Institute of Phonetic Sciences,
University of Amsterdam,
Proceedings 25 (2003), 207-209.

SUMMARY OF PH.D. THESIS DEFENDED IN 2002

PROMINENCE. ACOUSTIC AND LEXICAL/SYNTACTIC CORRELATES

author: Barbertje M. Streefkerk
promotor: Louis C.W. Pols
co-promotor: Louis F. M. ten Bosch
date of defence: October 8th, 2002

Summary

The purpose of this study was to explore the notion of prominence in spoken language. It concentrated on finding an operational definition of prominence, on giving a description of the linguistic and acoustical correlates of prominence, and on analyzing these correlates in terms of their contribution to prominence distinctions. Furthermore, this study was concerned with feature extraction, and with prominence prediction, either on the basis of linguistic features or on the basis of acoustic features.

In chapter 1 the notion of prominence is explained, the use of prominence in (speech) communication is illustrated and research questions are described.
In speech some parts are more prominent than others. This is a gradient property. In many languages prominence helps to structure the message e.g. the prominent parts are the important ones. In addition, prominence helps to increase the comprehensibility and the naturalness of speech.
The listener uses two information sources to perceive prominence levels: bottom-up information and top-down information. The listener uses cues from the speech signal such as speech segments being louder, being longer and being realized with a pitch movement (bottom-up information) to detect prominence. The expectation of prominence is built on the basis of his/her knowledge of the language (top-down information).
From a phonetic point of view prominence is closely related to the notion of pitch accent and lexical (word) stress. Prominence is a perceptual phenomenon and is intuitively clear to non-experts. Prominence can function as an interface between acoustics and aspects of structure e.g. in terms of 'given' and 'new' information. The prediction of prominence may also be useful in speech technology.
The research questions concentrate mainly on: 1) how to find an operational definition of prominence, 2) which are the linguistic determinants / correlates of prominence, and 3) which acoustic correlates can be found. The implementation part of this

research concentrates on the automatic extraction of features, on the analysis, and on the prediction of prominence on the basis of the pre-selected features.

In chapter 2 a perceptual definition of prominence is investigated. The read-aloud sentences of the Dutch Polyphone Corpus (telephone speech) are used as research material, which unavoidably contains a great deal of speaker variability, and which is typical for many speech-technology applications.
The prominence-marking task was made as easy as possible to the subjects, giving them as much freedom for their own interpretation of prominence as possible, and allowing listeners to label large amounts of data. It was decided to mark prominence in a binary rather than multi-valued way, because otherwise the task was too time consuming and multi-valued marks from each listener appeared not to be necessary since it was shown that the cumulative marks also provide gradient prominence information. However, the results of a pilot-experiment were not very convincing. It was concluded that listeners mark prominence at the word level more consistently than at the syllable level. Since the unit of a word is also more meaningful to naive listeners than syllables, the word was chosen as the unit to mark.

In this research prominence was made operational in the following way: ten listeners were asked to mark those word(s) that they considered were spoken with emphasis. The cumulative marks of listeners provided detailed information about the degree of prominence of each word. In such a way the 1,244 sentences of the training set were marked for prominence. One 'optimal' listener, just giving binary judgments only, marked the independent test set of another 1000 sentences. This relatively simple binary marking allowed for an annotation of word prominence for more than 4.5 hours of speech. The listeners were rather consistent (mean agreement expressed in Cohen's Kappa $\kappa = 0.50$) and reliable. Many of the inconsistencies could be attributed to shifts of the individual prominence detection thresholds. However, threshold shifts and differences occur, which influence the agreement measure negatively.

In chapter 3 linguistic correlates of prominence are described, analyzed and used as predictors for prominence.
Relationships between, on the on hand, (1) Part-of-Speech (e.g. Noun, Adverb, Article), (2) word length, (3) position of a word in the sentence, and (4) interdependency of Part-of-Speech categories such as Adjective-Noun combinations and, on the other hand, prominence are described and analyzed in detail. Word classes are ranked according to increasing prominence and word length appears to be related to prominence. In generally, the longer the words the more prominent they are. Nouns occurring in Adjective-Noun combinations tend to be less prominent than in all other combinations and the first content word in a sentence is more prominent than the content words occurring at other positions in the sentence.
Based on these relationships an algorithm was developed to predict prominence degrees. This gradient prominence prediction, especially in the middle part of the scale, is more problematic. However, the reduced binary prominence prediction is correct in 81% of the cases for the independent test set.
Concluding, one is able to select a simple set of automatically derived linguistic features, which predicts prominence with the same agreement as listeners do (Cohen's Kappa $\kappa = 0.62$), indicating that top-down information can provide enough to predict prominence accurately. However, some used linguistic relationships may be specific for this type of speech / text material.

Chapter 4 deals with the description and detailed analysis of the acoustic features of prominence. It concentrates on the feature extraction on the level of the individual word and does not take the neighboring words into account. This research suggests that the following features are useful for predicting prominence: (1) F_0 range per word, (2) F_0 range per syllable, (3) syllable duration, (4) vowel intensity, (5) median F_0 per syllable, (6) vowel duration, (7) normalized vowel intensity and (8) normalized

vowel duration. The above order gives also the ranking with respect to the features' ability to discriminate between prominent and non-prominent words:
It was striking that when the vowels were normalized for their intrinsic properties, such as intrinsic vowel duration and intrinsic vowel intensity, the discrimination was no better than using the unnormalized counterparts. It was hypothesized that the variability in the speech material used was too large to properly correct for intrinsic durational and other properties.

Chapter 5 deals with the question whether the selected set of analyzed acoustic correlates could be used as input features in order to recognize prominence. It was shown that apart from F_0, syllable duration (more than vowel duration) and vowel intensity were useful input features for a recognition device. Automatic extraction of acoustic features was performed in such a way that a binary neural net classification resulted in a best recognition rate of 79% correct on the independent test set. The agreement of the predicted prominence (Cohen's Kappa $\kappa = 0.50$) was at least as good as the mean agreement of the listeners. This result was achieved with only twelve input features. Gradient prominence prediction on a 10-point scale is more difficult and requires further research.

In the last chapter (chapter 6) all findings and conclusions are summarized. Naive listeners were able to mark prominent words in spoken sentences with some consistency and reliability. The results of this study showed that acoustic and linguistic correlates of prominence can be determined automatically and they can be used to predict prominence either on text or on the speech signal.
Prominence assignment of naive listeners is valuable because the determined acoustic correlates, related to bottom-up information, and linguistic correlates, related to top-down information, describe the perceptual notion of prominence. This research shows that the prediction of prominence by acoustic or linguistic features is undistinguishable from prominence assigned by naive listeners.