Periodic energy mass on head and edge tones in Maltese wh-constructions

Maria Lialiou[1], Aviad Albert[1], Alexandra Vella[2], Martine Grice[1]

[1]University of Cologne, Germany
[2]University of Malta, Malta

{mlialiou, a.albert, martine.grice}@uni-koeln.de, alexandra.vella@um.edu.mt

Abstract

This paper is concerned with the relation between tonal association and prosodic strength in different tone bearing positions in Maltese wh-words. In these words, tones are associated with the stressed syllable (head association) in indirect and quoted questions, but with the initial syllable (edge association) in direct questions. In a language that has pitch accents to cue prominence (a head prominence language according to Jun's typology), the initial syllable, if not stressed, would not typically cue prominence, but rather juncture. Using periodic energy mass as a measure of strength, and thus prominence, we found that mass enhancement is not conditioned by tonal association (either head or edge) but rather by the lexical stress. Whereas the present study shows that the word-initial H tone does not affect the relative prominence between the stressed syllable and the word-initial one, and thus does not cue prominence on the initial syllable, there is a potentially different prominence-cueing function of this early H peak. That is, for example a prominence cueing function at the word level (i.e., one which makes the entire word more prominent) driven by modality or pragmatic force.

Index Terms: Maltese, tonal association, pitch accent

1. Introduction

Maltese is a language with lexical stress and regular pitch accents, phrase accents with secondary association to postnuclear stressed syllables, and edge tones [1;2]. Thus, according to [3], Maltese, as a language that uses pitch accents to cue prominence, is a head prominence language. Interestingly, sentence modality conditions tonal association [e.g., 2;4]. Recent work ([4]), following from [5;6], using read speech data, investigated Maltese wh-words of varying syllable length in different positions (initial, medial, final) in interrogatives and declaratives. [4] showed that an early F0 peak, analysed as an H tone, is associated with the initial syllable of the wh-word – rather than with its lexical stress – when it is used in interrogatives (direct questions). In contrast, when the same wh-word is used in declaratives (indirect questions), a falling pitch accent is associated with its stressed syllable. In other words, tones in Maltese wh-words alternate in their association between the stressed syllable (head association) and the (typically) unstressed word-initial syllable (edge association) as a function of sentence modality. This alternation holds regardless of the metrical structure of the word or its position in the phrase, and seems to be pragmatically rather than phonologically conditioned in that the association of an H tone depends on whether a question is direct or indirect ([2;4]).

In a head prominence language, according to [3], the initial syllable, if not stressed, would not typically cue prominence.

Following [4], this study is concerned with the relation between tonal association and prosodic strength in the two different tone bearing positions in Maltese wh-words (left edge association or head association), aiming to explore whether there are any acoustic indications for a prominence cueing function of H tones associated to the left edge of wh-words. We refer to these tones as early H peaks.

2. Methods

2.1. Participants and data

This study analyses read speech data produced by 10 native Maltese speakers (5 male; 5 female). This data was collected as part of a larger corpus reported in the [4] study (for additional participant- and task-related information, see [4]).

2.2. Speech material and measurements

In this study, we explore Maltese simple (MIN /miːn/ 'who') and complex wh-words (ma' MIN /mɐˈmiːn/ 'with whom', min MINnhmom /miːnˈmɪnːɔm/ 'which one of them', ma' min MINnhom /mɐmiːnˈmɪnːɔm/ 'with which one of them'; capitalization indicates lexical stress) in the final position in three different conditions: direct questions (interrogatives), indirect questions (declaratives), and quoted questions (narrow focus declaratives). Table 1 provides an example of the speech material using the complex wh-word *ma' min minnhom* across the different conditions.

Table 1: *Speech materials' example*

condition	Example
direct question (interrogative)	*Mar j-għum ir-Ramla <u>ma' min minnhom</u>?*
	'With which one of them did he go swimming to Ramla?'
indirect question (declarative)	*U staqs-iet-ni, mar j-għum ir-Ramla <u>ma' min minnhom</u>.*
	'And she asked me with which one of them did he go swimming to Ramla.'
quoted question (narrow focus declarative)	*Iva, mistoqsija oħra li għandna bżonn insaqsu hija <u>ma' min minnhom</u>.*
	'Yes, another question we need to ask is with which one of them.'

