Working Papers 25 1983
Linguistics-Phonetics
Lund University, 179-194

HOW TO MAKE A TEXT PRODUCTION SYSTEM SPEAK

Bengt Sigurd

Several text production models implemented on computers are able to print grammatical sentences and coherent text, see e.g. Mann & Matthiessen (1983), Allén (1982), Sigurd (1983). It is an interesting task to make such a text production system speak, but few researchers have ventured to solve the additional problems caused by a verbalization system which is intended to speak in a natural way.

Trying to simulate human speech behaviour by computer throws light on the whole production model, particularly if the system is required to have some psychological reality by simulating spontaneous speech with its hesitation pauses, repetitions and errors.

This paper will show how even a simple synthesizer, Votrax, interfaced with the system Commentator, see Sigurd (1983), Fornell (1983) raises interesting questions about the phonetic representation of words, ways of packing sounds into prosodic units, pauses, mistakes in the execution processes and the place of sound laws in a model of a human speaker. The Votrax can pronounce some 60 American English sounds and offers 4 pitch levels; it cannot, of course, be expected to produce high quality Swedish. It is used here for explorative and experimental purposes as is the whole Commentator system.

The problem of producing naturally sounding speech is also attested in experiments with text-to-speech systems, see Carlson & Granström (1978). Such systems, however, have printed text as input which creates additional problems. They have to be able to derive a phonetic transcription from the printed version, which e.g. means pronouncing th, ng and sh, as single sounds, deriving stress in words such as export, which is very difficult and in fact requires comprehension of the text.

COMMENTATOR'S INPUT TO A SPEECH PRODUCTION SYSTEM

The general outline of the Commentator is presented in fig. 1 after Sigurd (1983). The input to this model of verbal production is perceptual data or equivalent coordinate values, e.g.information about persons and objects on a screen. These primary perceptual facts constitute the basis for various calculations in order to derive secondary facts and conclusions about movements and relations such as distances, directions, right/left, over/under, front/back, closeness, goals and intentions of the persons involved. The Commentator produces comments consisting of grammatical sentences making up coherent and well-formed (although often boring) text. Some typical comments are shown below.

A question menu, different for different situations, suggests topics leading to propositions which are considered appropriate under the circumstances and their truth values are tested against the primary and secondary facts of the world known to the system. If a proposition is found to be true, it is accepted as a protosentence and verbalized by various lexical, referential, syntactic and textual subroutines. If e.g. the proposition CLOSE (ADAM, EVE) is verified after measuring the distance between the two referents Adam and Eve and comparing it with the standard for closeness between human beings, the lexical subroutines try to find out how closeness should be expressed in the language (Commentator has mainly produced Swedish), referential subroutines determine whether pronouns could be used instead of the proper names of the persons and textual procedures investigate whether connectives such as <u>however</u>, <u>also</u> or perhaps contrastive stress should be inserted. In the printing version of Commentator all the words in the sentence are packed into a string before they are printed.

The Commentator can deliver words one at a time whose meaning, syntactic and textual functions are well-defined through the verbalization processes. For the printing version of Commentator these words are characterized by whatever markers are needed to get correct printing (spelling, word spaces, punctuation marks). The input to a speaking version of Commentator must be a phonetic transcription, but the details of this and whatever additional information is required in order to produce prosodic units with the proper intonation and stress patterns has to be discovered by empirical and experimental studies and work with models such as the one presented in this paper.

Lines	Component	Task	Result (sample)
10-35	Primary information	Get values of primary dimensions	Localization coordinates
100-140	Secondary information	Derive values of complex dimensions	Distances, right-left, under-over
152-183	Focus and topic planning expert	Determine objects in focus (referents) and topics according to menu	Choice of subject, object and instructions to test abstract predicates with these
210-232	Verification expert	Test whether the conditions for the use of the abstract predicates are met in the situation (on the screen)	Positive or negative protosentences and instructions for how to proceed
500	Sentence structure (syntax) expert	Order the abstract sentence constituents (subject, predicate, object); basic prosody	Sentence structure with further instructions
600-800	Reference expert (subroutine)	Determine whether pronouns, proper nouns, or other expressions could be used	Pronouns, proper nouns, indefinite or definite NPs
700-	Lexical expert (dictionary)	Translate (substitute) abstract predicates, etc.	Surface phrases, words
900	Sentence connection (textual) expert	Insert conjunctions, connective adverbs; prosodic features	Sentences with words such as <u>också</u> (too), <u>dock</u> (however)
1000	Phonological (pronunciation, printing) expert	Pronounce or print the assembled structure	Uttered or printed sentence (text)

Fig 1 Components of the text production model underlying Commentator

A SIMPLE SPEECH SYNTHESIS DEVICE

The fundamental difference between a printing and a speaking verbal production model must, of course, be that the words of the speaking version have to be coded in some phonetic transcription, which serves as instructions for a speech synthesis device simulating human articulation. The transcription depends on the speech synthesis system to be used.

