Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Prosodic grouping in Akan and the applicability of the iambic-trochaic law

Constantijn Kaland [1], Anjali Bhatara [2], Natalie Boll-Avetisyan[3], Thierry Nazzi[4]

[1]Institute of Linguistics, University of Cologne, Germany
[2]Independent researcher
[3]Cognitive Sciences Unit, Department of Linguistics, University of Potsdam, Germany
[4]Integrative Neuroscience and Cognition Center, CNRS - Université Paris Cité, France

ckaland@uni-koeln.de, bhatara@gmail.com, nboll@uni-potsdam.de, thierry.nazzi@parisdescartes.fr

Abstract

The current study investigates prosodic grouping in Akan (iso: aka), a Kwa language. Except for tone, the prosody of Akan is largely understudied. Even fewer studies have addressed the perception of prosody. The current study investigates how prosodic cues such as duration and intensity contribute to grouping in this language by a replication of a perception experiment used in earlier work. In this experiment, participants indicate whether auditory sequences are composed of repetitions of weak-strong or strong-weak sound pairs. The experiment was designed to test the predictions of the iambic-trochaic law (ITL). The ITL predicts that auditory sequences alternating in duration are grouped as iambs (weak-strong), whereas sequences alternating in intensity are grouped as trochees (strong-weak). In addition to the role of the acoustic cues of duration and intensity, the current study also tests how overall acoustic variability in the sequences affects prosodic grouping. The results for Akan show that duration differences lead to iambic grouping, as predicted by the ITL, but only when there is low acoustic variability in the sequences. No grouping effects for intensity were found. The results are interpreted and discussed in a typological context.

Index Terms: prosody, grouping, iambic-trochaic law, Akan, stress

1. Introduction

In spite of efforts to increase the diversity of the languages studied in linguistics, many languages of the world are still in need of detailed description and investigation, in particular regarding their prosody (e.g., [1]). Investigating previously underdescribed prosodic systems is important as perception research has shown that speech processing mechanisms are often shaped by the native language of the listener (e.g., [2]). In other words, it is challenging to find prosodic processing mechanisms that hold across all languages. This holds in particular for supposedly universal phonological accounts, such as the iambic-trochaic law (ITL; [3]). This law makes predictions about how listeners prosodically group sequences (repetitions) of sounds into smaller binary units (i.e. *feet*), depending on whether these sounds vary in duration (short vs. long) or intensity (soft vs. loud). That is, sound sequences varying in duration lead to the perception of weak-strong (iambic) subsequences, whereas sound sequences varying in intensity lead to the perception of strong-weak (trochaic) subsequences. Crucially, research has shown that languages differ in the extent to which these acoustic cues affect listeners' grouping preferences [4], nuancing a strict interpretation of the ITL. Not only do languages differ in the extent to which listeners are affected by acoustic cues ([5]; [6]), they also differ in whether the observed effects are ITL-conform or not ([7]; [8]). Thus, the ITL predictions only partially hold for French listeners (without lexical stress), and to a much larger extent for German listeners (with lexical stress) [9]. This variability is at least partially reflective of the prosodic differences between languages. In particular, word stress and prosodic phrasing affect listeners' grouping strategies. Thus, despite the limited crosslinguistic applicability of the ITL, prosodic grouping tasks that test its predictions give insight into language-specific speech perception mechanisms. This is a particularly important instrument to investigate languages for which little to no quantitative work on prosody has been carried out. In the current study, we investigate one of these languages: Akan, a language spoken in Ghana.

Akan (iso: aka) refers to the Kwa dialect group of Niger-Congo languages. Akan is spoken by 80% of the people in Ghana (approximately 9.6 million speakers) [10]. Akan dialects are mutually intelligible and Asante Twi is the most widely used and studied dialect. Akan has high and low lexical tones (H and L) that are used to distinguish word meanings as well as for grammatical functions (mainly verbal) [11]. At the phrase level, Akan tone is characterized by downstepping H-tones that follow L-tones. At the surface level, phrases tend to have a typical overall downtrend for this reason, which has a 'terracing' nature in the sense that the downtrend shows a stepwise lowering of f0 [12]. The syllable is analysed as the tone-bearing unit in Akan. Although there are no in-depth studies of word stress for this language, it is assumed that there is none ([11]; [13]).

