



## Pitch accents show a Perceptual Magnet Effect: Evidence of internal structure in intonation categories

Joe Rodd<sup>1,2</sup>, Aoju Chen<sup>3</sup>

<sup>1</sup> Max Planck Institute for Psycholinguistics, Nijmegen, NL; <sup>2</sup> Centre for Language Studies, Radboud University, Nijmegen, NL; <sup>3</sup> Utrecht Institute of Linguistics OTS, Utrecht University, NL

joe.rodd@mpi.nl, aoju.chen@uu.nl

### Abstract

The question of whether intonation events have a categorical mental representation has long been a puzzle in prosodic research, and one that experiments testing production and perception across category boundaries have failed to definitively resolve. This paper takes the alternative approach of looking for evidence of structure within a postulated category by testing for a Perceptual Magnet Effect (PME). PME has been found in boundary tones but has not previously been conclusively found in pitch accents. In this investigation, perceived goodness and discriminability of re-synthesised Dutch nuclear rise contours (L\*H H%) were evaluated by naive native speakers of Dutch. The variation between these stimuli was quantified using a polynomial-parametric modelling approach (i.e. the SOCoPaSul model) in place of the traditional approach whereby excursion size, peak alignment and pitch register are used independently of each other to quantify variation between pitch accents. Using this approach to calculate the acoustic-perceptual distance between different stimuli, PME was detected: (1) rated “goodness” decreased as acoustic-perceptual distance relative to the prototype increased, and (2) equally spaced items far from the prototype were less frequently generalised than equally spaced items in the neighbourhood of the prototype. These results support the concept of categorically distinct intonation events.

**Index Terms:** Categorical structure of intonation, phonology of intonation, Perceptual Magnet Effect, parametric modelling

### 1. Categorical vs. gradient intonation

The question of whether intonation events have a categorical mental representation has long been a puzzle in prosodic research [1]. Traditionally, researchers have attempted to answer this question by examining perception and production across boundaries of alleged intonation categories differing in peak height or peak alignment. Evidence that both supports and contradicts the categorical representation of certain intonation events has been reported for different languages.

For instance, using the categorical perception (CP) paradigm consisting of a forced-choice identification task and a discrimination task, [3] examined the difference between a “normal” high and “emphatic” high pitch accent in English and found an identification boundary but no discrimination peak [3]. When reaction time (RT) was measured during the identification task on a comparable stimuli set, slower reactions were found at the identification boundary,

strengthening a categorical interpretation of peak height in English intonation [4]. In Central Catalan, counter-expectational questions, narrow- and contrastive statements are all realised on L\*H + L%, with varying peak height. [5] combined the CP paradigm with the RT measurement and found that differences between the “question” meaning and either “statement” interpretation were represented categorically, but the two “statement” meanings were not categorically distinct. In Bari Italian, an S-shaped curve and RT peak were detected for the identification of information-seeking (L+H\*) vs. counter-expectational questions (on L+H\* with a higher peak F<sub>0</sub>), but no clear discrimination peak was observed [6]. In Spanish, similar categorical and gradient distinctions were found in utterance-initial pitch peaks (H1) between (lower) statements and (higher) non-statements [7].

Regarding peak alignment, [8] used a repetition experiment to test for categoricity between English L\*+H and L+H\*: participants were asked to repeat stimuli from a continuum that varied in peak alignment in 20ms steps. The repetitions fell into two categories, leading the authors to conclude that the peak alignment dimension is represented in a binary manner. However, using a CP-with-RT approach revealed no evidence of categorical perception of a similar stimulus continuum [4].

#### 1.1. The Perceptual Magnet Effect

Since examining perception across the boundaries of categories has presented evidence both supporting and contradicting a categorical account of intonation representation, it is desirable to consider other ways to answer the question of whether intonation is represented categorically. An alternative approach is to look for evidence of categoricity *within* a postulated category. The Perceptual Magnet Effect (PME) is such evidence of categoricity [9], [10].

PME is a pair of experimentally-observable “symptoms” that result from the presence of the internal structure of categories, e.g. colours and sounds. The first symptom is that the variability in density and activation of exemplars corresponds with a measurable preference for items in the centre of the category, with more peripheral items dis-preferred as members of the categories. We term this *gradient goodness*. The second is that the participants’ success in discriminating equally spaced pairs of items taken from the vicinity of the category centre is reduced compared to equivalent pairs of items taken from further away from the category centre. We term this *differential discriminability*.