The experimental system presented in this paper uses a Votrax speech synthesis unit (for a presentation see Giarcia, 1982). Although it is a very simple system designed to enable computers to deliver spoken output such as numbers, short instructions etc, it has some experimental potentials. It forces the researcher to take a stand on a number of interesting issues and make theories about speech production more concrete. The Votrax is an inexpensive and unsophisticated synthesis device and it is not our intention to achieve perfect pronunciation using this circuit, of course. The circuit, rather, provides a simple way of doing research in the field of speech production.

Votrax simulates the human vocal apparatus by means of a harmonic source (for vowels and voiced consonants) and a noise source (for voiceless consonants), supplemented with filters. Information about the sounds to be produced is given according to the LPC technique. Votrax (which is in fact based on a circuit named SC-01 sold under several trade names) offers a choice of some 60 (American) English sounds (allophones) and 4 pitch levels. A sound must be transcribed by its numerical code and a pitch level, represented by one of the figures 0,1,2,3. The pitch figures correspond roughly to the male levels 65,90,110,130 Hz. Votrax offers no way of changing the amplitude or the duration, but choice of pitch level as well as choice between long/short and stressed/unstressed sounds (allophones) can be made with some success.

Votrax is designed for (American) English and if used to synthesize other languages it will, of course, add an English flavour. It can, however, be used at least to produce intelligible words for several other languages. Of course, some sounds may be lacking, e.g. Swedish <u>u</u> and <u>y</u> and some sounds may be slightly different, as e.g. Swedish <u>sh</u>-, <u>ch</u>-, <u>r</u>-, and <u>l</u>-sounds.

Long, short and stressed, unstressed variants are found for some vowels, and if used with some ingenuity the inventory of sounds may serve fairly well for several languages provided the foreign accent and the robot voice are accepted.

English model words containing the sounds and information about the duration of the sounds are given in the manual, but this information is certainly not sufficient for the phonetician. The synthesis device seems to have been constructed for technical purposes without consulting a phonetician or linguist and the presentation of the device is very unsystematic from the point of view of a language professional.

Most Swedish words can be pronounced intelligibly by the Votrax. The pitch levels have been found to be sufficient for the production of the Swedish word tones: accent 1 (acute) as in and-en (the duck) and accent 2 (grave) as in ande-n (the spirit). (For details about the word accents see Gårding, 1977). Accent 1 can be rendered by the pitch sequence 20 and accent 2 by the sequence 22 on the stressed syllable (the beginning) of the words. Stressed syllables have to include at least one 2. Word tones and stress are necessary characteristics of Swedish words and must be given in the lexicon.

Words are transcribed in the Votrax alphabet by series of numbers for the sounds and their pitch levels. The Swedish word höger (right) may be given by the series 27,2,58,0,28,0,35,0, 43,1, where 27,58,28,35,43 are the sounds corresponding to h,ö:,g,e,r, respectively and the figures 2,0, etc after each sound are the pitch levels of each sound. The word höger sounds American because of the ö, which sounds like the (retroflex) vowel in bird, but it is assigned the proper accent 1 by the sequence 20. The 1 on one of the following unstressed syllables is introduced in order to get some variation. No detailed systematic studies of the effects of various combinations of sounds and pitches have been undertaken and the Votrax has not been evaluated properly by phoneticians to my knowledge.

The pronunciation (execution) of the words is handled by instructions in a computer program, which transmits the information to the sound generators and the filters. Some programming details will be given below, but the technical details are outside the scope of this paper.

PAUSES AND PROSODIC UNITS IN SPEECH

The spoken text produced by human beings is normally divided by pauses into parts of several words (prosodic units). There is no generally accepted theory explaining the location and duration of the pauses and the intonation and stress patterns in the prosodic units. Many observations have, however, been made, see Dechert & Raupach (1980).

