Ghitza [1] has focused on delta-oscillations, suggesting their particular role for prosodic segmentation. In his study, he used a modified Sternberg memory task wherein 10-digit numbers, grouped either regularly (2-2-2-2-2 pattern) or irregularly (3-3-2-2 pattern), had to be memorized. Participants were auditorily presented with these number strings, followed by a 2- or 3-digit probe. Probes could be either inside a chunk (i.e., **589** from the string 334 **589** 33 22, grouping congruency), be split between two successive chunks (i.e., **933** from the string 334 58**9** **33** 22, grouping incongruency) or not at all be contained in the string (i.e., **555**). The results of Ghitza's study demonstrated an effect of grouping congruency: Error rates significantly increased if probes were incongruent to the grouping pattern of the 10-digit strings. Furthermore, error rates also increased for grouping patterns whose cycle durations deviated from the delta-oscillation (>3 Hz).

While these results have provided a mechanistic explanation for the importance of delta-oscillations in speech segmentation, the precise functional relevance of the particular segmentation (or grouping) has rather been left unexplored. As delta-oscillations correspond to temporal windows which may comprise of up to two syllables, a possible functional role of delta could be the tracking of binary foot structure in speech comprehension. This assumption will be tested in the below-described experiment, attempting to answer the research question of whether delta-oscillations provide a foot-based segmentation of continuous speech, capitalizing on prosodic regularities and perhaps the trochaic bias in German [7, 8].

2. CURRENT STUDY

The rationale of this study is to explore the influence of prosodic foot structure on speech comprehension. Following Ghitza [1], we assume that the encoding of continuous speech information benefits from prosodic regularities, and in particular, foot structure. We hypothesize that strong-weak (trochaic) and weak-strong (iambic) prosodic patterns drive delta oscillations. Assumedly, segmenting the speech signal into delta-sized units (i.e., between 500-2000 ms) provides discrete information for optimized working-memory processing. We thus also use a modified Sternberg memory task, with the following important differences compared to Ghitza's study [1]: (1) 8-digit number strings, regular grouping (2-2-2-2); (2) prosodic manipulations to either conform to a trochaic (strong-weak) or an iambic (weak-strong) pattern; (3) cross-modal design, with auditorily presented 8-digit number strings and visually presented 4-digit probe strings. Using a cross-modal design, we hope to avoid that congruency effects stem merely from acoustic matching between the large string and the probe.

2.1. Methods

2.1.1. Stimuli

Stimulus construction was based on naturally spoken single numbers of German. A female native speaker of German pronounced the numbers from 0 to 9 in several renditions, which were recorded on a Windows PC with 44.1 kHz sampling rate and 16-bit amplitude resolution. The best token per number (based on comparability in intensity and pitch) was selected for further stimulus construction. The two-syllable number "7" was realized as monosyllable ([ziːm]). Subsequently, all natural tokens were lengthened to 250 ms using the overlap-add algorithm in PRAAT [9]. A silence of 50 ms was added to all renditions, resulting in 300 ms recordings of each number.

