6

GENERAL CONCLUSION AND DISCUSSION

Abstract

In this final chapter the various threads in the previous chapters are put together. We demonstrated that naive listeners are able to assign in a consistent way word prominence in individual Polyphone sentences, with little instruction to the listener. This enables prominence annotation for large speech databases by non-experts. Most differences between-listeners (reliability) and within-listeners (consistency) can be ascribed to level shift or level differences. These findings allow us to use an operational definition of prominence. Next, we have demonstrated that, on the basis of text input only, the prominence level of words can be predicted with a performance similar to that of naive listeners. The linguistic information used was limited to word class (POS), word length, Adjective-Noun combinations and the first content word in a sentence. We studied a selected set of acoustic correlates of prominence, all based on F_0, duration and intensity extracted at units not larger than a word. Finally, we used twelve acoustic features as input for a neural net classifier. A binary prominence classifier appeared to be correct in 79% of the cases, which is statistically indistinguishable from a consistent naive labeler. We can thus conclude that automatic prominence assignment is sufficiently powerful to simulate naive labelers for read-aloud declarative medium length sentences in Dutch.

6.1 Introduction

The perceptual concept of prominence, which is the amount of emphasis put on syllables and words that make them 'stand out' in their environment, can function as an interface between acoustics and linguistics. It is strongly related to information structure in terms of 'given' and 'new' information or in terms of focus. In our approach a perceptually defined concept of prominence is the basis for all further analyses. This perceptual phenomenon of prominence appears to be intuitively clear to non-experts. A proper modeling of prominence, either from the signal, or from text, may be useful for several applications in speech technology (e.g. in dialogue handling).
The purpose of this study was to explore various aspects of prominence. First, we developed a useful operational definition of prominence via of judgments of naive (native) listeners. We focused on analyzing various linguistic and acoustic correlates of prominence referring to bottom-up and top-down information, respectively. The final goal of this study was to find and extract those features that can best be used to predict prominence automatically.

6.2 Operational definition of prominence by naive listeners

On the basis of two pilot studies and a larger experiment concerning prominence assignment, an operational definition of prominence was formulated. Prominence is defined as the amount of emphasis attributed by a group of naive native transcribers. For the sake of validity we wanted to keep our listeners as 'naive' as possible as to the task to be accomplished. The marking of prominence in the way we did (binary marking on words by more than one listener) was a useful approach to mark large databases, leaving us with enough detailed information to analyze acoustic and linguistic correlates of prominence. However, the investigation of gradient prominence on an 11-point scale proved too difficult to handle.

In this part of our study, there appeared to be three main discussion items: 1) should we assign prominence at the word or at the syllable level, 2) should we use binary marks or marks on a gradient scale to indicate the degree of prominence and 3) how is the consistency and reliability of the listeners.

6.2.1 Word or syllable prominence marking

An advantage of word prominence assignment is, that it is much easier to perform than syllable-prominence assignment. Additionally, words are more meaningful elements for naive listeners than syllables are. A word is a unit of expression, which has a universal intuitive recognition and meaning by native listeners and speakers. This increases the validity and this makes the labeling intuitively more plausible.
A disadvantage of marking prominence on word level is, however, that one has no detailed information about the identity of the prominent syllable in polysyllabic

words. To circumvent this disadvantages, an alternative approach would be to mark the prominent syllable(s). Using this approach about 250,000 words of the ten-million-words *Corpus Gesproken Nederlands* (short CGN *Spoken Dutch Corpus, http://lands.let.kun.nl/cgn/,* Buhmann et al., 2002) are currently annotated for syllable prominence. This implies that detailed information about the prominence distribution within words, as realized in utterances, will become available. However, the labelers had to be trained to mark syllable prominence according to a detailed protocol. This increases the consistency, but decreases the simplicity and the applicability of the approach.
In our study, it was decided to ask the listener to mark word prominence rather than syllable prominence. This decision was partly based on the results of the pilot experiment as described in section 2.3, which showed that word prominence tends to be assigned more consistently than syllable prominence. For all analyses presented in this study this word-based approach appeared to be detailed enough. Lexical (word) stress was used as a next best approximation to identify the most prominent syllable in those words that were labeled as prominent. For further research it is interesting to investigate whether different strategies were used to mark prominence on the syllable level in comparison to the word level.

