Prominence. Acoustic and lexical/syntactic correlates

Streefkerk, B.M.

Citation for published version (APA):
1

INTRODUCTION

Abstract

This study investigates the acoustic and linguistic correlates of prominence. In this chapter the notion of prominence is explained and its use in language and communication is illustrated. Next, the research questions that will be dealt with in this study, will be identified.
1.1 Notion of prominence

This chapter is divided into five sections. In the first section, we will explain what we consider ‘prominence’ to be from a general viewpoint, as well as from a phonetic and a linguistic viewpoint. The things we want to know about prominence are dealt with in the second section. In the third section we will motivate why we want to investigate prominence and in the fourth section we will discuss how prominence will be investigated. The final section will present an outline of the present study.

1.1.1 General viewpoint

When we listen to speech some parts seem more prominent than others. In other words, we perceive specific parts of the speech signal as uttered with more ‘emphasis’ than other parts. This emphasis is called ‘prominence’.

Prominence is not a fixed property. It changes over time and is dependent on many linguistic, textual and acoustic-phonetic factors. Word groups, single words, syllables and even single phonemes can differ in prominence (Ladd, 1996; Sluijter, 1995; van Heuven, 1994). This difference in prominence is not a binary property but rather a gradient property (Terken, 1996; Rietveld & Gussenhoven, 1985).

In many languages, such as Dutch, English, German and French prominence is used primarily to structure a message i.e. to give emphasis to specific parts of the message. Prominence is just one of the ways in which the information structure of a message can be made more explicit; another way is phrasing. One could also change the word order such as with clitic pronouns in French (je, tu, il) that can never receive prominence and therefore give rise to structures such as c’est moi qui l’ai fait ‘it was I who did it’ (cf. *Je l’ai fait, I did it).

Structuring the message is not the only benefit of prominence; applying appropriately varying levels of prominence also increases the naturalness and the comprehensibility of speech.

A speaker uses prominence to mark those parts that are important in his message, and the listener uses (perceived) prominence in order to know which parts are of special interest for the perceived message. The listener combines bottom-up information from the speech signal with his expectation of prominence on the basis of his knowledge of the language (top-down information). In this study, we will concentrate on the perceived prominence.

1.1.2 Phonetic viewpoint

From a phonetic viewpoint the notion of prominence is not clearly defined. Pitch accent, sentence accent, stress, lexical (word) stress, word stress, reduction of vowels or syllables are all terms for which the definition may vary between linguistic models, but all are related to prominence. These terms often describe
nearly the same phenomena and therefore may lead to confusion. In this study we restrict ourselves to use two terms: pitch accent and lexical (word) stress.

Pitch is strongly related to $F_0$, referring to the periodicity in a harmonic complex. An increase in $F_0$ correlates with an auditory sensation of a higher pitch. Changes in $F_0$ closely correspond with perceived pitch movements. The intonation contour can be seen as consecutive pitch movements of which some can be associated with pitch accents. These phenomena are described on a more abstract level by using so-called intonation grammars of which two important ones are the IPO intonation grammar and the auto-segmental approach (TOBI). The IPO grammar, deals with subsequent rising and falling pitch movements (t Hart et al., 1990); the auto-segmental grammar (Silverman et al., 1992; Gussenhoven, 1984; Pierrehumbert, 1980) describes pitch movements in a functional and abstract way. (See for more information section 2.1.1.)

In many languages lexical (word) stress is a property of a syllable within the domain of a word and is generally defined in the lexicon for languages such as Dutch, German and English. Lexical (word) stress can be seen as a linguistic phenomenon (this will be described in the next subsection). However, if it concerns the acoustic realization of a word or its percept it is evidently more related to phonetic-acoustic properties of the speech signal.

The usual distinction in lexical (word) stress is between ‘stressed’ and ‘unstressed’ syllables, but a distinction of four degrees (1) ‘primary’, (2) ‘secondary’ (3) ‘tertiary’ and (4) ‘weak’ is also used. In acoustic phonetics lexical (word) stress is in many languages usually due to an increase in intensity of the stressed syllable, but increase in duration and $F_0$ changes may be involved as well (Lehiste, 1970; Fry 1958).

In this study we do not concentrate so much on the actual relationship between prominence on the one hand and lexical (word) stress and pitch accent on the other. We merely want to say that these phenomena are closely related to prominence and that prominence is a complex mix of several phenomena.

1.1.3 Linguistic viewpoint

From a linguistic viewpoint prominence is mainly concerned with focus and lexical (word) stress as a property of the lexicon. From a linguistic viewpoint the function of lexical word stress is probably a more indirect way to compartmentalize the mental lexicon of the listener (Cutler, 1984). If the listener knows the position of the lexically stressed syllable in words this may help to recognize the word more quickly in the appropriate sublexicon. The actual realization of lexical (word) stress in the acoustic signal is related to the prominence of the syllable. By placing
prominence on different syllables, the phoneme string can acquire widely differing meanings. This is shown in the following examples (prominent syllables are indicated by capital letters):

i. \textit{CAmeraatje} (little camera)

ii. \textit{kameRAAdje} (little companion)

iii. \textit{CAnon} (canon)

iv. \textit{KaNON} (cannon)

Only the perceived differences of prominence of the first versus the second or third syllable disambiguate between the different meanings of the two words in the examples above. The first syllable is perceived as more prominent in (i) and (iii) and the third / second syllable is perceived as more prominent in (ii) and (iv). This is related to lexical (word) stress.

We consider an utterance to have an ‘information structure’ that is related to the relative prominence of all the speech elements. Analyzing this information structure is complex and sometimes controversial, however, it is closely related to prominence.

In as far as information structure is concerned it is common to distinguish between ‘given’ and ‘new’ information (e.g. Halliday, 1967). ‘Given’ refers to information already supplied by the previous linguistic context whereas ‘new’ information has not been previously supplied.

In this context, some authors speak of ‘focus’. The speaker can highlight the information for the listener that is at the ‘focus’ of their communicative interest (Baart, 1987; Nooteboom & Kruyt, 1987; Gussenhoven, 1984; Ladd, 1980).

We will describe the following examples in terms of focus. Distinctions such as ‘broad’ and ‘narrow focus’ or ‘contrastive focus’ can be made. We decided to explain in our example sentences only ‘(narrow) focus’ (A) and ‘contrastive focus’ (B). The last example shows a mix of lexical and focal contracts (C), resulting in completely different meanings of the sentences.

\textbf{A)}

i. \textit{IK wil nog twee bloemen}

ii. \textit{Ik wil NOG twee bloemen}

iii. \textit{Ik wil nog TWEE bloemen}

iv. \textit{Ik wil nog twee BLOEMen}

Changes in prominence patterns can guide the attention of the listener to specific words. Different words are prominent in sentences (i) to (iv). A neutral translation is: \textit{I want two more flowers}. If the word \textit{ik} (I) is more prominent, the attention is
INTRODUCTION

guided to this word and for the listener it is clear that it is ‘me’ and not someone else who wants the flowers. In (ii) and (iii) there is a difference of meaning in the sentence; (ii) means I want two additional flowers and (iii) means I want exactly two more flowers, rather than four more. In (iv) the sentence transmits the information that the speaker wants flowers rather than something else.

When contrast is required, the speaker can highlight different parts of a sentence, which makes the contrast more recognizable.

B)  
v. Ik ga niet naar ZAANdam maar naar LEERdam (I do not go to Zaan dam but to Leerdam)
vi. Het is niet DE boek maar HET boek (it is not the (non-neuter) book, but the (neuter) book)

In (B-v) the first syllable of the names of two Dutch cities are put in contrast. Normally the lexical stress of these two names is located on the last syllable, so this is an example were two normally not lexically stressed syllables are more prominent than the lexically stressed one. The last example (vi) shows that the Dutch articles de and het can also be put into contrast and can be more prominent than the other words in the phrase.

C)  
Just as word meaning can change with alterations to syllable prominence, so can the meaning of a sentence alter as prominence is given to differing words.

i. uitsluitend VOOR instappen (only get on at front)
ii. uitsluitend voor INstappen (only for getting on)

iii. naTUURlijk(.) VOORkomen van bosbrand is wenselijk (of course the occurrence of forest fire is desirable)
iv. naTUURlijk(.) voorKOMen van bosbrand is wenselijk (of course the prevention of forest fire is desirable)
v. Natuurlijk VOORkomen van bosbrand is wenselijk (the natural occurrence of forest fire is desirable)
vi. Natuurlijk voorKOMen van bosbrand is wenselijk (the natural prevention of forest fire is desirable)

Examples (C-i) and (C-ii) are written on each door of the Amsterdam busses and streetcars except on the first carriage. The meaning of the whole sentence depends on differences in the allocation of prominence. If the word voor is more prominent than the word instappen, as in (i), incoming passengers should enter the streetcar
only at the front door. In the second reading (ii) *instappen* is more prominent; the meaning then is that this door should only be used to enter and not to exit.

Four different meanings are possible in examples (C iii-vi), because of lexical ambiguity as well as sentence ambiguity. In (iii) and (iv) there is a lexical conflict of different lexical meanings of the word *voorkomen*. Prominence on the first syllable means *occurrence*, but with prominence on the second syllable it means *prevention*. In (iii) and (iv) versus (v) and (vi) the difference in prominence of the first and the second word disambiguates the meaning of the sentence. *Natuurlijk* in (iii) and (iv) means *of course* and in (v) and (vi) *natural*. Thus four different meanings are possible, (iii) means *of course the occurrence of forest fire is desirable* and (iv) means *of course prevention of forest fire is desirable*. A different phrasal meaning is given, for (v) and (vi), (v) means *natural occurrence of forest fire is desirable* and a prominence pattern as in (vi) changes the meaning to *natural prevention of forest fire is desirable*.

The main topic in this study is not to describe the relationship between prominence and either focus, or given and new information (see for that for instance van Donzel, 1995). It also does not study prominence in terms of pitch accent, using the IPO intonation grammar (‘t Hart et al., 1990), or the auto-segmental theory (Gussenhoven, 1984; Pierrehumbert, 1980). We concentrate on prominence as such. These and other topics will be discussed and compared with the literature in the introductions to the appropriate chapters. Prominence gives access to the information structure and is closely related to concepts of pitch accent and lexical (word) stress.

1.2 Topic of investigation

The main topic in this study will be prominence itself. Prominence refers to the degree in which a phoneme, syllable, and / or a word is perceived to stand out from its environment. Prominence is therefore primarily a perceptual concept. On the one hand the listener uses variations in length, loudness and pitch as cues (bottom-up information) in order to signal relative prominence of a unit (Terken, 1996; Hermes & Rump, 1994; Rietveld, 1983; Lehiste, 1970). On the other hand the listener’s perception is biased by expectations. These expectations are based on knowledge of the language (top-down information). Examples of linguistic knowledge are the syntax of a language, the differences between content word and function word, Part-of-Speech information in general, position of a word in a sentence, and word frequency (Altenberg, 1987; Baart, 1987; Chomsky & Halle, 1968). The relationships are visualized in figure 1.1.

Our research questions fold into three parts:
1) How to find an operational definition of prominence?
2) What are the linguistic determinants / correlates of prominence (top-down)?
3) Which acoustic correlates contributes to the perception of prominence (bottom-up)?

The first question concerns the perceptual notion of prominence and how it should be defined. An operational definition is needed in order to label a database in terms of prominence. General questions arise here. Is it necessary to have experts listeners label prominence, or can naive native listeners do this labeling as well? On which unit (segment, syllable, word) should the judgments be given and should one use one or more subjects for this labeling task? Prominence is a relative and gradient phenomenon; should labeling thus be multi-valued or binary? If multi-valued what should be the range of the scale; a 10-point scale or a 4-point scale? Once labeled, what are the consistency and the reliability of the labelers? All these questions are discussed in chapter 2.

The second research question deals with the linguistic determinants / correlates of prominence. Lexical and syntactic features such as Part-of-Speech, word length and

Figure 1.1: Relation of prominence and linguistic and acoustic correlates.
position of a word in the sentence are related to pitch accents (Hirschberg, 1993; Baart, 1987) and therefore to perceived prominence. Do these features correlate with perceived prominence and if so, how? To what extent do these features contribute to the prediction of prominence purely on such linguistic information? Chapter 3 provides more information on this topic.

The third research topic concerns the acoustic correlates of perceived prominence. What are the acoustic correlates of prominence? From literature it is known that F₀ changes, and duration of vowels and / or syllables are related to lexical (word) stress and pitch accent. What is the relationship of these acoustic correlates to prominence, and to what extent are combinations of these features correlates of prominence? Can the prominence labels from listeners be ‘predicted’ by a classifier that is only using the acoustic signal as input? What is the contribution of a selected set of acoustic features to prominence classification? And to what extent do certain normalizations, for instance intrinsic vowel duration and speaking rate, contribute to a better prominence prediction. Chapter 4 and chapter 5 discusses these questions.

1.3 Usefulness of acoustic and lexical / syntactic correlates of prominence

From a general viewpoint the perceptual phenomenon prominence seems to have an important communicative function. Therefore it is interesting from a phonetic viewpoint to investigate prominence itself as a perceptual phenomenon and correlates of prominence. The following questions form an interesting topic. To what extent is prominence marking by naive listeners useful for our research? What is the consistency and reliability of listeners? Which acoustic correlates and linguistic determinants contribute to the perception of prominence and what is their effect if they are used to predict prominence solely from acoustic input features on the one hand and from linguistic correlates on the other?

Apart from dealing with the communication process between speaker and listener this study is also concerned with speech technology. This introduces specific limitations to the investigation method. For example, the acoustic features on which the automatic classification of prominence will be based must be derived from the speech signal in an automatic way without additional human (knowledge-based) correction or intervention.

Three applications for speech technology are briefly introduced here. The first application concerns prominence prediction for a Text-to-Speech system to increase the naturalness and intelligibility of synthetic speech. Most speech synthesis systems today use the notion of pitch accent and lexical (word) stress without using the degree of prominence. Predicting prominence for speech synthesis purposes has to be solely based on textual input. The prediction of prominence is not a unambiguous process and the location and degree of prominence is not explainable from textual information only. Reminiscent to Bolinger’s remark about humans not
being a ‘mind-reader’ (Bolinger, 1972), a computer is certainly not. This makes that most of pragmatic and semantic information can not play a role and will not be used in the experiments of the present study.

The second and third application are in the field of automatic speech recognition: a prominence indicator and an algorithm for disambiguating the meaning of sentences. Prominence can guide or alter the meaning of the sentence, as, for instance, illustrated in the example given before *Ik wil nog twee bloemen*. A prominence indicator or classifier can provide useful information for the speech recognition process, more specifically during the word search process. Knowing the degree of prominence can help to decide whether a word is important for communication. Prominence indication and sentence disambiguation can be based on acoustic input, as well as on information coded in the lexicon.

1.4 Research method

The approach we choose to investigate the perceptual notion of prominence and the related acoustic and linguistic correlates, imposes restriction on our research methods as well as on the choice of speech material to be investigated.
First, the speech material should constitute a sufficiently large speech corpus that is valid for speech technological applications.
Second, the prominence labeling should still be possible for such a large corpus, while leaving us with as much detailed labeling as possible. With respect to this constraint, the use of naive listeners is of special interest.
Third, feature extraction and prediction should be done automatically. However, we still want to control the feature extraction and want to describe and to analyze the features in an interpretable way. The prediction of prominence should also be controllable. Once knowing the ‘optimal’ features, rules should be formulated and used for prominence prediction, either on acoustic or linguistic input. The analysis of the linguistic and acoustic features should not be based on a purely statistical and / or brute force approach, because we want to control the individual features and their contribution to the prediction of prominence. This limits the prediction tools to simple ones. The prominence prediction should be tested on an independent test set in order to get an idea of the consistency of the prediction and the generalization capability of the prominence classifier.

The appropriate literature will be discussed in detail in the separate chapters.

1.5 Outline of this study

This study deals with the question which features in the speech signal and in the text (acoustic, linguistic) can be used to predict prominence automatically. To this end a
test and a training corpus were designed for classification. A group of listeners
judged the speech material used and marked all the sentences for prominence
(chapter 2). In chapter 3 the lexical / syntactic correlates are analyzed and are used
to predict prominence by using textual features. This mainly concerns correlates /
determinants such as word class, and the position of words in the sentence. Chapter
4 focuses on the acoustic correlates that are used to predict prominence. In contrast
to chapter 3, the emphasis in chapter 4 is on the acoustic features that can be
automatically derived from the speech signal. In addition, this study will attempt to
discover which acoustic features can be used for the automatic classification and
prediction of prominence (chapter 5). A general discussion as well as conclusions
and ideas for further research will be presented in the last chapter.