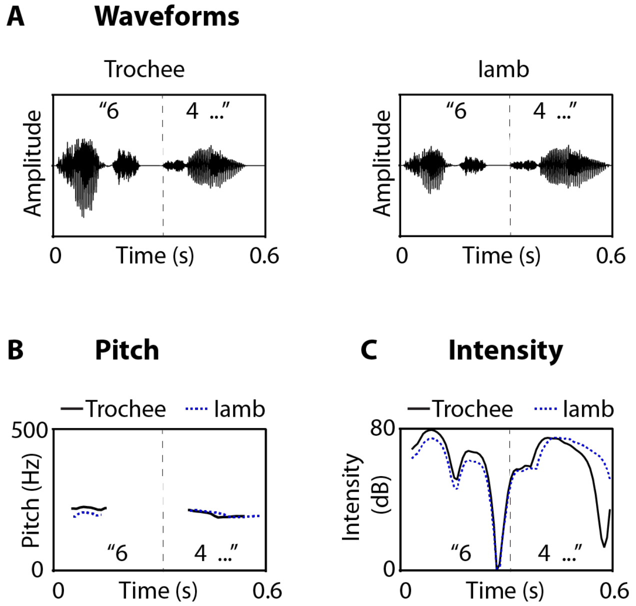


Figure 1: Illustration of stimulus construction. Naturally-produced single digits were modified to either elicit a strong-weak (trochee) or weak-strong (iamb) grouping (A). In odd-numbered positions in the 8-digit strings, pitch and intensity were enhanced (pitch: 20 Hz, intensity: 5 dB), yielding the strong-weak (trochaic) pattern. In even-numbered positions in the 8-digit strings, duration was enhanced (from 250 to 300 ms), yielding the weak-strong (iambic) pattern (B, C).

The 8-digit strings were then constructed by concatenating a random selection of the digits between 0 and 9 (without replacement). Strong-weak (trochaic) and weak-strong (iambic) prosodic patterns were achieved based on the insights of [8] (see Figure 1). In total, 100 8-digit strings with a trochaic and 100 8-digit strings with an iambic pattern were constructed. For half of the trochaic and iambic strings, a 4-digit number was selected as probe, being a proper subset (e.g., **4183** from 64183297). 50% of these probes were congruent with the foot structure of the 8-digit string, and 50% were not congruent (as in the above-example, corresponding to the split-chunk-condition of Ghitza [1]). For the other half of the trochaic and iambic strings, the 4-digit probe was not a proper subset (e.g., **4138** from 64183297). Note that the first digit pair of these probes was always contained in the larger string, congruent or incongruent with the foot structure, while the second digit pair provided the mismatching information. Thereby, a fully-crossed 2x2x2 design (foot: trochee/iamb; congruency: congruent/incongruent; match: yes/no) could be achieved.

2.1.2. Design

The 2x2x2 design (see Figure 2) determined the Sternberg memory task in which an auditory 8-digit string was followed by a visual probe that could be either contained in the larger string (match: yes) or not (match: no). In half of these cases, the probe was congruent or incongruent with the foot structure (grouping) of the larger string.

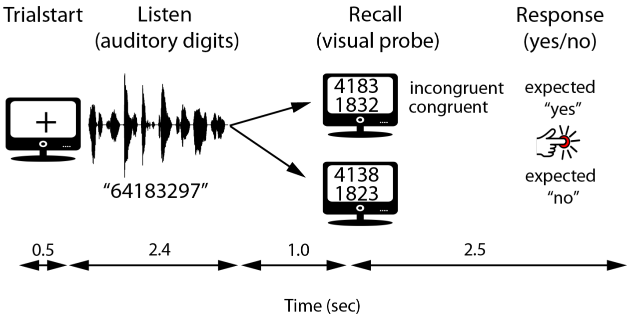


Figure 2: Trial structure. A trial started with a fixation cross displayed for 0.5 sec. The auditory 8-digit string was presented subsequently. After 1-sec pause, the probe was displayed on the screen. Participants had then to respond "yes" if the probe was contained in the string and "no" if it was not.

2.4. Participants

Participants (N=20) were native speakers of German (mean age: 27 yrs, SD: 9 yrs., 14 females, 5 males 1 diverse). They were right-handed and had normal hearing. Participants provided written informed consent before the experiment started. The study was approved by the Ethics Committee of the German Society for Linguistics (DGfS) and in accordance to the declarations of Helsinki. Participants received monetary compensation for their participation.

2.5. Procedure

The memory task was conducted in a shielded chamber because participants’ EEG was recorded in addition to their behavioural responses (data presented elsewhere). Participants were seated approximately 2 m in front of a computer screen (27-inch). A trial started with a fixation cross displayed in the middle of the screen for 0.5 secs. Subsequently, participants heard the 8-digit strings over loudspeakers (2.4 sec). After a pause of 1 sec, the probes were displayed in the middle of the screen and participants had 2.5 secs to give their responses on a computer keyboard, using the keys “f” and “j”. Response-to-key attribution to indicate a probe match- or mis-match was counter-balanced across participants. Each participant received a different randomization of the 200 trials. Trials were presented with the open-source software OpenSesame [10]. The experiment lasted for about 30 minutes.