The printing version of Commentator collects all letters and spaces into a string before they are printed. A speaking version trying to simulate at least some of the production processes cannot, of course, produce words one at a time with pauses corresponding to the word spaces, nor produce all the words of a sentence as one prosodic unit. A speaking version intended to have some psychological reality must be designed to be able to produce prosodic units including 3-5 words, cf Svartvik (1982) and lasting 1-2 seconds, see Jönsson, Mandersson & Sigurd (1983). How this should be achieved may be a called <u>chunking problem</u>. It has been noted that the chunks of spontaneous speech are generally shorter than in text read aloud.

The text chunks have internal intonation and stress patterns often described as superimposed on the words. That is why they may be called prosodic units. Deriving these internal prosodic patterns may be called the <u>intra-chunk problem</u>. We may also talk about the inter-chunk problem having to do with the relations e.g. in pitch, between succesive chunks.

For the purpose of simulating at least some of these features by Votrax we will touch upon these problems and discuss different ways to control chunking and pause placement.

As human beings need to breathe they have to pause in order to inhale at certain intervals. The need for air is generally satisfied without conscious actions. We estimate that chunks of 1-2 seconds and inhalation pauses of about 0.5 seconds allow convenient breathing. Clearly, breathing allows great variation. Everybody has met persons who try to extend the speech chunks and minimize the pauses in order to say as much as possible, or to hold the floor.

It has also been observed that pauses often occur where there is a major syntactic break (corresponding to a deep cut in the syntactic tree), and that, except for so called hesitation pauses, pauses rarely occur between two words which belong closely together (corresponding to a shallow cut in the syntactic tree). There is no support for a simple theory that pauses are introduced between the main constituents of the sentence and that their duration is a function of the depth of the cuts in the syntactic tree. The conclusion to draw seems rather to be that chunk cuts are avoided between words which belong closely together. Syntactic structure does not govern chunking, but puts constraints on it. Click experiments which show that the click is erroneously located at major syntactic cuts rather than between words which are syntacticly coherent seem to point in the same direction. As an illustation of syntactic closeness we will mention the combination of a verb and a following reflexive pronoun as in Adam avlägsnar+sig från Eva. ("Adam distances himself from Eva"), one of the sentences which Commentator often produces. Cutting between avlägsnar and sig would be most unnatural. All other places seem acceptable, although for different reasons. A cut before avlägsnar or Eva would seem to reflect the search for the proper word.

Lexical search, syntactic and textual planning are often mentioned as the reasons for pauses, so called hesitation pauses, filled or unfilled. In the speech production model envisaged in this paper sounds are generally stored in a buffer where they are given the proper intonational contours and stress patterns. The pronunciation is therefore generally delayed to allow context adjustments and various prosodic operations. This delay also offers ways of explaining speech errors. The length of the delay varies.

Hesitation pauses seem, however, to be direct (on-line) reflexes of searching or planning processes and at such moments there is no delay. Whatever has been accumulated in the articulation or execution buffer is pronounced and the system is waiting for the next word. While waiting (idling) some human beings are silent, others prolong the last sounds of the previous word or produce sounds, such as ah, eh, or repeat part of the previous utterence. Hesitation pauses may occur anywhere, but as has been observed they seem to be more frequent before lexical words than function words.

As to the internal prosodic patterns within chunks, it has been observed that sentence final sounds often are prolonged and that the final pitch is low in declarative sentences. Non-sentence final chunks may often also have final lengthening and rising pitch signaling incompleteness. As the Votrax only allows the manipulation of pitch, not duration, we will not go into more detail here (for details of the Nordic Languages see Gårding, Bruce & Bannert, 1978).

In the approach to be presented in detail below we will demonstrate a two buffer model and a chunking mechanism based on the length of the chunk and syntactic structure. It allows natural breathing and avoids unnatural cuts. After the presentation of this approach we will discuss variants and models where other factors may influence the final chunking and pausing in speech.

