

TITLE : **Synthetic Speech Output for PX.**

REFERENCE : 91092

DATE : February 1991

AUTHOR : Jérôme Chiabaut, Mike Kelly and André Vellino

ABSTRACT

A word-concatenation speech synthesis system based on rules for modifying prosodic features was built for PX. This paper outlines the method of Young and Fallside (1980) and describes the encoding and synthesis techniques (McAulay and Quatieri (1986) and Griffin and Lim (1988)) that were implemented. The results of word concatenation experiments are analyzed and we conclude that, in the context of PX, the disadvantages of concatenating prosodically modified recorded words for speech synthesis outweighs its advantages.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

Synthetic Speech Output for PX

Introduction

The range of applications for the Personal eXchange system (PX) [Kame 90] would be greatly increased if it included a high-quality, low-cost, text-to-speech facility (for a general overview of the current state of the art in text-to-speech see [O'Mal 90]). For example, Borynec describes an experimental PX application—COCOVOX—for reading electronic mail over the telephone (see [Bory 89]). The prototype, which used a DECTALK server on a LAN, demonstrated the versatility of PX as well as the usefulness of synthetic speech. However, users were dissatisfied with COCOVOX because of the unnatural quality of the speech generated by DECTALK. The lack of natural intonation and its dependence on special-purpose hardware drove us to explore alternative sources of synthetic speech in the context of PX.

One technique for making speech output more natural than synthetic speech which mimics the vocal tract is to concatenate speech segments that are recorded by human speakers. Given that the PX voice toolkit provides ready access to pre-recorded voice clips, the question arose whether a system based on concatenation of speech segments would be sufficient to meet the speech output needs for PX.

Units of pre-recorded speech that can be concatenated come in different sizes. Any one of phonemes, demi-syllables (di-phones), syllables, or words can be used depending on the intended application. For the intended PX applications—reading digits, announcing calendar appointments, customized "repair-shop" announcements (although *not* COCOVOX)—the vocabulary is limited to a relatively small number of words and a relatively restricted grammar (i.e. not *all* possible English sentences). For such purposes it is feasible to produce synthetic speech by concatenating pre-recorded *words*, whose intonation and timing characteristics are modified to make the phrases sound more natural [Youn 80], [Lenn 80].

The purpose of our experiment with word concatenation synthesis was two-fold. Firstly we wanted to implement (in software) a known, high quality speech synthesis technology in the context of PX. Secondly, we wanted to compare two different speech encoding schemes suitable for the modification of prosody (see [Cors 91]).

This report contains the following sections:

- Outline of the concatenation method

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

- Characteristics of the method
- Encoding techniques
- Rules for modifying prosody
- Experimental results
- Conclusions

Outline of the Method

The word concatenation method of Young and Fallside [Youn 80] consists of taking a sentence from a text file and concatenating the corresponding prerecorded words. The prerecorded words are encoded so that it is possible to modify their pitch and intonation (prosodic) characteristics according to rules that depend on the syntactic analysis of the sentence and the stress patterns of the syllables in the words. Figure 1 summarizes the essential components of the system.