Although PME has been attested in segmental sound categories (vowels: [9], [12], stops and liquids: [13], [14]), the

evidence for PME in intonation categories is thin on the ground, and at best partially conclusive. [15] found both PME symptoms in German H% and L% boundary tones. [16] found gradient-goodness in a set of Pisa Italian pitch accent stimuli that varied simultaneously in peak alignment and peak height from H\*L to H\*, but failed to find differences in discrimination performance between the prototype neighbourhood and the non-prototype neighbourhood.

PME is reliant on a concept of acoustic-perceptual distance, to define how far an exemplar is from the category centre, and to define the spacing between pairs of items. This distance metric should be derived from quantification of the acoustic variation that causes change in category identity. In the segmental domain, this is simple: formant frequencies characterise vowels, for instance, whilst voice onset time conveys the voicing distinction in stops. Intonation, as changes in  $F_0$  in time anchored to the segmental stream, is by definition multi-dimensional: changes in pitch scaling, peak- and valley alignment and accent duration all conceivably contribute to category identity. This multi-dimensional nature of intonation suggests that how the acoustic-perceptual distance between exemplars is quantified may be the cause for absence of differential discriminability in [16]. More specifically, [16] tested a CP-like continuum where both alignment and scaling were altered in each step between two targets, meaning only a very thin slice through the category space was examined, and any independent influence of alignment or scaling on identity was impossible to detect.

## 1.2. This investigation

This investigation tests for the presence of PME in Dutch L\*H pitch accent using the paradigm established in [9]. In contrast to [16], we adopt a parametric modelling approach to quantify acoustic-perceptual distance between variants of the pitch accent, and test many more stimuli. This allows us to respect the multi-dimensional nature of intonation, and capture a much broader range of variation in realisation than was possible in that investigation (more on this in section 2.1).

## 2. Method

Testing for PME requires a set of stimuli that vary systematically over a large section of the postulated category, including the area near the prototype and an area further from the prototype, to allow for testing of differential discrimination near and far from the prototype. We will first discuss the parametric modelling approach used to define the dimensions of the acoustic-perceptual space, then describe the arrangement of the stimuli in that space and their construction from the contours predicted by the model, and finally present details of the experiments.

### 2.1. Modelling approach

The method used to quantify pitch accent variation in this investigation is adapted from recent work by Reichel and colleagues [17–20]: the CoPaSul (contour, parametric, superpositional) intonation model. It models a linear global declination contour in the domain of the intonation-phrase, then uses a series of parametrically defined third-order polynomial functions, as [21] used, to stylise the residual movement in the domain of the accent group.

We adapted CoPaSul in two ways: by (1) removing the global contour, which models declination in connected speech, and is thus not relevant in our isolated stimuli, and (2) substituting CoPaSul's natural polynomials for orthogonal

polynomials. This change means that the parameters are not by default correlated with each other, as they are with natural polynomials, resulting in a round rather than ovoid exemplar cloud, making the calculation of acoustic-perceptual distance between exemplars and the placement of referents to test more straightforward. For convenience, we refer to the resulting model as Simplified, Orthogonal CoPaSul (SOCoPaSul).

SOCoPaSul characterises different intonation contours in terms of four parameters, which are the parameters of the third order Legendre orthogonal polynomial (Figure 1): a parameter controlling the local intercept (INTERCEPT), two inter-related parameters that control the rising or falling direction of the intonation contour and the peak alignment (CO1 and CO3) and a parameter controlling peak shape, from convex to concave (CO2). The interactions between parameter values create more complex shapes. To complete our characterisation of the prosodic properties of each pitch accent exemplar, we also add its duration as a fifth metric, to capture the interaction of duration with the other parameters. We propose that the five SOCoPaSul parameters are the dimensions of variation in the exemplar cloud for pitch accents, meaning that pitch accents form a five-dimensional exemplar cloud. To calculate the acoustic-perceptual distance between two exemplars, we calculate the euclidean distance in this five-dimensional space.

