Laboratory Phonology
Journal of the Association for Laboratory Phonology

Arnhold, A and Kyröläinen, A-J 2017 Modelling the Interplay of Multiple Cues in Prosodic Focus Marking. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 4, pp. 1–25, DOI: https://doi.org/10.5334/labphon.78

JOURNAL ARTICLE

Modelling the Interplay of Multiple Cues in Prosodic Focus Marking

Anja Arnhold[1,2] and Aki-Juhani Kyröläinen[3]

[1] University of Alberta, Department of Linguistics, Edmonton, CA

[2] University of Konstanz, Department of Linguistics, Konstanz, DE

[3] University of Turku, Department of French, Turku, FI

Corresponding author: Anja Arnhold (anja.arnhold@gmail.com)

Focus marking is an important function of prosody in many languages. While many phonological accounts concentrate on fundamental frequency (F_0), studies have established several additional cues to information structure. However, the relationship between these cues is rarely investigated. We simultaneously analyzed five prosodic cues to focus—F_0 range, word duration, intensity, voice quality, the location of the F_0 maximum, and the occurrence of pauses—in a set of 947 simple Subject Verb Object (SVO) sentences uttered by 17 native speakers of Finnish. Using random forest and generalized additive mixed modelling, we investigated the systematicity of prosodic focus marking, the importance of each cue as a predictor, and their functional shape. Results indicated a highly consistent differentiation between narrow focus and givenness, marked by at least F_0 range, word duration, intensity, and the location of the F_0 maximum, with F_0 range being the most important predictor. No cue had a linear relationship with focus condition. To account for the simultaneous significance of several predictors, we argue that these findings support treating multiple prosodic cues to focus in Finnish as correlates of prosodic phrasing. Thus, we suggest that prosodic phrasing, having multiple functions, is also marked with multiple cues to enhance communicative efficiency.

Keywords: Focus; givenness; prosody; Finnish; random forests; generalized additive mixed-effects

1. Introduction

The human voice conveys a variety of information, and prosodic variables such as F_0 and voice quality are involved in communicating linguistic as well as extra-linguistic information at the same time (see, Ladd, 1996, for an overview, especially pp. 36–38). For example, fundamental frequency (F_0) marks lexical tone in many languages, but is also one of the signals that conveys a speaker's emotional state. Moreover, prosody plays a substantial role in speaker recognition (Adami et al., 2003; Shriberg, 2007; Leemann et al., 2014) since it shows large inter-speaker variation (Dellwo et al., 2015), and is also crucial in managing turn-transitions in conversation (Beattie et al., 1982; Cutler & Pearson, 1986; Couper-Kuhlen & Ford, 2004). This paper deals with the signalling of distinctions in information structure, in particular focus and givenness, which is a central function of prosody in many languages (Gussenhoven, 2004; Jun, 2005; Jun, 2014). For example in English, when the sentence "Alex met with Taylor" answers the question "Who met with Taylor?", the word "Alex" is in narrow focus and will receive the most prominent accent of the sentence, realized with a higher F_0 range, longer duration and higher intensity, whereas the other words, which are contextually given, will be less prominent, with a smaller F_0 range, shorter duration and smaller intensity than in broad focus, e.g. when the same sentence

answers a question like "What's new?" (for more discussion and a formal definition of information structural concepts see Krifka, 2008). Although the best-researched prosodic cues to information structure are F_0, duration, and intensity (e.g., Eady & Cooper, 1986 on American English; Jun & Lee, 1998 on Korean; Xu, 1999; Wang & Xu, 2011 on Mandarin; Patil et al., 2008 on Hindi; Kügler & Genzel, 2012 on Akan; Peters et al., 2014 on different Germanic varieties; Genzel et al., 2015 on Hungarian), further effects of focusing appear for use and duration of pauses (Romøren & Chen, 2015) and, as has recently been shown, voice quality (Epstein, 2002; Ní Chasaide et al., 2011). However, the relative contribution of different cues is rarely analyzed. Here, we model multiple prosodic cues to focus simultaneously, test the significance of each cue while the others are taken into account, and investigate the relative importance of different prosodic cues to focus in Finnish.

Finnish is a Uralic language with lexical quantity oppositions for both vowels and consonants in most positions in the word (e.g. *muta* [muta] 'mud', *muuta* [mu:ta] 'another (partitive)', *mutta* [mut:a] 'but', *mutaa* [muta:] 'mud (partitive)', *muuttaa* [mu:t:a:] 'to change'). Primary stress always falls on the first syllable, but is less prominent than in Germanic languages, with duration being the only confirmed phonetic cue so far (Iivonen, 1998; Suomi et al., 2003; Ylitalo, 2009, p. 16; Arnhold, 2014a, pp. 130–139). A further prosodic difference is that intonation in Germanic languages is characterized by a set of contrasting pitch accents, which are used to express a range of pragmatic meanings (see e.g., Pierrehumbert, 1980; Grice et al., 2005; Gussenhoven, 2005, for accent inventories of English, German, and Dutch, respectively). One function of these accents is marking information structural distinctions. For example, in the utterance *Morgan drinks coffee and Jamie drinks tea*, *Jamie* and *tea* both contrast with words in the preceding utterance and will likely be prosodically prominent. However, *Jamie* normally carries a fall-rise accent, as it is the topic of the second clause (the clause conveys information about Jamie), whereas *tea* is realized with an F_0 fall, since it is the focus (out of a set of beverages, tea is identified as the one Jamie drinks; for more on this prosodic contour, see e.g. Jackendoff, 1972; Büring, 2003). By contrast, no inventory of contrasting accents has been suggested for Finnish. Instead, researchers generally agree that a uniform rise-fall appears on most content words except finite verbs, and is used in broad focus, narrow focus and for topics (Välimaa-Blum, 1993; Iivonen, 1998; Suomi et al., 2010, pp. 79–84).

Nevertheless, Finnish possesses various prosodic means of marking information structure. In broad focus, intonation shows a regular downward trend over the course of the utterance, with the peak of each rise-fall being lower than that of the preceding one. By contrast, words in narrow focus show a larger F_0 range, whereas the range is compressed for given words (Välimaa-Blum, 1993; Vainio & Järvikivi, 2007). Additionally, words in narrow focus have longer durations (Mixdorff et al., 2002; Suomi et al., 2003) and higher intensity peaks than words in broad focus, whereas given words have lower intensity peaks (Vainio & Järvikivi, 2007). Furthermore, Vainio et al.'s (2010) inverse filtering analysis of voice quality found a more breathy voice quality for words in narrow focus and a more tense voice quality for given words. Previous studies of the data set we analyze here have confirmed and supplemented further details to these findings, especially with respect to intensity and voice quality, and added information on a further prosodic correlate of information structure, the use of pauses. Thus, within one data set, significant effects of information structure appeared for each of the prosodic parameters, F_0 range, duration, intensity, pauses, and voice quality (see section 2 for details).

Here, we investigate the variability and interplay between these prosodic cues by using random forests modelling (Breiman, 2001) and generalized additive mixed models (GAMM, see Wood, 2006). These methods are particularly well-suited to analyzing the simultaneous effects of multiple cues as well as their relative weight. A further advantage is that

they do not assume that effects will be linear, which is often not the case for linguistic, and especially phonetic, data due to categorical perception of continuous variables (Liberman et al., 1957, and subsequent literature). Random forests are widely used across different scientific fields but in spite of their reliably high performance (see Fernández-Delgado et al., 2014), they are only rarely implemented in language-related studies. Examples of their use include modelling phonetic decision trees (Xue & Zhao, 2008), prosodic prominence, automatic speech recognition (Siohan et al., 2005; Xue & Zhao, 2006), and the variation between *was* and *were* in York English (Tagliamonte & Baayen, 2012). While previous analyses have investigated the effects of information structure on individual prosodic measures, here we use these statistical methods to flip the approach by testing the value of the prosodic measures as predictors of focus condition. This allows us to directly access their contribution as prosodic cues, enabling us to address four questions in particular. First, we will analyze how systematically focus conditions were marked prosodically across variation between speakers and items. Second, the analyses will indicate which acoustic measures constitute important prosodic cues to information structure when all phonetic measures are taken into account simultaneously. That is, the analysis will for example reveal whether including some of the cues makes other cues superfluous. Third, we will compare the relative importance of different cues. Fourth, whereas many statistical methods simply assume that the relationship between the dependent and independent measure(s) is linear, our random forests and GAMM modelling will allow us to glean detailed information about the association between them. Based on this, we will finally address theoretical modelling of these data, as well as the connection between multiple cues and multiple functions more generally.

After Section 2 describes the data and the analyzed variables, Section 3 will present the results of our random forests analysis, whereas Section 4 will deal with the GAMM analysis. Both sections will first introduce the statistical analysis methods in more detail before turning to their application to the present data. Section 5 provides a discussion and Section 6 a conclusion.

2. Data and analyzed variables

We provide a new analysis of the materials reported in Arnhold (2014a, 2014b, 2016), altogether 2841 words in 947 sentences. Materials were based on eight simple SVO sentences containing three disyllabic words. They included only open syllables, with half of the sentences containing mostly long vowels, the other half containing mostly short vowels. Seventeen participants produced all eight sentences in seven different focus conditions each: 1) Broad focus (e.g., "What did you see then?" – "Jani pushed a platform.") 2) Narrow information focus on the subject (e.g., "Who pushed the platform?" – "Jani pushed the platform."), 3) Narrow information focus on the verb (e.g., "What did Jani do with the platform?" – "Jani pushed the platform."), 4) Narrow information focus on the object (e.g., "What did Jani push?" – "Jani pushed a platform."), 5) Narrow corrective focus on the subject (e.g., "Did Otto push the platform?" – "Jani pushed the platform."), 6) Narrow corrective focus on the verb (e.g., "Did Jani polish the platform?" – "Jani pushed the platform."), 7) Narrow corrective focus on the object (e.g., "Did Jani push a bike?" – "Jani pushed a platform."; see Krifka, 2008, for a discussion of the information structural terms).[1] As the original analyses of this data set showed the same effects for corrective focus and information focus, we

[1] Note that standard Finnish does not have articles. Thus, the difference between the use of definite articles for given, i.e. previously mentioned, referents and the use of indefinite articles for new ones did not appear in the Finnish materials, making them identical across conditions (*Jani töni lavaa* 'Jani pushed a/the platform').

collapsed the two narrow focus types here.[2] Thus, each word appeared in one of three different information structural conditions (referred to in the following as 'focus conditions' for brevity): Broad focus, narrow focus (e.g., the subject for corrective focus on the subject), and given (e.g., the object and the verb in a sentence with narrow focus on the subject).