6.2.2 Binary or gradient prominence marking

Fant & Kruckenberg (1989), Portele & Heuft (1997), and Grover et al. (1997) used naive listener judgments to define prominence in a detailed way. In their studies, prominence judgments were given for every syllable in the sentence on a very detailed scale. Such a detailed prominence assignment was unpractical in the present study, because it would not allow the annotation of large acoustic databases. Moreover, such detailed information about prominence generally appeared to be unnecessary, even according to the above authors: they all considerably reduced the detailed prominence scales in their further analyses.
In the present study it was decided to ask from each listener a binary prominence mark instead of an n-points gradient prominence scale. Binary marking makes the annotation task easier to perform. This appeared to be a useful approach in our study, whereas the cumulative marks over a group of listeners provided a useful indication for the gradient degree of prominence.
As pointed out before problems remained with application of the cumulative 11-point scale. We have reduced this scale to four prominence classes by means of a hierarchical cluster analysis and even reduced these four classes to a binary prominence distinction. While ten listeners marked the training set, one 'normative' listener was selected to mark the test set. Further research is needed to find out what a useful and necessary detailed range of the prominence scale is. Related to this, questions arose such as what is the relationship of prominence and the linguistic phenomenon of lexical (word) stress in terms of 'stressed' and 'unstressed' syllables, and / or distinction of four degrees 'primary', 'secondary' 'tertiary' and 'weak'. The phonetic notion of pitch accent and the relationship of prominence needs further research to determine if words marked as highly prominent receive

always an accent-lending pitch movement, and whether words being marked as less prominent belong to e.g. 'secondary' stressed words.

6.2.3 Consistency and reliability

We found that listeners are rather consistent and reliable in their prominence judgments. On average the listeners agreed with one another with a Cohen's Kappa of $\kappa = 0.50$. Considering that one had the freedom to mark just one or several words for prominence per sentence, this is a reasonable degree of consistency. Most of the differences between- and within-listeners can be explained by having a different threshold for prominence marking or, in case of the within-listener differences a threshold shift for prominence marking. Training of the labelers and close instructions by the researchers may help to increase the agreement. However, we point out that we aimed at the smallest influence from the researcher as possible, because the interpretation of prominence (e.g. in terms of the number of prominence marks per sentence) has to be defined by the listener. The listener has been asked to 'mark those words that he / she perceived to be pronounced with emphasis', and had to confirm whether he / she understood the task prior to the annotation itself.
Although this has not been investigated, one must assume that the listeners all have used their own strategy for judging prominence on words. It would be interesting for further research to investigate whether the listeners use different strategies to mark prominence.

6.2.4 Concluding remarks

The growing number of publications on prominence shows the great interest in this topic, especially in combination with speech technology. At some point there were even suggestions to change TOBI into a TOBI-lite version in which the number of different pitch accent types would be reduced and a degree of prominence would be added (Wightman & Rose 1999; Wightman et al., 2000). These publications underline that there is a need to come to a good definition of prominence and how to use prominence.

6.3 Lexical / syntactic correlates of prominence

In chapter 3 we described linguistic correlates / determinants of prominence, how they were extracted and how they were used to predict prominence. These predictions were exclusively based on information that is derived automatically from the text. A detailed analysis of the linguistic determinants gives on the one hand insight into the relationships that allow automatic prediction of prominence. On the other hand the analysis gives insight into the linguistic information, namely the expectation of prominence (top-down information) the listener uses.

6.3.1 Individual correlates

After considering many options of the lexical and syntactic correlates of prominence, we drew up the following list of promising candidates:

A) word class;
B) word length;
C) Adjective-Noun combinations;
D) the first content word of the sentence;

A) Our first finding is that word classes can be ordered according to their ability to carry prominence. The following ranking of increasing prominence is found for the sentence material we used: Article, Conjunction, Pronoun, Auxiliary verb, Verb, Numeral, Adverb, Adjective, Noun and Negation. Note that Negations normally belong to the category of function words, which are considered to be less prominent, but which were marked as 'highly prominent' in our speech material. For the other word classes the overall group distinction between function word / content word remains.

B) Secondly, it was found that the word length, expressed in the number of syllables, is a useful correlate of prominence. The metrical weight, referring to the complexity and the length of a word, is related as well. However, we have not investigated whether another component, the 'complexity of a word', is also related to prominence.