Supplementing analysis from [4], we use the periodic energy mass metric (henceforth, mass) from the *ProPer* toolbox [7] to explore the underlying prosodic strength of the two different tone bearing positions in Maltese wh-words. Mass is the integral of duration and power, i.e., the area under the periodic energy curve [8;9]. Our data is designed to compare mass in cases where the wh-words exhibit a falling pitch accent on the stressed syllable (indirect and quoted questions), to cases where the wh-words bear an early H peak (direct questions). The

difference between indirect and quoted questions is that in quoted questions, the wh-words are in narrow focus. Figure 1 illustrates representative examples of *periograms* and periodic energy curves of *ma' min minnhom* in the three conditions as produced by a Maltese native speaker.
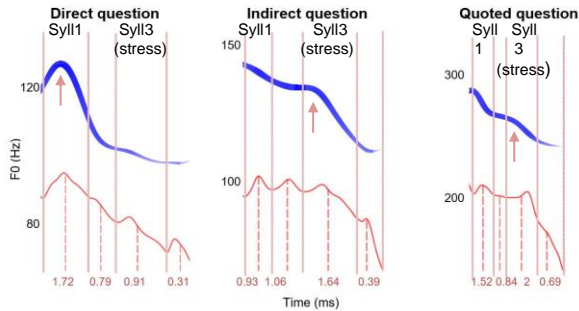

Figure 1: *Periograms and periodic energy curves of ma' min minnhom produced by a Maltese native speaker in direct, indirect, and quoted questions. The solid vertical red lines denote syllable intervals and the dashed vertical red lines depict the center of mass within each interval. Arrows reflect F0 inflection.*

Since it is as yet unclear whether the stressed syllable in direct questions bears a pitch accent (possibly a L*) or not, for the purposes of this experiment we refer to the stressed syllable in this condition as unaccented.

2.3. Statistical analysis

Data were analyzed in R [10]. Separate Bayesian linear mixed models were fitted for simple and complex wh-words, using the Stan modeling language [11] and the package *brms* [12].

For complex wh-words, we model mass values as a function of dummy-coded factors SYLLABLE (reference level "Syllable1"), and CONDITION (reference level "Direct") as well as their interaction. Further, the model includes by speaker and by wh-word random intercepts and slopes. For the intercept (syllable1 in direct questions), we define a weakly informative prior, normally distributed ($\mu = 1$, $\sigma = 4$). For regression coefficients, we define a weakly informative prior, normally distributed ($\mu = 0$, $\sigma = 0.5$). Lastly, for standard deviations of random effects, we use *brms*' default priors (a Student's t-distribution with $v = 3$, $\mu = 0$ and $\sigma = 20$).

For simple wh-words, we model mass values as a function of the dummy-coded factor CONDITION (reference level "Direct"). Further, the model includes by speaker random intercepts. For the intercept (direct questions), we define a weakly informative prior, normally distributed ($\mu = 1$, $\sigma = 4$). For regression coefficients, we define a weakly informative prior, normally distributed ($\mu = 0$, $\sigma = 1$). Lastly, for standard deviations of random effects, we use *brms*' default priors (a Student's t-distribution with $v = 3$, $\mu = 0$ and $\sigma = 20$).

For both models, four sampling chains ran for 4000 iterations with a warm-up period of 2000 iterations for each model, yielding 8000 samples for each parameter tuple. For all relevant coefficient means and differences between them, we report the expected values under the posterior distribution and their 95% credible intervals (CIs) as well as the posterior probability that a difference δ is greater than zero. If a hypothesis states that $\delta > 0$, there is compelling evidence for this hypothesis if zero is (by a reasonably clear margin) not included in the 95% CI of δ and the posterior $P(\delta > 0)$ is close to one [13]. There was no indication of convergence issues (for

the model on complex wh-words: 1 divergent transition after warm-up; all Raht = 1.0).

3. Results

We first, report results on complex wh-words, and then on the simple *min*. Tables of values derived from *ProPer* analysis as well as the scripts for the analyses are available at https://osf.io/zpdnc/ on the OSF platform.

3.1. Maltese complex wh-words

Figure 3 illustrates normalized mass values for word-initial and stressed syllables across wh-words and speakers as a function of direct, indirect, and quoted questions.

