Dept. for Speech, Music and Hearing

Quarterly Progress and Status Report

Lexical prediction for a text-to-speech system

Hunnicutt, S.

journal: STL-QPSR
volume: 26
number: 2-3
year: 1985
pages: 047-055

II. SPEECH SYNTHESIS

A. LEXICAL PREDICTION FOR A TEXT-TO-SPEECH SYSTEM

Sheri Hunnicutt

Abstract

A lexical prediction system is being developed at the Department of Speech Communication and Music Acoustics to be used with our speech synthesizer (Carlson, Granström and Hunnicutt, 1982). It is being developed in response to the observation that some non-speaking users of the synthesizer find communication with such an aid laboriously slow. One user commented that the other children were guessing the words she was typing before she finished, but that she had to finish typing them anyway for the speech to be correct. The lexical prediction system addresses this problem. It is implemented on an Eclipse computer in a test system with speech output from a connected synthesizer. Hardware implementation has begun while the system undergoes continued development and testing.

Lexical Data Files

The data base for lexical prediction is a set of lexical data files. These files have been constructed to reflect some of the human capabilities of accessing lexical items. This is accomplished by appealing to literature in the theory of lexical access, and using experimental results as a basis for lexical predication.

The largest data file is the "Two-letter Lexicon," a file partitioned according to the first two letters of a word and frequency-ordered. This file contains about 10,000 entries, the most frequent Swedish words according to the Allén corpus (Allén et al., 1970). A process has been devised to quickly convert to the required form any of our 10,000-entry lexicons used with the text-to-speech system (speech synthesizer). These lexicons exist for Swedish, English, German, Norwegian, Italian and French. In addition to the spelling and rank order of each word, in the Swedish lexicon the word's part-of-speech, or word class, is given. This information is accessed during grammatical decisions preceding prediction. A second file, the "First Choice Lexicon," contains the most frequent word beginning with each letter. This file is used to very quickly access this most frequent word as soon as its first letter is typed in. Two-word combinations are stored in a third file, the "Two-word Lexicon." These represent the 1500 most common sequences of two words according to a further study by Allen (Allén et al., 1975). A fourth file, the "One-letter Lexicon," contains all single letters, several of which can also be words.

The approach to lexical prediction followed here is supported by psycholinguistic literature. It is well documented, for example, that the initial sounds and the initial letter or letters of a word are "access points" for words. That is, a person can guess a word faster given its initial letter(s) than given medial or final letters (Marslen-Wilson and Welsh, 1978; Jakimik and Cole, preprint, among others). It has also been shown that a person's mental lexicon is, in some sense, frequency-weighted (Broadbent, 1967). That is, the more frequent a (content) word in a language, the faster the reaction time of a person hearing that word in classifying it as a word of the language (Bradley, 1978).

One further file, the "Subject Lexicon," exists which has space for up to 510 words and is initially empty. Each word typed is entered in this lexicon, and a count kept of the number of times it has been used. We plan to send a message to the user when this lexicon is full. The user will then be able to choose whether to incorporate these words in his/her permanent lexicon. A decision to incorporate these words will result in an automatic update of the frequencies of words in the Two-letter Lexicon. Words not previously occurring in this lexicon will be added to it, while the lowest frequency words drop out.

A second use for this file is reflected in its name, the Subject Lexicon. It is possible to set two frequency values for this lexicon, one value for word inclusion and another value for word prediction. A preliminary test indicated that words with rank greater than 1,000 (words not in the 1,000 most frequent) should be placed in the Subject Lexicon and given a temporary rank between 201 and 400. This rank for prediction allows only the 200 most frequent words, most of which are function words which can be expected to make up over 50% of a text (Kučera and Francis, 1967) to precede words with rank greater than 1000 but currently in usage. In this way, word frequency can be temporarily raised during a conversation about a specific topic. It is possible to save or empty this lexicon at any time.

This set of lexicons and the facility for updating frequencies and adding new entries will allow users to develop their own personal frequency-weighted vocabularies over an extended period of time. In a hardware implementation, lexicons should be easily changeable, permitting multiple users and multiple lexicons for a single user. It would be possible, for example, to have separate lexicons for use at home, at school and at work.