C) The third interesting relationship is found between Adjective-Noun combinations and prominence. The Noun in such a combination is generally less prominent than in other combinations.

D) The last striking finding is that the first content word is often a very prominent one in these read-aloud sentences. This may be specific for this type of speech material, but even then such a type of relationship could still be profitable to predict prominence. We did not investigate the possibility of testing the prominence prediction algorithm on utterances with an entirely different grammatical structure (main and sub clauses, questions). It would be interesting for further research to find out how this relationship behaves for other speaking styles and text types. E.g. in a dialogue situation there may be no need to mark the first content word of an utterance, because the contextual situation is clearer than in separate read-aloud sentences.

6.3.2 Prominence prediction on textual input

All these relationships were implemented into an algorithm to predict prominence. Our algorithm initially predicted the degree of prominence on a 4-point scale, which was later reduced to a 2-point scale. The final binary prominence prediction on an

independent test set appeared to be 81.2% correct. The prediction on the test set based on textual input agrees with a Cohen's Kappa of $\kappa = 0.62$ with the marks of the listener who showed the highest consistency on the training set. This agreement was better than the mean agreement between listeners ($\kappa = 0.50$). So, our algorithm for binary prominence prediction from text shows an agreement that is at least similar to that of between-listeners.

6.3.3. Discussion about lexical / syntactic correlates

6.3.3.1 Comparison to the literature

Comparing our results to other research it appeared that we have achieved similar results. A performance of 80-90% correct prediction of pitch accent placement is reported by Hirschberg (1993). She included automatically derived discourse information in her predictions. A result of 82.5% correct prediction of accent placement is reported by Ross & Ostendorf (1996). They used hand-labeled boundary information, but automatically derived Part-of-Speech tags and even topic information to predict accent placement. Vereecken et al. (1998) predict degrees of prominence on a 4-point scale for Dutch with a performance of about 80% correct.
The task to predict different degrees of prominence is difficult and complicated. This is also reflected in the literature mentioned in the introduction and especially in the research of Widera et al. (1997).

6.3.3.2 Method used

The method we used to analyze the linguistic correlates of prominence consisted of the following steps. First, we had a closer look at the linguistic data and tried to find dependencies. Second, we translated these dependencies into simple heuristic rules, and third, these rules were validated on an independent test set. All these steps were performed fully automatically. This makes our findings relevant in two ways a) they show that certain annotation tasks by humans can be simulated by algorithms with similar or better reliability and b) for speech technology applications. The optimization method we used is a heuristic one, although there are more sophisticated techniques available to analyze large databases. Some of these more probabilistic techniques such as classification trees or artificial neural networks may yield higher performance in predicting prominence. However, with such probabilistic techniques it is more difficult to extract specific knowledge about the relationship of perceived word prominence and lexical and syntactic correlates. Such extraction is explicitly possible from the heuristic rules derived in this study.

6.3.3.3 Useful for Text-to-Speech

When the amount of prominence for each word in a sentence could accurately be predicted it would greatly improve the intelligibility, the naturalness and the pleasantness of Text-to-Speech systems. Rules to predict different degrees of prominence, which are solely based on automatically derived textual information, would be very useful for more sophisticated Text-to-Speech systems. Because of the inter-subject differences (see section 2.4.1.2), a perfect synthesis of a prominence contour will be very difficult if ever possible. Furthermore, this is not even necessary, as only one good prominence prediction for all the words in a sentence is needed for speech synthesis purposes.

6.3.3.3.1 No context information available for 'our' read-aloud sentences

In our case, a disadvantage might be that the speech material is not uttered in context. Therefore it was impossible to determine 'focus' and / or 'given' and 'new' information. An advantage could be that the reading of these sentences is a default reading and that the material is also useful for certain technological applications. However, the speech material is designed for speech recognition; for speech synthesis only one professional example speaker is needed. From literature it is known that there are speaker-dependencies (especially gender) of perceived prominence of F_0 peaks (Gussenhoven & Rietveld, 1998). Since the speech material that we used contains a lot of different speakers, these speaker dependencies are averaged out in the analysis results.