Visual inspection of the mass values reveals consistent differences between word-initial and stressed syllables with the stressed syllables exhibiting higher averaged mass values compared to the word-initial ones across all conditions. This difference seems to increase from direct to indirect and quoted questions. Moving to word-initial syllables across direct and indirect questions, visual inspection reveals subtle differences. Both the density (violin plots) and the individual tokens (dots within violins) of mass on word-initial syllables show that word-initial syllables bearing an early H peak in direct questions exhibit increased mass compared to the unaccented word-initial syllables in indirect questions. Similar inspection of stressed syllables across direct and indirect questions does not reveal any visual difference. Focusing on the quoted questions, it is apparent that the overall mass of the wh-words is enlarged compared to the other conditions, with both word-initial and stressed syllables showing higher mass values compared to the same syllables in the other conditions.
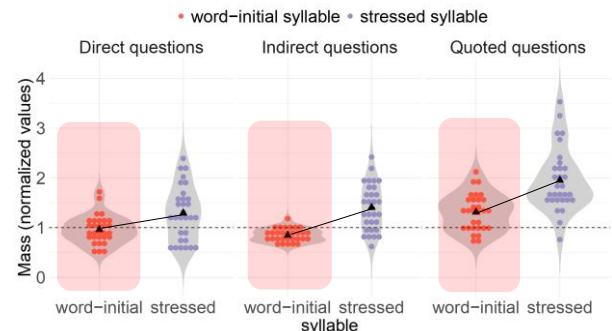

Figure 2: *Mass values (1 = average strength) of word-initial (pink) and stressed (purple) syllables (x-axis) across wh-words as a function of direct (left panel), indirect (middle panel), and quoted (right panel) questions. Violin plots depict the distribution of the data. Dots within violins illustrate individual data points for word-initial and stressed syllables, while black triangles depict mean values across wh-words and speakers.*

Let us now move on to the results of the model. Comparing mass on word-initial and stressed syllables, the model shows lower mass values on the word-initial syllables within each condition ($\delta_{direct} = -0.23$; $\delta_{indirect} = -0.44$; $\delta_{quoted} = -0.53$). It appears that the mean difference between word-initial and stressed syllables is smaller in direct questions, while this difference increases in indirect and further still in quoted questions. For direct questions, the 95% CI [−0.68, 0.29] includes zero by some margin and the posterior probability $P(\delta > 0)$ is 0.8 and thus below the 0.95 heuristic. The model therefore does not suggest a reliable difference between the

word-initial syllable marked by an early H peak and the stressed unaccented syllable, but only a trend towards higher mass on the latter than the former (evidence ratio = 4.09). For indirect questions, the 95% CI [−0.91, 0.09] includes zero by a small margin and the posterior probability $P(\delta{>}0)$ is 0.92. Whereas one could say that these values suggest a more than negligible degree of uncertainty, an overall, even if not overly strong, tendency towards higher mass on the stressed accented syllable compared to word-initial toneless syllable can be seen, and this is reflected in an evidence ratio of 12.07 and the facts that zero is only narrowly included in the 95% CI and that the posterior probability is quite high. Lastly, in quoted questions, although the 95% CI [−0,99, 0] includes zero on the margin and the posterior probability $P(\delta{>}0)$ = 0.95, the model still very strongly favours the interpretation of higher mass values on the stressed syllable as opposed to the word-initial syllable (evidence ratio = 18.46).

Moving on to comparisons across questions, we compare word-initial syllables bearing an early H peak (direct questions) to toneless word-initial syllables (indirect questions). The model estimates a modest increase in mass on the word-initial syllables bearing the early H peak (δ = 0.09). Although mass on word-initial syllables with an early H peak differs on average from syllables without an associated tone, this difference is only very subtle, thus the model does not suggest a reliable difference but only a weak trend towards increased mass on word-initial syllables with an H tone (CI [−0.11, 0.29], $P(\delta{>}0)$ = 0.81). Comparing toneless word-initial syllables in quoted questions to both word-initial syllables bearing an early H peak in direct questions and to toneless syllables in indirect questions, the model estimates higher mass values on the quoted toneless word-initial syllables, though only suggesting a trend towards increased mass on the toneless quoted word-initial syllables. This trend is more reliable when comparing toneless quoted syllables to toneless syllables in indirect questions than to word-initial syllables bearing an early H peak in direct questions (quoted vs. direct: δ = 0.29, CI [−0.15, 0.62], $P(\delta{>}0)$ = 0.9, evidence ratio = 8.93; quoted vs. indirect: δ = 0.38, CI [−0.11, 0.77], $P(\delta{>}0)$ = 0.92, evidence ratio = 11.48).