The rhythmic nature of Akan was studied in a speech cycling task [14]. In this task, participants repeated phrases a large number of times at a medium or high tempo, inducing a (more) rhythmic version of their productions. The impressionistic analysis of where in the phrases the rhythmical beats occurred led to a tentative conclusion that Akan could be stress-timed [14]. A similar conclusion was reached on the basis of Akan poem recitations, for which prominent beats were annotated, showing stable timings across various tempi [15]. It should be noted that the studies on rhythm in Akan both rely on impressionistic data by the authors themselves, being native speakers of a so called stress-timed language (English). It is therefore by no means clear whether the outcomes can be taken as a reliable account of Akan rhythm. In a study investigating rhythm using metrics to test the rhythm-class hypothesis (stress-, syllable-, or more-timed) found that Akan did not pattern with any of these classes [16]. Furthermore, without word-stress studies for this language, there is little to no solid basis for hypothesizing about the nature of prosodic grouping in Akan. Therefore, the need

10.21437/SpeechProsody.2024-210

for further research on this language is evident.

In the current study, we report a partial replication of a perception experiment that tested the validity of the ITL in French and German ([5]; [9]). These studies found that German listeners were more sensitive than French listeners to the acoustic grouping cues. Indeed, grouping performance was stronger/clearer in German than in French for both the duration and intensity cues ([5]; [9]). Moreover, when tested with non-linguistic sequences, as done here, variability in the sounds making up the sequences decreased grouping performance, leading to at-chance performance in the French-speaking adults [9]. This was taken as a reflection of the fact that German listeners are used to perceiving and processing word-level stress differences, which are absent in French ([5]). In the current study we investigate duration and intensity as prosodic cues to perceptual grouping in Akan in order to further test the extent to which the ITL applies and to shed a first light on prosodic processing mechanisms in native speakers of this language.

2. Methodology

This section describes the experimental methodology of this study, which was largely based on the one used in [9], who used both linguistic and non-linguistic stimuli, and found convergent results between the two sets. Therefore, the current study uses a non-linguistic subset of their stimuli, as they allow for easier crosslinguistic comparisons.

2.1. Participants

38 participants carried out the experiment. Participants' data was excluded when responses occurred before the the fade-in time of the stimuli (3 s, see below), as was the case for 14 participants. The remaining 24 participants were all native speakers of Akan (Asante Twi dialect; 9 female, 15 male, M age = 22, age range: 18-27). None of them had hearing problems and they were all multilingual and spoke English (22), French (2), Bono (2), Ewe (1), Sefwi (1) or Fante (1) in addition to Akan.

2.2. Stimuli

The sound sequences used in this study were generated from long-term average spectral and temporal structures, each taken from four different musical instruments (see [17] and [9] for further details). The long-term average spectra were taken from a bassoon, cello, clarinet and a trumpet. The temporal structures, in particular onset structures, were taken from a violin, trombone, English horn, and an alto saxophone. The combination of spectra and temporal structures was called a 'chimera', and one for each of the possible combinations (4 x 4) was generated. The default average intensity of each of these chimeras was set to 60 dB, duration to 260 ms, and f0 to 200 Hz. The final 5 ms were a raised cosine fade out to 0 dB. In this way, the chimeras reflected the characteristics of a 'syllable', i.e. the part that represented the building block for a sequence.

High variability sequences consisted of varying chimeras. They were generated by concatenating all of the 16 chimeras, and repeating the result once, such that the total length of a default sequence was 32 x 260 = 8320 ms. Low variability sequences consisted of 32 repetitions of a single chimera, i.e. the combination of trombone temporal structure and cello spectrum. During the first three seconds a sequence was faded in whilst masking white noise was fading out (both raised cosine). This was done in order to mask the start of the sequence. Note that the sequences had no pause between the chimeras (the ones used in the 'no pause' condition in [9]).

Intensity contrasts were obtained by amplifying every other chimera in the sequence by either 2, 4, 6 or 8 dB. Duration contrasts were obtained by lengthening every other chimera in the sequence with either 50, 100, 150 or 200 ms. For each sequence there were two prominence variants; one variant started with the weak chimera (either soft or short) and the other variant started with the strong chimera (either loud or long). This was done, in addition to the white noise fading, to prevent listeners' prosodic grouping choices from being biased by the prominence of the initial chimera in the sequence. In other words, listeners could not identify the actual start of the sequence. In control sequences (either high or low variability) there was no such intensity or duration contrast between the chimeras. The high variability controls were randomly selected from the ones in [9], for the low variability ones there was only one variant.