6.3.3.3.1 Translation into acoustic properties

A complicating factor for Text-to-Speech systems is the need to translate the different degrees of prominence into proper acoustic values. This is, however, not our main concern in this study. As Hermes (1991) reports, a falling pitch movement is generally perceived as less prominent than a rising pitch movement with the same excursion size measured in ERB's. A first-order normalization for this finding might be possible, but we have to keep in mind that the exact relationship might be complex. Another problem is the detection of when the pitch movement starts as compared to the onset of the vowel. A pitch movement starting late in the vowel has a different effect on the perceived prominence than a pitch movement starting very early (Hermes, 1995).

6.3.3.3.2Testing the prosody

For speech synthesis purposes, actual testing of the prediction of degrees of prominence with the rules developed and described in chapter 3 will be difficult. We made some preliminary attempts in a pilot experiment, but were not very successful. The different degrees of prominence must first be properly translated into acoustic correlates that cannot be limited to certain pitch movements only. This must then be

implemented into an existing synthesizer in order to compare it with a default algorithm. Much research will have to be done concerning this problem.

6.4 Acoustic correlates of prominence

Acoustic features concerning F_0, duration and intensity have been used by us to discriminate between non-prominent and prominent words. A detailed analysis of possible acoustic features, which are extracted from a unit not larger than a word, is performed. The current analysis concentrates on the individual unit (vowel, syllable and word) and did not investigate relative features e.g. looking at the previous or following unit(s). We came to the conclusion that prominence (as assigned by naive listeners) is reflected in the acoustic speech signal of an individual unit. The listener perceived variations in duration, intensity and F_0 and could use it as bottom-up information contributing to the prominence of a word.

6.4.1 Individual acoustic features

Several selected features based on F_0, duration and intensity at syllable, word or sentence level can automatically be extracted from the speech signal. From the twelve acoustic features, the most distinctive single feature for binary prominence prediction appeared to be the range of F_0 measured in semitones. The F_0 range per word showed a better ability to discriminate prominent and non-prominent words than the F_0 range per syllable. The scores for correct prominent / non-prominent classification were 72% and 69% on word and syllable level, respectively. In this study it is found that syllable duration is also a powerful feature for prominence prediction; even a better one than vowel duration. Without any corrections for intrinsic vowel duration and/or the number of phonemes, binary prominence classification using only syllable duration gave about 71% correct. On the basis of vowel duration as a single feature, a correct binary prominence prediction of 66% was reached. Vowel intensity used as the only input feature to classify prominent and non-prominent words gave results of about 67% correct. This result indicates that vowel intensity is also an important cue for prominence.

6.4.2 Prominence prediction on acoustic input

As observed in the previous section, several individual features showed good results. However, a neural-net classification with all twelve selected features performed better. On the independent test set an appropriately trained neural net performs at 79% correct classification. Comparing the agreement of, on the one hand, the predicted prominence marks between those of the listener in the test set (Cohen's Kappa $\kappa = 0.57$, see table 5.3) with, on the other hand, the average agreement between listeners on the training data (Cohen's Kappa $\kappa = 0.50$), it can be concluded that the neural net prediction is at least as good as the naive listeners' performance.

6.4.2.1 Complexity

From the analyses presented in chapter 5 it also became clear that the relationship between the acoustic features and prominence is not simply linear, but sometimes rather complex. This conclusion justifies and probably partly explains the fact that the use of a hidden layer substantially improves the recognition performance of the neural networks, in comparison with an LDA. What, however, the structure is of this complexity needs further investigations because of the multitude of confounding factors. E.g. how long a given syllable had to be in combination with which changes in pitch is an interesting research question.

6.4.3. Discussion about acoustic correlates

Various points of our approach that may require further discussion are: firstly a comparison to the literature, secondly the method used, thirdly the HMM-alignment, fourthly the applied normalization and lastly, the strictly separate use of linguistic and acoustic features.

6.4.3.1 Comparison to the literature

Just as for the prediction of prominence with textual input, our prediction method with acoustic input seems to produce results comparable to those reported in the literature. The statistical and / or brute force method used by the authors mentioned below aimed at high recognition rates whereas our approach aimed at different perspectives of acoustics and classification providing insight in phonetic questions as well. Our prominence classification results were achieved using solely twelve selected acoustic input features. The comparison is not completely fair as it concerns different methods for training and testing and differences in speech material used. Kießling (1996) reported a recognition rate of 83% correct accent classification for the VERBMOBIL-speech data while using also textual features such as identity of the vowel. Kompe et al. (1995) report classification rates of 95.6% correct for the ELRA corpus. This corpus contains read-aloud sentences with a simple grammatical structure. Wightman & Ostendorf (1994) reached 83% correct using hand-labeled boundary features and Silipo & Greenberg (1999) classified stressed and unstressed syllables with 80% and 77% correct, respectively.

6.4.3.2 Method used

Just as in section 6.3, where we discussed the lexical / syntactic approach, we will now discuss the acoustic approach. Following a detailed analysis of some of the most promising acoustic correlates (correlates that were promising on an individual basis), we selected those features that seemed to have the greatest potential to predict prominence using a multidimensional classification. This pre-selection reduced the number of features for the classification task. In this study it was decided to use simple feed-forward neural networks for predicting prominence. As

said before, sophisticated techniques such as CART, and more complex self-learning neural networks are available, but we decided to put emphasis on a full understanding of the classification process. Within the classification process both the individual features as well as various combinations were tested for their contribution to discriminate prominence.

6.4.3.3 HMM-alignment

In our approach a human-made orthographic transcription was available to do the HMM-alignment. This alignment is used to automatically obtain the segmentation (in terms of phoneme and word boundaries) of the utterances. In realistic speech recognition situations such a precise and correct word-level transcription of what has been said is generally not available. The only available transcription in such a case is the result of automatic speech recognition, which introduces additional errors into the automatic alignment. Furthermore, the automatic HMM-alignment we used introduces, by its very nature, errors and problems to the resulting segmentation. The automatic segmentation used in this study was based on Dutch standard pronunciation, although the pronunciation of some speakers showed regional variants. We did not investigate the precise consequences of this in any detail, as in speech technology in general one has to cope with similar conditions.

6.4.3.4 Normalizations

Apart from possible segmentation errors ascribed to the assumption of a standard pronunciation, other segmentation errors are unavoidable in this automatic procedure. This may cause fewer problems for measurements in larger segments, such as syllables and words, than for measurements in smaller segments, such as vowels and consonants. Normalizations concerning intrinsic vowel duration and intrinsic vowel intensity were conducted at such small units, namely vowels. Maybe this fact and the large speaker variability explain why these normalizations on the level of small units did not substantially improve the ability to discriminate between prominent and non-prominent words. Analyses on sentence speaking rate were also conducted in this study. Speaking rate is reported to influence the duration of vowels and syllables. Although an effect of sentence speaking rate on vowel and syllable duration is indeed shown in chapter 4 (see section 4.2.2.3 and figure 4.14), the normalization for speaking rate that we applied, did not improve prominence classification (section 5.2.7). Further research is required to investigate the precies effect of normalizations. We tried overall-normalizations (intrinsic vowels duration, intrinsic vowel intensity, sentence speaking rate) with no effect on the prominence classification when used separately. However, the approach and the speech material may have triggered this negative result. Other research shows that putting the vowel / syllable / word in larger context (e.g. taking the previous and next syllable and /or word into account) increases the classification results. Taking into account these relative features in a detailed analysis may help to get more insight into the relative character of prominence.

6.4.3.5 Separate use of linguistic and acoustic features

The strictly separate use of linguistic and acoustic features may be disadvantageous for syllable duration. Syllable duration might also be influenced by linguistic information, such as word class. Content words often consist of more complex syllables than function words; this means that syllables of content words contain more phonemes. This higher complexity may result in longer syllable duration. So, there might be a relationship between syllable duration and the content word / function word distinction. Fant & Kruckenberg (1999) noticed that syllable duration was the most robust correlate of prominence. Furthermore, we have of course to keep in mind that pitch extraction is not always error free (octave errors). An example of such an error was demonstrated in chapter 4, figure 4.4.

6.5 Future research

In this section, first some suggestions for the use of other promising features of prominence are given. This concerns for instance the relationship between word prominence and spectral quality. Next, a few speech-technological applications are discussed and finally, a number of remarks are made about combining linguistic and acoustic features.