A TWO BUFFER MODEL OF SPEECH PRODUCTION

The approach proposed in this paper presupposes two buffer memories and trigger mechanisms for filling, emptying and reading these buffers. The buffers work in series and are called the current buffer and the execution buffer. The operation of such a system will be illustrated by using the length of the chunck required as the primary trigger. Words are assumed to be classified as stressed or unstressed and by this difference the text will be divided into chunks generally containing one or several stressed syllables and a greater number of unstressed syllables. These chunks are then given a rising or falling final pitch contour. A computer implementation of this text division method will be discussed in some detail below.

As only pitch, not intensity, is available in Votrax, pitch must be used to signal stress. Unstressed words are assigned pitch level 1 or lower, stressed words get 2 or higher on at least one segment. Words are assumed to be inherently stressed or unstressed as given in the lexicon and the division is assumed to coincide roughly with the division into lexical and grammatical (functional) words often mentioned. In the restricted vocabulary of Commentator the following illustrate lexically stressed words: Adam, Eve, vänster (left), höger (right), porten (the gate), nära (close, near), närma (sig) (approach), också (too), heller (either, neither). The following words are lexically unstressed: han (he), hon (she), honom (him), henne (her), den (it), det (it), om (if), i (in), till (to), och (and), men (but), är (is). Inherently unstressed words may become stressed, e.g. by contrastive stressing, and stressed words may become unstressed.

As described above the verbalization processes of Commentator produces one word at a time and for each word delivered to the first buffer (the current buffer) or the second buffer (the execution buffer) a number of variables may be checked. If a sentence termination is signaled, whatever has been accumulated in the execution buffer is executed (pronounced) and the pitch of the last segment is set at low. If the number of the segments in the chunk being accumulated in the execution buffer does not exceed a certain limit a new word is only stored after the others in the execution buffer. The duration of a sound in Votrax is 0.1 second on the avarage. If the limit is set at 15 the system will deliver chunks about 1.5 seconds, which is what we want. The length of the chunks can be preset in order to simulate different individuals, speech styles or speech disorders.

If the number of segments in the execution buffer exceeds the limit the system proceeds to find out whether there is a tight syntactic link between the last word and the following. Such links (syntactic coherence) is signaled through the process. If not the cut is made after the last word in the buffer and the buffer is pronounced with a rising pitch on the last sounds.

The short sentences produced by Commentator seem to require short chunks. If the limit is set at 15 we would get the following result (L= low pitch, H= high). The number of segments is also given.

1. ADAM Ä(R) TI(LL) VÄNSTER OM (H)(16)	Adam is to the left of
2. EVA.(L) (3)	Eve
3. HON Ä(R) TI(LL) HÖGER OM/	She is to the right of/
PÖGER OM PORTEN/HORTEN.(L)(19)	gight of the gate/the rate.
4. ADAM AVLÄGSNAR SIG (H)(16)	Adam is going away
5. FRÅN EVA.(L)(7)	from Eve.
6. HAN AVLÄGSNAR SIG(H)(15)	He goes away
7. FRÅN PORTEN OCKSÅ.(L)(14)	from the gate too.

The typical speech errors indicated by the slashes will be commented on below.

EXPLAINING SPEECH ERRORS AND SOUND CHANGE

Speech errors may be classed as semantic, lexical, syntactic or phonetic. Semantic errors can be explained in the Commentator model as errors in verifying a proposition, e.g. estimating the closeness when the proposition CLOSE (ADAM, EVE) is to be tested. Lexical errors can be explained as mistakes in picking up the address of a lexical item. Instead of picking up <u>höger</u> (right) the word <u>vänster</u> (left, a semiantonym) stored on an adjacent address is picked up and sent to the current buffer. Syntactic mistakes may occur as a result of mixing the contents of memories used for storing different constituents, forgetting structural conditions etc. Of course various other explanations, including Freudian associations may also have to be evoked.

Phonetic errors, which are of greatest interest in this paper may be context-free (spontaneous) or context-sensitive. Context-free errors consist of substitutions which do not depend on any features in the context. Context-sensitive errors are traditionally divided into 1. progressive, 2. regressive and 3. inversions (metathesis). These cases can be explained in our model if we assume the two buffers and some simple operational mistakes in handling them.

If a speaker says <u>pöger</u> instead of <u>höger</u> in the text illustrated we would have a regressive error. The <u>p</u> of <u>porten</u> is said to influence its phonotactic equivalent in the preceding stressed word of the text. As Merringer noted already in 1895, errors generally concern corresponding elements in stressed or unstressed syllables: the initial consonantal constituent of a stressed syllable is substituted for the initial consonant constituent of another stressed syllable etc. Substituting the initial parts of stressed syllables seems particularly frequent. If the words are coded accordingly our model, where long strings of sounds are stored before pronunciation, offers simple ways of explanation.