Here, we analyze measurements representing six potential prosodic cues to focus condition: 1) F_0 range, i.e., the distance between the F_0 maximum and the F_0 minimum of the word (in semitones, reference 50 Hz), 2) Word duration (in ms), 3) Intensity range, i.e., the distance between the intensity maximum of first syllable and the intensity minimum of the second syllable (in dB), 4) Duration of stretches with non-modal voice quality (e.g., creaky or whisper), as marked based on waveform, spectrogram and auditory impression (in ms), 5) The distance of the F_0 maximum from the beginning of the first syllable vowel (in ms), 6) Occurrence of pauses following the word. Since our analysis modelled all six prosodic cues at once, we had to remove words with a missing value for any of the cues. This led to a loss of 34% of the data, restricting the analysis to 1868 words.

The previous analyses examined each of the six measures separately and found significant effects of information structure on all of them (with the exception of 5, as explained below), which **Table 1** illustrates. F_0 maxima were higher in narrow focus and lower for given words than in broad focus, whereas following F_0 minima were lower in narrow focus and higher for given words than in broad focus. Thus, F_0 range was larger for words in narrow focus and smaller for given words compared to the broad focus condition. Here, we analyzed F_0 range directly to have only one measure of the F_0 component. Word durations were longer in narrow focus and shorter in given words than in broad focus. This effect was significant not only for word durations on a whole, but also for each of the individual segments. In Finnish, durations of segments within one word influence each other, with differences in phonological quantity leading to sub-phonemic adjustments of other segments (see Wiik & Lehiste, 1968; Lehtonen, 1970; Suomi, 2009, for details). Here, we only considered total word duration to simplify the analysis.

For intensity, the original analyses considered mean values for first and second syllables separately. Mean intensity of first syllables was higher in narrow than in broad focus,

	Narrow focus		**Broad focus**		**Given**	
Continuous variables	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
F_0 range (semitones)	6.21	2.59	3.53	2.14	2.6	1.68
Word duration (ms)	403.73	103.54	350.27	93.52	321.83	90.88
Intensity range (dB)	9.29	5.77	4.66	4.47	3.65	3.58
Duration of non-modal voice quality (ms)	51.61	81.5	25.94	63.69	18.93	52.9
Distance of F_0 maximum (ms)	59.73	53.27	80.94	107.6	87.61	119.27
Count variable	**Narrow focus**		**Broad focus**		**Given**	
Number of following pauses	59		2		16	

Table 1: Means and standard deviations (SD) for analyzed continuous variables and number of pause occurrences by information structural condition.

[2] An alternative model was fitted to the data where the response variable consisted of five levels, i.e., including the distinction between corrective and information focus. The results showed that the model was not able to discriminate between these two. Thus, a simplified model is reported here, where these levels are collapsed.

while mean intensity of second syllables was lower. For given words, mean intensity was significantly lowered on post-focal, but not on pre-focal words, resulting in an overall lower mean value. Like for F_0, we analyzed intensity range here to obtain one measure capturing the effects.

Non-modal voice quality appeared on longer stretches of words in narrow focus than in broad focus, but was mostly restricted to second syllables. For post-focal given words, non-modal voice quality frequently appeared on both syllables, but since word durations were shorter and pre-focal given words rarely showed non-modal realizations, the mean duration of stretches with non-modal voice quality was shortest for given words, as shown in **Table 1**.

The distance of the F_0 maximum from the beginning of the first syllable vowel was smaller in narrow focus than in broad focus and for given words, as illustrated in **Table 1**. Moreover, the location of the maximum was much more consistent in narrow focus, with the standard deviation being about half as large as the standard deviations of the other two conditions. Arnhold (2014a, 2014b) argued that F_0 maxima on words in narrow focus generally corresponded to realizations of tonal targets, whereas maxima on broad focus and especially given words frequently did not. Instead, they were often part of an F_0 movement to or from an F_0 peak on a neighbouring word.

Pauses were overall rare in the data, as is to be expected for three-word sentences. However, pauses were significantly more frequent following words in narrow focus. Thus, pauses after subjects appeared mostly in subject focus condition, whereas pauses after verbs mostly occurred in sentences with narrow focus on the verb. Information structure did not significantly affect the occurrence of pauses preceding a word. Thus, **Table 1** shows the absolute number of pauses following words in the three information structural conditions. The numbers are sums of pauses following subjects and verbs. Objects, being utterance-final, were always followed by silence, and could thus not be evaluated. Therefore, we analyzed pause occurrence as a categorical variable with three levels here: Pause, no pause, and irrelevant. Note that assigning NA values to objects instead would have meant loss of all data for objects, since the analyses removed all data rows with NA values to consider all prosodic cues at once. For this reason, it was also impossible to analyze pause duration. Subjects and verbs not followed by a pause would be assigned a pause duration of zero, but giving the same value to all objects would be conceptually inaccurate and misleading, while assigning them NA values would again result in a loss of data.

In addition to these six prosodic measures, we included two further categorical variables in our analyses: Position and first syllable vowel quantity. Position was in effect correlated with grammatical role, as all subjects occurred sentence-initially (position 1), all verbs sentence-medially (position 2), and all objects sentence-finally (position 3) in accordance with unmarked word order. The original analyses of this data set found several effects of position on prosodic realization, most notably a reduction of F_0 and intensity over the course of the utterance and final lengthening, as previously reported by Lehtonen (1974), Myers and Hansen (2007), and Nakai et al. (2009, 2012). Furthermore, non-modal voice quality was also significantly more frequent in final position, in line with previous reports of creaky, breathy, or voiceless realizations in final positions (Lehtonen, 1970, p. 45; Iivonen, 1998; Myers & Hansen, 2007; Nakai et al., 2009) and Ogden's (2001, 2004) finding that creaky voice has a turn-yielding function in spontaneous interaction.

As mentioned above, half of the sentences contained mostly long vowels and the other half only short ones, although vowel quantity could only be systematically manipulated for first syllable vowels. This was due to lexical and morphological restrictions, e.g., very few Finnish words end with a long vowel in nominative case, whereas the partitive marker *-a/ä* always contributes an additional mora, frequently causing word-final long vowels.

In the present data, words with long (quantity 2) first syllable vowels had significantly longer durations than words with short (quantity 1) first syllable vowels. As argued in the previous publications on this data set, other smaller effects of vowel quantity can be traced back to this difference.

3. Random forests modelling

3.1 Method

Analyzing data from a production study, we used two methods that flipped the analysis by treating focus condition as the dependent variable and using the prosodic measures as predictors. This section introduces the first method, random forests analysis. Random forests were originally developed by Breiman (2001) and are built from a large collection of classification and regression trees (CART, see Breiman et al., 1984). CART are similar to a regression-based analysis in that the model aims to predict a response variable given a set of predictors. In our analyses, the model tried to predict a word's focus condition (narrow focus, broad focus, or given) from the six prosodic cues listed in **Table 1** above. In contrast to regression, CART are a nonparametric method; they do not make any assumptions about the distribution of the underlying population. Instead, the distribution is estimated from the data. This makes them particularly suitable for data which may be inherently nonlinear, as is often the case in linguistics. The likelihood of a particular value of the response variable is estimated with a series of binary splits of the data, which are based on the values of the predictors. For example, in the case of the variable gender, the data could be split into two partitions consisting of data points associated with men and women, respectively, if one of these partitions simultaneously contained more occurrences of one of the focus conditions. In the standard CART, a binary split is made if the subsequent partitioning produces a purer division of the data with respect to the levels of the response variable. That is, the variation within the resulting two partitions has to be smaller than the variation within the complete data set. CART continue through the whole data set in this manner, recursively partitioning it into increasingly more homogenous sets (see Breiman et al., 1984, p. 104; Strobl et al., 2009, p. 8).

In random forests analysis, two additional layers of randomness are introduced during the model fitting. First, each tree in the forest is only grown on a subset of the available data that is randomly sampled with replacement. This subset of the data is commonly referred to as in-bag data, consisting of approximately 2/3 of the whole data. The remaining data (1/3) are referred to as out-of-bag (OOB). The accuracy of the model predictions is based only on the OOB data. Random forests by default divide the data set into a training and a testing set (in-bag and OOB, respectively) during the model fitting process. Measures of model performance are only based on the testing. This procedure can safeguard against overly optimistic model performance because a given tree has never seen the OOB data (Hastie et al., 2009, pp. 592–593). Breiman (2001) offers evidence that the OOB error is a good, unbiased estimation of model accuracy. Second, a further layer of randomness is introduced with respect to the predictors used during the splitting procedure. Only a selection of predictors is available during a particular split in a given tree, although the performance of random forests has shown not to be overly sensitive to it. This tuning parameter is commonly referred to as mtry, i.e., the number of predictors considered at each split. Because only a subset of predictors is available during a split, random forests are suitable in situations where there are more predictors than data points, the so-called small n large p problem. Additionally, this procedure also alleviates some issues related to collinearity, e.g., when predictors are correlated with each other. This is also a common situation in language studies and, for example, can create issues when carrying out a regression analysis.

Given that a large collection of trees is used in random forests, model predictions are based on voting. Each tree in the forest votes for the likelihood of the outcome. In our case, the votes are for the different levels of the response variable, focus marking. For a particular data point, random forests predict the level of the response variable which received the majority proportion of the votes. Thus, random forests are a flexible method that can handle a number of different scenarios in terms of analysis. They can be used for regression or classification tasks and, importantly, can be applied to data where the response variable is either binary, consisting of two categorical levels, or–as in the present case–multinomial/polytomous, i.e., consisting of multiple categorical levels.