2.3. Design

In total, 108 stimuli were presented to participants, half of which in each variability condition. For each level of variability, 6 control, 24 intensity and 24 duration sequences were presented. For each acoustic cue, half of the sequences started with weak prominence (12), the other half with strong prominence (12). There were three variants for each of the four levels of acoustic contrast, randomly picked from the stimulus set in [9].

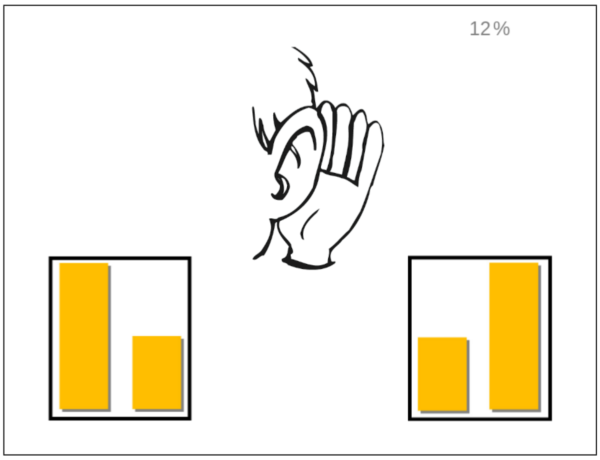

Figure 1: *Screenshot of the experiment displaying long-short bars on the left (trochaic response) and short-long bars on the right (iambic response).*

2.4. Setup and procedure

The experiment was run using PsyToolkit ([18]; [19]), which generates html-pages for each stimulus. For each auditory stimulus, which was played automatically, two pictures were displayed. Each picture consisted of two vertical bars; one short and one long, displayed next to each other (Figure 1). In one picture, the short bar appeared on the left and the long bar on the right (short-long; corresponding to iambic perception), whereas in the other picture the long bar appeared on the left and the short bar on the right (long-short; corresponding to trochaic perception). Participants' task was to listen to the stimulus and click on the picture that corresponded with how they perceived the grouping. They were instructed to give an answer as soon as they could, that is, they did not have to wait until the stimulus had been fully played. Once participants had made their choice, their answer was saved and a 500 ms pause occurred before automatically playing the next stimulus and re-displaying the respective pictures. The two pictures were displayed in two orders; iamb-left, trochee-right and vice versa, which was determined randomly at the start of the experiment for each participant. The display order did not change during the experiment.

Before the start of the experiment, participants received a questionnaire about their language background. This was done to assess their level of native language proficiency and multilingualism. Thereafter, they received instructions about the experimental task. After the first instructions they completed a training round consisting of four stimuli with the most extreme acoustic contrasts from the low variability set: either 200 ms for duration or 8 dB for intensity, each of them in both orders (strong first, weak first). This was done to provide participants with the stimuli that were hypothesized to be the clearest acoustically, in order to make their choice for iamb or trochee the easiest. Note that with more subtle acoustic contrasts, potentially inaudible to participants, the idea of the task would not have been clear to them from the start.

The variability conditions were assigned to participants in two block orders; either high-low or low-high. The order was determined randomly at the start of the experiment for each participant. The experiment was run in offline mode due to an unstable internet connection, thus collecting the responses on a local computer instead of a web-server. For each stimulus, block order, picture display order, wave file name (consisting of: variability, cue, initial chimera, contrast level), picture chosen and reaction time relative to the start of the stimulus were saved.

2.5. Statistical analysis and model selection

Generalized linear mixed model (GLMM) analyses were run on the binary responses (0 for iambs, 1 for trochees). Four models were run to check for effects against chance level (using the explicit intercept syntax `response ~ -1 + factor ...`), where each model had a single different fixed factor, either *cue* (successive differences contrast coded, three levels: intensity, control, duration), *variability* (treatment coded, two levels: low, high), *sequence order* (three levels: control, strong first, weak first) or *block order* (two levels: high first, low first). In each of the four models participant and item were random intercepts. The results of these four models are reported here as they led to the selection of factors that were included in the final model described below. Significant effects were only found for *cue* (duration: b = -0.27, SE = 0.09, z = -3.06, p < 0.01; intensity: b = 0.20, SE = 0.09, z = 2.33, p < 0.05) and *variability* (low variability: b = -0.24, SE = 0.10, z = -2.39, p < 0.05), indicating that *sequence order* and *block order* did not affect the responses. Based on these results, the interaction of *variability* and *cue* (coding as above) was then included in a new final model, with the binary responses as response variable (model was optimized for convergence): `glmer(response ~ var / cue + (1|pp) + (1|item), data = df.akan, contrasts = list(cue=contr.sdif),family = binomial, control = glmerControl( calc.derivs=FALSE))`.