The Algorithm

Given an initial letter (or letters) by the user, the word it introduces is predicted based on word frequency and a simple phrase structure grammar. Typing the first letter of a word results in accessing the most frequent word beginning with that letter from the First Choice Lexicon. Successive predictions are made from the Two-letter Lexicon if predicted letters are overwritten by the user. When a complete word (word plus space) has been typed, or when a word has been predicted and accepted, the Two-word Lexicon is consulted to see if that word occurs in it as an index word. If found, the word following it (the most frequent) is automatically predicted without its initial letter being typed. This word may be accepted, or it may be rejected by typing the first letter of the next word desired. In the latter case, the process begins again, with the most frequent word beginning with that letter being predicted from the First Choice Lexicon.

There are three ways to conclude a word. Two of the characters that have been chosen to indicate word conclusion, the "line feed" and "escape" characters, are picked arbitrarily and can easily be altered. The "line feed" character indicates that the latest prediction is accepted unconditionally, and types a space after the predicted word. This is the normal mode of accepting a prediction. An example, typing the word "telefon," is shown in the top part of Fig. 1. The user first types the letter "t," and the prediction of the word "till" ("to"), which is the thirteenth most frequent word, is made from the First Choice Lexicon. The user then types the letter "e," and the most frequent word beginning with the sequence "te," i.e.,"tekniska" ("technical") with rank 601, is predicted from the Two-letter lexicon. The letter "l" is then typed by the user, and the desired prediction, "telefon," (rank 4047) appears. It is accepted by typing the "line feed" character, and a space is added after the word.

The "escape" character accepts the latest prediction, but does not type a space, positioning the cursor at word-end so that characters may be deleted or added. These deletions and additions produce further predictions, as well. A common use of this facility is the addition of suffixes, before which a morph-final letter may be deleted or changed. In the example in the middle of Fig. 1, we see that in typing the word "möjligheter" ("possibilities"), the "escape" character is used to position the cursor at the end of the predicted word "möjligt" ("possible"). When the final "t" is erased, the desired word is automatically predicted. This facility is also useful in compounds, which Swedish has in abundance. This use is demonstrated also, in the word "presidentvalet" ("presidential election"). After "president" is predicted and the cursor is positioned word-finally, the letter "v" elicits the correct root, "valet."

Terminating a word with a space indicates rejection of predictions and the end of a user-typed word, as in normal typing. A common example of this type of word conclusion is shown in the bottom part of Fig. 1. The user desires to type the preposition "i" ("in"). The letter "i" is typed, and the prediction, from the First Choice Lexicon, is "inte" ("not"), with rank 15. This prediction is rejected and the word "i" completed by simply typing a space.

Note: Predicted characters are underlined;
Letters typed by user are in parentheses

1.) Line Feed: Unconditional acceptance of latest prediction

User Types	Prediction	Rank Order
T	(T)ILL	13
E	(TE)KNISKA	601
L	(TEL)EFON	4047
Line Feed	(TEL)EFON space	

2.) Escape: Conditional Acceptance of Latest Prediction;
Pointer positioned at word end

User Types	Prediction	Rank Order
M	(M)ED	11
Ö	(MÖ)JLIGT	275
Escape	(MÖ)JLIGT^	
Delete	(MÖ)JLIGHETER	363
----------	-------------------	------------
P	(P)Å	9
R	(PR)OBLEM	218
E	(PRE)CIS	644
S	(PRES)IDENT	694
Escape	(PRES)IDENT^	
V	(PRES)IDENT(V)ALET	4302

3.) Space: End of User-typed word; all predictions rejected

User Types	Prediction	Rank Order
I	(I)NTE	15
Space	(I Space)	

Fig. 1. Three Ways to Conclude a Word

Testing

The prediction system is currently being tested on a 4567-word text to determine the contributions of particular data files and other contributions of the algorithm. The text is a transcription of one person's (a teenager's) communication via a personal communicator, i.e., a sequence of typed one-sided conversations. One change has been made to the text: each sentence was originally terminated by a carriage return; a sentence-final punctuation mark has been inserted before each such carriage return.

The text has been divided into eight parts. All averages and sums reported have therefore been calculated using a weighting factor. The size of the Subject Lexicon was set at 200. The rank for word inclusion in this lexicon was 1000, and the rank for word prediction, 200. The tests have been automated. This implies, in particular, that conditional acceptance of predictions is not possible. A test of the entire text gave the following results (Table 1):

Words partially or fully predicted	80%
Savings in keystrokes	26%
Savings in letters	34%

Table 1: Prediction in a 4567-word communicator text

In this test, all data files were used, i.e., the 10,000-word Two-letter Lexicon, the First Choice Lexicon, the 1500-word Two-word Lexicon and a 200-word Subject Lexicon. The Subject Lexicon was cleared before beginning each of the eight text files. The grammar was not employed.

It is encouraging that some help is provided in the typing of 80% of the words. This means that the user can generally expect to be spared the typing of entire words. The savings in actual keystrokes, however, is considerably lower, only 26%. One-third of all letters are predicted. The 8% discrepancy between savings in keystrokes and savings in letters lies, firstly, in the fact that one character (line feed) is necessary to accept each predicted word, nullifying the savings in predicting the word-final space, and, secondly, that two characters (period and carriage return) terminate each sentence. This carriage return character could, of course, be omitted, but otherwise, possibilities for improvement all lie in the category "savings in letters."

Looking, then, at the present savings in letters, we may inspect the contribution of the various lexical data files (see Table 2 below):

Two-letter Lexicon (10,000 entries)	17%
First Choice Lexicon (27 entries)	13%
Two-word Lexicon (1500 entries)	2%
Subject Lexicon (200 word spaces)	2%
	34% total

Table 2: Contribution of Various Lexicons
Savings in Letters

The Two-letter Lexicon accounts for predictions that lead to one-half of the total savings in letters. Most of the words

Ranks	Percent of words in text
1-1000	73.1%
1001-2000	6.1
2001-3000	2.0
3001-4000	1.7
4001-5000	1.7
5001-6000	1.0
6001-7000	.5
7001-8000	.9
8001-9000	.7
9001-10000	.2

New Words: 11.9%

Table 3: Rank Composition of 4567-word communicator text

accepted from this large lexicon (73.1%) have rank order between one and one thousand, as seen in Table 3. This figure differs by only four percent from the figure commonly quoted for typed text (Kučera and Francis, 1967), 68.9%. Upon inspection of the higher-ranking (lower-frequency) words in the Two-letter Lexicon, however, it is clear that these words are common. A user might rightfully expect these words to be part of the predictable vocabulary. This is an obvious argument against limiting the size of the Two-letter Lexicon based on efficiency considerations alone.

The First Choice Lexicon, containing only 27 entries, makes a substantial contribution as well; 13% of the savings in letters come from words in this lexicon. Looking at the entire text, we see that only ten words from the First Choice Lexicon, comprising at least 1.0% of the text each, make up 21% of the total text when taken together. The most frequently occurring word is "jag" ("I"), as is common in speech. These ten words, listed in frequency order in the text are:

Jag, det, är, på, inte, och, en, till, med, som.

(I, the/it, is/are, on, not, and, a, to, with, which/as).

The Two-word Lexicon and Subject Lexicon contribute only 2% each to savings in letters typed. Once again, one can easily argue to keep them, even so. For example, in the first text file, seven of the 51 occurrences of "jag" were followed by "har" ("have"). This combination was obtained by typing only "J" followed by 2 line feeds ((J)AG HAR), one line feed after each of the two words is predicted. This is the type of prediction a user could well expect.

In the case of the Subject Lexicon, the small savings is a function of the number of times more than once that each word (with rank greater than 1000) occurs in each of the eight text files. This lexicon appears to be the one from which most substantial improvements in prediction can come. As a ceiling on lexical contributions to prediction in these eight texts, each text was run a second time with the Subject Lexicon constructed on the first run. In the second run, 12.8% more predicted words were accepted, and 34.1% more predicted letters were accepted. Since all keystroke savings are in letters saved, our expected ceiling on savings would be 34.5% on keystrokes (an increase of 8.8% from present results) and 45.6% on letters (an increase of 11.6% from present results). Any further savings would have to come from sources other than lexical data files.

One such source, a precedence-type grammar, has been implemented and tested. Unfortunately, it was found to cause as many mistakes as it corrected. Further grammatical work is evidently required.

Another source of possible improvement seemed to lie in changing the Subject Lexicon less frequently, allowing new words to accumulate so that they might be used for prediction in later text. Processing the entire 4567-word text with the same Subject Lexicon, however, provided only a small contribution, a 2.1% improvement in acceptance of predicted letters and a 0.3% improvement in acceptance of predicted words. This result suggests that the 200 word spaces may not be sufficient to retain a word until it is used again and attains a more permanent status. Or the amount of text might have been excessive. Another factor contributing to the Subject Lexicon's small usefulness may be the rank at which a word is either included in the Subject Lexicon or predicted from it. These factors will be tested further.

Other Prediction Systems

There are several similar prediction systems already in use which have been developed for English in recent years, each of them making contributions to the effort of helping users of augmentative communication systems achieve a faster input rate. One such system is "Speedkey," an expansion algorithm developed at the Trace R&D Center at the University of Wisconsin in Madison, USA (Kelso and Vanderheiden, 1982; Vanderheiden, 1984). This technique uses abbreviations beginning with one to three letters and ending with a digit. It is claimed that this method increases input on a standard keyboard by 200-300%. Speedkey is quite efficient. It is necessary, however, to learn the codes.

A prediction system for English which resembles the lexical prediction system reported here in its adaptive nature is one developed at the University of Dundee in Scotland (Arnott, Pickering, Swiffin and Battison, 1984). It employs several interchangeable context-specific 1,000-word dictionaries which are adapted on-line to the user's input. The original 1,000-word dictionary can be read in from a text file or generated by on-line usage. Operation of the system involves a split display screen with an area in which the 10 most frequent words beginning with the input character(s) appear. A switch initiates scanning and accomplishes selection.

The MicroDEC II, a combined environmental control unit and text prediction system, was developed in the Northwestern University Rehabilitation Engineering Program in Chicago, Illinois, USA (Heckathorne, Leibowitz and Strysik, 1983). The system uses a combination of digram letter prediction and lexicons based on work by Gibler (Gibler, 1981; Gibler and Childress, 1983). A 500-word core element similar to the Two-letter Lexicon is used together with a 200-word learning element similar to the Subject Lexicon; new words in the learning element are added to the core element. The initial two letters of a word are selected from a digram-based display. Word selections are made from a scanned 8-word display. Testing using a 1000-word core element and the 200-word learning element yielded text-to-lexicon matches in 71-76% of the words, and a text generation efficiency of 2.99-2.87 inputs (switch engagements) per letter, a 24% improvement over digram prediction alone. (Substituting the 500-word lexicon lost 4% in efficiency.) The text generation rate for subjects studied was around 3%-10%, however, since the tradeoff between switch activation movement time and visual search time must also be considered.

References

Allén, S. et al., (eds.) (1970): Nusvensk Frekvensordbok, Vol. I, Almqvist & Wiksell, Stockholm.

Allén, S. et al., (eds.) (1975): Nusvensk Frekvensordbok, Vol. II, Almqvist & Wiksell, Stockholm.

Arnott, J.L., Pickering, J.A., Swiffin, A.L. and Battison, M. (1984): "An Adaptive and Predictive Communication Aid for the Disabled Exploits the Redundancy in Natural Language," Proc. of the 2nd International Conf. on Rehabilitation Engineering, Ottawa, pp. 349, 350.

Bradley, D. (1978): "Computational Distinctions of Vocabulary Type," doctoral thesis, Dept. of Psychology, M.I.T., Cambridge, USA.

Broadbent, D.E. (1967): "Word-Frequency Effect and Response Bias." Psychological Review, 74, pp. 1-15.

Carlson, R., Granström, B. and Hunnicutt, S. (1982): "A Multi-Language Text-to-Speech Module," Conference Record, 1982 IEEE-ICASSP, Paris, France.

Gibler, C. (1981): "Lingistic and Human Performance Considerations in the Design of an Anticipatory Communication Aid," Ph.D. dissertation, Northwestern University, Chicago, Illinois.

Gibler, C.D. and Childress, D.S. (1983): "Adaptive Dictionary for Computer-based Communication Aids," Proc. of the 6th Annual Conf. on Rehabilitation Engineering, pp. 165-167.

Heckathorne, C.W., Leibowitz, L. and Strysik, J. (1983): "MicrodecII - Anticipatory Computer Input Aid," Proc. of the 6th Annual Conf. on Rehabilitation Engineering, San Diego, California, pp. 34-36.

Jakimik, J. and Cole, R. (forthcoming): "Sound and Spelling in Spoken Word Recognition." To be published.

Kelso, D.P. and Vanderheiden, G.C. (1982): "Ten-Branch Abbreviation Expansion for Greater Efficiency in Augmentative Communication Systems, Proc. of the 5th Annual Conf. on Rehabilitation Engineering, Houston, Texas, p. 3.

Kučera, H. and Francis, W.N. (1967): Computational Analysis of Present-Day American English, Brown University Press, Providence, Rhode Island.

Marslen-Wilson, W. and Welsh, A. (1978): " Processing Interactions and Lexical Access during Word Recognition in Continuous Speech." Cognitive Psychology, 10, pp. 29-63.

Vanderheiden, G.C. (1984): "A High-Efficiency Flexible Input Acceleration Technique: Speedkey," Proc. of the 2nd Int.Conf. on Rehabilitation Engineering, Ottawa, pp. 353, 354.