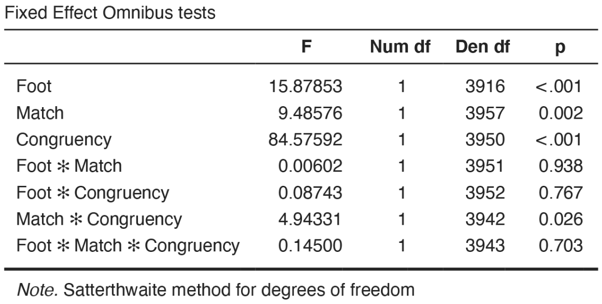
2.7. Statistics

Reaction times (starting from the onset of probe presentation) were used as dependent variable in a Linear Mixed Effects (LME) model, with *subject* (participant ID) and *item* (Trial) as random effects (intercept only). The fixed effects comprised *match* (yes/no), *foot* (trochee/iamb) and *congruency* (congruent/incongruent). Accuracy was analysed as logistic dependent variable (0: incorrect, 1: correct detection) in a Generalized Mixed Effects model with the same effect structure as in the reaction time model. Additionally, participants’ response sensitivity was modelled with *d’* in a Linear Mixed Effects model with *subject* (participant ID) as random effect (intercept only) and the fixed effects *foot* and *congruency*. Statistical analyses were calculated in jamovi [11].

3. RESULTS

The LME model on reaction times showed main effects of *match, foot* and *congruency*. Reaction times were faster for mismatching compared to matching probes (t=-3.08, p<0.01), for trochaic compared to iambic digit strings (t=-3.98, p<0.001) and for congruent compared to incongruent probes (t=-9.2, p<0.001). The interaction of the effects *match* and *congruency* was based on larger differences between congruent and incongruent probes when the probe was contained in the larger string than when it was not.

The GME model on accuracy revealed main effects for *match* and *congruency*. Probes not contained in the larger digit strings and probes congruent with the foot structure of these larger strings were responded to more accurately (z=19.8, p<0.001; t=6.75, p<0.001).



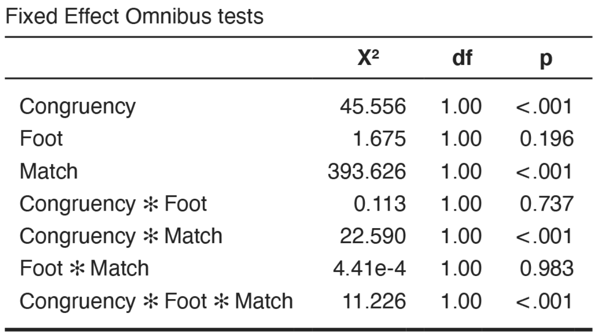
Fixed Effect Omnibus tests

	F	Num df	Den df	p
Foot	15.87853	1	3916	<.001
Match	9.48576	1	3957	0.002
Congruency	84.57592	1	3950	<.001
Foot ✻ Match	0.00602	1	3951	0.938
Foot ✻ Congruency	0.08743	1	3952	0.767
Match ✻ Congruency	4.94331	1	3942	0.026
Foot ✻ Match ✻ Congruency	0.14500	1	3943	0.703

Note. Satterthwaite method for degrees of freedom

Table 1: Results of the LME model on reaction times.

The interaction of the effects *match* and *congruency* was based on larger accuracy differences between incongruent and congruent probes when the probes were contained in the larger strings than when they were not. The three-way interaction of *match*, *congruency* and *foot* resulted from a stronger congruency effect in matching probes when the large digit strings had trochaic compared to iambic foot structure. Accuracy dropped almost to chance level when probes were contained in the trochaic 8-digit strings but were incongruent with the foot structure.



Fixed Effect Omnibus tests

	X^2	df	p
Congruency	45.556	1.00	<.001
Foot	1.675	1.00	0.196
Match	393.626	1.00	<.001
Congruency ✻ Foot	0.113	1.00	0.737
Congruency ✻ Match	22.590	1.00	<.001
Foot ✻ Match	4.41e-4	1.00	0.983
Congruency ✻ Foot ✻ Match	11.226	1.00	<.001

Table 2: Results of the GME model on accuracy.