In terms of our model the regressive mistake may be explained as a case where a sound of the buffer is substituted or picked up too soon for pronunciation.

If a speaker says horten instead of porten in the text illustrated, we would say that he has made a progressive error. In terms of our model the speaker has then replaced a later consonant in the buffer.

If a person says pöger om horten, he is said to have produced an inversion of h and p. In terms of our model such a mistake may come about if the two mistakes mentioned are both at work.

Models of speech errors often assume mistakes in reading buffer memories, although they do not generally describe the procedure as detailed as we have done, which in fact allows us to simulate speech errors by a computer program. The serial two buffer model seems to have advantages compared to the competing parallel buffers asumed by Baars (1980). Although several processes may go on at the same time,cf Kempen & Hoenkamp (1982) it seems less natural to assume that the same activity e.g. pronunciation is carried out tentatively with two (or several) processes using several buffer memories, whose contents are then sometimes erroneously mixed.

Most explanations of speech errors assume an unconscious or a conscious monitoring of contents of the buffers used during the speech process. This monitoring may also result in sound changes as the speaker may want to adjust his pronunciation to a new norm. There are several places in the Commentator where sound changes may be introduced. Some changes may correspond to changes in the hardware (Votrax). Changes may be brought about in the execution buffer, deleting unstressed vowels, etc. Changes may be brought about in the execution buffer as a result of monitoring comparing its contents to norms in ways resembling word processing systems which apply automatic spelling correction. Sound changes may be brought about by changing the phonetic representation of one or several words in the lexicon (lexical diffusion).

The chunking shown above often attaches unstressed words to the end of stressed words. It is natural to associate this enclitic buffer procedure with enclitic processes characteristic of language change. Our 2-buffer model attaches unstressed pronouns, auxilliaries and particles to the preceding word (deleting word boundaries). Historical enclitic processes often consist of attaching such function words, reducing their shape and sometimes turning them into inflectional morphemes. The definite articles are well-known examples from the Nordic languages. The development of the passive in the Nordic languages is also a result of enclitic attachment of the reflexive pronoun to the verb: kallas<kallask<kalla-sik<kalla sig. In our sample text nära den may result in the enclitic form nära-n and vänster om in vänstrom. The place and domain of enclitic phonological reduction or deletion rules seems to be the executive buffer.

The idea of packing words in a buffer before they are pronounced is reminiscent of an interesting proposal put forward by Lindblom et al. (1976) in order to explain the inverse relation between segment duration and number of segments in the utterance. Lindblom assumes that segments are packed in a buffer where they are compressed before being uttered. The more segments inserted in the buffer the more they are compressed. This holds for all segments except the last, which keep their inherent duration and therefore seem to be longer than the others. According to Lindblom one should therefore talk about non-final shortening instead of final lengthening.

Lindblom associates his buffer with short term memory, a central concept in psychology. The capacity of his buffer of pronunciation is assumed to be a few syllables which, however, seems to be below the size generally assumed for short term memory.

COMPUTER IMPLEMENTATION

The ideas presented above have been implemented on a microcomputer as an addition to the Commentator program. The Commentator program will not be presented here, as it can be found elsewhere, Sigurd (1983) and Fornell (1983). We will limit ourselves to presentation of the fragment relevant to pronunciation (see Appendix).

The words are represented as data, illustrated in lines 3001-3014. The first figure gives the number of items, and the following pairs of numbers are the code numbers of the speech sounds according to the Votrax manual and the pitch required for that sound (0-3). The pitch (stress) pattern of a word can be identified by reading every other figure starting from the third.

Lines 2005, 2020 illustrate how the lexical numbers given by the Commentator to the variable U make different data lines available for reading. Lines 2050-2070 read the sound and pitch numbers into the current buffer (B1). Line 2064 raises the pitch of a pitch segment, if the contrastive flag (C6) has been set previously by the verbalizing program. Line 2068 sets the flag if a full stop (62) has been entered into B1. Line 2063 shows how a sound change L1>L2 may be introduced in the system.