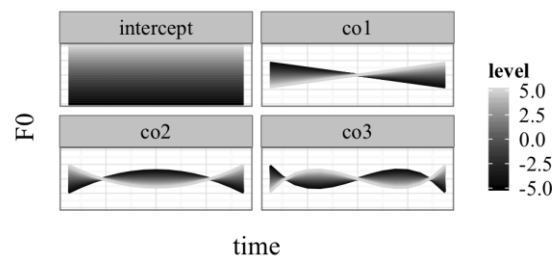


Figure 1: *Parameters in SOCoPaSul. This figure shows, in panel, the effect on the contour shape of changing each parameter from a high value (light) to a low value (dark), whilst holding the values of all the other parameters constant.*

### 2.2. Stimuli

This investigation examines isolated L\*H (rise) pitch accents on the Dutch one-word utterance “mi”. Their isolation means that there is also a high boundary tone (H%) present marking the end of the phrase. Following [22], we assume that the variation created in our stimuli did not change the (consistently high) identity of the boundary tone, but does influence the perception of the pitch accent. We thus refer to the stimuli as “pitch accents” for the remainder of the paper.

To create our stimuli, we started by using SOCoPaSul to characterise [23]’s natural productions of L\*H. This gave us an exemplar cloud with variable density, varying considerably in all five SOCoPaSul dimensions. These productions were subsequently rated for prototypicality by native speakers of Dutch [24]. The raters showed clear preference for the items near the centre. We then selected the location of the highest rated item as our first ‘referent’, the prototype referent. From near the edge of the cloud of natural productions, we selected two points in space to serve as potential non-prototype referents (the inside-limit non-prototype referents). These were equally far from the prototype. To evaluate the impact of extreme pitch register, out of the range of the natural productions of L\*H, we additionally created two further

referent points, by shifting the non-prototype referents further from the prototype in the INTERCEPT SOCoPaSul dimension (the outside-limit referents, see Figure 2). In total, five referents were determined, one prototype referent and four non-prototype referents, i.e. two inside-limit ones and two outside-limit ones.

Around each referent, we created a pattern of "neighbouring" points, arranged in a star-burst pattern, so that there were neighbours that were close to the referent, and neighbours that were further from the referent. Around the prototype, we created two further layers of neighbours that were located further from the referent, which were used in experiment 1 only.

The coordinates of all the points were then passed to SOCoPaSul, to create a pitch contour for each of the referents and neighbours. PSOLA resynthesis in Praat [25] was used to apply these contours to the best rated item from [24]. This gives us a very large set of subtly varying L\*H contours.

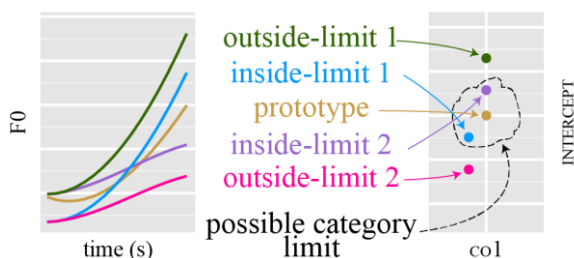


Figure 2: The contour shape of each of the referents, and their arrangement (diagrammatic, neighbours not shown).

### 2.3. Experiment 1: Gradient goodness

Ten participants (students, L1 Dutch, six females, mean age: 22;2) rated the perceived goodness of the resynthesised stimuli on a five point equal-appearing interval (EAI) scale in a computerised experiment in which the stimuli were presented over headphones. The key instruction was "determine how typical the rising melody of each example sounds in Dutch", the two ends of the EAI scale were labelled "bad example" and "good example".

All participants rated the prototype-referent stimuli. They each additionally rated one of the non-prototype referent sets (either two or three participants per set), meaning that each participant rated 1117 items in total. The experiment was spread over five twenty-minute blocks with breaks, conducted in two sessions on sequential days.

### 2.4. Experiment 2: Differential discrimination

Another fifteen participants (students, L1 Dutch, twelve females, mean age: 21;1) performed a computerised experiment (30 minutes) where they listened to pairs of stimuli and reported whether or not they perceived them as different, by pressing computer keys labelled "yes" or "no". The key instruction was "determine whether you hear a difference between the two examples".

Each participant was tested on pairs of items consisting of: (1) the referent and a comparator taken from the neighbours closest to the referent (small difference *test trials*, 41%), or (2) the referent and a comparator taken from the neighbours slightly further from the referent (moderate difference *test trials*, 41%) or (3) the referent repeated (*control trials*, 18%).