The LME model on *d'* showed a main effect of congruency (F(1,57)=38.13, p<0.001), based on higher *d'* for congruent than for incongruent probes (t=6.17, p<0.001, see Figure 3).

4. DISCUSSION

The results of the memory task suggest that the prosodically induced grouping of numbers is beneficial for memory encoding, particularly if the grouping adheres to a trochaic (strong-weak) pattern. Furthermore, the congruency effect, replicating the "split-chunk" finding of Ghitza [1], provides evidence that foot structure plays an important role during encoding and/or retrieval from memory.

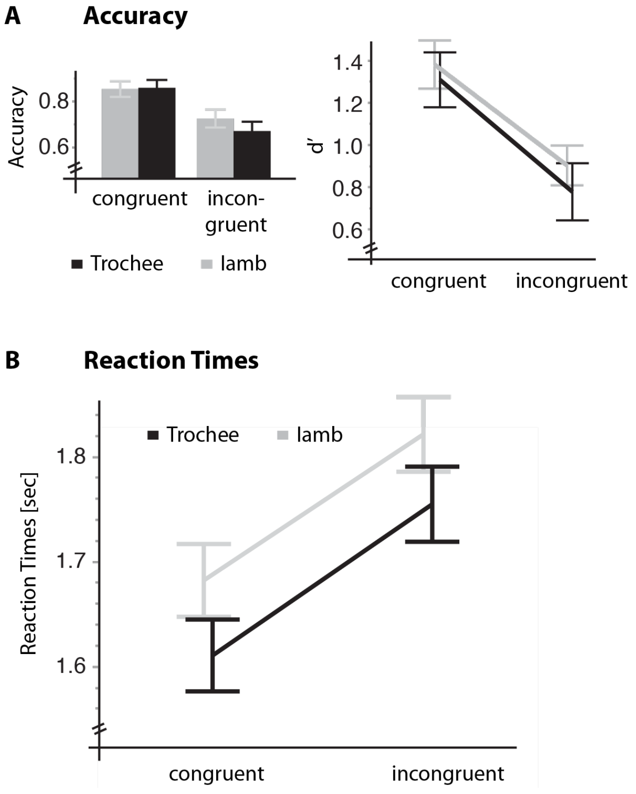


Figure 3: Summary of results. A. (left): Accuracy (proportion correct responses), pooled across matching/mismatching probes. (right) The *d'* analysis corroborates the accuracy pattern, with higher *d'* for congruent than incongruent probes. B. Reaction times differed between trochaic and iambic digit-string patterns and between congruent and incongruent probes, with probes congruent to trochaic digit-strings being responded to fastest. Error-bars/whiskers indicate standard errors of the means.

Considering the experimental memory task as a proxy to speech comprehension, the aforementioned research question can be answered as follows.

In addition to Ghitza's findings [1], the present study's results suggest that the encoding of continuous speech information benefits from prosodic regularities in terms of foot structure. Segmenting the speech signal into delta-sized units (i.e., between 500-2000 ms) seems to provide chunks for optimized working-memory processing, supporting [12] who showed that delta-sized units are more related to meaningful utterances compared to theta-sized units that relate to syllable units. In order to test whether the type of foot structure plays a role for the memory task, we compared the performance for trochees and iambs that varied with regard to the position of the prosodic prominence as well as the acoustic properties establishing prominence, following the iambic/trochaic law [13]. As can be seen in Figure 3B, reaction times were significantly shorter for trochaic patterns, irrespective of the congruency of the 4-digit probe with the 8-digit sequence. According to the oscillation models, we would have expected that both disyllabic structures drive delta oscillations and facilitate the chunking of all congruous sequences and their memorability. The significant advantage of trochees over iambs, however, speaks in favour of strong-weak segmentation strategies and sensitivity to specific prosodic cues (pitch and intensity) as has previously been observed in first language acquisition [14] and adult processing [15] (but see [7] for iambic and trochaic grouping preference in German monolinguals and French-German bilinguals). The latter study suggests that future research has to include languages with iambic preferences or languages in which word stress patterns are less important than in German or English, as is the case in French, for instance. The inclusion of such languages enables us to test whether working memory abilities can be improved by chunks of the shape of any language-specific prosodic pattern or – for languages without any word stress pattern – does not play a role in the recall of sequences at all. Moreover, future research has to also speak to the (alleged) universality of the Iambic/Trochaic law [8, 15-17] as well as to (possibly) universal constraints on working memory [18, 19].