Moving to stressed syllables, the model estimates only a slight difference between stressed unaccented syllables (direct questions) and stressed syllables bearing a falling pitch accent (indirect questions) (δ = −0.12). In this respect, the model does not provide indications for a reliable difference between stressed unaccented syllables and stressed accented syllables, but only a trend towards lower mass values on stressed unaccented syllables (CI [−0.34, 0.08], $P(\delta{>}0)$ = 0.87). Comparing stressed unaccented syllables (direct questions) to quoted stressed syllables bearing a falling pitch accent, the model estimates lower mass on the stressed unaccented syllables providing in this case strong evidence for this difference (δ = −0.59, CI [−0.93, −0.14], $P(\delta{>}0)$ = 0.97). Lastly, for the comparison between stressed syllables bearing a falling pitch accent (indirect questions) and quoted stressed syllables with the same accent (δ = −0.46), although the 95% CI [−0.87, 0.03] includes zero by a very narrow margin and the posterior probability $P(\delta{>}0)$ is 0.94, the model still very strongly favours the interpretation that the stressed accented syllable in indirect questions have lower mass values compared to the same syllable in quoted questions (evidence ratio = 15.77).

3.2. Maltese simple wh-word *min*

The morpholygically simple *min* is treated separately from the other wh-words because, with only one syllable, the initial and stressed syllable coincide. Figure 3 depicts normalized mass values of *min* across speakers as a function of direct, indirect, and quoted questions.

Visual inspection of the mass values on *min* reveals consistent differences across conditions. It is apparent that mass is considerably stronger for *min* in the quoted, as compared to any other, condition. Moving to direct questions, in which *min* is marked by an early H peak, and indirect questions, in which *min* bears a falling pitch accent, we can see, both from the density (violin plots), the individual tokens (dots within violins), and the mean (black triangles), that *min* exhibits greater mass when it is in direct as compared to indirect questions.
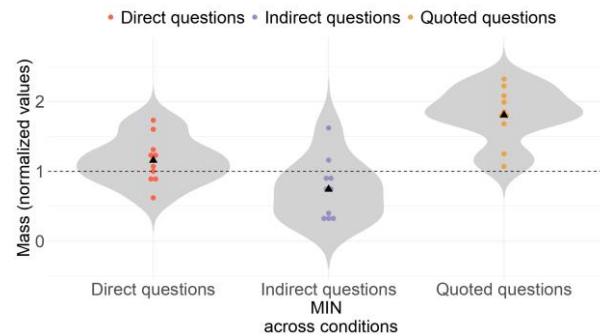

Figure 3: *Mass values (1 = average strength) of min in direct (left), indirect (middle), and quoted (right) questions (x-axis). Violin plots depict the distribution of the data. Dots within violins illustrate individual data points, and black triangles depict mean values across speakers.*

The results of the model, when comparing quoted *min* to that in direct or indirect questions, provide strong evidence for higher mass on quoted *min* (δ quoted vs. direct = 0.64, CI [0.43, 0.84], $P(\delta{>}0)$ = 1; δ quoted vs. indirect = 1.05, CI [0.84, 1.26], $P(\delta{>}0)$ = 1). Further, the model estimates higher mass values for *min* in direct compared to indirect questions (δ = 0.41), providing strong evidence for stronger mass on *min* when marked by an early H peak than when it bears a falling pitch accent (CI [0.2, 0.62], $P(\delta{>}0)$ = 1).

4. Discussion and Conclusions

Unlike previous research on acoustic correlates of prosodic prominence, our work explores how periodic energy mass, as a refined measure related to prosodic strength, is modified in the two different tone bearing positions (head versus edge) in Maltese wh-words, aiming to shed light on whether there are acoustic indications of a prominence cueing function of early H peaks in direct questions. We compare the strength of syllables bearing this early H peak in direct questions to syllables bearing a clear (falling) pitch accent in indirect and quoted questions.

Our results provide subtle rather than strong evidence of prosodic enhancement as a function of tonal association. Let us start with the quoted condition. In this condition, mass is undoubtedly greater across the whole wh-word compared to the other conditions, while word-internally, the model favours the interpretation that the stressed syllable bearing a falling pitch accent exhibits reliably greater mass compared to the toneless word-initial one. Remember that the quoted wh-words are in narrow focus which is presumably related to the boosted mass we observe here. Thus, while narrow focus enhances the overall mass of the quoted wh-words, internally the relations are

retained reflecting the greater prosodic strength of the stressed accented syllable as compared to the toneless word-initial one.

Consider that in the quoted condition there are three possible factors that could affect mass/ prosodic enhancing (i.e., lexical stress, a tonal event, narrow focus), thus making the comparison to the other two conditions (both of which have broad focus) difficult and unbalanced. Therefore, for the rest of the discussion, we will leave aside the condition of quoted questions.