Two blocks were conducted, one where the referent was the prototype (180 trials) and one where the referent was one of the four non-prototypes (180 trials). Block order and presentation order within each block and in each trial were counterbalanced. Four participants were tested on each of the within-limit non-prototype referents and the outside-limit 2 non-prototype, three participants were tested on the outside-limit 1 non-prototype referent.

## 3. Results

### 3.1. Gradient goodness

To test for gradient goodness, linear regression models were fitted, predicting the rated goodness of each stimulus by its Euclidean distance from the prototype. Each model used a different combination of SOCoPaSul dimensions to calculate the distance from the prototype. All the plausible models were constructed and evaluated in an "all-subsets" approach. Additionally, we measured conventional metrics typically used in the literature to characterise pitch accents (i.e. pitch register, excursion size, valley alignment and duration) for each of our stimuli, and combined these to create further models to test in the same way.

We found that none of the models using the conventional metrics of pitch accent variation accounted for the variation in the goodness rating as successfully as the best of the models incorporating the SOCoPaSul parameters, i.e. the model which excluded CO3.

The best-fitting model, depicted in Figure 3, has a very clear, statistically significant downward pattern (model  $R^2 = .37$ , estimate = -0.576,  $t = -23.844$ ,  $p \leq .01$ ): items closer to the prototype received significantly higher ratings than those further from it, in line with the gradient goodness symptom.

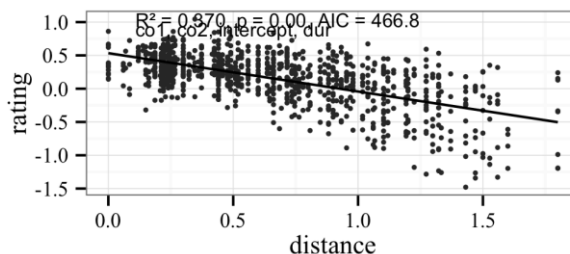


Figure 3: The model that explains the most variation, as evaluated by  $R^2$ , characterises the distance from the prototype in (CO1, CO2, INTERCEPT, duration) space.

### 3.2. Differential discrimination

The differential discrimination symptom of PME is present if participants exhibit more generalised trials ("misses": test trials in which the difference was not detected, [9]) in the prototype condition (represented with dashed lines in Figure 4) than in the non-prototype condition (represented with solid lines in Figure 4). We found this typical pattern in two of the four groups of participants, i.e. those tested on the within-limit non-prototypes. Additionally, increasing the difference (from + to ++ in Figure 4) between the comparator and the referent reduced generalisation in two of the within-limit non-prototype conditions and one outside-limit non-prototype condition.

The outside-limit non-prototypes exhibited the opposite pattern, with greater generalisation in the non-prototype than in the prototype. This suggests that abnormal pitch register interrupts discrimination similarly to prototypicality. Since all participants were tested on the same set of prototype-referent test items, the differences between the groups on these items (dashed lines in Figure 4) suggests individual differences in their inherent sensitivity to subtle contrasts such as these. To take individual differences into account, we subsequently conducted logistic mixed-effects modelling on the whole dataset. The outcome variable was binary (generalised, coded 1, or not generalised, coded 0). The fixed factors were the prototypical status of the referent and distance of the comparator from the referent, the random factor was participant nested within group. The most predictive model (selected on BIC, Table 1) contained only the main effects of the fixed factors, prototypicality and difference from the referent, no interaction term was present. It indicated that non-prototypicality significantly decreased generalisation, and that increased difference from the referent significantly increased generalisation across the four groups of participants.

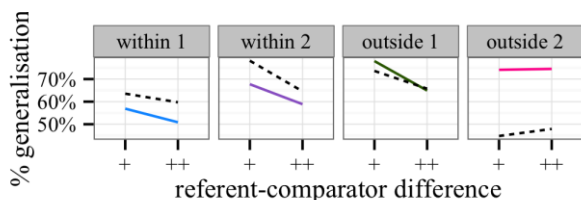


Figure 4: Misses (generalisation) in the prototype (dashed) and non-prototype (solid) conditions.

	estimate	SE	z	p
(intercept)	-0.57	0.243	-2.346	≤.05
+ non-prototype	-0.316	0.068	-4.674	≤.01
+ moderate difference	0.271	0.068	4.010	≤.01

Table 1: The best fitting model of differential discrimination.