Line 2075 checks whether the sentence has ended. If not the content of the current buffer is passed on to the executive buffer by the subroutine 2080-2090. If the number of segments in the execution buffer (Ö) exceeds the preset limit (C9) the variable marking syntactic adherence has to be checked (2078-9). If cutting is allowed the execution buffer is sent to pronunciation. Lines 2100-2115 place sound numbers in S and pitch numbers in P. Line 2117 adds a switch-off signal (63). If the last segment number is found to be 62 (sentence termination, full stop) the pitch of the last sound is lowered, otherwise raised (2120). Line 2122 shows another place for sound change. Lines 2200-2228 handle the pronunciation through the interface and the Votrax.

Speech errors may be simulated by copying a sound of one buffer into the other buffer, interchanging within a buffer or reading mistakes as described before.

The resulting pronunciation gives intelligible words and prosodic chunks, but the interchunk and intrachunk prosodic quality is far from perfect. The prosodic chunks and the pauses may be given variable duration simulating more or less hesitant speakers. The pitch levels may be varied and different triggers set during the verbalization process may be used to control pitch. After some experimentation we know that it is easy to simulate many kinds of speech disorders. But it is indead very difficult to simulate normal speech.

ACKNOWLEDGEMENT

I am endebted to Gösta Bruce for discussions of some of the issues of this paper. Gustav Jönsson och Bengt Mandersson have assisted in interfacing the hardware, Jan Fornell in designing the software.

APPENDIX

```
2000 IF U%>20% THEN U%=U%-20% : GOSUB 2020 : RETURN
2001 IF U%>10% THEN U%=U%-10% : GOSUB 2010 : RETURN
2005 ON U% RESTORE 3001,3002,3003,3004,3005,3006,3007,3008,3009,3010 : GOSUB 205
0 : RETURN
2010 ON U% RESTORE 3011,3012,3013,3014,3015,3016,3017,3018,3019,3020 : GOSUB 205
0 : RETURN
2020 ON U% RESTORE 3021,3022,3023,3024,3025,3026,3027,3028,3029,3030,3031 : GOSU
B 2050 : RETURN
2050 READ N% : REM FILL B1(CURRENT BUFFER)
2060 FOR J%=1% TO N% STEP 2%
2062 READ B1%(J%),B1%(J%+1%)
2063 IF B1%(J%)=L1% THEN B1%(J%)=L2%
2064 IF C6%=1% THEN B1%(J%+1%)=B1%(J%+1%)+1% : C6%=0% : REM INCREASE PITCH IF CO
NTRAST VARIABLE SET (C6%)
2068 IF B1%(J%)=62 THEN P1%=1 : REM FLAG IF STOP IN B1
2070 NEXT J% : READ A$ : S5$=S5$+A$
2072 I=RND : IF R5<I THEN B2%(1)=B1%(1) : REM  REGRESSIVE SPEECH ERROR SIMULATIO
N
2074 GOSUB 2080 : REM B2 FILL
2075 IF P1%=1 THEN GOSUB 2100 : P1%=0 : RETURN : REM PRONOUNCE IF STOP IN B1
2078 IF Ö%<C9 THEN RETURN : REM IF NOT TOO LONG CHUNK RETURN
2079 IF A1%=1 THEN A1%=0% : RETURN ELSE GOSUB 2100 : RETURN : REM IF SYNTACTIC A
DHERENCE RETURN ELSE CHUNK
2080 FOR J%=1% TO N% : REM FILLS B2 FROM B1 WHICH IS EMPTIED
2082 Ö%=Ö%+1% : B2%(Ö%)=B1%(J%) : B1%(J%)=0%
2085 NEXT J% : B1%=0%
2090 RETURN
2100 FOR J%=1% TO Ö% STEP 2 : REM READS NUMBERS INTO SOUND(S) AND PITCH(P) REGIS
TERS
2105 V%=V%+1%
2106 S%(V%)=B2%(J%) : P%(V%)=B2%(J%+1) : B2%(J%)=0% : B2%(J%+1%)=0%
2115 NEXT J% : Ö%=0%
2117 I=RND : IF I>F9 THEN S%(V%+1%)=63 : REM SWITCHOFF UNLESS FLOORHOLDER
2120 IF S%(V%)=62 THEN P%(V%-1%)=0 ELSE P%(V%-1%)=2 : REM LOWERING PITCH AT END
OF SENTENCE(62) ELSE RAISING
2122 IF S%(V%)=L1% THEN S%(V%)=L2%
2200 OUT 1,133 : FOR J%=1 TO V%+1% : REM UTTAL
2210 K%=INP(0) : IF K%<>254% THEN 2210
2215 OUT 0%,S%(J%),2%,P%(J%)
2217 IF S9=1 THEN ; S%(J%)":"P%(J%),
2220 OUT 3,1 : OUT 3,0 : OUT 3,1
2225 NEXT J% : B2%=0%
2228 V%=0% : FOR Z=1 TO P5% : NEXT Z : RETURN
3001 DATA 8,14,2,53,2,30,1,49,0 ,   " BÅDA" :   REM LEXIKON
3002 DATA 6,27,0,50,1,13,0," HAN "
3003 DATA 8,21,2,30,3,50,1,12,0  ," ADAM "
3004 DATA 6,27,0,22,1,13,0 ," HON "
3005 DATA 6,6,2,15,3,50,1,"  EVA "
3006 DATA 2,0,0," ÄR "
3007 DATA 4,42,0,9,0," TILL "
3008 DATA 10,27,2,58,0,28,0,35,0,43,1," HÖGER "
3009 DATA 18,21,2,15,2,24,1,47,0,28,0,31,0,13,1,50,0,43,1," AVLÄGSNAR "
3010 DATA 4,31,0,6,1," SIG "
3011 DATA 8,29,0,43,1,52,1,13,0," FRÅN "
3012 DATA 14,13,2,47,2,43,0,12,1,50,0,31,0,42,1," NÄRMAST "
3013 DATA 14,15,2,2,0,13,0,31,1,42,0,35,1,43,1," VÄNSTER "
3014 DATA 12,13,2,47,1,43,0,12,1,50,0,43,1," NÄRMAR "
```