We now turn to our results concerning word-initial syllables and stressed syllables across direct and indirect questions. Remember that in direct questions the word-initial syllable, which is lexically unstressed, is marked by an early H peak, while the stressed syllable is assumed to be unaccented. Inversely, in indirect questions, the word-initial syllable is unstressed and toneless, while the stressed syllable bears a falling pitch accent. In similar fashion, the monosyllable *min*, which is lexically stressed in both conditions, is marked by an early H peak in direct questions, while in indirect questions, it bears a falling pitch accent. In short, our data does not provide unambiguous support in favour of mass enhancement in complex wh-words neither when the word-initial syllable carries an early H peak (direct) compared to toneless word-initial syllables (indirect) nor when the stressed syllable is marked by a pitch drop (indirect) compared to stressed unaccented ones (direct). In both cases, we only see a subtle tendency of mass enhancement on the word-initial syllable marked by the early H peak as well as a subtle tendency for mass reduction of the stressed unaccented syllable compared to the stressed accented one. Hence, mass appears to remain rather constant, and generally unaffected by the rising or falling intonational events on the wh-word, while the stressed syllable appears to preserve its prosodic strength in all cases (i.e., regardless of being associated with a pitch accent or not).

These results are in line with [4] where the authors report no durational enhancement of stressed syllables marked by a falling pitch accent compared to stressed unaccented syllables. Likewise, they found no durational enhancement of word-initial syllables marked by an early H peak compared to toneless word-initial syllables in complex wh-words. Remember that mass reflects an integration of duration and periodic energy. In parallel work where we pursue comparable comparisons exploring duration and intensity in isolation as a function of tonal event, duration appears to be the more stable dimension that reliably favours stressed syllables (as shown in [4]) such that most of the subtle differences in the trends we encounter stems from the power dimension. Thus, mass values mostly replicate the main trends found for duration in [4]. [4] suggests that a plausible reason of lack of durational enhancement (and thus in our case, of mass enhancement) as a function of tonal association may be that Maltese possibly does not exhibit as much accentual lengthening as English.

Let us now move to the case of the monosyllable *min*. *Min* is the only wh-word for which our model provides strong evidence in favour of prosodic enhancement when it is marked by an early H peak (direct) as opposed to when it bears a falling pitch accent (indirect). This result makes it tempting to speculate that the prosodic enhancement related to the early H peak points towards a prominence cueing function. However, we cannot claim this with certainty because *min* is always lexically stressed, and thus we cannot tease apart the contribution of stress. An alternative explanation that can be entertained for this is that there could be a L* accent on the stressed syllable in the direct questions. For *min*, the presence

of a H edge tone and a L* accent tone would result in tonal crowding onto the one syllable. The increased mass could therefore be the result of a cumulative effect of tonal association.

Except for the case of the monosyllable *min*, the present study does not provide evidence for a prominence cueing function linked to early H peaks, although it does not conclusively rule out this possibility either. This is so because mass, which is driven to a large extent by duration, is affected by the steady duration patterns of lexical stress in Maltese, the role of which in the intonational system of Maltese is not yet clear [cf. 4]. Nonetheless, what is still interesting in Maltese is that this alternation between head and edge association, which is important for pragmatic meaning marking, is typologically rare [2]. Precisely because the current understanding of prosodic typological diversity across languages cannot easily predict such an alternation, one could argue that this early H peak instead of having an association to the left edge of the wh-word (ωH as proposed by [4]) could possibly be attributed to a shift in the stress position from the penultimate to the initial syllable of the word. For example, English words undergo a stress shift under conditions of stress clash (e.g., thirTEEN vs. THIRteen MEN, cf. [14]). In such contexts, [14] reported that stress shift results in early accent placement. Although an interpretation on these lines is plausible, the present study rules it out on a number of counts: (1) In Maltese, the early H peak is not phonologically conditioned as in English, but rather pragmatically conditioned. In other words, the presence of the early H peak depends on the sentence modality or the pragmatic force and not on following/ preceding words and their stress patterns. (2) Remember that in the current study we investigated wh-words of varying syllable length. To put it another way, in the wh-words under investigation different numbers of syllables occur between the word-initial syllable (i.e., left word boundary) and the syllable that carries the main lexical stress [see section 2]. (3) Our results show that, when there is an early H peak, the mass of the stressed syllable remains stable (and prosodically strong), regardless of its distance from the early H peak. (4) We find that the word-initial syllable, whether marked by the early H peak or not, exhibits on average lower mass than the stressed syllable. Duration results in [4] corroborate our claim. Like our results, in [4] the presence of an early H peak was not found to affect the duration of the stressed syllable despite the number of the syllables interposed between the early H peak and lexical stress. (5) Finally, the word-initial syllable, with or without an early H peak on it, was found to be consistently shorter than the lexically stressed syllable.