## 4. Discussion

The results of the goodness experiment demonstrate quite clearly that the gradient-goodness symptom of PME is present in the Dutch L\*H pitch accent. The discrimination results of participants tested on the within-limit non-prototypes demonstrate that the differential discrimination symptom of PME is also present. Thus, PME is present in the Dutch L\*H pitch accent. Notably, the evidence gathered here appears not to be as strong as that reported for vowels [9]. More specifically, the generalisation rate in our materials is much higher than in [9]; the general discrimination accuracy was also much lower in this investigation than in [9]. These differences suggest that it may be more difficult for listeners to detect differences between tokens of pitch accents than in tokens of vowels.

The model using the SOCoPaSul parameters to characterise acoustic-perceptual distance was more successful in explaining the variation in the goodness rating than the model using the conventional quantifications of contour shape variation. This suggests that there is merit in such a parametric approach to stylising intonation contours, and for interpreting the parameters as the dimensions of perceptual space.

The discrimination performance on the outside-limit non-prototypes and the finding that duration and pitch register

variation are predictive of rating in the goodness experiment calls into question the long-standing consensus that pitch register and duration are “normalised out” as speaker-dependent variation [26]. Instead, these results support a mental representation of pitch accents incorporating these two phonetic dimensions of variation as phonologically salient.

The exclusion of the CO3 parameter from the best-fitting model of goodness rating is surprising: it suggests that variation in this dimension was not perceived as salient by the listeners. However, the fully specified model was only marginally less predictive, arguably within the margin of error of the ratings. The CO3 parameter, the cubic function, can be considered to control the degree of deviation from the overall curve trend at the extremities of the contour in the context of the isolated L\*H accents modelled here. It may have more influence in defining the shapes of more complicated accents involving both a rising and falling component, e.g. H\*LH.

After this paper’s original submission, we repeated the discrimination sub-test under different conditions in order to compare discrimination in all of the non-prototype neighbourhoods in a within-participant design. Participants were tested individually in a much longer session (90 minutes, compared to 30 in the original experiment) in a sound-attenuated booth. We found no difference in performance between blocks of items from the neighbourhood of the prototype and blocks of items from within or outside the postulated category boundary. The absence of the PME discrimination symptom in this situation suggests that the very extensive exposure in a laboratory environment may have led to the listeners’ capitalising on an exclusively auditory strategy to detect differences between the stimuli, instead of a linguistic strategy.

## 5. Conclusions

The two experiments in this investigation together demonstrate that both symptoms of PME are present in Dutch L\*H pitch accents. This is additional evidence that the mental representation of intonation events such as pitch accents is categorical in nature, reinforcing the conclusions of previous investigations examining pitch accents with the CP paradigm [3], [4], and boundary tones with both the CP and PME paradigm [15]. The results of our recent follow-up highlight how delicate this phenomenon is. We suggest that, although the current method shows promise in future research as a means to investigate the categoricity of other pitch accents in Dutch and intonation events in other languages, considerable care is necessary to avoid an auditorially-driven approach to the task.

## Acknowledgements

This paper was based on the first author’s Utrecht University MA thesis ‘A perceptual magnet effect in pitch accents: Parametric modelling’. The research was funded by a Talent grant awarded to the second author, from the *Dynamics of Youth* Utrecht University strategic theme. We thank Louis ten Bosch for useful comments on experimental methodology at an early phase of the preparations and on the completed MA thesis, Uwe Reichel for comments on the orthogonal-polynomial stylisation approach adopted, and Huub van den Bergh and Tom Lentz for advice on statistical matters and on the interpretation and presentation of the results.