3. Results

Table 1 shows the mean (trochaic) responses for each level of the acoustic cues. It can be seen that, in the control condition, participants chose iambs almost equally as often as trochees, whereas the number of trochaic responses is higher for intensity cues and lower for duration cues. The level of acoustic difference did not consistently increase the trochaic responses for intensity nor consistently decrease them for duration in the high variability condition. In the low variability condition, however, an overall trend is visible in that the trochaic responses increased with the increasing magnitude of the intensity differences and decreased with the increasing magnitude of the duration differences.

Table 1: *Mean responses (0=iambs, 1=trochees) per level of acoustic cue and variability.*

Variability	Cue	Level	Response
high	control		0.53
	intensity (dB)	2	0.61
		4	0.56
		6	0.53
		8	0.53
	duration (ms)	50	0.54
		100	0.51
		150	0.43
		200	0.53
low	control		0.49
	intensity (dB)	2	0.50
		4	0.53
		6	0.56
		8	0.56
	duration (ms)	50	0.42
		100	0.31
		150	0.32
		200	0.31

Table 2: *Results of the final GLMM.*

Factor	b	SE	z	p
(Intercept)	0.13	0.09	1.50	0.13
var:low	-0.32	0.10	-3.27	$<$ 0.01
var:high*cue:ctrl-int	-0.09	0.19	-0.49	n.s.
var:low*cue:ctrl-int	-0.21	0.19	-1.13	n.s.
var:high*cue:dur-ctrl	-0.13	0.19	-0.72	n.s.
var:low*cue:dur-ctrl	-0.62	0.19	-3.25	$<$ 0.01

The results of the final GLMM (Table 2) furthermore reveal a significant effect of *variability* in that overall fewer trochaic responses were obtained in the low than in the high variability condition. The interactions between *variability* and *cue* indicate that duration in the low variability items had a significant effect in that there were fewer trochaic responses compared to the control condition.

Inspection of the random effects per participant revealed a large amount of individual variation (see Figure 2). The deviant responses were impressionistically checked against the participants' language background. However, no pattern could be discovered indicating that the second language(s) of participants consistently led to a certain type of response.

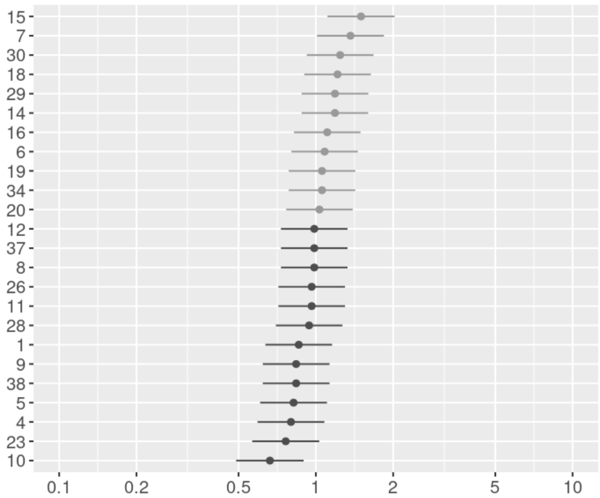

Figure 2: *Estimates of the random effects per participant (gray scale indicates whether estimate is below (dark) one or above (light) one.*

4. Conclusions

The results of the current study show that duration differences in non-linguistic auditory sequences lead to iambic grouping in Akan. The duration effect was only found for sequences with low acoustic variability. No grouping preferences were found for high acoustic variability and no effects at all were found for intensity. The outcomes thus partially confirm the predictions of the ITL in that long duration is associated with group-final position, thus leading to more iambic responses compared to no duration differences (controls). The Akan results in this study match the iambic grouping bias due to duration differences as reported for English (e.g., [20] [21]), French and German ([5]; [9]), and Italian [22]. Although the predictions of the ITL were only met for duration, the trends observed for intensity differences (more trochaic responses) showed an ITL-conform pattern.

It remains to be discussed what the outcomes of the current experiment tell us about the prosody of Akan. This is particularly challenging given the available literature on other languages. The duration-based iambic grouping preference was not consistently shown for the languages for which it was tested (e.g., [4]). This was taken as an indication that the ITL predictions for duration are not universally applicable. The intensity-based trochaic grouping preference was shown more consistently across languages and species [20]. In light of these findings, the current results for Akan thus seem to indicate that the duration effect relates to a characteristic that is typical for this language. It is not a priori clear which feature this could be, but it would corroborate impressionistic reports on Akan rhythm, favouring a major role for temporal characteristics (e.g., [14], [15]). It should furthermore be noted that although the stimuli were speech-like, their generalization to prosody is more limited than when stimuli with entirely natural speech would have been used.