In its most concise form, our main result can therefore be formulated as follows: in Maltese (at least in complex) wh-words, mass enhancement is not conditioned by tonal association (neither head nor edge) but rather by lexical stress. We found that the lexically stressed syllable preserved its prosodic strength, i.e., was always prosodically strong, independently of whether it bore a falling pitch accent or had level pitch (which may or may not be the reflex of an accent). Moreover, the present study showed that an early H peak on the word-initial syllable did not affect the relative prominence between the stressed syllable and this syllable. An H tone is thus unlikely to be used to cue prominence on the initial syllable. It is nonetheless possible that the early H peak has a different prominence-related function, such as a prominence cueing function at the word level (i.e., to make the entire word more prominent in relation to other words in the utterance) [15;16]. Future research is planned to address this matter.

5. Acknowledgements

6. References

[1] A. Vella, "Prosodic structure and intonation in Maltese and its influence on Maltese English," University of Edinburgh, Doctoral dissertation, 1995.

[2] A. Vella and M. Grice, "Association of Tones in Maltese," ms under revision for S.-A. Jun, Ed. *Prosodic Typology III*, Oxford: Oxford University Press.

[3] S.-A. Jun, "Prosodic typology: by prominence type, word prosody, and macro-rhythm", in *Prosodic Typology II: The phonology of intonation and phrasing*, 2nd ed., S.-A. Jun, Ed. Oxford: Oxford University Press, 2014, pp. 520-539.

[4] M. Grice, A. Vella, and A. Bruggeman, "Stress, pitch accent, and beyond: Intonation in Maltese questions," *Journal of Phonetics,* vol. 76, 100913, Sep 2019.

[5] A. Vella, "The phonetics and phonology of wh-question intonation in Maltese," in *Proc. 16th International Congress of Phonetic Sciences,* Saarbrücken, Germany, 2007, pp. 1285-1288.

[6] A. Vella, "Alignment of the 'early' HL sequence in Maltese falling tune wh-questions," in *Proc. 17th International Congress of Phonetic Sciences*, Hong Kong, China, 2011, pp. 2062-2065.

[7] A. Albert, F. Cangemi, T. M. Ellison, and M. Grice, "ProPer: PROsodic analysis with PERiodic energy," Jan 2022. [Online]. Available: osf.io/28ea5.

[8] A. Albert, F. Cangemi, and M. Grice, "Using periodic energy to enrich acoustic representations of pitch in speech: A demonstration," in *Proc. 9th International Conference on Speech Prosody*, Poznan, Poland, 2018, pp. 1-5.

[9] F. Cangemi, A. Albert, and M. Grice, "Modelling intonation: Beyond segments and tonal targets," in *Proc. 19th International Congress of Phonetic Sciences,* Melbourne, Australia, 2019, pp. 572-576.

[10] R Core Team, R: *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021.

[11] B. Carpenter, A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell, "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, vol. 76, no. 1, pp. 1-32, Jan 2017.

[12] P.C. Buerkner, "Brms: An R package for Bayesian multilevel models using Stan," *Journal of Statistical Software*, vol. 80, no. 1, pp. 1-20, Aug 2017.

[13] M. Franke and T. B. Roettger, "Bayesian regression modeling (for factorial designs): A tutorial," Jul 2019. [Online]. Available: psyarxiv.com/cdxv3.

[14] S. Shattuck-Hufnagel, M. Ostendorf, and K. Ross, "Stress shift and early pitch accent placement in lexical items in American English," *Journal of Phonetics*, vol. 22, 4, pp. 357-388, Oct 1994.

[15] M. Grice, "Commentary: The autosegmental-metrical model of intonational phonology," In *Prosodic Theory and Practice*, S. Shattuck- Hufnagel and J. Barnes, Eds., The MIT Press, 2021, pp. 66-75.

[16] A. Arvaniti, "Author response to the commentary: Prosodic typology and the handling of variability in intonation," In *Prosodic Theory and Practice*, S. Shattuck- Hufnagel and J. Barnes, Eds., The MIT Press, 2021, pp. 76-82.