3.2 Model fitting and results of random forests

In this study, we make use of the classical random forests algorithm implemented in the R package randomForest (Liaw & Wiener, 2002; R core team, 2015). Random forests were fitted to the data predicting the response variable focus condition (levels: narrow focus, broad focus, and given). The response variable was modelled as a function of F_0 range, word duration, intensity, the location of the F_0 maximum (relative to the beginning of the first syllable vowel), the duration of stretches with non-modal voice quality, the occurrence of a pause after the word, the position of the word in the sentence, first vowel quantity, speaker gender, trial number, subject, and lexical item. The default parameters were used first and the value of mtry, i.e., the number of predictors available for splitting, was tuned using the built-in estimation procedure. The results indicated that a value of 2 for mtry might be optimal. The final tuning parameter considered in this study concerns the number of trees included in the forest, controlled with the parameter ntree. By default, 500 trees are used to grow the forest. Several models were fitted by incrementally increasing the value of ntree in steps of 500 up to 2000 trees while recording the OOB error. The OOB error did not drastically change after 1000 trees. Based on this tuning procedure, a final model was fitted to the data where mtry was set to 2 and ntree to 1000. We will refer to this fitted model as the final random forests model.

The final random forests model has a high classification accuracy for the three different focus conditions, with 78.32% correct predictions. Thus, the results presented here clearly illustrate two points. First, the predictors used in the model are related to focus marking in production. Second, random forests are capable of learning, at least, certain aspects of focus marking based on this input.

As the response variable is polytomous, a measure of goodness-of-fit based solely on classification accuracy does not inform us about how the three different focus condition categories are learnt from the data. To inspect how well the model can distinguish the three conditions from each other, a confusion matrix is provided in **Table 2**. The rows in the table correspond to the observed labels in the data (i.e., experimentally induced focus conditions) and the columns to the model predictions. Correctly classified instances are located on the diagonal and misclassifications off-diagonal. Additionally, the error of a particular class is provided in the class-wise error column. The class-wise errors illustrate that both given (accuracy of 96%) and

	Broad focus	Narrow focus	Given	Class-wise error
Broad focus	10	63	235	0.968
Narrow focus	6	447	60	0.129
Given	1	40	1006	0.039

Table 2: Confusion matrix for the random forests model. Rows correspond to the observed categories and the columns to predicted ones along with the class-wise error. Correctly classified instances are located on the diagonal.

narrow focus (accuracy of 87%) are well separated from the other conditions by the model. In the case of narrow focus, the model accuracy is extremely high, considering that these are estimated accuracies for unseen data (OOB). However, the class-wise error associated with broad focus shows that this particular type is effectively not learnt by the model, with an error rate of 97%.

Although there may be several reasons for the poor performance of the model with respect to predicting broad focus, the distribution of the errors points to the simplest possible explanation, because the vast majority of broad focus data points are predicted to be instances of given ($n = 235$). Namely, the model goes with the majority class, i.e., given. If by contrast the distribution of broad focus was scattered across given and narrow focus more evenly, there might be also be room for a linguistically motivated reason.

Finally, it is possible that the random forests model reported above simply learnt the distributional properties present in the current data and not a more generalizable pattern. If this were the case, the model would display a poor performance on unseen data, i.e., overfitting to the current data. To evaluate the degree of possible overfit, the data were randomly split into training and test sets. The training set covered 80% of the data and the test set covered the remaining 20% of the data while the proportions of the response variable were approximately the same across the sets. A model was first fitted to the training data using the same parameters as reported above and then the predictions for the unseen test set were recorded. This procedure was carried out 1000 times (Efron & Tibshirani, 1993). The results showed an average classification accuracy of 0.78 (95% CI [0.76, 0.8]). Additionally, we also calculated accuracies for the individual levels associated with the response variable: broad focus (0.26, 95% CI [0, 0.07]), narrow focus (0.87, 95% CI [0.80, 0.93]) and, finally, given (0.96, 95% CI [0.92, 0.98]). These results were very similar to the ones reported above and indicated that the model is unlikely to overfit the data. Additionally, our models' classifications were better than what would be obtained with a naive statistical classifier which always chooses the most frequent outcome. In the present data set, given words occurred most often ($n = 1047$ out of 1868). The classification accuracy of a naive classifier always predicting givenness would only be 56%. In sum, the results presented here offer evidence that random forests are not only suitable for modelling experimental data containing repeated measures, but also offer good performance.

3.3 Relative variable importance

The way in which variable importance is estimated with random forests differs substantially from more traditional analyses in language studies, such as regression. In this study, we used permutation variable importance, which relies on differences in predictive accuracy of the model. The logic behind permutation variable importance is as follows: if a given predictor is associated with a particular level of the response variable, for example with narrow focus, randomly permuting the values of the predictor reduces the association between the response (narrow focus) and the predictor. In estimating this type of importance, the permuted predictor is used along with the other predictors to predict the response variable for the OOB data. Only an important variable is associated with a difference in the predictive accuracy of the model when it is permuted. Based on this, Breiman (2001) proposed a variable importance measure where the importance of a predictor is understood as the difference in prediction accuracy before and after permutation, averaged over all trees (see Strobl et al., 2008, for an alternative, but computationally more demanding, implementation). Lunetta et al. (2004) have shown that this type of variable importance outperforms standard univariate methods such as Fisher's exact test.

The estimated relative variable importance based on the final random forests model is represented with a dot plot in **Figure 1**, where predictors closer to or at zero are estimated to have a minimal contribution in predicting the response variable. The results suggest that the predictors can be sorted into three groups with respect to their importance in predicting focus condition. The first group consisted of only the F_0 range, which was the most important predictor of focus condition. Random permutation of the values for this predictor noticeably decreased the accuracy of the model's prediction of focus condition. The figure further suggests that word duration, intensity, and the location of the F_0 maximum were likewise important predictors of focus condition, although less important than F_0 range, forming a second group. The third group consisted of all other predictors, containing voice quality, the occurrence of pauses, and all non-acoustic predictors. These predictors were least important in the model. In particular, vowel quantity, speaker gender, trial, and lexical item were associated with a mean decrease in accuracy very close to zero and thus had little importance as predictors of focus condition.

This suggests that acoustic cues, especially F_0 range, were the best predictors of focus condition in the data, whereas other variables were less informative as predictors. The ranking within the fourth group also implies that of the two covariates, the position of the word in the sentence was more intertwined with the prosodic marking of information structure than first vowel quantity, since it was the better predictor. Similarly, the

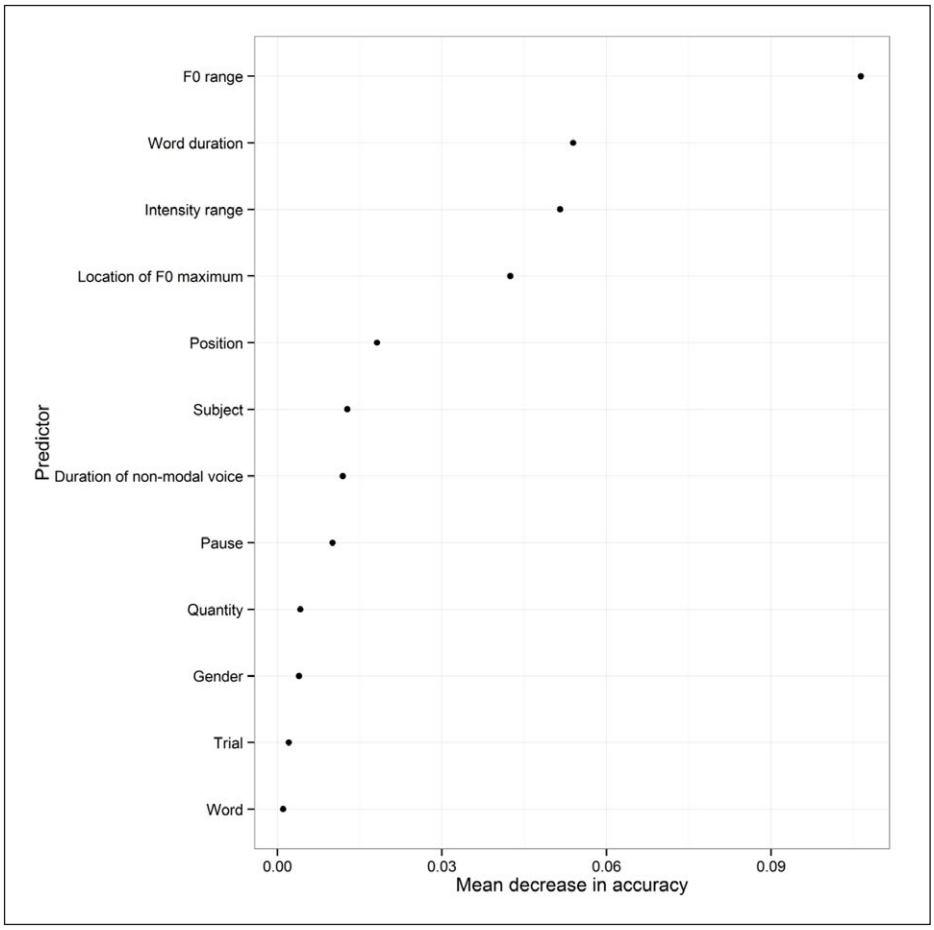

Figure 1: Estimated relative variable importance for the random forests with all predictors. The predictors are in a descending order based on their estimated relative importance. Higher values of mean decrease in accuracy are estimated to contribute more to the classification accuracy of the model.

analysis indicates that variation between individual speakers was larger than variation between genders, trials, or lexical items.

While this analysis directly ranked the relative importance of different acoustic cues in predicting the different focus conditions, it does not show their functional form. We will address this issue in the following section.