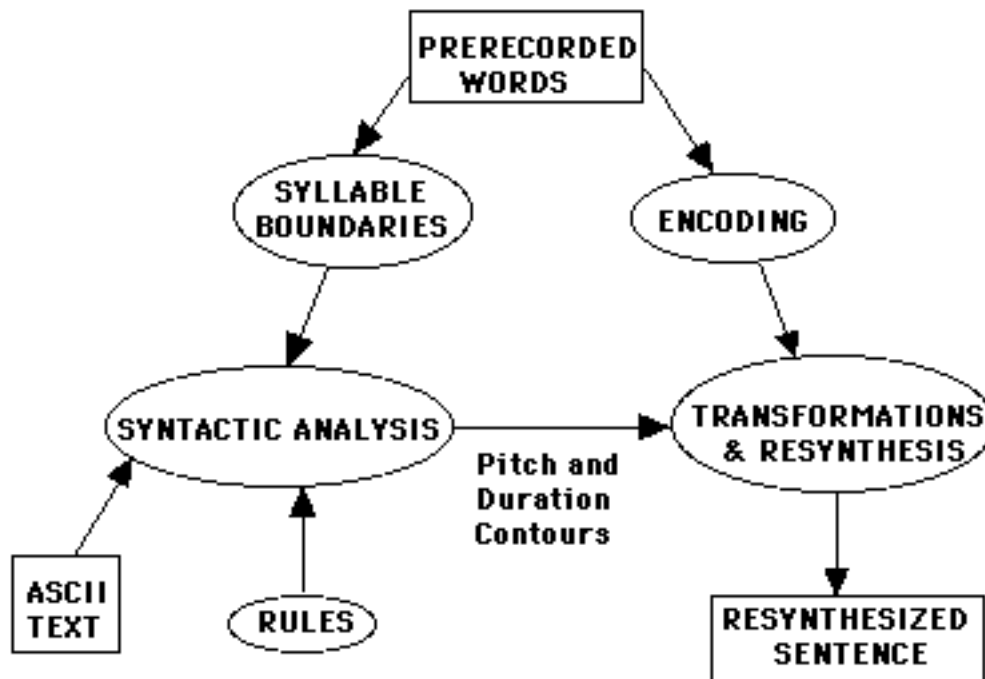


Figure 1: Speech synthesis by word concatenation

A lexical entry for a group of words consists of a file with a labeled voice clip for each word (in PCM) and a record of the end of each syllable within a word. These are determined automatically by a syllable boundary detector.

Computing Research Laboratory, Bell-Northern Research
 P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
 Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

Synthesis by Concatenation

Systems based on the concatenation of phonemes (DECTALK) [Klat 82] or di-phones (ORATOR) [Macc 87], although of lower speech quality, have the advantages over word-based synthesis that they require much less storage and can accept arbitrary phrases as input. On the other hand, they have the disadvantage that the recognition rates are less high than recognition rates for pre-recorded words.

The choice of words as the unit of prosodic modification (rather than di-phones or syllables) was motivated primarily by their availability in the PX environment. Using PX, users can record their own lexicon for their personal applications and the modification of prosodic features would still enable the listener to identify the original speaker. Provided the domain vocabulary is limited and the end-user application does not require the synthesis of arbitrary text, the word concatenation method seems appropriate.

Characteristics of the Method

The method outlined above requires the following:

A recording for each word in the sentence

Obviously, each word in the textual sentence must have a corresponding recorded word. In fact, even with very restrictive grammars, some words, verbs most notably, will require the recording of several utterances (am, is, are). Moreover, all the recordings in the lexicon should be made by the same person, and preferably normalized for pitch, amplitude, and rate of articulation.

A set of rules to modify the prosodic features of the recorded utterances

It is necessary to have rules that determine, from a syntactic analysis of the sentence, which changes to make to the prosody of the recorded words. These rules, taken here from Young and Fallside [Youn 80] are determined largely by trial and error. Unfortunately, there seems to be no sound theoretical basis for these rules and their improvement is possible only by experimentation.

Syntactically correct sentences

Since the Young and Fallside method depends on the successful syntactic analysis of sentences, it is not possible, without a sophisticated natural-language front-end, to synthesize ungrammatical phrases. Indeed, this method was originally designed for synthesizing data obtained from a computer database, whose grammatical characteristics are fixed. Thus, this method cannot deal with arbitrary text.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

A dictionary of the syllable stresses for each word

To generate the best intonation contour from the Young and Fallside rules, the dictionary information should contain not only the decomposition of words into syllables, but also the stress information for each syllable (primary, secondary and null) as well as their durations.

A good encoding/resynthesis technique

Changing the prosodic features (intonation and timing) of recorded words requires an encoding and resynthesis technique that preserves voice quality under transformations. The section below outlines our study of two such techniques.

Encoding Techniques

A speech signal is usually modelled as the output of an excitation (glottal vibrations) passing through a time varying filter (vocal tract). In most representations, the excitation can only be of two kinds, depending on whether the signal is voiced or unvoiced. This either/or model results in synthesized speech which is often buzzy and unnatural sounding, though intelligible, because a large fraction of the speech is neither completely voiced nor completely unvoiced. LPC (Linear Predictive Coding) is a good example of a technique which produces such distortions. Attempts have been made to improve the quality of the synthesized speech by modelling the excitation more accurately: MultiPulse Excited LPC and Residual Excited LPC in particular were developed for that reason. Residual Excited LPC has been successfully used for speech synthesis by word concatenation [by, for example, Stephen Eady at Speech Technology Research Ltd., University of Victoria].