## References

- [1] C. Gussenhoven, "Discreteness and gradience in intonational contrasts," *Language and speech*, vol. 42, no. 2-3, pp. 283–305, 1999.
- [2] J. B. Pierrehumbert, "The phonology and phonetics of English intonation," PhD thesis, Massachusetts Institute of Technology, 1980.
- [3] D. R. Ladd and R. Morton, "The perception of intonational emphasis: Continuous or categorical?" *Journal of Phonetics*, vol. 25, no. 3, pp. 313–342, 1997.
- [4] A. Chen, "Reaction time as an indicator of discrete intonational contrasts in English," in *EUROSPEECH*, 2003, pp. 97–100.
- [5] J. Borrás-Comes, M. del M. Vanrell, and P. Prieto, "The role of pitch range in establishing intonational contrasts," *Journal of the International Phonetic Association*, vol. 44, no. 01, pp. 1–20, 2014.
- [6] M. Savino and M. Grice, "The perception of negative bias in Bari Italian questions," in *Prosodic categories: Production, perception and comprehension*, S. Frota, E. Gorka, and P. Prieto, Eds. Springer, 2011, pp. 187–206.
- [7] P. Prieto, "The search for phonological targets in the tonal space: H1 scaling and alignment in five sentence-types in Peninsular Spanish," *Laboratory approaches to Spanish phonology*, pp. 29–59, 2004.
- [8] J. B. Pierrehumbert and S. A. Steele, "Categories of tonal alignment in English," *Phonetica*, vol. 46, no. 4, pp. 181–196, 1989.
- [9] P. K. Kuhl, "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Perception & psychophysics*, vol. 50, no. 2, pp. 93–107, 1991.
- [10] F. Lacerda, "The perceptual-magnet effect: An emergent consequence of exemplar-based phonetic memory," in *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 1995, vol. 2, pp. 140–147.
- [11] E. Rosch, "Cognitive reference points," *Cognitive psychology*, vol. 7, no. 4, pp. 532–547, 1975.
- [12] O. Aaltonen, O. Eerola, Å. Hellström, E. Uusipaikka, and A. H. Lang, "Perceptual magnet effect in the light of behavioral and psychophysiological data," *The Journal of the Acoustical Society of America*, vol. 101, no. 2, pp. 1090–1105, 1997.
- [13] K. Davis and P. K. Kuhl, "Tests of the perceptual magnet effect for American English/k/and/g/," *The Journal of the Acoustical Society of America*, vol. 95, no. 5, pp. 2976–2976, 1994.
- [14] P. Iverson and P. K. Kuhl, "Tests of the perceptual magnet effect for American English/r/and/l/," *The Journal of the Acoustical Society of America*, vol. 95, no. 5, pp. 2976–2976, 1994.
- [15] K. Schneider, G. Dogil, and B. Möbius, "German boundary tones show categorical perception and a perceptual magnet effect when presented in different contexts," in *INTERSPEECH*, 2009, pp. 2519–2522.
- [16] B. Gili Fivela, "Meanings, shades of meanings and prototypes of intonational categories," in *Prosody and meaning*, P. Prieto and G. Elordieta, Eds. De Gruyter Mouton, 2012, pp. 197–237.
- [17] U. D. Reichel and R. Winkelmann, "Removing micromelody from fundamental frequency contours," 2010.
- [18] U. D. Reichel, "The CoPaSul intonation model," *Elektronische Sprachverarbeitung*, pp. 341–348, 2011.
- [19] U. D. Reichel, A. Markó, and K. Mády, "Parameterization and automatic labeling of Hungarian intonation," *Proceedings of Speech Prosody*. Dublin, 2014.
- [20] U. D. Reichel and K. Mády, "Comparing parameterizations of pitch register and its discontinuities at prosodic boundaries for Hungarian," in *Proc. Interspeech 2014*, 2014, pp. 111–115.
- [21] E. Grabe, G. Kochanski, and J. Coleman, "Connecting intonation labels to mathematical descriptions of fundamental frequency," *Language and speech*, vol. 50, no. 3, pp. 281–310, 2007.
- [22] C. Gussenhoven, "Transcription of Dutch intonation," in *Prosodic typology and transcription: A unified approach*, S.-A. Jun, Ed. Oxford University Press, 2006.
- [23] A. Chen, A. Chen, R. Kager, and P. Wong, "Rises and falls in Dutch and Mandarin Chinese," *Proceedings of the Fourth International Symposium on Tonal Aspects of Languages*, 2014.
- [24] A. Chen, "Prototypes of tonal categories in tone- and non-tone languages," in preparation.
- [25] P. Boersma and D. Weenink, "Praat: Doing phonetics by computer" [computer program]. Version 5.4.15; University of Amsterdam, Amsterdam, 2015.
- [26] D. R. Ladd, "Intonational phonology". Cambridge, CUP, 2008.