3.4 Functional form of the estimated effects with random forests

In order to investigate the functional relationship between a given predictor and focus marking, the four top-ranked predictors are visualized in **Figure 2** using partial dependency plots. Similar to a regression analysis, each of the predictors is visualized separately while all the other predictors are held statistically constant. The narrow focus is given on the y-axis, measured here as log-odds and mean centred. Positive values indicate a preference for narrow focus and negative values a dispreference (i.e., a preference for givenness), whereas values around the zero line indicate that a word with these acoustic features had an equal likelihood of being focused or given. Because broad focus was not learnt from the data, its partial effects are of no particular interest (see Section 3.2). Instead, the model learnt a binary distinction between words in narrow focus and given ones. In this respect, the partial dependency plots in **Figure 2** resemble logistic regression (compare to **Figure 3** in Section 4.3).

In **Figure 2**, the upper left panel illustrates the partial effects of F_0 range. For words with an F_0 range close to zero, the narrow focus condition received very few votes, but the proportion of narrow focus words increased almost linearly with an increase in F_0 range. At around five semitones, the line in the figure crosses zero and then flattens out,

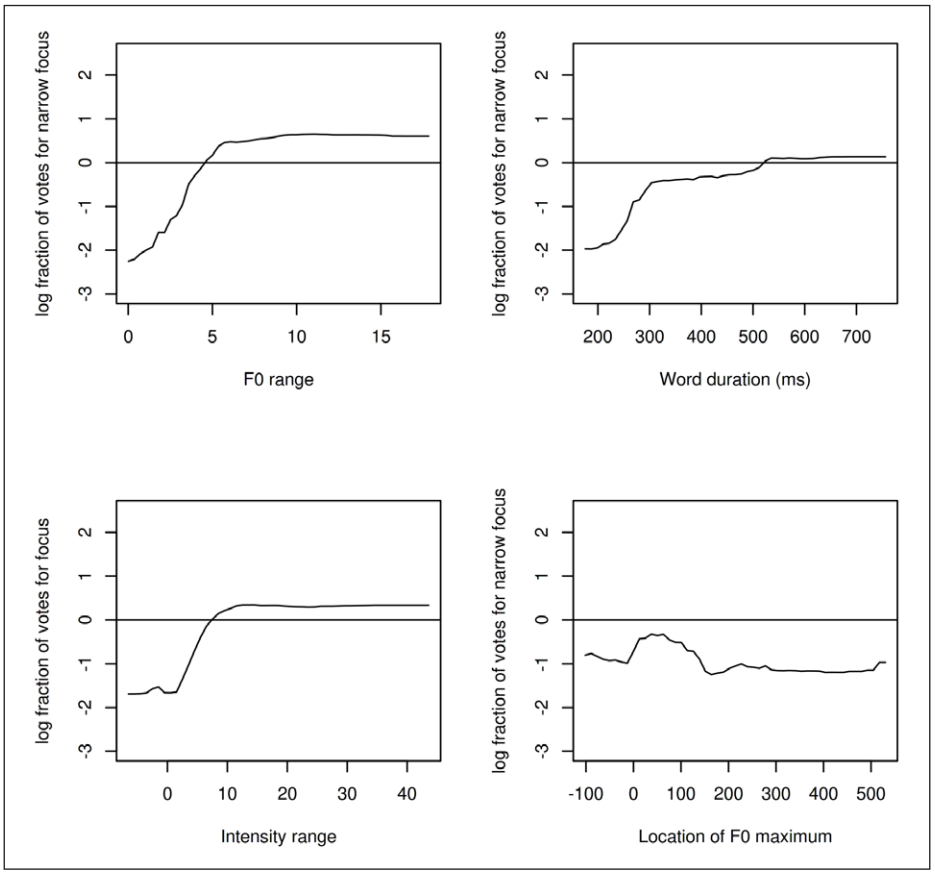

Figure 2: Partial dependency effects for narrow focus as estimated with random forests. The y-axis is given on logit scale and it is centred to have a mean of zero over the data distribution.

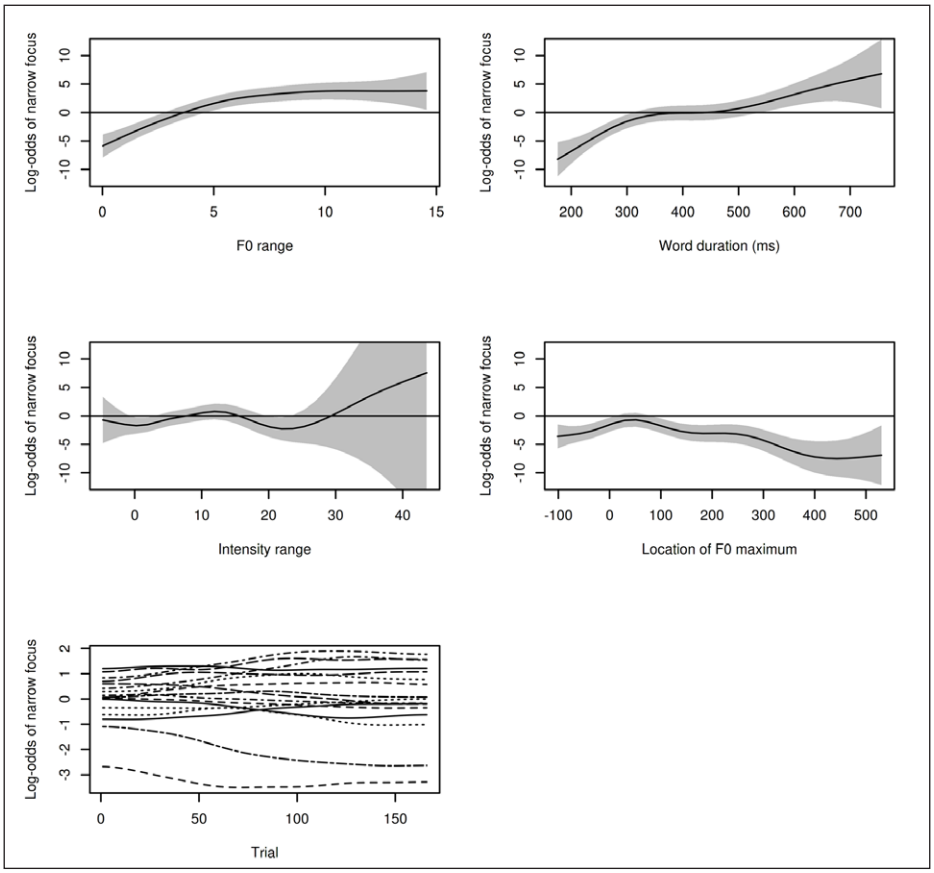

Figure 3: Partial effects for narrow focus as estimated with GAMM. The y-axis is given on log-odds scale and it is mean centred.

demonstrating that a further increase in F_0 range did not increase log-odds for the narrow focus condition any further. This means that small F_0 ranges were disassociated with narrow focus and instead associated with given words, with smaller ranges being more likely to appear in given condition. By contrast, above a certain threshold, higher F_0 ranges were not more strongly associated with narrow focus.

A similar picture emerged for word duration (upper right panel) and intensity range (lower left panel). For duration, short values around 200 ms were strongly associated with givenness and increased duration correlated with increased votes for the narrow focus condition, reflecting the finding that durations of given words were shorter and those of focused words were longer in the data. Interestingly, the line shows two plateaus, one starting around at 300 ms and a slightly higher one starting around 540 ms. This could have been due to the additional influence of first vowel quantity on duration. Note that the mean duration of narrow focus words with short first vowels at 372 ms was only slightly above that of given words with long first vowels at 359 ms, while narrow focus words with long first vowels had an average duration of 448 ms. Partial effects of intensity were more similar to those of F_0 ranges. Small intensity ranges led to an increase in log-odds for the given condition, whereas larger values were more predictive of narrow focus.[3] The line visible in the figure flattens out around 12dB, suggesting that words with an intensity range above this value were generally focused, but that a higher range was not more strongly associated with focus above this threshold.

[3] Negative values for intensity range visible in the figure are due to the fact that for these words, maximum intensity of the first syllable vowel was lower than minimum intensity for the second syllable vowel. Thus, these words showed a slight rise in intensity instead of the usual fall.

A different picture emerges for the partial effects of the location of the F_0 maximum measured relative to the beginning of the first syllable vowel. Log-odds for narrow focus were relatively stable across different values of this measure, with the exception of a bump between 0 and 120 ms. Only for values in this range, the log-odds for narrow focus came close to zero. This indicates that words with an F_0 maximum closely following the beginning of the first syllable vowel were almost equally likely to be narrowly focused or given. For words with either earlier or later peaks, the log-odds for narrow focus remained further below zero, signalling that these words were more likely to be given. This fits the observation that the location of the F_0 maximum was relatively stable in the narrow focus condition, but varied strongly for given words in the data set. The implication is that, with respect to this cue, givenness was not marked directly but was rather the absence of marking narrow focus. This contrasts with the findings for F_0 range, intensity, and duration, where small values were associated with given condition and large values with narrow focus, making it difficult to say which marking is primary.

3.5 Summary

Random forests modelling of the data indicated that the two information structural conditions narrow focus and given could be predicted with a high accuracy from the acoustic cues, while taking the co-variates position and quantity, as well as lexical item and subject-related factors into account. Thus, narrow focus and givenness were consistently marked prosodically. By contrast, the model was unable to learn to predict the broad focus condition.

The best predictor of information structure was F_0 range, with small ranges predicting givenness and large ranges predicting narrow focus. Word duration and intensity range were likewise important predictors with higher values being associated with narrow focus and lower ones with givenness. Interestingly, the relation between prosodic cues and information structural condition was only partially linear, so that a rise above a certain threshold did not further increase the predictive power. The location of the F_0 maximum, which was likewise an important predictor, showed a different pattern: Only a small range of values led to the prediction of narrow focus, meaning maxima were consistently located early after the beginning of the first syllable vowel for narrow focus words, whereas earlier as well as later values were strongly associated with givenness. Of the acoustic cues, voice quality and the occurrence of a pause following the word were the least important predictors of information structure.

