

## THE EXPRESS REAL TIME FORMANT SYNTHESISER

D.A.Sinclair\*.

### ABSTRACT

A series/parallel hybrid formant synthesiser is described that is capable of real time synthesis. The system is based upon an IBM Personal Computer AT with attached PIE coprocessor and is employed as the synthesis vehicle in the IBM UKSC EXPRESS text-to-speech system. The synthesiser may be readily configured in a variety of different modes of operation: cascade alone, parallel alone, cascade/parallel, with or without voicing filters, spectral shaping filters and lip radiation filters. This paper will review the synthesiser design criteria and describe the extensive synthesis and analysis facilities available in the current implementation.

### INTRODUCTION

The synthesiser described here was originally intended as the sound generation element of the EXPRESS text-to-speech system (ref 1). It has however evolved into a more general tool suitable for many speech applications involving the generation of synthetic stimuli. The following design criteria were applied in developing the present synthesiser:

1. hybrid cascade/parallel formant synthesis architecture
2. highly configurable
3. real time performance when used with attached DSP coprocessor
4. access to internal synthesiser nodes
5. immediate signal analysis and display facilities, extendable by the user
6. flexible and efficient user interface
7. adaptable to a range of Personal Computer configurations
8. usable either standalone or as an integrated component of EXPRESS

The synthesiser model is a cascade/parallel formant synthesiser formulated largely in the traditional fashion and consequently draws on the proposals and ideas embodied in the synthesisers associated with Fant, Holmes, Klatt, Ove, Ainsworth, and others (ref 2).

Formant synthesisers have generally been constructed using either a cascade connection of resonators, a parallel connection of resonators, or a combination of the two. The cascade arrangement has the significant advantage that, for non-nasal sonorants, the relative amplitudes of the formant peaks in the synthesiser transfer function are automatically set at the correct level by setting the correct bandwidths for the formants. However a cascade arrangement cannot be used satisfactorily for the generation of fricatives and of plosive bursts - ie those sounds in which the source is above the larynx. In the parallel configuration it is possible to model the spectral characteristics of nasals and non-sonorants, however it can be difficult to produce the correct transfer function for some vowel like sounds. These problems are largely associated with difficulties in controlling the interference between the skirts of adjacent formants, the transfer function gain around dc and the overall spectral tilt of the synthesised spectrum. It is possible to use a purely parallel configuration, with suitable care, to produce a synthetic version of an utterance that is virtually indistinguishable from the original. The problems associated with the single cascade or parallel path can be largely overcome by combining the two - at the cost of increased complexity. With the advent of fast DSP coprocessors and the increasing power of Personal Computer systems the additional effort involved in computing both paths is no longer significant burden. The synthesiser described here adopts this hybrid

\*Speech Research, IBM UK Scientific Centre, Athelstan House, St Clement Street, Winchester SO23 9DR.

approach, but provides appropriate configuration controls that allow the synthesiser to be operated in either a simple cascade, simple parallel or hybrid cascade/parallel mode.

## SYNTHESISER ARCHITECTURE

The major synthesiser elements are shown in Figure 1. These include:

1. three voicing sources: (a) periodic voiced excitation; (b) random excitation used for aspiration; and (c) random excitation used for frication. Each excitation source has its own amplitude control and, if required, is passed through its own excitation spectral shaping filter,
2. a cascade formant branch: this may contain up to 10 cascade formant resonators with variable centre frequencies and bandwidths,
3. a parallel formant branch this may contain up to 10 parallel formants with variable centre frequencies, gains and optionally variable bandwidths,
4. a parallel excitation filter: this is intended to prevent any of the skirts in the parallel branch affecting the dc gain of the synthesiser,
5. a lip radiation filter: this provides the frequency transfer function associated with the lip radiation transfer function, and,
6. a large number of configurations switches which may be altered to permit operation of the synthesiser either in a series-only, parallel-only or series-parallel mode and which provide detailed control over the poles and zeros of the various fixed filters in the synthesiser.

The user controls the synthesiser by the use of a formant parameter input file, a configuration data input file (a simple example of which is shown in Figure 2) and sequences of command inputs.

