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.
The Interpolation Between the Two utterances is a "Spectral Envelope Interpolation" method. This method avoids and 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 Structure. An example: Given is the utterance 'MAX'. Its unvoiced/voiced/pause structure is:

\[
P V V P U
\]

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

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

\[
P V V P U
\]

example: MAX.

The program interpolates now between the first segment (Pause in the example) of one utterance and the first segment (Pause in the example) of the other utterance, the second segments (voiced segment in the example), the third and so on. The program cannot interpolate between unvoiced/voiced/pause segments of different speakers. Between these parameters the program can interpolate.

The Interpolation Between the Two utterances is performed with the procedure "Spectral Envelope Interpolation". 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 described 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:
singsmachine, 9. DECUS Muenchen e. V. Symposium, Stuttgart (1983)
APPENDIX:

Formula 1 [2]:

\[ (2) \quad U_p(n) = \sum_{l=0}^{N-1} U_p(l) \cdot e^{-i(2\pi n/N)l} \]

for \( n = 1, \ldots, N \)

Discrete Fourier Transformation

Formula 2:

\[ Q = N + FF \cdot (M - N) \]

where:

\( FF \) ... factor of interpolation

\( 0.0 \leq FF \leq 1.0 \)

Formula 3:

\[ X_f(n) = W_{lp}(n) + FF \cdot (W_{2q}(n) - W_{lp}(n)) \]

for \( n = 1, \ldots, Q \)

where:

\( FF \) ... factor of interpolation

\( 0.0 \leq FF \leq 1.0 \)

Formula 4 [2]:

\[ (3) \quad X_f(n) = \frac{1}{N} \sum_{l=0}^{N-1} X_f(l) \cdot e^{i(2\pi n/N)l} \]

for \( n = 1, \ldots, Q \)

Discrete Inverse Fourier Transformation

Subroutine ENVINT

\( (\text{SPECIN,ANZIN,SPECOU,ANZOU}) \)

C "SPECTRAL ENVELOPE INTERPOLATION METHOD"

C Input:

C SPECIN(I), I=1,,ANZIN

C \ldots Spectral Lines Input

C Output:

C SPECOU(I), I=1,,ANZOU

C \ldots Spectral Lines Output

DIMENSION SPECIN(1),SPECOU(1)

INTEGER*2 ANZIN,ANZOU

XIN=2*3.1459265/ANZIN \ l angle input

XOU=2*3.1459265/ANZOU \ l angle output

QQ2=ANZOU/2 \ lANZOU/2 spectral lines

DIN=0.0 \ lIncrement counter IN

DOU=XOU \ lIncrement counter OUT

IND=1 \ lSpectral line counter

SPECOU(1)=SPECIN(1)

DO 1 I=1,QQ2

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

CONTINUE

RETURN

END

Se 73.3.4