ARIPUC 22, 1988, p. 139-143

A DANISH TEXT-TO-SPEECH SYSTEM USING A TEXT NORMALIZER BASED ON MORPH ANALYSIS*

BJÖRN GRANSTRÖM**, PETER MOLBÆK HANSEN, AND NINA GRØNNUM THORSEN

A Nordic cooperative project has been started to develop a text-to-speech device for the Nordic languages. The development is based on the system originally created in Stockholm. Language specific features have necessitated modifications of the original structure. For Danish, this primarily involves the inclusion of a morph based "text normalizing component". This paper presents the construction and function of the system and also discusses some preliminary use of the device.

I. INTRODUCTION

Speech synthesis has been a major line of research in our two departments for several decades. In Sweden, this effort has resulted in a multi-lingual text-to-speech system (ref. 1), commercially available through Infovox AB. A joint effort within the project "A Nordic text-to-speech system", financed by The Nordic Committee on Disability, is aimed at making this device available to the handicapped in the Nordic countries.

Although the Nordic languages are mutually intelligible, Danish poses some special problems for a text-to-speech system because the relation between the standard orthography and pronunciation is rather complicated. To tackle this we have included a unique

---

*) Contribution for European Conference on Speech Technology, Edinburgh, September 1987 (edited by J. Laver & M.A. Jack), vol. I, 21-24

**) Department of Speech Communication and Music Acoustics, Royal Institute of Technology (KTH), Stockholm.

component in the system that transforms words into an idealized normalized orthography. This is accomplished through a morphological analysis based on a set of moderately large morph lexica. With a limited set of rules the result is transformed to a phonetic transcription, including stress.

In a phonetic rules component, special care has been taken to realize the prosodic structure of Danish which differs considerably from standard Swedish or Norwegian. There are also many other differences in structure such as the ample use of "stød", a kind of creaky voice unknown in the other Nordic languages but which corresponds roughly to the tonal word accent I in Swedish and Norwegian.

II. MORPHOLOGICAL ANALYSIS AND NORMALIZATION

The inclusion of a normalization component (NC) is a deviation from the general philosophy of the KTH system which is rule based. The work done by the NC corresponds with the work performed by certain early rules in the rule components for the other languages. There was a practical and a theoretical motivation behind the establishment at the Institute of Phonetics, University of Copenhagen (IPUC) of a lexicon based NC for Danish. The practical reason was that Danish orthography corresponds very badly with pronunciation and that the number of rules in a rule based conversion system, and also the degree of arbitrariness of most of such rules, would have been prohibitive. For a more detailed description of the peculiarities of Danish orthography, see (ref. 2). The theoretical reason was that the use of a morph lexicon should reflect more closely the human process of reading, since we generally know the morphemes of our native language and only feel the necessity of relying on some sort of mental rule system when occasionally we come across a hitherto unknown word in a text. The NC does two things: it supplies the correct morphological boundaries, and it normalizes the spelling of individual morphemes. A few examples illustrate this: A word like *dal* ('valley') is pronounced with a long vowel, whereas *tal* ('number') is pronounced with a short vowel. The NC will identify these words correctly and output them as DAL and TALH, respectively, thus assigning to them a notation which is consistent - but not identical - with a phonetic transcription. The symbol H represents an abstract consonant phoneme which will prevent the vowel from being lengthened and the consonant L from receiving the stød. A word like *kvindeemancipationen* ('the emancipation of women') represents a more complex case: It is analysed by the NC as consisting of the morpheme sequence KVIND (a native root), + E (a native suffix) + E (a latin prefix) + MAN + CIP (latin roots) + AT + ION (latin suffixes) + EN (a native ending), and it is output as KVIND#mEO#iE#fMANCIPATION#pEON. The symbol sequences #m, #i, #f, and #p represent various morphological boundaries with different phonological effects. No boundaries are inserted between MAN and CIP nor between AT and ION because such boundaries would not supply any information

relevant to pronunciation. The symbol sequence E0 represents the vowel schwa. A word like *hund* ('dog') is output as HU6ND. The sequence U6 represents a particular, abstract phoneme which resists an otherwise general phonological rule of Danish which lowers high vowels before homosyllabic nasals.