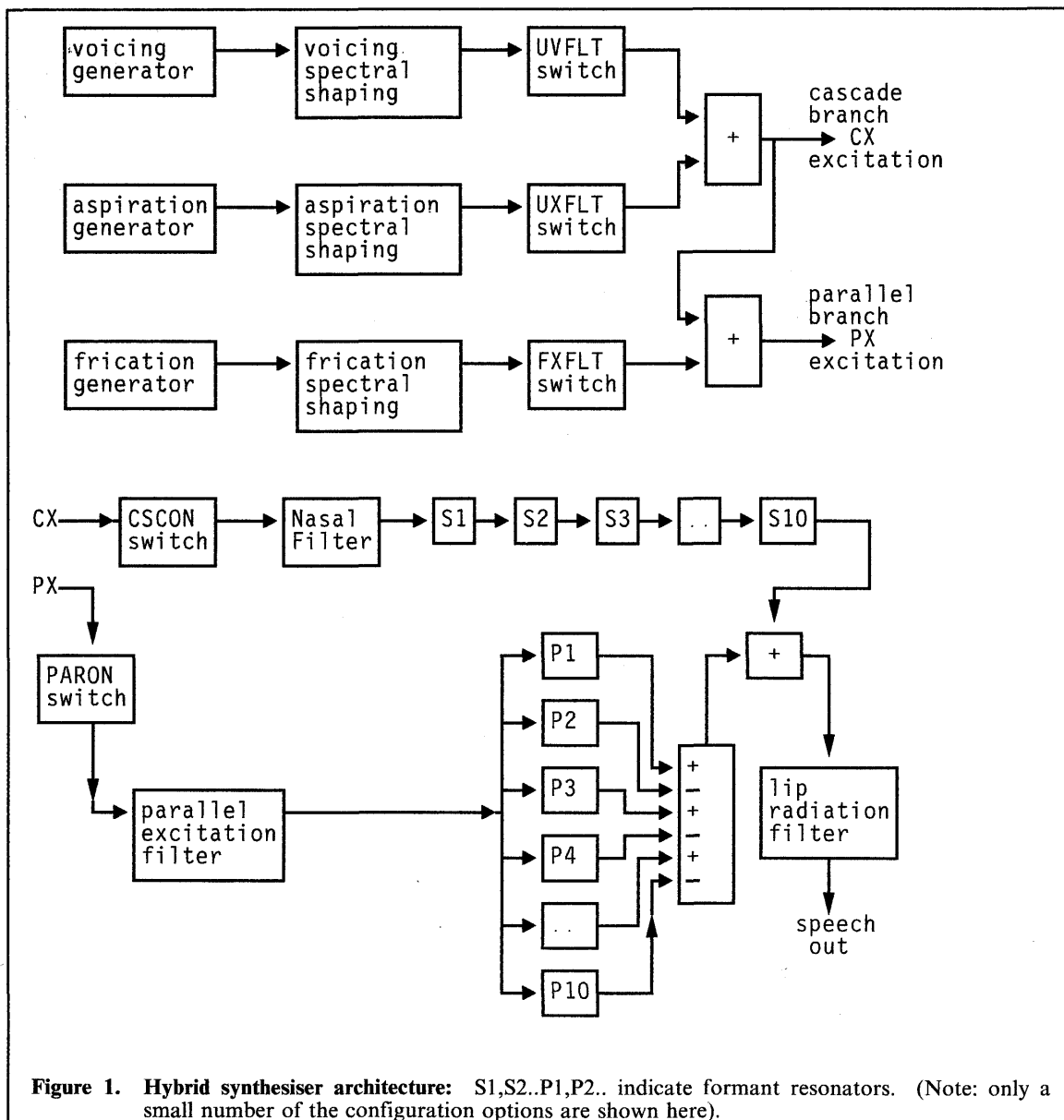
The synthesiser is implemented as a collection of software routines that run under the IBM PC Disk Operating System v3.1 on an IBM PC/XT or PC/AT. To obtain real time synthesis performance it is necessary to attach a PIE DSP coprocessor system (ref 3). A PC/AT with a PIE system provides real time synthesis of at least 6 parallel and 6 cascade formants, at a speech sampling rate of 20 KHz, and with a formant parameter frame rate of 100 Hz. If a PIE system is not present, the PC will perform the synthesis itself. The PC and PIE synthesiser implementations produce identical results.