REFERENCES

Allén, S. (ed) 1983. Text processing. Nobel symposium. Stockholm: Almqvist & Wiksell

Baars, B.J. 1980. The competing plans hypothesis. In: Dechert & Raupach (1980), pp 39-49

Carlsson, R. & B. Granström. 1978. Experimental text-to-speech system for the handcapped. JASA 64, p 163

Ciarcia, S. 1982. Build the Microvox Text-to-speech synthesizer. Byte 1982:Oct

Dechert, H.W. & M. Raupach (eds) 1980. Temporal variables in speech. The Hague: Mouton

Enkvist, N-E. (ed) 1982. Impromptu speech: symposium. Åbo: Åbo akademi

Fornell, J. 1983. Commentator, ett mikrodatorbaserat forsknings-redskap för lingvister. Dep of Linguistics. Lund (mimeo-graphed)

Fromkin, V.A. (ed) 1973. Speech errors as linguistic evidence. The Hague: Mouton

Fromkin, V.A. (ed) 1980. Errors in linguistic performance: slips of the tongue, ear and hand. New York: Academic Press

Gårding, E. 1977. The Scandinavian word accents. Travaux de l'institut de linguistique de Lund. Lund: Liber

Gårding, E., G. Bruce & R. Bannert.(eds) 1978. Nordic Prosody. Travaux de l'institut de linguistique de Lund. Lund: Liber

Jönsson, K-G, B., Mandersson & B. Sigurd. 1983. A microcomputer pausemeter for linguists. In: Working Papers 24. Lund. Department of linguistics

Kempen, G & E. Hoenkamp. 1982. Incremental sentence generation. Proc. ninth Intern. Conf, on computational linguist. Prague

Lindblom, B., B. Lyberg & K. Holmberg. 1976. Durational patterns of Swedish phonology. Do they reflect short-term motor memory processes? Department of Linguistics. Stockholm

Mann, W.C. & C. Matthiessen. 1982. Nigel:a systemic grammar for text generation. Information sciences institute. USC. Marina del Ray. ISI/RR-83-105

Meringer, R. 1895. Versprechen und Verlesen: Eine psychologisch-linguistische Studie. Stuttgart: Goeschen

Sigurd, B. 1982. Text representation in a text production model. In: Allén (1982)

Sigurd, B. 1983. Commentator: A computer model of verbal production. Linguistics 20-9/10 (to appear)

Svartvik, J. 1982. The segmentation of impromptu speech. In Enkvist, N-E. (1982)