AUTOMATIC RECOGNITION OF FOCUS ACCENT IN GERMAN

Robert Bannert
Department of Linguistics and Phonetics
Lund University

Speech recognition systems of the future, to a growing extent, have to be able to process information of various kinds. In my paper I would like to present a new project[1] where an attempt is made to integrate the interplay of semantic and tonal information in a dialogue system. Here I will confine myself to the tonal component.

When a speaker and a hearer communicate by means of a dialogue, it is done in a given world and under certain given linguistic constraints. Both partners of the dialogue react according to certain rules of the game. Therefore it is assumed that, among other things, the semantic focus in a response can be predicted as a consequence of semantic and logic rules of the dialogue. Semantic focus is mainly signalled phonetically by a tonal peak, in German associated with the pitch accent of a word (or syllable).

From a semantic point of view, a sentence may have one focus or several foci and the domain of focus may vary, it may be wide or narrow. The interplay between the semantic aspect of focus and its tonal manifestion in German are not quite understood yet. However, it seems to be totally clear that narrow focus is to be found not only in normal, "neutral" intonation, - the one that is used when plain facts are communicated -, but also in contrast and emphasis. Each sentence contains one or more pitch accents that reflect semantic focus. These accents mark those words of the sentence that are most important semantically and I shall call them focus accents.

In our dialogue system to be modelled, the recognition of focus accent is one component out of three. It is the link between spectral speech recognition and the semantic rules of the dialogue game. The Fo-values of the sentences of the dialogue are processed using an algorithm in order to find the focus accent or accents. The sentences used in the dialogue show a prosodic structure

[1]The project is called "Modellbildung für die Auswertung der Fokusintonation im gesprochenen Dialog" and is carried out at the Research Group for Artificial Intelligence and Speech Signal Processing at the Department for Advanced Information and Communication Technologies, Fraunhofer Institute for Industrial Engineering (IAO) in Stuttgart.

corresponding to a prosodic utterance which may consist of one or several prosodic phrases. To start with, the semantic component will use simple statement sentences and their negations like "Johannes liebt Susanne (nicht)" (John loves Susanne/John does not love Susanne). The sentences of the dialogue are pronounced by a speaker, recorded, sampled and analysed for Fo.

The algorithm for recognizing focus accent consists of two steps. First all those changes of Fo are to be found that correspond to pitch accents. Fo-differences spanning a certain period of time are picked after having taken into consideration the total intonation of the whole prosodic utterance. In the second step, all the pitch accents that were picked are evaluated in order to find the focus accent or accents. At this level, the pitch accents are compared with each other using criterea in the total Fo-course of the prosodic utterance that lie in the time and Fo-dimension as well. To start with, the system will be constrained to speakers whose basic characteristics of their intonation are known. Later on these basic tonal facts will be established automatically by speaking some appropriate model sentences.

The projection of focus accent onto the time axis, i.e. its temporal position in the utterance, will be marked in an adequate way. This information about the temporal projection of focus accent or accents is then passed on to the dialogue system via the spectral speech recognition component which is not part and parcel of the project but is available in our department. The information about which word is in semantic focus acts as feed back and verification for the further semantic development of the dialogue.

References

Bannert, R. 1083. Some phonetic characteristics of a model for German prosody. Lund Univerity, Department of Linguistics and Phonetics, Working Papers 25, 1-34.

Bannert, R. 1985a. Towards a model for German prosody. Folia Linguistics XIX, 321-341.

Bannert, R. 1985b. Fokus, Kontrast und Phrasenintonation im Deutschen. Zeitschrift für Dialektologie und Linguistik 52, 289-305.

Lea, W.A., Medress, M. F. and Skinner T. E. 1975. A prosodically guided speech understanding strategy. IEEE Transactions on Acoustics, Speech, and Signal Processing. Vol. ASSP-23, 30-38.

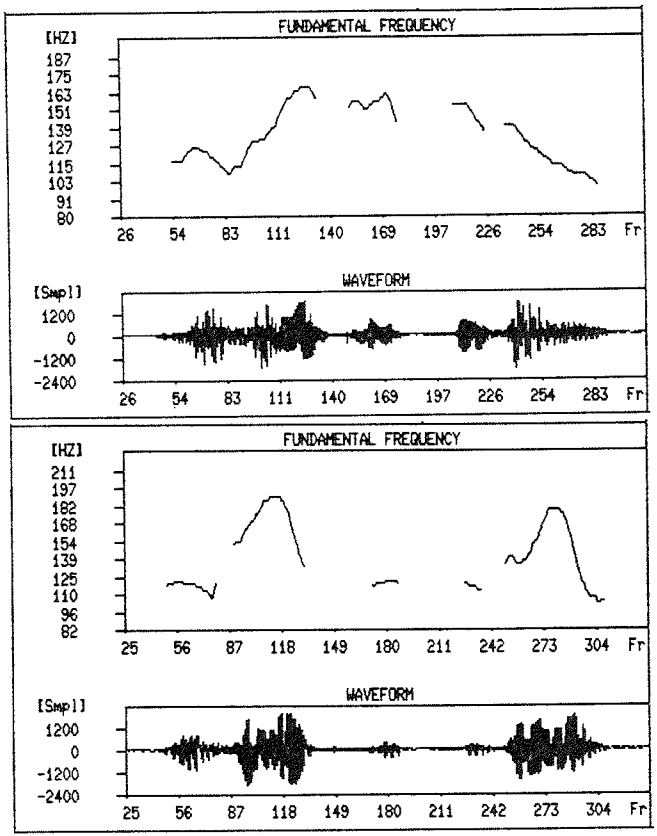

Fig. 1: Fo-curves of the utterance "Johannes liebt Susanne " containing two equal pitch accents. Hat pattern (Brücke) above, peak accents below.

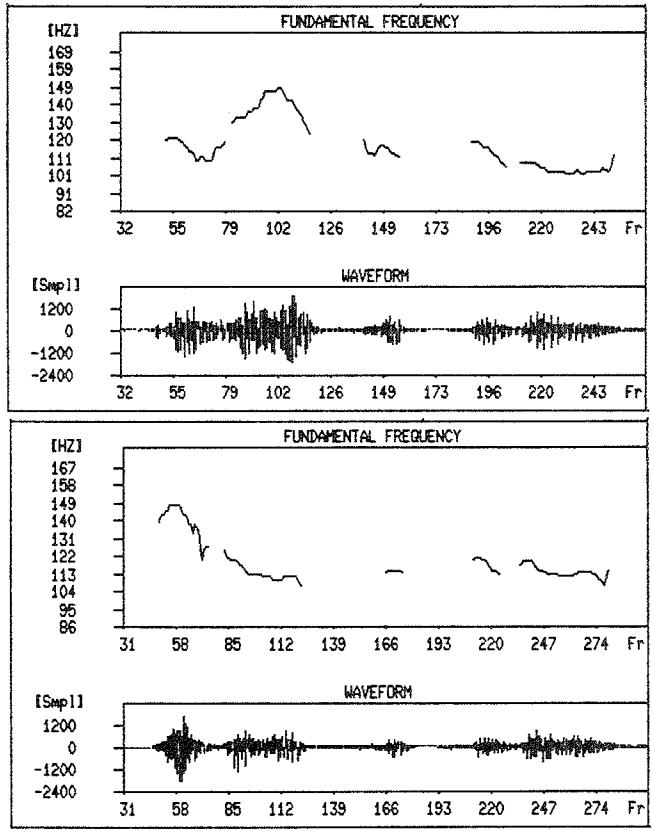

Fig. 2: Fo-curves of the utterance "Johannes liebt Susanne" above containing one accent (on Johannes). Peak accent above, low accent below.

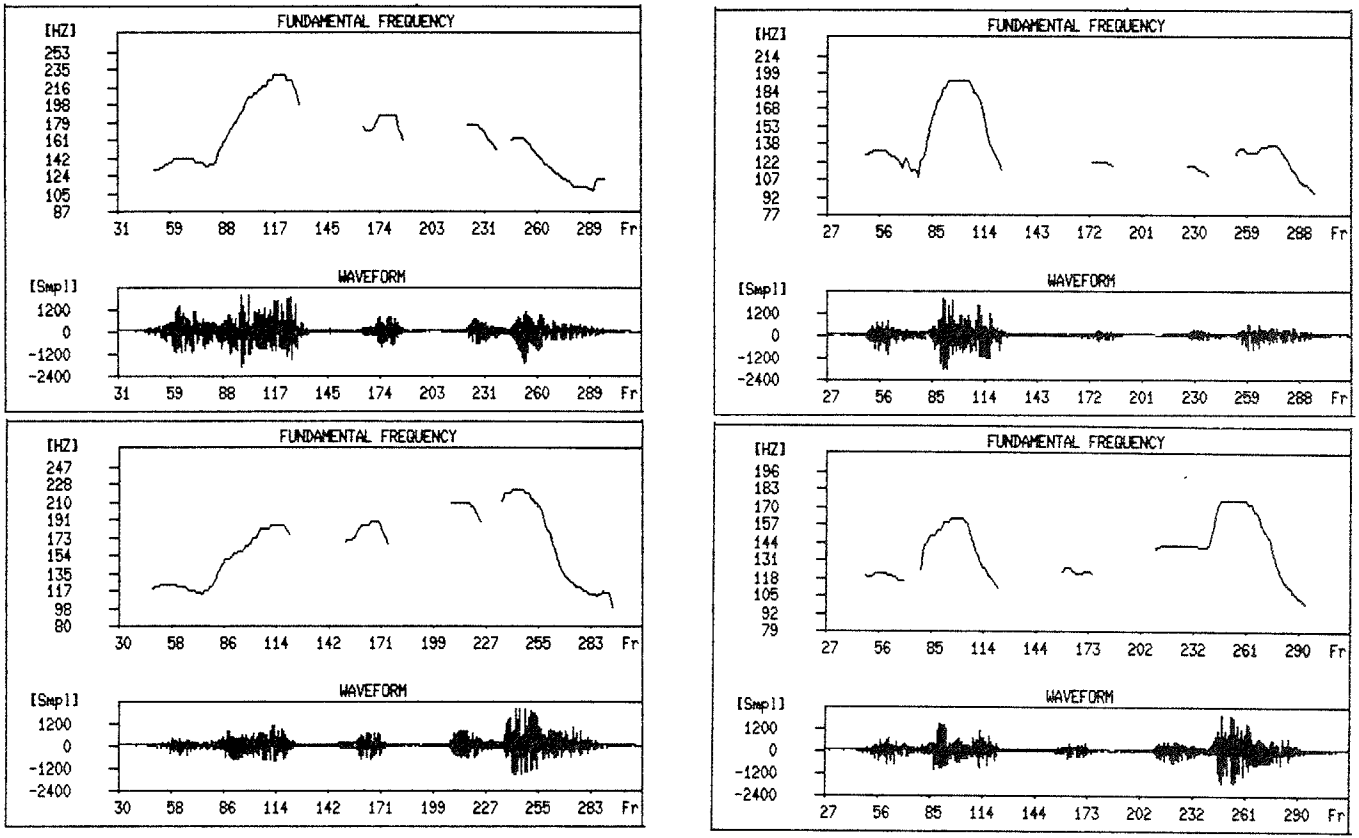

Fig. 3: Fo-curves of the utterance "Johannes liebt Susanne" with two unequal pitch accents involving contrast of emphasis on the strongest one (above on Johannes, beolw on Susanne). Left: hat pattern (Brücke), right: peaks.