5. CONCLUSION

Taken together, the congruency effect for trochees shows that delta-sized units are only advantageous for segmentation and processing if the processing units exhibit initial prominence.

6. REFERENCES

[1] O. Ghitza, 2017. Acoustic-driven delta rhythms as prosodic markers, *Lang Cogn Neurosci,* 32, 545-561

[2] L. Henke and L. Meyer, 2021. Endogenous Oscillations Time-Constrain Linguistic Segmentation: Cycling the Garden Path, *Cereb. Cortex,* 31, 4289-4299

[3] B. Zoefel and R. VanRullen, 2015. EEG oscillations entrain their phase to high-level features of speech sound, *Neuroimage,* 124, 16-23

[4] A. Baddeley, 1992. Working memory, *Science,* 255, 556-559

[5] N. Ding and J. Z. Simon, 2014. Cortical entrainment to continuous speech: functional roles and interpretations, *Frontiers in Human Neuroscience,* 8,

[6] A.-L. Giraud and D. Poeppel, 2012. Cortical oscillations and speech processing: emerging computational principles and operations, *Nat. Neurosci.,* 15, 511-517

[7] N. Boll-Avetisyan, A. Bhatara, A. Unger, T. Nazzi, and B. Höhle, 2020. Rhythmic grouping biases in simultaneous bilinguals, *Bilingualism: Language and Cognition,* 23, 1070-1081

[8] M. J. Crowhurst and A. T. Olivares, 2014. Beyond the Iambic-Trochaic Law: The joint influence of duration and intensity on the perception of rhythmic speech, *Phonology,* 31, 51-94

[9] *PRAAT: Doing phonetics by computer (version 6.1.24)*. (2020). Institut for Phonetic Sciences, Amsterdam.

[10] S. Mathôt, D. Schreij, and J. Theeuwes, 2012. OpenSesame: An open-source, graphical experiment builder for the social sciences, *Behavior Research Methods,* 44, 314-324

[11] D. J. Navarro and D. R. Foxcroft. 2019. *Learning Statistics with Jamovi: A Tutorial for Psychology Students and Other Beginners.* .

[12] Victor J. Boucher, Annie C. Gilbert, Boutheina Jemel; The Role of Low-frequency Neural Oscillations in Speech Processing: Revisiting Delta Entrainment. *J Cogn Neurosci* 2019; 31 (8): 1205–1215.

[13] B. Hayes, 1985. Iambic and trochaic rhythm in stress rules, *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society,* 429-446

[14] K. Johnson and P. W. Jusczyk, 2001. Word segmentation by 8-month-olds: When speech cues count more than statistics, *J. Mem. Lang.,* 44, 548-567

[15] D. M. de la Mora, M. Nespor, and J. M. Toro, 2013. Do humans and nonhuman animals share the grouping principles of the Iambic - Trochaic Law?, *Atten. Percept. Psychophys.,* 75, 92-100

[16] M. J. Crowhurst, 2020. The iambic/trochaic law: Nature or nurture?, *Language and Linguistics Compass,* 14, e12360

[17] A. Revithiadou, 2004. The iambic/trochaic law revisited: Lengthening and shortening in trochaic systems, *Leiden Papers in Linguistics,* 1, 37-62

[18] K. Schulze and S. Koelsch, 2012. Working memory for speech and music, *Annals of the New York Academy of Sciences,* 1252, 229-236

[19] J. Segawa, M. Masapollo, M. Tong, D. J. Smith, and F. H. Guenther, 2019. Chunking of phonological units in speech sequencing, *Brain Lang.,* 195, 104636

Acknowledgments:

We wish to express our thanks to Sharon Stürmer and Michelle Morgenstern for their invaluable help with data collection.