4. Generalized additive mixed-effects modelling

4.1 Method

In recent years, generalized additive modelling (GAM) has gained popularity and has been applied to a variety of different linguistic data such as modelling reaction times (Baayen, 2010), event-related potentials (Tremblay & Newman, 2015), accentedness ratings (Porretta et al., 2015) and dialectal variation (Wieling et al., 2014), among others. In addition to this, GAM has been extensively used in ecology (see Zuur et al., 2009). GAM does not assume a linear functional form and similar to a mixed-effects model it is possible to include random effects such as random intercepts and slopes (Baayen et al., 2008), extending the GAM into a generalized additive mixed model (GAMM). In addition to these random effects, factor smooths can be included in GAMM. A factor smooth is an interaction between a numeric predictor and a factor, allowing to fit to the data wiggly lines for each level of the factor (see Baayen et al.,

2016). In this manner, it is for example possible to account for differences between subjects across trials.

4.2 Model fitting and results of GAMMs

Given the results presented in Section 3.4, we fitted a GAMM to the data in order to offer accumulative evidence for the highly nonlinear functional form estimated for the four main predictors in the data. However, as shown in Section 3.2, the data presented in this study appears to be primarily driven by the difference between narrow focus and givenness. Therefore, we simplified the response variable for the GAMM and only consider the difference between given and narrow focus. Furthermore, we take into account only the four predictors that were estimated to be the most important ones by random forests in order to offer cumulative evidence for their functional form (see Section 3.3).

We fitted the GAMM to model the probability of getting narrow focus as a function of F_0 range, word duration, intensity range, and location of F_0 maximum. Additionally, we included in the model a factor smooth for trial and subjects, and random intercepts for words. The estimated parameters of the fitted model are given in **Table 3**.

For the smooth functions, the column labelled as "Edf" in **Table 3** indicates the estimated degrees of freedom. When they are equal to 1, the effect is estimated to be approximately linear. The estimated effects clearly illustrate that the functional forms of these predictors are estimated to be highly nonlinear. Thus, we will visually inspect them in Section 4.3 and compare them to the results obtained with random forests.

4.3 Functional form of the estimated effects with GAMM

Figure 3 illustrates the relationship between the acoustic cues and focus condition as estimated by GAMM. As for the partial effects of random forests modelling shown in **Figure 2**, positive values on the y-axis indicate an association with narrow focus, while negative values signal an association with givenness. The grey shades around the smooth represent confidence intervals. Overall, the effects estimated with GAMM were similar to those estimated by random forests modelling. For F_0 range, low values were again predictive of givenness, while high values were associated with narrow focus. Compared with **Figure 2**, the line in the top left panel of **Figure 3** crosses zero earlier, but again flattens out around 5 semitones, indicating F_0 ranges above this value were equally predictive of narrow focus.

The estimated partial effect of word duration in GAMM differed from random forests. Specifically, an increase in word duration above around 500 ms was more strongly

A. Parametric coefficient	**Estimate**	**Std. Error**	***z* Value**	***p* Value**
Intercept	−2.34	5.86	−4.21	< .001
B. Smooth terms	**Edf**	**Ref. df**	**Chi. sq**	***p* Value**
F_0 range	3.86	4.79	196.86	< .001
Word duration (ms)	4.81	5.86	100.02	< .001
Intensity range	5.71	6.82	32.11	< .001
Location of F_0 maximum	6.27	7.39	66.14	< .001
Factor smooth: Trial and Subject	25.83	152	131.91	< .001
Random effect: Word	6.4	7	51.83	< .001

Table 3: Estimated effects for the GAMM reporting a parametric coefficient (Part A), along with estimated degrees of freedom (Edf), reference degrees of freedom (Ref. df), chi squared values and *p* values for smooths and random effects (Part B).

positively associated with narrow focus. Given that the GAMM and random forests models are not exactly the same, future studies are required in order to tease apart the role of duration on focus marking. Furthermore, the line in the top right panel of **Figure 3** does not flatten out, but continues to rise above this value, indicating that the model estimated words with longer durations as more likely to be in narrow focus. However, note the increased confidence interval for very large durations.

For intensity, GAMM results agreed with random forests modelling in predicting that words with a range below 7dB were given while predicting words with larger ranges to be focused. In contrast to the flattening of the line visible in **Figure 2**, the line in the middle left panel of **Figure 3** displays a dip and then a further rise for higher intensity ranges. However, the wide confidence intervals for very high intensity ranges show that these predictions were based on very few observations.

For the location of the F_0 maximum, partial effects of GAMM were very similar to random forests results. Again, only words with a maximum located between 0 and 120 ms after the beginning of the first syllable vowel had a close to equal probability to be narrow focus or given, whereas all earlier and later locations were associated with givenness.

In addition to the four significant acoustic predictors, **Figure 3** also illustrates partial effects of trial for all subjects. Two lines are clearly lower than the others, signifying that the model predicted all words uttered by these two participants to be given. This suggests that these subjects did not mark focus condition clearly and, as the further lowering of the lines indicates, used even less acoustic marking over the course of the experiment. For all other subjects, lines were closer together and flatter, as is to be expected from the random distribution of focus conditions across trials. Expectably, some participants showed slight effects of fatigue (falling lines), while others showed learning effects of clearer focus marking later in the experiment (rising lines). The GAMM took this by-participant variation into account when estimating the effects of the acoustic predictors.

4.4. Summary

GAMM analysis confirmed F_0 range, word duration, intensity range, and the location of the F_0 maximum as important predictors of focus condition. It suggested that they were highly significant acoustic cues in the present data set. Importantly, the model arrived at a similar estimation of the functional forms as the random forests modelling, only indicating slightly different partial effects estimations for higher duration and intensity values.

5. Discussion

We used two complementary methods of statistical modelling, random forests and GAMM, to investigate the marking of focus condition by multiple prosodic cues. These analyses allowed us to address four questions: 1) How systematically were focus conditions marked prosodically across variation? 2) Which acoustic measures constituted important prosodic cues to information structure when all phonetic measures were taken into account at the same time? 3) What was the relationship between the multiple prosodic cues to focus condition, in particular: What was their relative importance? 4) What was the relationship between focus conditions and acoustic measures, i.e., what was the functional shape of the predictors? We will discuss these questions and the implications of our respective findings in turn, before considering the generalizability of our methods and results to other types of data and languages. Finally, we will address the question of theoretical modelling of the data and the question of multiple cues more generally.

Regarding the first question, our analyses flipped the analysis of the production data and investigated whether focus conditions could be predicted from the acoustic measurements. The methods of data modelling effectively assessed systematicity across variation between

speakers and items. GAMM modelling found several phonetic measures to be significant predictors of focus condition, indicating a systematic relationship between focus condition and prosodic cues. At the same time, it expectably suggested significant differences between items and between speakers, particularly regarding the way participants' behaviour changed over the course of the experimental session. Random forests modelling measured systematicity as learnability. The random forests model was able to learn to predict the focus condition of words in narrow focus and of given words with great accuracy. This suggests that across variation between individual speakers and lexical items, prosodic marking of focus and givenness was remarkably systematic. By contrast, the model was not able to recognize the broad focus condition based on the acoustic cues. Thus, narrow focus and givenness were consistently cued by prosodic characteristics, whereas the broad focus condition was not. This fits the idea that broad focus is the default, unmarked information structure, whereas prosody is adjusted to mark both narrow focus and givenness in different ways.

As for the second question, the models took six prosodic measures into account at the same time, finding four of them to be important predictors of focus condition: F_0 range, duration, intensity, and the location of the F_0 maximum. Thereby, models indicated that words in narrow focus were generally characterized by a larger F_0 range, longer duration, larger intensity range, and mostly had F_0 maxima localized shortly after the beginning of the first vowel. Givenness, by contrast, was generally cued through a smaller F_0 range, shorter word duration, smaller intensity range, and wider distribution of locations of F_0 maxima. These results confirmed previous analyses of the same data set separately modelling the effect of information structural condition on the acoustic measures. The present analyses were a closer approximation of language processing, where the listener encounters the raw acoustic signal as a whole and has to filter out the relevant cues. They further complement the separate analyses of the phonetic measures by establishing them as prosodic cues in their own right. Thus, while the previous analyses showed that several prosodic measures were influenced by focus condition simultaneously, this finding did not unambiguously establish that all of them were important prosodic cues. At least two related alternative interpretations could be considered. First, some of the prosodic measures could have been influenced not (only) by information structure directly, but also by other prosodic measures. For example, downtrends in F_0 and intensity often co-occur and some researchers have argued that the two measures are inextricably linked and/or determined by the same speech production mechanisms (e.g., Lieberman, 1967; Trouvain et al., 1998), which would render analyzing both of them unnecessary, whereas others have found them to be independent to some extent (Hird & Kirsner, 2002; see Strik & Boves, 1995, for further discussion). Second, some prosodic measures, even if independently affected by information structure, could have been superfluous as predictors. That is, even if focus condition significantly affected several prosodic measures, one or two of them could have been enough to predict focus condition when flipping the analysis. Our modelling showed that at least four prosodic cues were highly important in predicting focus condition when taken into account at the same time. Thus, even with F_0 already taken into account, additionally considering intensity and other prosodic cues significantly improved the prediction of focus condition. These results suggest that although prosodic focus marking in Finnish is redundant in one sense—it employs several prosodic cues in parallel—these cues are not superfluous, as leaving them out would decrease predictive accuracy.