6.5.1 Promising features of prominence

A possibly promising correlate, such as the distance between two prominent words in a sentence (speech rhythm, stress clash), was not investigated. We do not exclude the possibility, however, that another prominent word may be required to follow the first prominent word after a period of time in a sentence. Information concerning the text structure, for example, whether or not words convey given / new information or boundary information, as used in Wightman & Ostendorf (1994), could not be assigned and used in our analyses. Pragmatic and semantic information could not be derived automatically and was thus not available, and is most probably less relevant for isolated sentences. For future research on the improvement of speech synthesis based on prominence, the use of such more elaborate information about text structure could be helpful.
Spectral quality was also excluded from our analyses. We took this decision since consistent spectral properties are difficult to measure automatically. Further research will be required to measure spectral quality reliably and to learn more about its relationship with prominence. One of the issues to be solved is a phoneme-based normalization of the spectral quality measure.

6.5.2 Speech-technological applications

In this study several suggestions have been made about how prominence could be used in speech-technological applications. The suggestions include: the improvement of the intonation of speech synthesis, a word-by-word prominence

indicator, and sentence disambiguation. However, such implementations require further research.

Predicting prominence from textual input could help to improve speech synthesis. In the presently popular and very promising approaches of (variable) unit-based concatenative speech synthesis (Klabbers, 2000; Stöber et al., 1999; Wightman et al. 2000), the prediction of prominence (from text) is of crucial importance for the pleasantness and naturalness of the synthesized speech signal. The text-derived prominence labels will have to be matched with the prominence labels in the annotated speech database that contains the segments to be concatenated. A search algorithm is supposed to select the optimally matching segments and to concatenate them. So, a translation from prosodic labels to acoustic parameters is not always needed: the (larger) speech units in the database already contain most of that information, although sometimes additional signal adaptations are required. However, for diphone synthesis a translation from prominence labels to acoustical parameters will almost always be necessary. How exactly prominence labels that are predicted from text can be translated into acoustic features is beyond the scope of this study and needs further research. The exact relationship between prominence and information retrieval (focus, contrast, topic, etc.) has not been investigated either.

As mentioned earlier in this study, prosody is so far hardly used in present day speech recognition. We suggested two applications: a word-by-word prominence indicator and an instrument to disambiguate the meaning of an ambiguous sentence. Although, sentence disambiguation has already been a topic for research (Batliner et al., 1998, for instance within the German Verbmobil project), a running prominence indicator for speech recognition is a new idea that still awaits application. Ida & Yamasaki (1998) show improvements for keyword spotting as used in speech recognition based on prosodic information. Knowledge about or estimation of the prominence of a word during the recognition process can provide islands of reliability or can point out the importance of a word. Wang & Seneff (2001) and van Kuijk & Boves (1999) used lexical stress determined through lexical look-up to improve speech recognition, but the improvement they could achieve was negligible. Another example of using prosody is given in Taylor et al. (1998). They described how prosody helped to constrain speech recognition in a dialogue environment. Positive results in this area have been reported in recent papers by Hirschberg and Swerts (1998), and by Wang (2001).

6.5.3 Combination of linguistic and acoustic information

Combining acoustic and linguistic features may improve prominence prediction as shown by Vereecken et al. (1998). In speech synthesis only textual information is available, whereas in speech recognition only acoustical information is available. However, for the annotation of a speech corpus, usually both acoustical and textual information are available. For the large 10 million words CGN Corpus it is the intention to annotate prominence of 250,000 words by hand. Automatic labeling

procedures can substantially help to consistently annotate large databases. These procedures can also be useful to improve the quality of concatenative speech synthesis.
Linking the linguistic and the acoustic features can lead to more reliable recognition / prominence classification rates, and furthermore it will also be very useful for automatic annotation. Further tests and further research will be required to investigate whether automatic annotation will be possible, especially if it concerns a lot of different speaking styles and different speakers. However, the detailed analyses of prominence as presented in this study, as well as the acoustic and linguistic correlates, hopefully do provide much information and various suggestions for further improvements of speech technology.

Form a more scientific viewpoint it is interesting to know more about the recognition process of prominence in general. That prominence is reflected in the speech signal of individual units was shown in the present study. How far the listener uses this bottom-up information to match his expectations of prominence on the basis of linguistic knowledge of his languages (top-down) is an interesting research topic.
That prominence is reflected in the lexical and syntactic knowledge was also shown in the present study. However, to which end the listener uses 'our' correlates and how the resulting expectation is matched with the bottom-up information related to the speech signal is beyond the scope of this study, however, very interesting for future research.

This study underlines that prominence is reflected in the acoustic and the linguistic domain, and that a binary prominence prediction with a selected set of relatively simple features can lead to a performance similar to that of naive listeners.