In recent years, McAulay and Quatieri [McAu 86a], and Griffin and Lim [Grif 88] have independently proposed encoding techniques that depart from the traditional LPC framework. They both try to model the excitation more precisely, and propose original representations of the speech signal. McAulay and Quatieri [McAu 86b] show how to use their method to do speech transformations, and the method described by Griffin and Lim is known to offer such possibilities [c.f. synthesis of announcements by Brian Doherty *et. al.* BNR-RTP (3B72)].

Both techniques were implemented in Pascal on the Macintosh. [Cors 91] gives a detailed analysis of the two methods as well as an evaluation of their merits with respect to their quality in pitch and timing transformation, their effectiveness for data-compression, and their respective complexity.

The two encoding techniques produce speech that is almost indistinguishable from the original. Furthermore, if the change of pitch is not larger than 20% the quality of the synthesized speech is excellent.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

Unfortunately, when the pitch is modified by a larger amount, artifacts start to appear. This problem can be corrected by keeping several recordings of the same utterance at different values of the pitch and using the one with the pitch contour that is closest to the target pitch.

A straight-forward implementation of those technique produces intermediate representations of the speech which require more storage than the PCM encoding of the original signal. Griffin and Lim's method has the potential of doing a modest amount of compression (a factor of 2, perhaps) while retaining the same quality of synthesized speech. In practice, this means that each utterance will require several kilobytes of storage.

Both synthesis methods rely on the time domain synthesis of a family of sinusoids with time varying frequency and amplitude. This is computationally very expensive and cannot be done in real time on a MacII. On the other hand, a fixed-point DSP like Texas Instrument's TMS320C25 or Motorola's DSP56001 should run the synthesis part of the algorithm in real-time.

Rules for Modifying Prosody

The rules for modifying the prosodic features of a sequence of recorded words were taken from [Youn 80]. The system is given a grammar (in the form of Prolog DCG rules [Dahl 89]) for the language from which it constructs a parse tree for the sentence to be uttered. Given the parse tree, it decomposes the utterance into intonationally significant word groups. The syllables of the words in each word group are then categorized into one of four tone groups: pre-head, head, nucleus, and tail according to rules that place the stress and accent on the syllables. Given the length of each syllable in the utterance, a timing contour is generated that stretches or shrinks the syllable durations. Similarly, an intonation contour is generated for the tone group based on four primitive syllable contours.

The algorithms for determining word group boundaries and assigning stress marks to the syllables is detailed in [Youn 80] p245-248. A word group in a sentence is a group of words belonging to a part of speech (determined by the parse tree), whose last element is a content word and whose weighted sum of the syllables' lexical stress is less than or equal to some constant. The stress marking rules determine whether or not a syllable belongs to the pre-head, head, nucleus or tail, depending on the stress-context of a given syllable (whether, for example, it is preceded by one unmarked stress-1 syllable or two stress-0 syllables). The appropriate pitch contours are then selected and passed, along with the timing contours, to the re-synthesis module.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

The generation of timing contours is done analogously. The purpose of the timing analysis is to stretch and shrink the length of the syllables to ensure that the interval between each stressed syllable—the *foot*—is a constant for the utterance. This is done first by finding the average foot duration and then by adjusting the syllable durations so that all the feet in the utterance are as close as possible to the average.

These rules were formulated in Prolog to facilitate the interactive modification of their details as experiments were being performed. We believed initially, that the major difficulty with this method was in fine-tuning these empirically-based rules. But, as is evident from [Youn 80], the rules that are proposed there are rather crude.

However, it turns out that even the best possible intonation contours produce inconclusive results. One of the main problems, even with good intonation contours, are the effects of co-articulation. If the words are recorded in a context they may include co-articulation that is not required in a different sentence, or conversely, there may be some co-articulation required in the sentence that is not present in the recording.