Interestingly, two of the prosodic dimensions on which the previous analyses found significant effects of information structure were not confirmed as important predictors of focus condition by our analysis: the use of pauses and non-modal voice quality. Pauses were rare in the analyzed data set, which consisted of simple short sentences. This may

have made them an unreliable predictor: Although most occurring pauses followed words in narrow focus, most words in narrow focus were not followed by a pause, meaning that the presence of a pause was only useful for identifying a small subset of narrow focus words. We believe that it would be worthwhile to investigate pause occurrence as a continuous, instead of a categorical, variable in future studies. Regarding voice quality, it is important to recall that we excluded all words with missing values for any of the analyzed measures, in order to be able to model all prosodic cues at once. Since non-modal voice quality makes reliably measuring F_0 impossible, this meant that all words realized completely with non-modal voice quality were removed from the data set, since they were missing the value for F_0 range. The previous analyses indicated a significant increase of non-modal realizations for second syllables of words in narrow focus and for both syllables of following given words. In line with this, we removed 19% of given words due to missing F_0 range values, but only 2% of words in broad focus and none of the words in narrow focus. Altogether, we removed 36%, 24% and 37% of given, broad focus, and narrow focus words, respectively. This obscured the connection between non-modal voice quality and (post-focal) givenness, hampering the success of voice quality as a predictor of focus condition in our models. Thus, while it is possible and indeed likely that non-modal voice quality is a prosodic cue to information structure, our analyses were unable to confirm this without sacrificing the goal of modelling all potential cues at once.

Turning to the third question, our analyses suggested that not all prosodic cues to focus were equally important. Random forests modelling indicated that F_0 range was the most important cue to focus condition, followed at some distance by word duration, intensity range, and the location of the F_0 maximum. The differences between the latter three predictors were relatively small. Interestingly, this ranking differs from the results of Niemi's (1984) studies on acoustic markers of prominence ('stress' in his terms). He studied the distinction between the noun phrase *musta rastas* 'black thrush', where the first syllables of both words receive prominence, and the compound *mustarastas* 'blackbird', where only the first syllable is marked as prominent, but not the third one. In native Finnish speakers' productions, duration was the most important cue to this distinction, followed by F_0, which was in turn more important than overall intensity (duration $\geq F_0 >$ overall intensity). By contrast, in the productions of the same items by native speakers of American English, F_0 was more important than both duration and intensity ($F_0 >$ duration, intensity). Niemi confirmed this difference in the relative importance of duration and F_0 with further studies, for example asking participants to circle the most prominent syllable in nonsense words with manipulated prosody, concluding that English is more melodious whereas Finnish is more dynamic. Our findings are clearly at odds with Niemi's, although given the different methodologies, it is difficult to pinpoint the reason. In the present data, word duration was strongly affected by two other factors, vowel quantity and sentence position, which could have hampered its potential as a predictor of focus condition.[4] Another possibility is that the difference is due to the fact that Niemi investigated other kinds of prosodic prominence, but did not manipulate information structure. Thus, focus marking in particular could be characterized by an outstanding importance of F_0 among several important cues.

[4] F_0 also showed effects of these factors, but whereas final lengthening reduced the size of the durational effect of focus in final position, in terms of F_0, the effect of focus was even larger later in the sentence where F_0 range was reduced. Regarding quantity, the previous accounts argued that its effects on F_0 were due to its effects on duration, i.e. that longer durations provided more space to reach tonal targets whereas shorter durations lead to compression or undershoot.

Regarding the fourth question, random forests and GAMM modelling were able to provide more details regarding the relationship between focus condition and acoustic measures. Analyses of partial effects for the four significant prosodic cues suggested non-linear functional forms for all of them, confirming the advantages of using random forests and GAMM analysis. Functional forms were very similar for F_0, duration, and intensity. For these three prosodic cues, random forests and GAMM analysis suggested a positive linear correlation up to a certain threshold, i.e., the higher the value of the acoustic measure, the more likely the word was to be in narrow focus instead of being given. Above the threshold value, however, the relationship was not linear anymore. Partial effects of the random forests models indicated that above the threshold, all values of the three acoustic measures were equally predictive of narrow focus. Partial effects of the GAMM showed the same picture for F_0 range, whereas for duration and intensity, more variation appeared, especially for higher values associated with larger confidence intervals. Overall, it is noteworthy that threshold values emerged, which were mostly constant across both methods of analysis. These results fit the assumption that focus is a linguistic category, which is also marked categorically. Thus, a word can be in narrow focus or not, but it cannot be "more focused." This is in line with the way that focus has been discussed in the literature, whether as a feature that can be projected (e.g., von Stechow & Uhmann, 1986) or categorically marking the relevance of alternatives (Rooth, 1992). For givenness, however, the data are compatible with a gradient notion as for example advocated by Prince (1981) and Gundel et al. (1993; contra e.g., Schwarzschild, 1999). In this vein, one could interpret the linear rise of the lines for low values in the partial effects plots as indicating that for example a word with a small F_0 range was predicted to be given and a word with an even smaller range was even more given (with different degrees of givenness perhaps corresponding to given, inferable, uniquely identifiable etc., as discussed in the literature). Such an interpretation would of course need to be confirmed with further studies by systematically manipulating degrees of givenness.

For the fourth important predictor of focus condition, the location of the F_0 maximum, both methods agreed in revealing a very different functional form: No value for the distance between the F_0 maximum and the beginning of the first syllable vowel was strongly predictive of narrow focus, but values between 0 and 120 ms were about equally associated with narrow focus and givenness. This reflects the previously reported finding that the majority of F_0 maxima appeared on first syllable vowels for words in narrow focus, whereas their location was much more variable for broad focus and especially given words (see **Table 1**). In other words: F_0 maxima on given words could appear anywhere within the word, whereas in focus, maxima were reliably realized on the first syllable vowel.

Next, let us discuss how our methods and findings could be expected to generalize to other types of data and languages. Random forests analysis and GAMMs are in principle applicable to all kinds of data and are used in a wide range of fields, as mentioned above. In fact, they are particularly well-suited to dealing with data and effect types that are a challenge for other methods of statistical analysis, for example non-linear effects and unbalanced data sets, which appear often in linguistics. Importantly, random forests are, at least partly, able to handle collinear predictors. These types of predictors are known to be problematic for regression models, for example (Harrell, 2001). As large-scale corpora are becoming increasingly more readily available, this makes random forests suitable for modelling big data. On the other hand, GAMMs are capable of modelling complex random structures as illustrated in this study. This is especially important not only for experimental studies but also for corpus-based studies on spontaneous conversational data. Thus,

it would also be possible to use the methods presented in this study to analyze prosodic focus marking in other types of data. The only prerequisite for this is that focus conditions can be objectively and unambiguously identified in the data. For random forests, the model's classifications need to be compared to the true classifications to obtain a measure of model accuracy. For GAMM modelling, it is likewise necessary to know the focus condition of the word that a particular phonetic measurement came from to model the relationship between the prosodic measure and the focus condition. In the present data set, the information structure of the produced utterances and the focus condition of all the words in them were controlled through context questions. For less controlled data, especially spontaneously produced conversations, it is considerably more difficult to objectively determine information structure. For example, Calhoun et al. (2005) provide guidelines for manual annotation of information structure for a subset of the Switchboard corpus of spontaneous speech, but argue that these annotations have to be at least partially based on prosodic realizations. This would of course constitute a problem for an analysis like ours, since prosodic cues and focus condition would not be independent factors (but see, e.g., Baumann et al., 2004; Poesio, 2004; Dipper et al., 2007, for annotation guidelines without recurrence to prosodic realization). How our present findings generalize to other data sets is of course an empirical question. We have above discussed various reasons why our analyses may have systematically underestimated the importance of some prosodic cues to focus condition, e.g., voice quality. As mentioned above, perception experiments would be necessary to determine with more certainty how listeners use these cues when determining the information structure of an utterance. Importantly, the relative ranking of the predictors presented in this study offers fully testable predictions that can be used in future studies based on different types of data. While it is likely that the relative importance of cues differs between languages with different prosodic systems (recall, e.g., Niemi's, comparisons between Finnish and English cited above, 1984), we think that the most important and most generalizable outcome of our analyses is the fact that multiple prosodic cues are employed in parallel.

This leads to the question of how to theoretically model parallel prosodic cues. Analyzing the effects of information structure on each prosodic measure in the present data set separately, Arnhold (2014b) suggests an account that links effects on all the measures to prosodic phrasing. Here, we confirmed that several prosodic cues function as important predictors of focus condition in Finnish at the same time. The fact that these cues act in parallel suits the hypothesis that they have a common source. In fact, assuming that focus indeed affects phrasing, which is marked by several prosodic cues, it would be more surprising to find only one of these prosodic features to predict focus condition. Further, accounting for prosodic focus marking in terms of prosodic phrasing specifically offers a functional explanation of the existence of multiple parallel cues. There is evidence that the presence of multiple cues improves signal detection in animal communication (Rowe, 1999) and is crucial in language learnability (Christiansen et al., 1998), whereas the strategy of using only a subset of available cues has for example been associated with autism spectrum disorders (Rieth et al., 2015). A maximally robust marking seems particularly important for prosody given its many functions. Prosodic phrasing is not only influenced by information structure, but it is also known to be shaped by syntactic structure and semantics (Nespor & Vogel, 1986; Selkirk, 2000). A failure to convey these aspects of meaning can seriously impact communicative efficiency, as attested by misunderstandings specifically arising in written conversations, as well as by the emergence of written language equivalents of prosodic structuring like the use of punctuation to signal phrase boundaries. While prosody sometimes, but not always disambiguates syntactic ambiguities (Shattuck-Hufnagel & Turk, 1996; Kjelgaard & Speer, 1999; Snedeker & Trueswell,

2003), it is of fundamental importance in first establishing constituents during language acquisition (Hawthorne & Gerken, 2014). Prosody is also crucial in structuring conversation, e.g., by signalling the speaker's wish to hold the floor or yield their turn (Geluykens & Swerts, 1994; Caspers, 1998; Koiso et al., 1998). In addition to its linguistic functions, prosody signals a wide range of other information, such as the emotional state of the speaker, the understanding of which is likewise essential to human interaction. Thus, the use of multiple redundant cues, far from being inefficient, is in fact to be expected given the variety of functions that prosody performs.

6. Conclusion

In this article, we reanalyzed data from a production experiment on prosodic focus marking in Finnish. Arnhold (2014a, 2016) found that information structural manipulations affected several prosodic measures. Here, we directly investigated their role as prosodic cues to focus condition. Using two advanced statistical methods, random forests and GAMM modelling, allowed us to analyze all prosodic cues simultaneously, and to model non-linear effects.

