Disentangling emphasis from pragmatic contrastivity in the English H* ~ L+H* contrast

Amalia Arvaniti, Stella Gryllia, Cong Zhang, Katherine Marcoux

Radboud University, Netherlands

amalia.arvaniti@ru.nl, stella.gryllia@ru.nl,cong.zhang@ru.nl, katherine.marcoux@ru.nl

Abstract

English H* and L+H* indicate new and contrastive information respectively, though some argue the difference between them is solely one of phonetic emphasis. We used (modified) Rapid Prosody Transcription to test these views. Forty-seven speakers of Standard Southern British English (SSBE) listened to 86 SSBE utterances and marked the words they considered prominent or emphatic. Accents (N = 281) were independently coded as H* or L+H* using phonetic criteria, and as contrastive or non-contrastive using pragmatic criteria. If L+H* is an emphatic H*, all L+H*s should be more prominent than H*s. If the accents mark pragmatic information, contrastivity should drive responses. Contrastive accents and L+H*s were considered more prominent than non-contrastive accents and H*s respectively. Individual responses showed different strategies: for some participants, all L+H*s were more prominent than H*s, for others, contrastive accents were more prominent than non-contrastive accents, and for still others, there was no difference between categories. These results indicate that a reason for the continuing debate about English H* and L+H* may be that the two accents form a weak contrast which some speakers acquire and attend to while others do not.

Index Terms: English, H*, L+H*, pragmatics, Rapid Prosody Transcription

1. Introduction

An ongoing debate in the study of English intonation relates to the status of H* and L+H*. According to [1] and [2], H* and L+H* are phonologically distinct and encode new and contrastive information respectively: H* is the nuclear accent in broad focus declaratives and L+H* the nuclear accent in narrow focus declaratives. Ladd [3], however, argues that H* and L+H* are realizations of the same accentual category that varies under emphasis, with L+H* reflecting more emphatic renditions. If so, then L+H* and H* represent the ends of an emphasis continuum which is, by definition, paralinguistic [4].

Research so far has not provided a definite answer to this disagreement. First, it is possible that some English varieties lack the H* ~ L+H* contrast. This is reported for Minnesotan English [5], and British English (with speakers of a variety of regional accents) [6], [7] (cf. [8] for similar results for German). Further, even when the contrast is present, the distinction between the functions of H* and L+H* is not always clear; e.g., Watson *et al*. [9, pp. 1232] found that although in American English L+H* "creates a strong bias toward contrast referents […] H∗ is compatible with both new and contrast referents."

Perception studies have yielded mixed results as well. In an early study, Ladd & Morton [10] hypothesized that speakers would use a continuum if asked to rate utterances for emphasis, since emphasis is a paralinguistic gradient, but they expected them to respond in a categorial fashion when asked whether the utterances indicated a routine or unusual event (H* and L+H* respectively). The participants, however, responded to both questions in a gradient fashion, possibly because the routine-unusual dimension forms a paralinguistic continuum similar to that expected for emphasis.

The above findings suggest that it is not always easy to disentangle paralinguistic from linguistic aspects of intonation, particularly when participants are asked to make metalinguistic decisions. To avoid this issue, we carried out a Rapid Prosody Transcription (RPT) study [11], [12], [13]. In RPT, participants mark the words they perceive as prominent and researchers examine the correlation between p-scores (the percentage of participants who marked a given word as prominent) and phonetic and pragmatic correlates, such as average F0, duration, pragmatic importance and the like.

Here, we modified RPT to fit our research goals: we focused on accents already categorized as H* or L+H* based on phonetic criteria and independently categorized as contrastive or non-contrastive based on pragmatic criteria. We hypothesized that if RPT responses are based on the degree of emphasis encoded by each accent, and if L+H*s are emphatic versions of H*s, then L+H*s would be generally marked as more prominent than H*s independently of whether they were used contrastively or not. If, however, prominence is based on pragmatic importance, then all contrastive accents, independently of their phonetics should be marked as prominent. Should both phonetics and pragmatics play a part in determining prominence responses, the results would indicate that contrastiveness and emphasis are two independent dimensions encoded by these intonation categories.

2. Methods

2.1. Participants

We report data from 47 SSBE speakers out of 85 that we recruited through Prolific (https://www.prolific.co/). Selection criteria available on Prolific were used to ensure the right demographic was targeted. The participants whose data are reported were 19–47 years of age (mean = 33.77, S.D. = 7.7) and had been brought up in monolingual English-speaking households (some participants reported some knowledge of languages other than English that they had studied at school). No speaker reported any history of speech or hearing disorders.

As an extra precaution, questions similar to those used to screen participants on Prolific were included at the end of the task, excluding seven participants who answered that they did not have a SSBE accent. The responses of another 25

participants who did not complete the task were also discarded. Finally, six participants were excluded because they did not understand the task and marked as prominent words unlikely to be so, such as function words. Exclusion was based on the following criteria: (i) in utterances with five words or less, these participants marked all words as prominent; (ii) in utterances with more than five words, they marked more than 85% of the words as prominent; (iii) responses had to fit criteria (i) and (ii) in more than 10% of a participant's responses for the participant to be removed. The exclusion of these six participants did not result in substantive changes in the results. Details can be found in the OSF repository for this paper (https://osf.io/wm7bc/).

2.2. RPT Stimuli

The stimuli were 86 utterances selected from the data of 8 SSBE speakers (5 F, 3 M), aged 18–54 years of age (mean = 29.25, S.D. = 12.28), who had been recorded for a production study.

Twenty-two of the utterances were extracted from read text (a fable and a news item) and sixty-four from spontaneous monologues and dialogues (see (1)-(2)). The monologues were produced using *story cubes*, a game in which speakers create a story based on throwing dice with various icons on them and incorporating the items from the icons into their story. The dialogues came from map tasks [14] and discussions on the possible function(s) of a number of unusual objects [15].

(1) *I don't have a pink flamingo*

(2) *Well, they're tweezery things of some kind*

The total number of words in the 86 utterances was 879. The utterance length was 3–24 words (Mean: 10.2; SD: 4.7), and included 5–34 syllables (Mean: 13.3; SD: 6.5). The duration range was 0.52–6.8 s (mean: 2.6, SD: 1.4). The number of utterances per speaker was as follows: SP01, 11; SP02, 13; SP03, 10; SP04, 12; SP05, 2; SP06, 16; SP07, 5; SP08, 17.

2.3. Categorization of the stimuli

The utterances used in the study were selected with a view to testing the prominence ratings of words with high and rising accents used in contrastive and non-contrastive contexts. To this end, the stimuli were annotated by the first author both for the phonetic identity and pragmatics of the accents in question [cf. 16]. The annotations were checked by the second author.

Phonetically, high and rising accents were categorized into H* and L+H*. Accents were annotated as L+H* if they showed a deliberate F0 dip at the onset of the accented syllable. The dip had to be at a low level in the utterance's range (i.e., it should not be the result of high and rising F0). Accented syllables that started with a voiceless onset were annotated as L+H* if it could be ascertained that the preceding coda had low F0. When this was not possible, the accent was annotated as H*. Following MAE ToBI conventions [17] and in the absence of guidelines specific to SSBE, accents in absolute utterance initial position were classified as H*. Accents followed by uptalk were not considered in the analysis. Although we acknowledge that these accents could be particularly prominent, they were not included because it would not be possible to separate the effect of the accent from that of the final rise.

In terms of pragmatics, the accents were annotated as *contrastive* or *non-contrastive* based on context (rather than the perspective of the speaker or the potential addressee). To avoid circularity, the pragmatic annotation was done using orthographic transcripts, so it did not take into consideration the shape or auditory impression of the accent, only its role in discourse. *Contrastive* was used for accents that highlighted one item out of a small set of (inferred or mentioned) alternatives in discourse. Focus particles, such as *just*, *only*, and similar items were also marked as contrastive, as were negative expressions (e.g., *don't* in *I don't know*). Finally, items were considered contrastive if they entered in an explicit parallelism in discourse (e.g., in *I want tea as well as coffee*, *tea* and *coffee* would be considered contrastive). All other high and rising accents were considered non-contrastive. This pragmatic classification cuts across the categories *new*, *given*, *topic*, and *focus*, as it is possible to contrastively accent both new and given items, as well as topics and foci, and accent type does not appear to vary based on these pragmatic distinctions [18].

The two classification schemas resulted in 287 target accents in the 86 utterances. Six of these were removed from analysis due to technical issues with the platform used for the study which, for words that appeared more than once in an utterance, did not allow us to determine which instance of the word the participant had reacted to in that utterance. In terms of the phonetic and pragmatic categories discussed above, the remaining 281 accents were distributed as shown in Table 1.

Table 1: *Distribution of examined accents by pragmatic and phonetic category.*

	Phonetics		
Pragmatics	H*	L+H*	**Total**
Contrastive	50	58	108
Non-contrastive	139	34	173
Total	189	92	281

2.4. Phonetic differences between H* and L+H* stimuli

The categorization of the accents was further examined by inspecting the average smoothed and normalized contours of the accented syllables. As shown in Fig. 1A, the accents were phonetically distinguished: syllables categorized as carrying a L+H* showed a rise-fall, while those categorized as carrying a H* were high and slightly falling. The pragmatic categorization into contrastive and non-contrastive accents did not yield substantial differences in shape and only modest differences in scaling in the expected direction (Fig. 1B). The interaction between pragmatic and phonetic classification is shown in Fig. 1C: both contrastive and non-contrastive H*s were falls, with contrastive H*s being scaled higher. Interestingly, the scaling relationship was reversed for contrastive and non-contrastive L+H*s, with the latter showing higher scaling than the former. This is a point to which we return in the discussion.

Additionally, we statistically analysed the normalized (z-scored) duration of the accented syllables, and their RMS amplitude, normalized by dividing the RMS of the syllable by that of its utterance. Accent, pragmatics, and their interaction were included as fixed factors with speaker as random intercept. For duration, the results showed that none of the fixed factors were significant ($\beta_{accent:\ L+H*}$ = 0.31, t = 1.66, p = 0.10; $\beta_{pragmatics:\ contrastive}$ = 0.17, t = 1.06, p = 0.29; $\beta_{accent:\ L+H* \times pragmatics:\ contrastive}$ = -0.24, t = -0.89, p = 0.37). For amplitude, the model did not converge with the random intercept, so it was removed and a linear model was run. Both accent and pragmatics were significant, but not their interaction ($\beta_{accent:\ L+H*}$ = 0.32, t = 4.61, p < 0.001; $\beta_{pragmatics:\ contrastive}$ = 0.14, t = 2.33, p < 0.05; $\beta_{accent:\ LH \times pragmatics:\ contrastive}$ = -0.19, t = -1.9, p = 0.06). In short, syllables with L+H* had higher RMS amplitude than syllables with H*; the same applied to accented

syllables in words categorized as contrastive relative to non-contrastive ones.

2.5. Procedure

The RPT study ran on ROLEG, an online platform developed by the Centre for Language Studies at Radboud University. After giving consent, participants were given instructions and presented with two practice trials. The study, which started after these, was self-paced and included four prompts for self-paced breaks. In each trial, participants would first hear an utterance while seeing it in writing on screen; then they heard a second repetition during which they could respond to the written transcript. The transcript did not include punctuation or capitalization; however, apostrophes were retained in contractions and possessives, and proper nouns and the pronoun *I* were capitalized. The second presentation of the orthographic version of each utterance included a checkbox to the right of each word. Participants were instructed to click on all the buttons next to words they heard as "prominent, stressed, highlighted, important, or emphasised". Participants were told they could select as many words as they saw fit, but had to select at least one in order to proceed to the next trial.

At the end of the task, participants were presented with the same list of terms they saw in the instructions (*prominent*, *stressed*, *highlighted*, *important*, *emphasised*) and were asked to choose those that best represented the criteria they used to complete the task. We hypothesized that *highlighted* would be ambiguous between phonetics and pragmatics but we anticipated that *important* would be chosen by participants who focused on meaning, and *prominent*, *stressed* and *emphasised* by participants who focused on auditory impression.

2.6. Processing of responses and statistical analysis

The responses were put into a binary data frame (1 = the item was selected by the listener). For statistical analysis, these responses were used as the dependent variable in Binomial Generalized Linear Mixed-Effects Models built using the lme4 package [19] in R [20]. The fixed effects included accent (H*, L+H*, as determined by the phonetic annotation), pragmatics (contrastive, non-contrastive, as determined by the pragmatic annotation), and their interaction. Treatment coding was used since the data frame was a binary dataset and both categorical variables had two levels. The random effects included random intercepts for speaker, listener, and the accented word. The responses were then counted and the *prominence score* or *p-score* of each word carrying H* or L+H* was calculated. The p-scores are used in the figure for illustration purposes.

3. Results

3.1. Prominence-scores

The p-scores by accent and pragmatics are presented in Fig. 2. Statistical analysis showed that both accent and pragmatics affected p-scores: overall, the mean p-score for H* was 36.1%, and for L+H*, 62.5%; for contrastive accents, the p-score was 60.2% and for non-contrastive accents, 35.1%. As shown in Table 2, the interaction between the two factors was not statistically significant. Contrastive items accented with L+H* were considered the most prominent (69.7%), while non-contrastive items accented with H* were considered the least prominent (31.4%); non-contrastive L+H*s and contrastive H*s were considered equally prominent (50.2% and 49.2% respectively).

The results are also presented as density plots in Fig. 3 to better visualize the relationships between categories. As shown, the p-scores for non-contrastive H*s are concentrated on one end of the distribution, and those of contrastive L+H*s at the other end, while the overlap between them is moderate. In contrast, contrastive H*s and non-contrastive L+H*s form entirely overlapping distributions that span the entire p-score range with all values being almost equally likely.

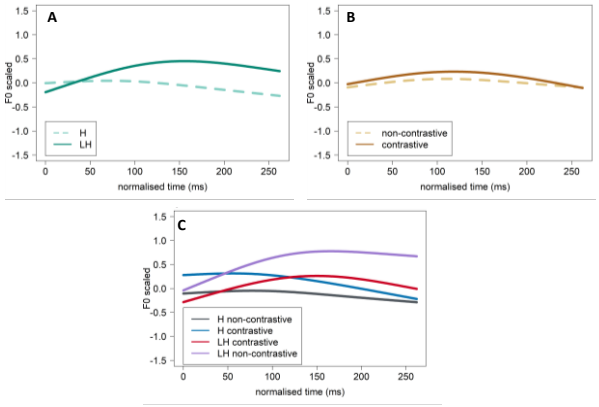

Figure 1. *Smoothed and normalized average F0 curves of accented syllables by phonetic category in A, by pragmatic category in B, and by pragmatic and phonetic category in C.*

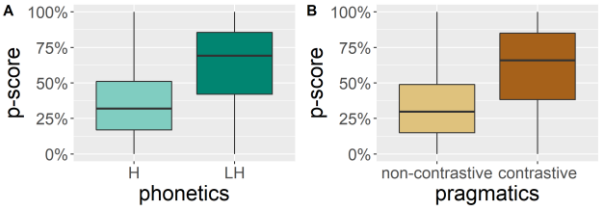

Figure 2: *p-scores by accent in A and pragmatics in B.*

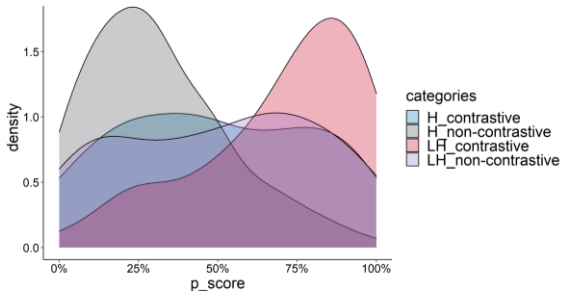

Figure 3. *Density plot of p-scores separately for the four combinations of accent type and pragmatic category.*

Table 2: *Results of the statistical analysis of binary PRT responses (for details see section 2.6).*

	Estimate	SE	z value	Pr(>|z|)
(Intercept)	-1.37	0.29	-4.72	< 0.001
Accent:LH	1.15	0.32	3.60	< 0.001
Pragmatics:Contrastive	1.12	0.28	4.02	< 0.001
Accent:LH × Pragmatics:Contrastive	0.20	0.46	0.43	0.669

3.2. Prominence criteria and p-scores

We also considered the terms participants selected as their criteria during the task and compared these to their responses. The preferred terms were *prominent*, *stressed* and *emphasised*. They were selected in equal measure by practically all participants: 45/47 selected *emphasised*, 43 *stressed*, and 42 *prominent*. *Important* and *highlighted* were selected only by 20 and 27 participants respectively.

These preferences suggest participants *thought* they were selecting items that stood out auditorily. However, inspection of the individual responses indicates that the criteria stated by participants did not quite hold. As Fig. 4 shows, some participants paid attention to auditory cues and generally marked L+H* accents as more prominent, independently of pragmatics (e.g., SL115), others were more sensitive to pragmatic distinctions (e.g., SL118), and still others barely distinguished the categories at all (e.g., SL101).

4. Discussion and Conclusions

The study provided novel data pertaining to the H* ~ L+H* contrast in English, SSBE in particular. Differences in the distribution of responses to the four categories were most revealing. They showed that while prototypical non-contrastive H* and contrastive L+H* had very low and very high p-scores respectively, the intermediate categories, contrastive H* and non-contrastive L+H*, elicited almost identical wide-ranging responses (Fig. 3). These distributions support the contention of Ladd [3] that H* and L+H* are the two ends of a continuum [cf. 21]. At the same time, accents that fall somewhere in between these extremes are not differentiated by the listeners in terms of prominence. This is consistent with the finding that SSBE speakers do not seem, according to [7], to use H* and L+H* to consistently encode the pragmatic distinction argued by Pierrehumbert & Hirschberg [2].

This interpretation is supported by the average contours of the stimuli, and the amplitude results. The amplitude differences between contrastive and non-contrastive accents suggest that means other than F0 may be enlisted to encode a pragmatic distinction, presumably in conjunction with linguistic means, such as deaccentuation of following non-contrastive material.

On the other hand, the F0 results paint a different picture: as discussed in 2.4, F0 was somewhat higher for contrastive than non-contrastive H*s but the reverse was the case for L+H*s. This could indicate that some phonetically L+H* accents are used to encode contrastivity, while others are used for emphasis. This was certainly the case in the stimuli from *story cubes*: speakers presented indisputably new information but frequently used exaggerated accents. These effects deserve thorough investigation in production but for now point to the need to carefully distinguish between pragmatic contrast and paralinguistic emphasis when designing studies and interpreting their results. The range of individual variation in RPT responses is an additional issue that deserves a closer look: it is clear that participants do not use the same criteria to mark prominence and are not fully aware of the criteria they use.

In conclusion, the results suggest a preference for linking each accent with a certain pragmatic function, but this is no more than a tendency. If researchers focus on the tails of the distribution of the accents' realization properties, they are likely to see the accents are forming distinct categories, but if they focus on the entire distribution, a continuum emerges. Finally, these differences in analysis may also reflect acquisition patterns among our participants, hinting at a weak contrast that some SSBE speakers acquire and associate with pragmatic distinctions while others lack.

5. Acknowledgements

The financial support of the European Research Council through grant ERC-ADG-836253 to Amalia Arvaniti is hereby gratefully acknowledged. We thank Chris Cummins and Hannah Rohde for helpful discussions on the annotation criteria for pragmatic categories; any mistakes and shortcomings remain ours. We thank Kathleen Jepson for help with the preparation of the stimuli.

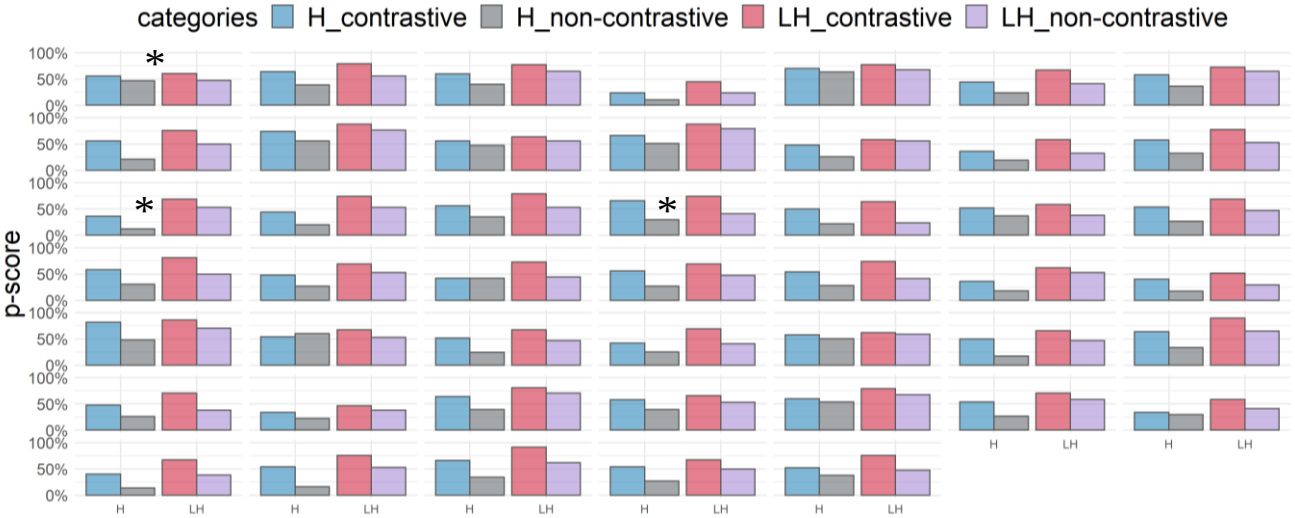

Figure 4: *p-scores by accent and pragmatic category (x axis) separately for each participant; from left to right, in row 1, participants SL101-SL107; in row 2, participants SL108-SL114; in row 3, participants SL115-SL121; in row 4, participants SL122-SL128; in row 5, participants SL129-SL135; in row 6, participants SL136-SL142; in row 7, participants SL143-SL147. Starred participants are discussed in section 3.2.*

6. References

[1] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," Doctoral Thesis, MIT. Distributed 1988, Indiana University Linguistics Club., 1980.

[2] J. B. Pierrehumbert and J. B. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," *Intentions Commun.*, pp. 271–311, 1990.

[3] D. R. Ladd, *Intonational Phonology*, 2nd ed. Cambridge University Press, 2008.

[4] D. L. Bolinger, *Intonation and Its Parts: Melody in Spoken English*. London: Arnold, 1986.

[5] A. Arvaniti and G. Garding, "Dialectal variation in the rising accents of American English," in *Papers in Laboratory Phonology 9*, J. Cole and J. H. Hualde, Eds. Berlin, New York: Mouton de Gruyter, 2007, pp. 547–576.

[6] D. R. Ladd and A. Schepman, "Sagging transitions" between high pitch accents in English: Experimental evidence, vol. 31, no. 1. 2003.

[7] K. Jepson, C. Zhang, G. Lohfink, K. Marcoux, and A. Arvaniti, "H* and L+H* in English and Greek", paper presented at *PaPE 2021, Barcelona, 21-23 June 2021.*

[8] L. E. de Ruiter, "Information status marking in spontaneous vs. read speech in story-telling tasks - Evidence from intonation analysis using GToBI," *Journal of Phonetics*, vol. 48, pp. 29–44, 2015, doi: 10.1016/j.wocn.2014.10.008.

[9] D. G. Watson, M. K. Tanenhaus, and C. A. Gunlogson, "Interpreting pitch accents in online comprehension: H* vs. L+ H," *Cogn. Sci.*, vol. 32, no. 7, pp. 1232–1244, 2008

[10] D. R. Ladd, and R. Morton, "The perception of intonational emphasis: continuous or categorical?", *Journal of Phonetics*, vol. 25, pp. 313-342, 1997.

[11] J. Cole, Y. Mo, and M. Hasegawa-Johnson, "Signal-based and expectation-based factors in the perception of prosodic prominence", *Journal of Laboratory Phonology*, vol. 1, pp. 425-452, 2010.

[12] S. Baumann, and B. Winter, "What makes a word prominent? Predicting untrained German listeners' perceptual judgments", *Journal of Phonetics*, vol. 70, pp. 20-398, 2018.

[13] S. Riesberg, J. Kalbertodt, S. Baumann, and N. Himmelmann, "Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay", *Laboratory Phonology*, vol. 11(1), pp. 1-35, 2020.

[14] A. H. Anderson *et al.*, "The HCRC map task corpus," *Lang. Speech*, vol. 34, no. 4, pp. 351–366, 1991.

[15] J. Edlund, J. Beskow, K. Elenius, K. Hellmer, S. Strömbergsson, and D. House, "Spontal: A Swedish spontaneous dialogue corpus of audio, video and motion capture," in N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, and D. Tapias (Eds.), *Proceedings of the Seventh Conference on International Language Resources and Evaluation* (LREC'10), Valetta, Malta, 2992–5, 2010.

[16] F. Cangemi, and M. Grice, "The Importance of a Distributional Approach to Categoriality in Autosegmental-Metrical Accounts of Intonation", *Laboratory Phonology* 7(1), p.9, 2016. doi: https://doi.org/10.5334/labphon.28

[17] A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux, *Transcribing Prosodic Structure of Spoken Utterances with ToBI* (MIT Open Courseware). http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-911-transcribing-prosodic-structure-of-spoken-utterances-with-tobi-january-iap-2006/index.htm, 2006.

[18] V. Molnár, "Contrast from a contrastive perspective", *Language and Computers* vol. 39, pp. 147-162, 2002.

[19] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Software1*, vol. 1, no. 1, pp. 1–48, Oct. 2015, [Online]. Available: https://www.jstatsoft.org/v067/i01.

[20] R Core Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2020, [Online]. Available: http://www.r-project.org/.

[21] A. Arvaniti, "Crosslinguistic variation, phonetic variability, and the formation of categories in intonation," in *Proceedings of ICPhS 2019*, Melbourne, Australia, Aug. 2019, https://assta.org/proceedings/ICsPhS2019/