Even if co-articulation effects are ignored, though, it is apparent that a greater variety of intonation contours is required for application at the resynthesis step. This implies that the rules need to be a lot more sophisticated which, in turn, requires a considerable amount of experimentation.

Experiments

The synthesis technique described above was applied to a small set of examples to demonstrate its capabilities for producing natural sounding English speech. One set of experiments was performed using a phone number as the target utterance. For these experiments, whole words (digits) were concatenated to form a seven-digit phone number. However, given the grammatical simplicity of the utterance, the prosodic modifications for the phone number were not rule-driven.

A second set of experiments was performed using a simple sentence as the target utterance. These experiments follow more closely the rule-driven method described in the previous section: syllables are concatenated to form the sentence and prosodic information is added to this sequence of frames based on a parsing of the text of the sentence.

Production of Phone Numbers

The phone number produced in this experiment is 515-1515. The following methods were used to produce utterances:

- 1) The phone number was spoken as a whole to provide a basis for comparison.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

2) The phone number was formed by concatenating the digits 5 and 1 alternately (without applying any prosodic contours). A pause was added between the 3rd and 4th digits to add some naturalness.

3) The phone number was formed by concatenating the digits 5 and 1 alternately and adding appropriate pitch contours at each digit position. A pause was also added between the 3- and 4-digit segments.

The contours used in part 3) of this experiment were determined from examination of naturally spoken phone numbers [c.f. Brian Doherty *et. al.* BNR-RTP (3B72)]. There are four distinct contours used in the sequence, shown in Figure 2. The range of pitch variation for these contours is $\pm 5\text{-}10\%$ (approximately 5-10 Hz), depending on the position of the contour. The *rise* and *fall-rise* contours have a range of +5 and -5 Hz, respectively. The *fall* contour causes the trailing '5' to decrease by 10 Hz over the duration of the word.



Figure 2: Phone Number Contours

The phone numbers produced using this technique are quite natural sounding and are substantially better than those produced by straight concatenation. There are some audible artifacts in the reconstruction — particularly where relatively large changes in pitch have been affected — but the improvement achieved by applying the pitch contour outweighs the negative effects of this distortion.

Production of English Sentences

The main subject of these experiments is the sentence which is shown in the Young and Fallside paper [Youn 80]. It is 'The actual consumption for the Droitwich zone is twenty-nine point two four megalitres'. The parsing of this sentence is shown in Figure 3.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

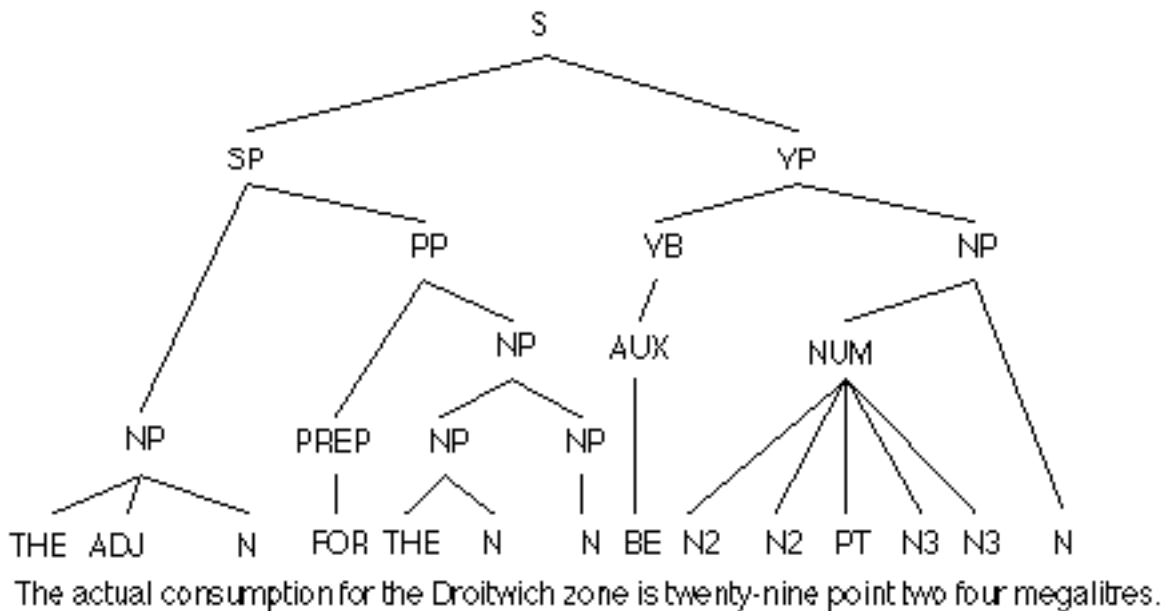


Figure 3: The Parse Tree of the Test Sentence

The following methods were used to form this sentence:

- 1) The sentence was spoken as a whole to provide a basis for comparison.
- 2) The sentence was formed by concatenating the necessary syllables (always whole words). Pauses were added at various points in the sentence to increase the naturalness.
- 3) The sentence was reconstructed using words stored in a lexicon and the pitch and duration contours extracted from the naturally spoken sentence. This experiment was performed to judge the effect of using words from a pre-defined lexicon (most spoken in isolation, others extracted from word sets) on the sound of the sentence. It also provides a standard against which the results of 4) can be compared; in principle it represents the best result that could be achieved by a rule-based production of the prosodic contours for the sentence.
- 4) The sentence was reconstructed using words stored in a lexicon and pitch and duration contours produced using the Young and Fallside rules. The next section describes how the lexicon was formed, and the contours that were applied during the reconstruction.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

The Lexicon

The lexicon consists of (PCM) recordings of each of the words used in the test sentence. In order to be used in the reconstruction, the syllable boundaries of the words have been marked. Following is a list of the words in the lexicon, showing their syllables separately:

The (thee)	ac_tu_al	con_sump_tion	for	the
Droit_wich	zone	is	twen_ty	nine
point	two	four	me_ga_li_tres.	

These words were recorded with as flat a tone as possible, and at as constant an energy (over all of the words) as possible (these attributes were not normalized later).

Figure 4 is a diagram of the loudness contour of the word 'consumption' with the syllable boundaries marked. The horizontal axis in this diagram is time measured in centiseconds.

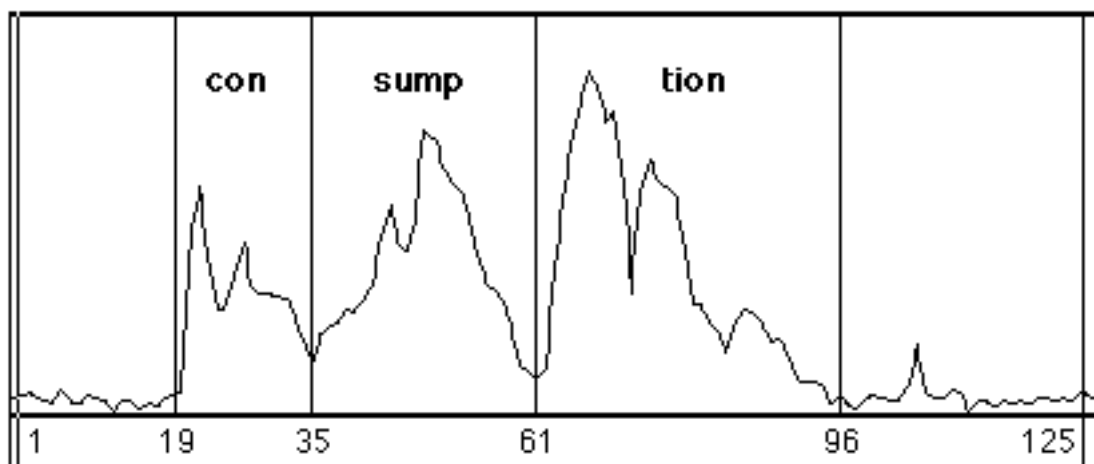


Figure 4: Loudness Contour and Syllable Markings for 'Consumption'

The loudness contour and the syllable boundary positions are stored as resources associated with each clip in the lexicon voice file. The syllable boundaries from this file are used by the rule manipulation program to determine time-scaling values.