Results indicated that all relationships between focus condition and prosodic cues were non-linear, confirming the advantages of applying these statistical methods. The analyses further showed that at least F_0 range, word duration, intensity, and the location of the F_0 maximum cued the distinction between narrow focus and givenness with high consistency. The random forests analysis further suggested that among the prosodic cues, F_0 range was the most important one. By contrast, our modelling did not confirm the role of post-focal pauses and non-modal voice quality as important cues to focus, although both prosodic variables were significantly affected by information structure. We hypothesized that this may have been connected to a limitation of our methods, since considering all potential cues required eliminating all tokens with missing values for any of the measures. Further studies, especially perception experiments, are required to clarify this issue.

Importantly, including all potential prosodic cues into the same analysis indicated that multiple prosodic cues acted as significant predictors of focus condition. As all of these cues improved prediction accuracy, none of them was a redundant addition to the most important predictor, F_0 range. For the present data set, we suggest that an account focusing on prosodic phrasing is well-suited to account for multiple prosodic cues to focus in Finnish. More generally, this research further confirms the importance of investigating the co-occurrence of multiple cues in investigations of prosody, as well as the importance of accounting for multiple cues theoretically. We argue that far from making communication inefficient, the use of multiple cues increases reliability and is therefore to be expected, especially given the multiple functions prosody performs.

Competing Interests

The authors have no competing interests to declare.

References

Adami, A. G., Mihaescu, R., Reynolds, D., and Godfrey, J.J. 2003. Modeling prosodic dynamics for speaker recognition. International Conference on Acoustics, Speech, and Signal Processing Proceedings (ICASSP), (Vol. *4*) (pp. IV-788). DOI: https://doi.org/10.1109/ICASSP.2003.1202761

Arnhold, A. 2014a. Finnish prosody: Studies in intonation and phrasing (Doctoral dissertation). Goethe-University Frankfurt am Main. Available from http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/34798.

Arnhold, A. 2014b. Finnish as a phrase language. Unpublished manuscript, University of Alberta.

Arnhold, A. 2016. Complex prosodic focus marking in Finnish: Expanding the data landscape. *Journal of Phonetics*, *56*, 85–109. DOI: https://doi.org/10.1016/j.wocn.2016.02.002

Baayen, R. H. 2010. Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon, 5*(3), 436–561. DOI: https://doi.org/10.1075/ml.5.3.10baa

Baayen, R. H., Davidson, D. J., and Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., van Rij, J., de Cat, C., and Wood, S. N. 2016. Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. arXiv preprint retrieved from https://arxiv.org/abs/1601.02043. Accessed January 10, 2016.

Baumann, S., Brinckmann, C., Hansen-Schirra, S., Kruijff, G.-J., Kruijff-Korbayová, I., Neumann, S., Steiner, E., Teich, E., and Uszkoreit, H. 2004. The MULI project: Annotation and analysis of information structure in German and English. In: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC) (pp. 1489–1492).

Beattie, G. W., Cutler, A., and Pearson, M. 1982. Why is Mrs Thatcher interrupted so often? *Nature*, *300*(5894), 744–747. DOI: https://doi.org/10.1038/300744a0

Breiman, L. 2001. Random forests. *Machine Learning, 45*(1), 5–32. DOI: https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. 1984. *Classification and regression trees*. New York: Chapman and Hall.

Büring, Daniel. 2003. On D-trees, beans, and B-accents. *Linguistics and Philosophy, 26*(5), 511–545. DOI: https://doi.org/10.2307/25001898

Calhoun, S., Nissim, M., Steedman, M., and Brenier, J. 2005. A framework for annotating information structure in discourse. In: Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky (pp. 45–52). Stroudsburg, PA, USA: Association for Computational Linguistics. DOI: https://doi.org/10.3115/1608829.1608836

Caspers, J. 1998. Who's next? The melodic marking of question versus continuation in Dutch. *Language and Speech*, *41*(3–4), 375–398. DOI: https://doi.org/10.1177/002383099804100407

Christiansen, M. H., Allen, J., and Seidenberg, M. S. 1998. Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*(2–3), 221–268. DOI: https://doi.org/10.1080/016909698386528

Couper-Kuhlen, E. and Ford, C.E. (Eds.). 2004. *Sound patterns in interaction: Cross-linguistic studies from conversation*. Amsterdam/Philadelphia: John Benjamins Publishing. DOI: https://doi.org/10.1075/tsl.62

Cutler, A. and Pearson, M. 1986. On the analysis of prosodic turn-taking cues. In: Johns-Lewis, C. (Ed.), *Intonation in discourse* (pp. 139–155). London: Croom Helm.

Dellwo, V., Leemann, A., and Kolly, M-J. 2015. Rhythmic variability between speakers: Articulatory, prosodic, and linguistic factors. *The Journal of the Acoustical Society of America, 137*(3), 1513–1528. DOI: https://doi.org/10.1121/1.4906837

Dipper, S., Götze, M., and Skopeteas, S. 2007. *Information structure in cross-linguistic corpora*. Potsdam: Universitätsverlag Potsdam.

Eady, S. J. and Cooper, W. E. 1986. Speech intonation and focus location in matched statements and questions. *The Journal of the Acoustical Society of America*, *80*(2), 402–415. DOI: https://doi.org/10.1121/1.394091

Efron, B. and Tibshirani, R. J. 1993. *An introduction to the bootstrap*. New York: Chapman & Hal. DOI: https://doi.org/10.1007/978-1-4899-4541-9

Epstein, M. A. 2002. Voice quality and prosody in English (Doctoral dissertation). University of California, Los Angeles.

Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. 2014. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research, 15*(1), 3133–3181.

Geluykens, R. and Swerts, M. 1994. Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication*, *15*(1–2), 69–77. DOI: https://doi.org/10.1016/0167-6393(94)90042-6

Genzel, S., Ishihara, S., and Surányi, B. 2015. The prosodic expression of focus, contrast and givenness: A production study of Hungarian. *Lingua*, *165*, 183–204. DOI: https://doi.org/10.1016/j.lingua.2014.07.010

Grice, M., Baumann, S., and Benzmüller, S. 2005. German intonation in autosegmental-metrical phonology. In: Jun, S-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0003

Gundel, J. K., Hedberg, N., and Zacharski, R. 1993. Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274–307. DOI: https://doi.org/10.2307/416535

Gussenhoven, C. 2004. *The Phonology of tone and intonation*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511616983

Gussenhoven, C. 2005. Transcription of Dutch intonation. In: Jun, S-A. (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 118–145). Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.003.0005

Harrell, F. E., Jr. 2001. *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer. DOI: https://doi.org/10.1007/978-1-4757-3462-1

Hastie, T., Tibshirani, R., and Friedman, J. R. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. Second edition. New York: Springer. DOI: https://doi.org/10.1007/978-0-387-84858-7

Hawthorne, K. and Gerken. L. 2014. From pauses to clauses: Prosody facilitates learning of syntactic constituency. *Cognition, 133*(2), 420–428. DOI: https://doi.org/10.1016/j.cognition.2014.07.013

Hird, K. and Kirsner, K. 2002. The relationship between prosody and breathing in spontaneous discourse. *Brain and Language*, *80*(3), 536–555. DOI: https://doi.org/10.1006/brln.2001.2613

Iivonen, A. 1998. Intonation in Finnish. In: Hirst, D. and Di Cristo, A. (Eds.), *Intonation systems: A survey of twenty languages* (pp. 311–327). Cambridge: Cambridge University Press.

Jackendoff, R. 1972. *Semantics in generative grammar*. Cambridge, MA: MIT Press.

Jun, S-A. (Ed). 2005. *Prosodic typology: The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199249633.001.0001

Jun, S-A. (Ed.). 2014. *Prosodic typology II. The phonology of intonation and phrasing*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/acprof:oso/9780199567300.001.0001

Jun, S.-A. and Lee, H.-J. 1998. Phonetic and phonological markers of contrastive focus in Korean. In: Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP 98) (pp. 1295–1298).

Kjelgaard, M. M. and Speer, S. R. 1999. Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153–194. DOI: https://doi.org/10.1006/jmla.1998.2620

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., and Den, Y. 1998. An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. *Language and Speech*, *41*(3–4), 295–321. DOI: https://doi.org/10.1177/002383099804100404

Krifka, M. 2008. Basic notions of information structure. *Acta Linguistica Hungarica, 55* (3–4), 243–276. DOI: https://doi.org/10.1556/ALing.55.2008.3-4.2

Kügler, F. and Genzel, S. 2012. On the prosodic expression of pragmatic prominence: The case of pitch register lowering in Akan. *Language and Speech, 55*(3), 331–359. DOI: https://doi.org/10.1177/0023830911422182

Ladd, D. R. 1996. *Intonational phonology*. Cambridge: Cambridge University Press.

Leemann, A., Kolly, M.-J., and Dellwo, V. 2014. Speaker-individuality in suprasegmental temporal features: Implications for forensic voice comparison. *Forensic Science International*, *238*, 59–67. DOI: https://doi.org/10.1016/j.forsciint.2014.02.019

Lehtonen, J. 1970. *Aspects of quantity in standard Finnish*. Jyväskylä: Gummerus.

Lehtonen, J. 1974. Sanan pituus ja äännekestot [Word length and segment duration]. *Virittäjä*, *78*. 152–160.

Liaw, A. and Wiener, M. 2002. Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358–368. DOI: https://doi.org/10.1037/h0044417

Lieberman, P. 1967. Intonation, Perception and Language (Doctoral dissertation). Department of Modern Languages and Linguistics, MIT.

Lunetta, K. L., Hayward, B. L., Segal, J., and Van Eerdewegh, P. 2004. Screening large-scale association study data: Exploiting interactions using random forests. *BMC Genetics*, *5*(1), 32. DOI: https://doi.org/10.1186/1471-2156-5-32

Mixdorff, H., Vainio, M., Werner, S., and Järvikivi, J. 2002. The manifestation of linguistic information in prosodic features of Finnish. Proceedings of Speech Prosody 2002. Aix-en-Provence. 511–514.