In order to draw more solid conclusions, Akan tone should also be considered. The current study did not test pitch differences, which supposedly contribute to trochaic (high-low) grouping along with intensity [23]. The interaction between lexical tone and (phrase) intonation in Akan is reported to be minimal [11]. The commonly observed downtrend and phrase-final L contribute to the perception of finality and enhances the effects of final lengthening and reduced intensity phrase-finally [11]. The effects in the current study seem to indicate that Akan listeners pay attention to duration as a cue to finality. The question remains how pitch relates to this finding. A plausible option is that final L tones in Akan are perceptually weak, due to reduced intensity and their occurrence toward the bottom of the speaker's pitch range. Lengthening (duration) could thus be the most reliable prosodic cue to finality in Akan phrases (but see discussion in [24]; [25]).

The literature has questioned to what extent the ITL makes straightforward predictions for word stress in any given language [20]. That is, stress is often realised by an enhancement of duration, intensity and/or pitch, i.e. with increased values for any of these compared to unstressed counterparts. Consider a disyllabic word with stress on the first syllable (trochee). If this syllable is lengthened, a long-short sequence would occur. Thus, a strict interpretation of the ITL is problematic in the context of word stress, given its prediction that duration differences should be grouped as short-long. The vast amount of languages with penultimate stress [26] and duration being the most common cue [27] show that a direct link between the ITL and word stress does not hold (see also [4] for a similar argument). A recent study also argued against the universal applicability of the ITL by showing that the underlying principles for parsing the incoming speech signal are grouping *and* prominence [28]. A nuanced reinterpretation of the ITL was proposed in that intensity can cue initiality or stress, whereas duration can cue finality or stress [21].

It is furthermore important to observe that the duration effect for Akan only holds for sequences with low acoustic variability. In previous work, it was argued that the ability to consistently group high variability sequences may depend on listeners' experience with stress contrasts (i.e. as in German and not in French in [9]). That is, high variability sequences are easier to process when listeners can rely on abstract stress representations. Note, that both duration and intensity had effects in German listeners' responses in [9], whereas in Akan the observed effects are much more subtle. The Akan results pattern with those from the French group in [9] in the sense that an ITL-conform effect was only found for low variability. The amount of ITL-conform responses was also higher for duration than for intensity in the French listeners (both in [5] and in [9]). On the basis of this parallel with previous results, the Akan outcomes would favor a fixed-stress or no-stress analysis of this language. Without further acoustic and perceptual evidence these observations remain speculation, however.

5. Acknowledgements

Research for this paper was funded by the German Research Foundation (DFG) – Project-ID 281511265 (SFB-1252 "Prominence in language"). The authors thank Mareike Philip for carrying out the perception experiment.

6. References

[1] N. P. Himmelmann and D. R. Ladd, "Prosodic Description: An Introduction for Fieldworkers," *Language Documentation & Conservation*, vol. 2, no. 2, pp. 244–274, 2008.

[2] A. Cutler, *Native listening: language experience and the recognition of spoken words*. Cambridge, Mass: The MIT Press, 2012.

[3] B. Hayes, *Metrical stress theory: principles and case studies*. Chicago: University of Chicago Press, 1995.

[4] M. J. Crowhurst, "The iambic/trochaic law: Nature or nurture?" *Language and Linguistics Compass*, vol. 14, no. 1, 2020.

[5] A. Bhatara, N. Boll-Avetisyan, A. Unger, T. Nazzi, and B. Höhle, "Native language affects rhythmic grouping of speech," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3828–3843, 2013.

[6] M. J. Crowhurst and A. Teodocio Olivares, "Beyond the Iambic-Trochaic Law: the joint influence of duration and intensity on the perception of rhythmic speech," *Phonology*, vol. 31, no. 1, pp. 51–94, 2014.

[7] M. Molnar, M. Carreiras, and J. Gervain, "Language dominance shapes non-linguistic rhythmic grouping in bilinguals," *Cognition*, vol. 152, pp. 150–159, 2016.

[8] J. R. Iversen, A. D. Patel, and K. Ohgushi, "Perception of rhythmic grouping depends on auditory experience," *The Journal of the Acoustical Society of America*, vol. 124, no. 4, pp. 2263–2271, Oct. 2008.