The syllable boundaries are found using a minima-picking algorithm which scans through the loudness contour and selects the most likely set of syllable boundaries for a word. The constraints on this search are that a minimum must be of a certain depth in order to be classified as a syllable boundary, and that two syllable boundaries must be separated by a given minimum amount. The thresholds used are

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

interdependent; in effect, the algorithm imposes a minimum value on the depth-interval product for any pair of neighbouring syllable boundaries. The syllable boundaries placed by this algorithm can be manually adjusted if they are not deemed correct by a human listener.

The Pitch Contours

The pitch contours used in 4) are derived from the descriptions given in [Youn 80]. There are four contours in all required for the reconstruction of the given sentence. Each is used for a given set (or single) syllable type in the sentence. The four contours, and their uses are shown in Figure 5.

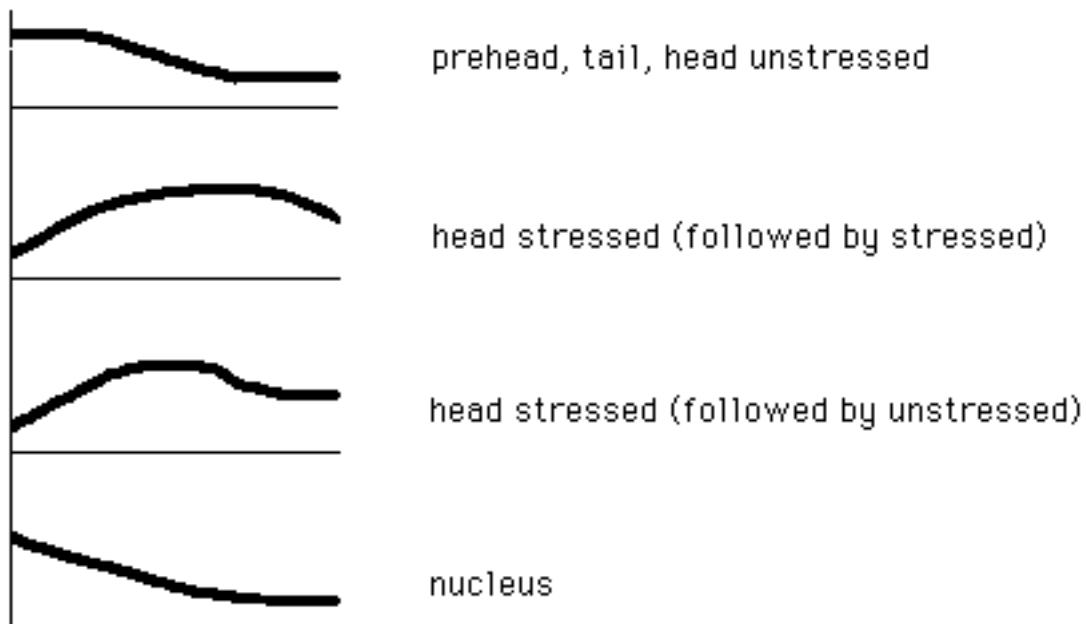


Figure 5: Syllable Contours

Each of these is scaled to result in approximately 5 Hz difference between highest and lowest points, and to match the duration of the syllable to which it is being applied. Recall that the 'head' superimposes a falling slope onto the contours applied to each syllable, to cause a generally falling pitch as the nucleus is approached. (Two other nucleus contours, and 4 corresponding tail contours, are described in [Youn 80], but were not used in this experiment.)

The sentence created using concatenation alone (without prosodic modification) is noticeably stilted and sounds very unnatural. It is sufficient to transmit information but listening to it is not at all pleasing.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

The sentence reconstructed using contours from the naturally spoken sentence is much better than the simple concatenation case, but is still noticeably artificial. The overall contour of the sentence is very close to that of the naturally spoken sentence (by construction), but its differences from that sentence serve to point out the deficiencies of the lexicon. Efforts were made to form words for the lexicon which have a flat pitch and suitable duration for in-sentence positioning. The results of this experiment indicate that the sound of the result is much more sensitive to variations in the lexicon words than was previously suspected.

The sentence formed using the rule-generated contours is, as expected, intermediate between straight concatenation, and modification using the natural contour. Although it is more natural sounding than the straight concatenation, it shows some noticeable variations from normal speech. The variations are due to deficiencies in the lexicon, and oversimplifications in the rule base used to form the contours. The shapes of the contours themselves may also be incorrect.

Co-articulation Problems

One of the audible problems with word concatenation is its inability to handle co-articulation. For a system based on word-concatenation, the simplest solution to this problem is to include co-articulated words in the vocabulary. The result, however, is to dramatically increase the size of the vocabulary since all the probable combinations of words would have to be pre-recorded, defeating the purpose of a synthetic speech system.

Conclusions

Whereas the primary purpose of [Lenn 80] was to test the algorithms for prosodic modification, our experimentation also aimed at comparing coding techniques. The coding technique in [Lenn 80] was linear predictive coding (LPC) whereas we compared the G&L vs Q&M methods. The comparison shows that, for digital telephone voice quality, the two techniques are indistinguishable for the range of changes made to the original recordings.

In general terms our investigations confirm the results from previous studies, namely that modifying the prosodic features of recorded words produces better quality speech than simple word concatenation. All things considered, we have come to the conclusion that in the context of PX, a general purpose text-to-speech system such as DECTALK is preferable to a word-concatenation system with modified prosodic features. The reasons for this are as follows.

A text-to-speech system that uses a word-concatenation technique suffers a number of rigid constraints. It must not only have a recorded lexicon of the entire

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

vocabulary, which requires large amounts of storage, but also a well defined grammar and all the lexical stress markings for all the syllables of all the words in the lexicon. The quality of the recordings should be studio-quality to offset the impact of encoding and resynthesis and each of the recorded words in the lexicon should also be normalized for pitch, intensity and duration to maximize the naturalness of the speech. Since syllable boundary detection is not always accurate some of this work must be done manually. All these things require a substantial amount of effort.

On the other hand, if we allow the use of similar quantities of information about the lexicon—e.g. a phonetic dictionary—to optimize the output of a general purpose text-to-speech system like DECTALK the result is, in our opinion, very good quality.

The prosodic rules are too simplistic and produce unnatural sounding speech. This is a problem with general text-to-speech systems as well as synthesis by word concatenation. More work is required in that area.

It may be worth repeating the concluding remarks from [O'Mal 90]:

Unrealistic expectations for dramatic future improvement in text-to-speech technology sometimes arise from an unsophisticated view of the complex linguistic information involved. Improvements will continue at the same slow, steady pace that has produced incremental progress in accuracy, intelligibility and naturalness over the past three decades.

References

- [Bory 89] J. Borynec "Report on Text To Voice" August 1989-BNR-CRL internal report.
- [Cors 91] I. Corset ."A Comparison of Two Encoding Techniques for Speech Transformations" January 1991, CRL internal report 91090
- [Dahl 89] H. Abramson and V. Dahl *Logic Grammars*, Springer-Verlag, Berlin 1989.
- [Kame 90] R. Kamel, K. Emami and R. Eckert "PX: Supporting Voice in Workstations" *IEEE Computer* (August 1990) vol.23, No.8 pp. 73-80
- [Klat 82] D. H. Klatt "The Klattalk Text-to-Speech System" *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (1982)* pp. 1589-1592.
- [Macc 87] M. Macchi, C. Kamm, L. Streeter "Expanding the template inventory for concatenative speech synthesis" *Speech Tech '87* pp. 159-161.
- [O'Mal 90] M. H. O'Malley "Text-to-Speech Conversion Technology" *IEEE Computer* (August 1990) vol.23, No.8 pp. 17-25

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.

[Youn 80] S. J. Young and F. Fallside "Synthesis by rule of prosodic features in Word Concatenation Synthesis" *International Journal of Man-Machine Studies* (1980) 12 pp. 241-258.

Computing Research Laboratory, Bell-Northern Research
P.O. Box 3511, Station C, Ottawa Canada K1Y 4H7
Telephone 613-763-3841, Fax 613-763-4222

© Bell-Northern Research 1990

The data herein are not to be used or disclosed without the consent of The Computing Research Laboratory. This note is a working paper intended for limited circulation and discussion. No departmental or corporate approval or commitment is implied, unless such approval or commitment is expressly given in a covering document.