Thus, the inventory of distinct symbols which may be output from the NC is considerably larger than both the number of letters in the alphabet and the number of phonemes needed to represent Danish speech. The problems of integrating the NC as such into the system have been few and small, since its output, i.e. the input to the rule system, is of the same type as orthographic input, namely a sequence of ASCII characters.

III. THE PHONOLOGICAL RULE COMPONENT

The rule language developed at IPUC is of the same SPE type as the one used in the KTH system, and the rule component of the IPUC system could therefore be translated into the notation of the present system. However, owing to certain technical differences between the input scanning routines and feature interpreting procedures of the two systems, this translation could not be done in a simple rule-by-rule fashion. One difficulty arises from the fact that in the IPUC system a segment is identifiable exclusively by its feature composition, whereas in the KTH system a segment is identifiable by its symbolic representation. Another main difference lies in the way a string of segments is scanned by a rule. In the KTH system the context is matched left to right, whereas the IPUC system starts the match at the structure to be changed and then matches the left and right context. This has necessitated the reformulation of certain IPUC rules.

IV. THE PHONETICS OF STANDARD DANISH

The vowel system is rich, with ten vowel phonemes /i e ɛ a y ø œ u o ɔ/ which may be either short or long (the difference is phonological). Generally, the long and short vowels have identical phonetic quality, except for /a(:), o(:), ɔ(:)/, and due to a language specific variation in some of the vowels with the phonetic context, a total of 18 distinct phonetic vowel qualities must be distinguished: [i e ɛ æ a a- ɑ y ø œ ɶ Œ u o ɔ ɒ ʌ] plus [ə].

The consonant system is correspondingly restricted: /p t k b d g f s h v ð j m n ŋ l r/. It differs from the consonantal systems of the closest germanic neighbours in the realization of, particularly, the stop series. Firstly, an opposition between /p t k/ versus /b d g/ is found only initially in syllables containing a full vowel (i.e. not /ə/). Secondly, the manifestation of the contrast is one of aspiration only. I.e. /p t k/ are unvoiced aspirated, and /b d g/ are unvoiced unaspirated. Both series are lenes, rather than fortes.

Stress is free and phonemic on the surface, although stress placement can to a large extent be predicted from the syllabic structure and the morphology, e.g. *billigst*, *bilist* ['bilisd, bi'lisd] ('cheapest, motorist'). Phonetically, stress is signalled mainly through fundamental frequency variation, but (full) vowel quality and (longer) duration also contribute to the identification of stressed syllables.

A complication in the phonology and phonetics is the Danish "stød", a kind of creaky voice whose occurrence, like stress, is to a large extent predictable, but on the surface, stød versus non-stød distinguishes words, like *Møller*, *møller* ['møl'ʌ, 'mølʌ] (a proper name, miller).

Standard Danish intonation can be decomposed into (and synthesized from) the following components: a text contour, an utterance or sentence contour, which may be decomposed into a succession of two or more phrase contours, a stress group pattern, and a stød movement. These are all speaker controlled. Involuntary variations arise due to intrinsic properties of the segments. These components are hierachically organized so that components of smaller temporal scope are superposed upon and subordinate to components of larger temporal scope. Sentence intonation is signalled by the overall global course of the intonation contour, rather than by a special local (final) movement. Standard Danish lacks an obligatory sentence accent, which makes it prosodically rather simpler than, e.g., Swedish and English, see further (ref. 3). The synthesis incorporates the utterance, stress group and stød components, and further adds certain segment conditioned fundamental frequency variations.

Below is an illustration of the parameters and the acoustic output of the utterance *Hvem har en søster der hedder Kamma?* ['vɛm' 'hɑ:' en 'søsdʌ dɑ heðʌ 'kɑma] ('Whose sister is called Kamma?'). The corresponding transcription in the text-to-speech system looks like this: [V'ÄMQ H'A3:Q EN S'ÖSDÅ2 DA3 HEDHA2 K'A3MA1].

V. CONCLUDING REMARKS

At this stage in the project we have made no formal evaluation of the system. Preliminary versions of the program have been used in a project aiming at a Danish workstation for visually impaired persons. Several imperfections still exist in both the text analysing part and the phonetic realization part of the system. However, the output is unmistakably Danish and judged useful in a variety of applications. One improvement that is still needed is a somewhat faster and more reliable NC. It is a problem with the current implementation of that component that it makes quite a few wrong choices in cases of ambiguous input. We are at present developing a new version with a better performance.

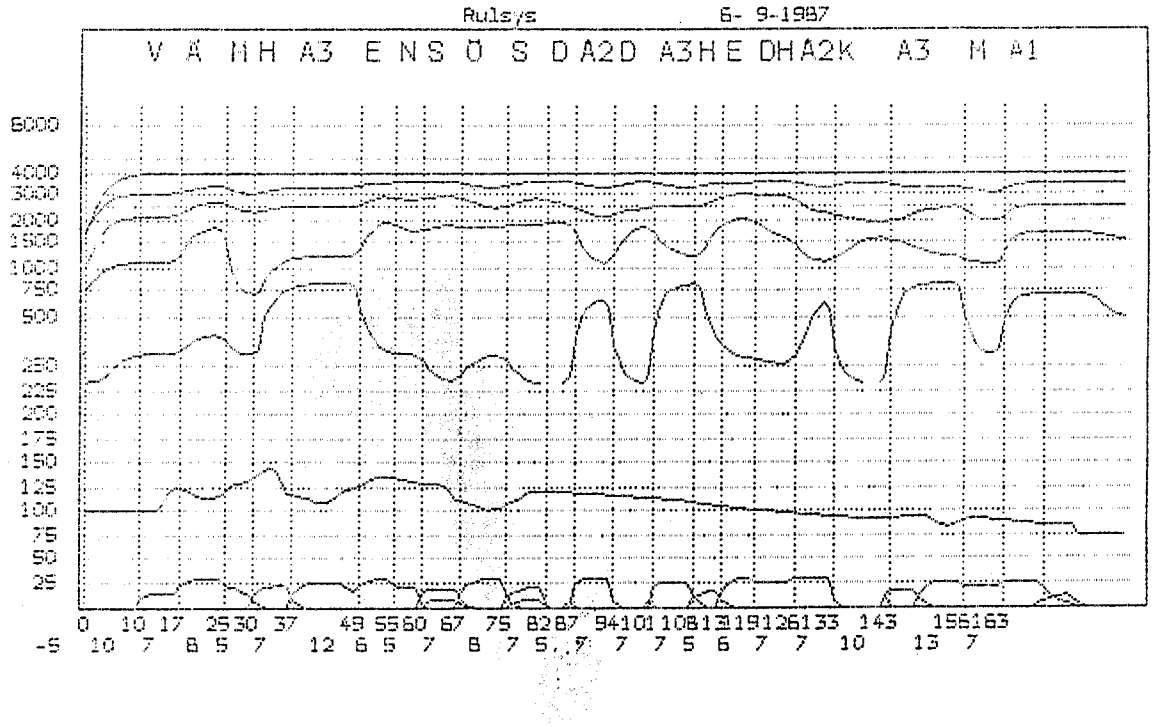

Figure 1

The Danish utterance 'Hvem har en søster der hedder Kamma?' as produced by the text-to-speech system. The parameters are, from the bottom up: Diverse amplitude parameters, fundamental frequency, and the four lowest formants.

REFERENCES

Carlson, R. and Granström, B. 1986: "Linguistic processing in the KTH multilingual Text-to-Speech system", *Conference Record, IEEE-ICASSP, Tokyo*

Molbæk Hansen, P. 1983: "An orthographic normalizing program for Danish", *Ann. Rep. Inst. Phon., Univ. Copenhagen 17*, p. 87-109

Thorsen, N. 1983: "Standard Danish sentence intonation - phonetic data and their representation", *Folia Linguistica 17*, p. 187-220.