[9] A. Bhatara, N. Boll-Avetisyan, T. Agus, B. Höhle, and T. Nazzi, "Language Experience Affects Grouping of Musical Instrument Sounds," *Cognitive Science*, vol. 40, no. 7, pp. 1816–1830, 2016.

[10] D. M. Eberhard, G. F. Simons, and C. D. Fennig, *Ethnologue: Languages of the World*, 23rd ed. Dallas: SIL International, 2020.

[11] F. Kügler, "Tone and intonation in Akan," in *Intonation in African Tone Languages*, L. J. Downing and A. Rialland, Eds. De Gruyter, 2016, pp. 89–130.

[12] G. N. Clements, "The Description of Terraced-Level Tone Languages," *Language*, vol. 55, no. 3, p. 536, 1979.

[13] F. A. Dolphyne, "Tone and grammar in Akan: the tone of possessive constructions in the Asante dialect," in *The phonological representation of suprasegmentals: studies in African languages offered to John M. Stewart on his 60th birthday*, ser. Publ. in African languages and linguistics (PALL), K. Bogers, H. van der Hulst, and M. Mous, Eds. Dordrecht: Foris Publ., 1986, vol. 4, pp. 35–49.

[14] J. Anderson, "Preliminary to preliminary: Speech rhythm in Akan (Twi)," in *African Linguistics Across the Discipline*, 2nd ed., J. Anderson, C. R. Green, and S. G. Obeng, Eds. Bloomington: IULC Publications, 2009, vol. 10, pp. 133–143.

[15] T. M. Purvis, "Speech rhythm in Akan oral praise poetry," *Text & Talk - An Interdisciplinary Journal of Language, Discourse Communication Studies*, vol. 29, no. 2, pp. 201–218, 2009.

[16] N. Boll-Avetisyan, P. O. Omane, and F. Kügler, "Speech rhythm in Ghanaian languages: The cases of Akan, Ewe and Ghanaian English," in *Speech Prosody 2020*. ISCA, May 2020, pp. 586–590.

[17] T. R. Agus, C. Suied, S. J. Thorpe, and D. Pressnitzer, "Fast recognition of musical sounds based on timbre," *The Journal of the Acoustical Society of America*, vol. 131, no. 5, pp. 4124–4133, 2012.

[18] G. Stoet, "PsyToolkit: A software package for programming psychological experiments using Linux," *Behavior Research Methods*, vol. 42, no. 4, pp. 1096–1104, 2010.

[19] ——, "PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments," *Teaching of Psychology*, vol. 44, no. 1, pp. 24–31, 2017.

[20] M. Crowhurst, "Iambic-Trochaic Law Effects among Native Speakers of Spanish and English," *Laboratory Phonology*, vol. 7, no. 1, p. 12, Oct. 2016.

[21] M. Wagner, "Two-dimensional parsing of the acoustic stream explains the Iambic–Trochaic Law." *Psychological Review*, vol. 129, no. 2, pp. 268–288, Mar. 2022.

[22] A. Langus, S. Seyed-Allaei, E. Uysal, S. Pirmoradian, C. Marino, S. Asaadi, Ö. Eren, J. M. Toro, M. Peña, R. A. H. Bion, and M. Nespor, "Listening natively across perceptual domains?" *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 42, no. 7, pp. 1127–1139, Jul. 2016.

[23] M. Nespor, M. Shukla, R. Van de Vijver, C. Avesani, H. Schraudolf, and C. Donati, "Different phrasal prominence realizations in VO and OV languages," *Lingue e linguaggio*, vol. 7, Jan. 2008.

[24] S. Genzel, "Lexical and post-lexical tones in Akan," doctoralthesis, Universität Potsdam, 2015.

[25] S. Genzel and F. Kügler, "Production and perception of question prosody in Akan," *Journal of the International Phonetic Association*, vol. 50, no. 1, pp. 61–92, Apr. 2020.

[26] R. Goedemans, J. Heinz, and H. Van der Hulst, "StressTyp2: Home Page," Feb. 2020.

[27] M. K. Gordon and T. Roettger, "Acoustic correlates of word stress: A cross-linguistic survey," *Linguistics Vanguard*, vol. 3, no. 1, p. 20170007, Aug. 2017.

[28] E. Moghiseh, M. Sonderegger, and M. Wagner, "The iambic-trochaic law without iambs or trochees: Parsing speech for grouping and prominence," *The Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 1108–1129, Feb. 2023.