## THE USER INTERFACE

With a view to maximising the flexibility of the system, the user is provided with an interface using full screen menus and function keys, and also with a command language interface which offers the following facilities:

- load, edit and save formant/configuration data
- plot/print and graphically edit formant parameters
- synthesis using the current formant and configuration data
- recall, play and save one or more waveforms
- plot/print waveform
- plot/print spectral sections, cepstrally smoothed spectral sections
- plot/print spectrograms
- manipulate formant data numerically
- review and modify configuration parameters
- issue conventional DOS commands
- invoke a sequence of commands stored in a file

Perhaps most important amongst these features is the ability to change the configuration or formant parameters rapidly and hear the result of the changes immediately, as well as analysing the spectral content of the regenerated speech. An example of a command sequence is shown in Figure 3.



### CONCLUSIONS

A flexible and highly configurable formant synthesiser has been demonstrated. The synthesiser is based upon an IBM Personal Computer augmented by a DSP coprocessor. Its key features are its real time performance, its flexibility and its extensive facilities for examining and manipulating the synthesiser control data and the resultant synthetic speech. We are currently using this synthesiser to evaluate the effectiveness of the intonation, stress, phonetic and cosegmentation components of the EXPRESS text to speech system.

### ACKNOWLEDGEMENTS

Many people have contributed in a variety of ways to the material presented here. A special mention is due to Peter Alderson, Gary Calder, Steve Elliott, Andrew Faulkner, Geoff Kaye, Brian Pickering, Gideon Shichman, and Richard Wright. Their help is gratefully acknowledged.

## REFERENCES

1. R A Sharman, Designing an Experimental Text-to-Speech Synthesiser, Proc Inst of Acoustics, Speech and Hearing, Vol 8: Part 7, p355-362, (1986)
2. G Bristow, Electronic Speech Synthesis (Granada, 1984)
3. G Shichman, Personal Instrument (PI) - A PC Based Signal Processing System, IBM J of Res and Dev, (March 1985)

selout	0	selector for synthesiser output node
sampint	100	sampling interval in microseconds
frmlen	100	length of frame in samples
adbscl	20	data offset for all dB conversions
vxon	1	enables the voicing source when 1, disables it when 0
uvratio	0	ratio of amplitude of unvoiced to voiced excitation
pcratio	60	ratio of excitation of parallel to cascade branch
cscscl	0	scaling factor applied to the cascade branch (in dB)
ncform	1	number of resonators in the cascade branch
bp1	100	fixed bandwidth of parallel formant 1 (in Hz)
bp2	100	fixed bandwidth of parallel formant 2 (in Hz)
varbnd	0	selects fixed (0) or variable (1) parallel bandwidths
vxflt	0	include voicing filter if non zero
fgf	0	centre freq. of voiced spectral shaping pole (in Hz)
bgf	100	bandwidth of voiced spectral shaping pole (in Hz)
vximpk	100	absolute peak impulse response of voicing filter pole (%)
paron	0	include parallel branch if non zero
parflt	1	include parallel branch spectral shaping filter if non zero
bpz	100	bandwidth of single real zero in parallel branch filter (Hz)
lipflt	1	include lip radiation filter if non zero
blz	0	bandwidth of single real zero in lip radiation filter (Hz)

**Figure 2. Synthesiser configuration file format and default values:** Only a subset of the 50 or so configuration variables are shown here.

```

* load the configuration and formant parameter files for the utterance
* 'good morning how are you' and synthesise the utterance
load configuration goodmorn
load formants goodmorn
synthesise
* display the cepstrum of the speech section starting at sample 3000
* and take a hard copy of the plot
display cepstrum from 3000 print
* turn off the aspiration source and resynthesise the utterance
set uxon 0
synthesise
* attempt to convert this male data into a female speaker
* raise the fundamental frequency by 100 Hz
f0 = f0 + 100
* increase the formant frequencies by 20%
f1 = f1 * 1.2
f2 = f2 * 1.2
f3 = f3 * 1.2
f4 = f4 * 1.2
synthesise
* plot spectrogram of modified speech
display spectrogram

```

**Figure 3. An example of a command sequence:** This command sequence generates male and female versions of the same utterance and displays various (graphical) spectral analyses of the generated speech. The lines starting with an '\*' are comments and indicate the functions performed at each stage of the command sequence.