A New Program for Manipulation of Natural Speech : ---- Interpolation Between Two Natural Utterances ----

Maximilian Hadersbeck

Institut für Phonetik und sprachliche Kommunikation
der Universität München
Schellingstr. 3, 8000 München 40, F.R.G.

ABSTRACT

For phonetic experiments it is very important to be able to manipulate distinctive features in natural utterances by not loosing the natural sound of the utterance.
In this paper the description of a computer program is delivered which allows interpolation between two natural utterances by means of Spectral Envelope Interpolation. The program produces high quality synthetic utterances, where speech parameters like pitch, intensity and formant structure of one speech utterance can be adopted to the same parameters of another utterance.
You can produce a natural sounding utterance continuum starting with the unmanipulated initial utterance towards a final utterance with the desired degree of manipulation. With this kind of manipulation you can for example change speaker identity, sentence intonation and stress of a natural utterance.

INTRODUCTION

This program is a new solution to interpolation between two natural utterances. The program produces higher quality speech utterances than a former program, developed by Simon [1], [2] at this institute.
Simon describes the program as following :

> "Contours of speech parameters such as pitch, intensity or formant structure can partially or totally be imposed on utterances, thus changing speaker identity, sentence intonation, stress or other psychophysical parameters. All manipulation can be done in discrete steps from the unmanipulated initial utterance to the final utterance with the desired degree of manipulation." (Simon (1984))

The program which I developed allows speech utterances to be manipulated in the same way as Simon described, but I introduced a new interpolation method in this program. The interpolation method is a means of "Spectral Envelope Interpolation" :
The spectra of the two utterances are calculated pitch-synchronously and according to the mode of interpolation and number of discrete interpolation steps, out of the two spectra a new spectrum is geometrically developed. From this spectrum the time signal of the new speech utterance is produced by a means of Inverse Discrete Fourier Transformation.
A continuum between two utterances can be developed by interpolating between the following parameters :

Spectral and Intonation Interpolation
The spectral shape and the intonation of speaker one's utterance is interpolated towards the spectral shape and intonation of speaker two's utterance.

Spectral Interpolation (the intonation of utterance one remains unchanged)
It is interpolated between the spectral shape of the two utterances. All the utterances of the continuum have the intonation of utterance one.

Intonation Interpolation (the spectrum of utterance one remains unchanged)
It is interpolated between the intonation of the two utterances. The spectral shape of the utterances is not changed.

THE PROGRAM [1] :

Logarithmic Interpolation :
The user can choose between a linear and a logarithmic interpolation method.
The logarithmic interpolation method has the advantage that the steps from the intonation of the first utterance towards the second are not simply analytically defined, but they are fitted to the properties of listener's speech perception

Spectral Envelope Interpolation :

The interpolation method between every single pitch period of the two utterances is a "Spectral Envelope Interpolation" method. This method avoids some disadvantages of the interpolation method, introduced by Simon [1].
One disadvantage was distortion in the synthetic speech signal if the two utterances are very different in their F0 structure.

Input and Output :

The inputs of the program are the two speech utterances, with a description of the unvoiced/ voiced/ pause structure. The next input is the number of discrete interpolation steps between the two utterances and the mode of interpolation. The last input is the kind of interpolation : Linear interpolation or logarithmic interpolation.

The outputs of the program are a continuum of natural sounding utterances with their unvoiced/ voiced/ pause description in accordance to the input.

Restrictions on the Phonetic Structure of the Two Utterances :

The two utterances must have the same phonetic unvoiced/ voiced/ pause structure.
An example : Given is the utterance "MAX". Its unvoiced/ voiced/ pause structure is :

```
 P   V   V   P   U
/    M   A       KS /
```

with P for pause, V for voiced, U for unvoiced.

The second utterance must have the same unvoiced/ voiced/ pause structure:

```
 P   V   V   P   U
```

for expample : "MIX"

The program interpolates now between the first segment (Pause in the example) of utterance one and the first segment in utterance two, the second segments (voiced segment in the example), the third and so on.
The program cannot interpolate between utterances with different phonetic structure. The only parameter which can vary between the utterances is the information within the coinciding segments :
The segments can have totally different length, F0 Curve, energy distribution, they can be spoken for example from different speakers. Between these parameters the program can interpolate.

The Interpolation Between the Two utterances Within the Speech Segments.

Pause- Segments :

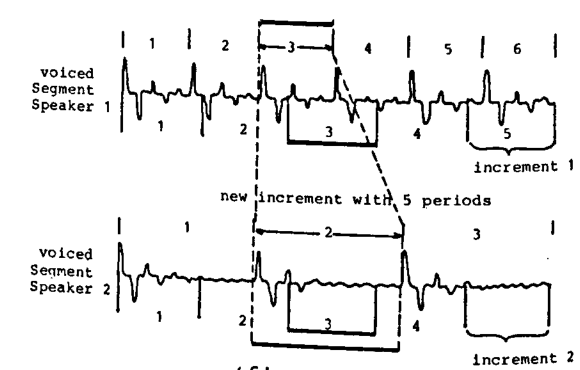
To interpolate between Pause segments, the speech datas are transformed with small modifications. The duration of the Pause segments are adopted to the interpolation step.

Unvoiced- Segments :

According to the interpolation step, the duration of the unvoiced segment is adopted. The speech data for the new utterance are calculated from the time signal of the two utterances, without any spectral transformation.

Voiced- Segments :
Determination of The Actual Pitch Period :

In the first step the program counts the number of pitch periods within the voiced segment for both utterances. We call the number of pitch periods in utterance one IANZ1, those in utterance two IANZ2. According to the interpolation step, the number of pitch periods for the new utterance (IANZX) is calculated. (see Formula 2 in Appendix)
Now two increments are developed to move through the voiced segment : INCR1 (utterance one) and INCR2 (utterance 2). With that increment you can step through the voiced segment of utterance one and utterance two, meeting IANZX pitch periods in the segments of utterance one and in utterance two.



(figure 1)

An example: (look at figure 1)

Utterance one has 6 pitch periods and utterance two has 3 pitch periods.
The interpolation step demands 5 pitch periods for the new utterance.
Now you calculate an increment for speaker one and two, to step through the utterances to meet 5 pitch periods in utterance one and utterance two.
With that increment you step through the voiced segments, and the current increment pointer determines the pitch periods in utterance one and utterance two, between which the "Spectral Envelope Interpolation" is performed.
In the example an interpolation is performed between pitch period number 3 of utterance one and pitch period number 2 of utterance two, to calculate the new pitch period number 3.

The "Spectral Envelope Interpolation" :

The aim of this method is, to calculate out of two pitch periods, with their F0 value, a new pitch period which has an F0 value in accordance to the interpolation step.
I decided to take the Discrete Fourier Transformation and Discrete Inverse Fourier Transformation to reach this aim [2].

The Algorithm :

Step 1 :
Calculate the N spectral lines U1p(n) of the time signal u1p(n), n=1,...,N in utterance one within pitch period p

Calculate the M spectral lines U2q(n) of the time signal u2q(n), n=1,...,M of utterance two within pitch period q (See Formula 1 in the Appendix).

Step 2 :
Calculate the number Q of neccesary spectral lines in the new period according to the interpolationstep.
(See Formula 2 in the Appendix)

Step 3 :
Calculate the Q spectral lines W1p(n) for utterance one out of the N spectral lines U1p(n) via the "Spectral Envelope Interpolation"

Calculate the Q spectral lines W2q(n) for utterance two out of the M spectral lines U2q(n) via the "Spectral Envelope Interpolation".
(See Subroutine ENVINT in Appendix)

Step 4 :
Calculate the new spectral Lines Xf(n), n=1,...,Q via an interpolation between W1p(n) and W2q(n).
(see formula 3 in the Appendix)

Step 5 :
Calculate the time signal for period f xf(n), n=1,...,Q for the new utterance out of the Q spectral lines Xf(n)
(see formula 4 in the Appendix)

IMPLEMENTATION :

The program is implemented in the High Level Language FORTRAN 77 on a Digital Equipment Corporation Computer PDP11/73 [3].
The main emphasize in the program lies on a very good readable form. Its subroutines are clearly defined and a programmer can understand very easy the work of the routines.
Because of this, the computation time is rather long : On a PDP 11/73 it takes for example 15 minutes to interpolate between two utterances of 5 seconds length.

SUMMARY :

I have descibed in this paper a computer program, which is able to produce a continuum of natural sounding utterances. The utterance continuum starts with an unmanipulated utterance towards a manipulated utterance. The kind of manipulation can be chosen manually via program input.

ACKNOWLEDGEMENTS :

The work was sponsored by DIGITAL EQUIPMENT CORPORATION GmbH Munich, FRG with a complete PDP11/73 computer system, including Software and Analog/Digital periphery [3] [4].

REFERENCES :

[1] Simon, Th. : Manipulation of natural speechsignals according to the speech parameters of different speakers, Forschungsberichte des Instituts fuer Phonetik und Sprachliche Kommunikation der Universitaet Muenchen (FIPKM) 17 (1983) page 233-245
[2] Rabiner L. Gold B. :Theory and Application of Digital Signal Processing, Prentice Hall (1975)
[3] Hadersbeck M. :Sprache und Klang Eine MICRO 11/73 als Speechprocessingmachine, 9. DECUS Muenchen e. V. Symposium, Suttgart (1986)
[4] Hadersbeck M. :Digitale Sprachverarbeitung unter Micro RSX, 10. DECUS Muenchen e. V. Symposium, Berlin (1987)

APPENDIX :

Formula 1 [2] :

$$(2)\quad Up(n) = \sum_{l=0}^{N-1} up(l) \cdot e^{-i(2\pi/N)ln}$$

for $n=1,\ldots,N$

Discrete Fourier Transformation

Formula 2 :

$$Q = N + FF * (M - N)$$

where :
FF ... factor of interpolation ($0.0 <= FF <= 1.0$)

Formula 3 :

$$Xf(n) = W1p(n) + FF * (W2q(n) - W1p(n))$$

for $n=1,\ldots,Q$

where :
FF ... factor of interpolation ($0.0 <= FF <= 1.0$)

Formula 4 [2] :

$$(3)\quad xf(n) = 1/N \sum_{l=0}^{N-1} Xf(l) \cdot e^{i(2\pi/N)ln}$$

for $n=1,\ldots,Q$

Discrete Inverse Fourier Transformation

```
      Subroutine ENVINT
                 (SPECIN,ANZIN,SPECOU,ANZOU
C "SPECTRAL ENVELOPE INTERPOLATION METHOD"
C  Input :
C  SPECIN(I), I=1,,,ANZIN
C       ....  Spectral Lines Input
C  Output :
C  SPECOU(I), I=1,,,ANZOU
C       ....  Spectral Lines Output

   DIMENSION SPECIN(1),SPECOU(1)
   INTEGER*2 ANZIN,ANZOU

   XIN=2*3.1459265/ANZIN   ! angle input
   XOU=2*3.1459265/ANZOU   ! angle output
   QQ2=ANZOU/2   !ANZOU/2 spectral lines
   DIN=0.0       !Increment counter IN
   DOU=XOU       !Increment counter OUT
   IND=1         !spectral line counter
   SPECOU(1)=SPECIN(1)

   DO 1 I=1,QQ2
2  IF (DOU .GE.  DIN
     .AND.  DOU .LT.  DIN+XIN) GOTO 3
   IND=IND+1
   DIN=DIN+XIN
   GOTO 2
3  GRAD=(DOU/DIN)/XIN
   SPECOU(I+1)=
   (SPECIN(IND+1) - SPECIN(IND))*GRAD +
         SPECIN(IND)
   DOU=DOU+XOU
1  CONTINUE

   RETURN
   END
```