Myers, S. and Hansen, B. B. 2007. The origin of vowel length neutralization in final position: Evidence from Finnish speakers. *Natural Language & Linguistic Theory*, *25*(1), 157–193. DOI: https://doi.org/10.1007/s11049-006-0001-7

Nakai, S., Kunnari, S., Turk, A. E., Suomi, K., and Ylitalo, R. 2009. Utterance-final lengthening and quantity in Northern Finnish. *Journal of Phonetics*, *37*(1), 29–45. DOI: https://doi.org/10.1016/j.wocn.2008.08.002

Nakai, S., Turk, A. E., Suomi, K., Granlund, S., Ylitalo, R., and Kunnari, S. 2012. Quantity constraints on the temporal implementation of phrasal prosody in Northern Finnish. *Journal of Phonetics*, *40*(6), 796–807. DOI: https://doi.org/10.1016/j.wocn.2012.08.003

Nespor, M. and Vogel, I. 1986. *Prosodic phonology*. Dordrecht: Foris Publications.

Ní Chasaide, A., Yanushevskaya, I., and Gobl, C. 2011. Voice source dynamics in intonation. In: Lee, W-S. and Zee, E. (Eds.), Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS 2011), 1470–1473.

Niemi, J. 1984. *Word level stress and prominence in Finnish and English. Acoustic experiments on production and perception*. Joensuu: University of Joensuu. (Joensuun yliopiston humanistisia julkaisuja/University of Joensuu. Publications in the humanities 1)

Ogden, R. 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association, 31*(1), 139–152. DOI: https://doi.org/10.1017/S0025100301001116

Ogden, R. 2004. Non-modal voice quality and turn-taking in Finnish. In: Couper-Kuhlen, E. and Ford, C. E. (Eds.), *Sound patterns in interaction: Cross-linguistic studies from conversation* (pp. 29–62). Amsterdam: John Benjamins. DOI: https://doi.org/10.1075/tsl.62.05ogd

Patil, U., Kentner, G., Gollrad, A., Kügler, F., Féry, C., and Vasishth, S. 2008. Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics, 1*(1), 55–72.

Peters, J., Hanssen, J., and Gussenhoven, C. 2014. The phonetic realization of focus in West Frisian, Low Saxon, High German, and three varieties of Dutch. *Journal of Phonetics*, *46*, 185–209. DOI: https://doi.org/10.1016/j.wocn.2014.07.004

Pierrehumbert, J. B. 1980. The phonology and phonetics of English intonation (Doctoral dissertation). Cambridge: MIT.

Poesio, M. 2004. Discourse annotation and semantic annotation in the GNOME corpus. In: Proceedings of the 2004 ACL Workshop on Discourse Annotation (pp. 72–79). Stroudsburg, PA, USA: Association for Computational Linguistics. DOI: https://doi.org/10.3115/1608938.1608948

Porretta, V., Kyröläinen, A.-J., and Tucker, B. V. 2015. Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception, & Psychophysics*, *77*(7), 2438–2451. DOI: https://doi.org/10.3758/s13414-015-0916-3

Prince, E. 1981. Toward a taxonomy of given-new information. In: Cole, P. (Ed.) *Radical Pragmatics* (pp. 223–256). New York: Academic Press.

R Core Team. 2015. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Rieth, S. R., Stahmer, A. C., Suhrheinrich, J., and Schreibman, L. 2015. Examination of the prevalence of stimulus overselectivity in children with ASD. *Journal of Applied Behavior Analysis*, *48*(1), 71–84. DOI: https://doi.org/10.1002/jaba.165

Romøren, A. S. H. and Chen, A. 2015. Quiet is the new loud: Pausing and focus in child and adult Dutch. *Language and Speech*, *58*(1), 8–23. DOI: https://doi.org/10.1177/0023830914563589

Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116. DOI: https://doi.org/10.1007/BF02342617

Rowe, C. 1999. Receiver psychology and the evolution of multicomponent signals. *Animal Behaviour*, *58*(5), 921–931. DOI: https://doi.org/10.1006/anbe.1999.1242

Schwarzschild, R. 1999. Givenness, avoidF and other constraints on the placement of accent. *Natural Language Semantics*, *7*(2), 141–177. DOI: https://doi.org/10.1023/A:1008370902407

Selkirk, E. O. 2000. The interaction of constraints on prosodic phrasing. In: Horne, M. (Ed.), Prosody: Theory and experiment. Studies presented to Gösta Bruce (pp. 231–261). Dordrecht, Boston and London: Kluwer. Retrieved from http://link.springer.com/chapter/10.1007/978-94-015-9413-4_9, DOI: https://doi.org/10.1007/978-94-015-9413-4_9

Shattuck-Hufnagel, S. and Turk, A. E. 1996. A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–246. DOI: https://doi.org/10.1007/BF01708572

Shriberg, E. 2007. Higher-level features in speaker recognition. In: Müller, C. (Ed.), *Speaker Classification I*. Lecture notes in artificial intelligence vol. 4343, (pp. 241–259). Berlin Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-540-74200-5_14

Siohan, O., Ramabhadran, B. and Kingsbury, B. 2005. Constructing ensembles of ASR systems using randomized decision trees. Proceedings of International Conference on Acoustics Speech and Signal Processing (ICASSP), I–197–I–200.

Snedeker, J. and Trueswell. J. 2003. Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, *48*(1), 103–130. DOI: https://doi.org/10.1016/S0749-596X(02)00519-3

Strik, H. and Boves, L. 1995. Downtrend in F0 and Psb. *Journal of Phonetics*, *23*(1–2), 203–220. DOI: https://doi.org/10.1016/S0095-4470(95)80043-3

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis. A. 2008. Conditional variable importance for random forests. *BMC bioinformatics*, *9*(1), 1–11. DOI: https://doi.org/10.1186/1471-2105-9-307

Strobl, C., Malley, J., and Tutz. G. 2009. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. *Psychological Methods*, *14*(4), 323–348. DOI: https://doi.org/10.1037/a0016973

Suomi, K. 2009. Durational elasticity for accentual purposes in Northern Finnish. *Journal of Phonetics*, *37*(4), 397–416. DOI: https://doi.org/10.1016/j.wocn.2009.07.003

Suomi, K., Toivanen, J., and Ylitalo, R. 2003. Durational and tonal correlates of accent in Finnish. *Journal of Phonetics*, *31*(1), 113–138. DOI: https://doi.org/10.1016/S0095-4470(02)00074-8

Suomi, K., Toivanen, J., and Ylitalo, R. 2010. *Finnish sound structure. Phonetics, phonology, phonotactics and prosody*. Oulu: University of Oulu.

Tagliamonte, S. A. and Baayen, R. H. 2012. Model, forests and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178. DOI: https://doi.org/10.1017/S0954394512000129

Tremblay, A. and Newman, A. J. 2015. Modeling nonlinear relationship in ERP data using mixed-effects regression with R examples. *Psychophysiology*, *51*(1), 124–139. DOI: https://doi.org/10.1111/psyp.12299

Trouvain, J., Barry, W. J., Nielsen, C., and Andersen, O. 1998. Implications of energy declination for speech synthesis. Proceedings of the Third ESCA/COCOSDA Workshop on Speech Synthesis (SSW3), 47–52.

Vainio, M, Airas, M., Järvikivi, J., and Alku, P. 2010. Laryngeal voice quality in the expression of focus. Proceedings of Interspeech, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26–30, 2010. 921–924.

Vainio, M. and Järvikivi. J., 2007. Focus in production: Tonal shape, intensity and word order. *The Journal of the Acoustical Society of America*, EL55–61. DOI: https://doi.org/10.1121/1.2424264

Välimaa-Blum, R. 1993. A pitch accent analysis of intonation in Finnish. *Ural-Altaische Jahrbücher 12*, 82–94.

von Stechow, A. and Uhmann S. 1986. Some remarks on focus projection. In: Abraham, W. and de Meij, S. (Eds.). *Topic, focus, and configurationality* (pp. 295–320). Amsterdam/Philadelphia: John Benjamins. DOI: https://doi.org/10.1075/la.4.14ste

Wang, B. and Xu, Y. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. Journal of Phonetics, *39*(4), 595–611. DOI: https://doi.org/10.1016/j.wocn.2011.03.006

Wieling, M., Montemagni, S., Nerbonne, J., and Baayen, R. H. 2014. Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692. DOI: https://doi.org/10.1353/lan.2014.0064

Wiik, K. and Lehiste, I. 1968. Vowel quantity in Finnish disyllabic words. In: Ravila, P. (Ed.), Congressus Secundus Internationalis Fenno-Ugristarum, Helsingiae habitus 23–28. VIII. 1965. Pars I, (pp. 569–574). Helsinki: Societas Fenno-Ugrica/ Suomalais-Ugrilainen Seura.

Wood, S. N. 2006. *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman & Hall/CRC Press.

Xue, J. and Zhao, Y. 2006. Random forests-based confidence annotation using novel features from confusion network. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), I–1149–I–1152.

Xue, J. and Zhao, Y. 2008. Random forests of phonetic decision trees for acoustic modeling in conversational speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *16*(*3*), 519–528. DOI: https://doi.org/10.1109/TASL.2007.913036

Xu, Y. 1999. Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*, 55–105. DOI: https://doi.org/10.1006/jpho.1999.0086

Ylitalo, R. 2009. *The realisation of prominence in three varieties of standard spoken Finnish*. Oulu: University of Oulu, 2009. (Acta Universitatis Oluensis, B Humanoiara 88)

Zuur, A. F., Ieno. N., Walker, N. J., Saveliev, A. A., and Smith, G. M. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer. DOI: https://doi.org/10.1007/978-0-387-87458-6

How to cite this article: Arnhold, A and Kyröläinen, A-J 2017 Modelling the Interplay of Multiple Cues in Prosodic Focus Marking. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 8(1): 4, pp. 1–25, DOI: https://doi.org/10.5334/labphon.78

Published: 13 March 2017

]u[*Laboratory Phonology: Journal of the Association for Laboratory Phonology* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS