BMGC08-1: Musical dictionary with voice access to MP3 archive

```
Project: BMGC08-1.
Area: Digital media processing.
Supervisor: Barry Cheetham
Difficulty grading: SH/INF=F, BM=F, CM=F.
Max number of students who can do this project: 1
```

How many times have you had a tune in your head and wondered what it is called? For people who can read music there is a host of ‘musical dictionaries’ in the central library in Manchester. For those who can’t, there is an amazing musical dictionary that only needs to know the progression of the first 10 notes or so to identify the theme. You have to specify the opening of Beethoven's 5th symphony, for example, as "S(ame) S S D(own) U(p) S S D". There is no timing or rhythm required, just the frequency progression. Also the number of notes is generally fixed.

This method of coding themes is known as ‘Parson’s code’ and previous final year projects have shown that it can be used in a computer-based dictionary with real time voice input. The user sings the theme into a microphone and the application identifies the individual notes and their changes in frequency, to produce the coding. The computer then compares the note progression with the contents of a data-base and displays or speaks the name of the tune. If an MP3 recording of the tune is accessible, the program can then play it automatically. The dictionary may be therefore used to provide convenient access to the music in your own private MP3 collection .

A previous student investigated the use of the ‘Levenshtein’ or ‘edit’ distance as the basis of comparison thus allowing for the inevitable errors from the singer or the computer producing occasional wrong, missing or extra notes. Modified versions of Parson’s code have been devised by previous students and the means of establishing a data-base of themes has been investigated. It has been shown that Parson’s code may be improved upon by:

(a) introducing ‘long’ and ‘short’ notes as some indication of timing and rhythm.
(b) varying the number of notes that the singer is required to produce.

Three possible innovations may be investigated: the introduction of timing/rhythm, the use of a variable length code and/or the provision of a version for use over the Internet or the telephone.

Attending Comp30291 'Digital Media Processing' in the first semester and learning MATLAB may be an advantage.

Main objectives:

(i) A survey of previous projects and commercial and other applications available, for example, on the Internet.

(ii) The construction and testing of a prototype musical dictionary with Parson’s (or other) coding of a data-base of themes. Key-board input could be used for this initial testing.

(iii) A program for inputting a sung theme from a microphone, segmenting this into individual notes and analyzing each note to determine its fundamental frequency. Standard Parson’s (or other) coding may then be produced. A different form of coding may be investigated and a challenge may then be the requirement for key transposition which is eliminated by Parson’s code.

(iv) An implementation of one or more of the innovations mentioned above and investigation of its/their effectiveness.

EQUIPMENT: PC with sound I/O

BMGC08-2: Speaking or singing voice trainer

```
Project: BMGC08-2.
Area: Digital media processing.
Supervisor: Barry Cheetham
Difficulty grading: SH/INF=F, BM=F, CM=F.
```

Max number of students who can do this project: 2
Speaking or singing voice trainer

These are two possible projects concerned with the human voice as used for speaking and/or singing. A study of the mechanisms by which humans speak, sing, convey emotion and generally communicate reveals an anatomical process that, in outline, is fairly simple to understand and can be modelled by a computer program. An understanding of sound perception and the mechanism of the ear allows the computer modelling of voice to be made even more effective. Mobile phones use such a model to produce the speech you hear and this allows speech to be transmitted at low bit-rates to preserve radio spectrum. However, an understanding of the human voice can have much wider applications than bit-rate compression.

Mastering some of the basic principles of human speech and its perception is the goal of the first part of each of these projects. Attending Comp30291 'Digital Media Processing' in the first semester and learning MATLAB would be a definite advantage.

The two possible projects are as follows:

'<u>Speaking voice trainer</u>'. Voice training as used by actors and politicians to enhance their presence and persuasiveness may be studied and a computerised voice analysis package may be used to help this process. A famous example is the training Mrs Margaret Thatcher, British Prime Minister from 1979 to 1990, received to lower the pitch of her voice to give herself greater 'gravitas' and authority. Voice experts teach that the voice must 'resonate' deep within the body (as with Winston Churchill) rather than within the mouth (David Beckham) or the nose (Michael Cain). Measuring the pitch of the voice by computer is an interesting and achievable goal, and the detection of body resonance may be possible by examining some Fourier spectral graphs.

The inspiration for the speaking voice trainer came from a comedy program on BBC7 called " I'd know that voice anywhere " presented by Alister McGowern (First broadcast 19 Jan 2000 but repeated many times). As well as raising the points mentioned above, it discusses ventriloquism and features Rory Bremner and others who can mimic other voices. 'Voice morphing' and the creation of cartoon voices was not mentioned but is related. Anyone interested in these aspects of voice analysis and/or synthesis could adapt the 'speaking voice trainer' in this direction. Downloading the BBC7 program would be a good start (and a good laugh).

The main aim of the ‘speaking voice trainer’ project is to develop an application that provides feedback to a person wishing to train his or her speaking voice. Specific objectives are as follows:

(i) To consult professional publications on speaking voice training and, if possible, to contact staff in the University or elsewhere.
(ii) To study the physical mechanisms responsible for voice production.
(iii) To develop tools for analysing sections of the voice to determine pitch, intonation and other characteristics.
(iv) To test the analysis tools by applying them initially to famous recordings, for example of Margaret Thatcher, Winston Churchill and Michael Cain.
(v) To develop the tools into a prototype ‘speaking voice trainer’ providing a suitable ‘graphical user interface’ to the user and convenient feedback from the analysis.
(vi) A suggestion is that the user be asked to repeat a phrase or sentence in a particular style and be informed how close he/she is to the style.
(vii) The usefulness of the application must be evaluated with some subjective testing.

The willingness to come to grips with some basic digital signal processing techniques and to do some high level language programming with GUIs is needed.

<u>'Singing voice trainer'</u>. This project has much in common with the speaking voice trainer and would be ideal for somebody who sings or has sung in a choir. The aim is to develop and test a computerised 'singing voice trainer' and to try it out on a real singer to determine whether it is of any use at all. The objectives are as follows:

(i) To develop software that checks that a singer is singing the right notes at the right time. It could have a copy of the score (with words), a few bars would be sufficient to begin with, and make a recording of a person's attempt to sing the words to the notes. Segmenting the person's recording into individual notes presents a challenge. Then the program could 'Fourier analyse' the frequencies sung and check that they are correct.

(ii) The timing may now be carefully analysed, checking for late or early entries and the singing of quavers as crotchets, for example (a common mistake).

(iii) The software could then go further, for example by looking out for the dreaded early or late consonant. Since consonants like 's' and 'sh' are unvoiced and are therefore not periodic (the waveform is more random or noise-like) they are unable to carry a pitch. Vowels carry pitch as they are periodic. Mistimed consonants (especially 's's) are an anathema to conscientious choral directors and singers must practise to get this right.

(iv) Resonance and projection are as important with singing as with speaking, if not more so, and an analysis of these aspects of voice recordings could be made.

(v) Professional singers maximise sound volume by modifying vowels to make vocal tract resonances coincide with pitch harmonics and it would be fascinating to be able to analyse this process and maybe help a singer to emulate it.

(vi) The prototype computerised 'singing voice trainer' should be provided with a ‘graphical user interface’ to allow it to be tried it out on a real singer to determine whether it has any potential.

BMGC08-3: An application for teaching Westerners to recognize and reproduce tones in Chinese

```
Project: BMGC08-3.
Area: Digital media processing.
Supervisor: Barry Cheetham
Difficulty grading: SH/INF=F, BM=F, CM=F.
```

Max number of students who can do this project: 1

For many Westerners studying Chinese, the recognition and reproduction of the four basic tones in Mandarin, and the many more in Cantonese, are a source of great difficulty. Some ‘wai(4)-guo(2)-ren(2)’ simply cannot discern the differences as they are so attuned to other aspects of intonation. Reproducing the tones can be an even more difficult problem. Imagine the frustration of a poor old Englishman who has learnt his vocabulary only to find that no Chinese person has a clue what he is saying because the tones are all wrong.

This project is for a Computer Scientist with knowledge of Chinese, preferably Mandarin, and an interest in developing interactive educational software which includes sound analysis and reproduction.

Mastering some of the basic principles of human speech and its perception will be required and attending Comp30291 'Digital Media Processing' in the first semester and learning MATLAB would be an advantage.

The main aim of the project is to develop an application that demonstrates the use of Mandarin tones to a Westerner, analyses his/her attempts to reproduce Chinese words and sentences correctly, and provides feedback and suggestions for corrective action. Specific objectives are as follows:

(i) Consult professional publications on the teaching of Chinese to foreigners and, if possible, to contact language teaching staff in the University or elsewhere to gain from their experience. The teaching methods adopted by Chinese teachers of foreigners and the computer software packages currently available should be considered and references on automatic speech recognition as applied to tonal languages should be researched as background.

(ii) Study the physical mechanisms responsible for voice production.

(iii) Develop tools for segmenting recorded sections of spoken Mandarin to isolate single words. Then devise the means of measuring the fundamental frequency of the voice at suitable intervals to determine how the pitch of the voice varies within each word.

(iv) Given a 'pin-yin' representation of the segment of spoken Chinese, the application should then be enabled to impose the standard symbols over the vowels to indicate the tones (right or wrong) as detected by the application. These can be compared with the correct tones and imperfections in either the software or the speaker can be identified.

(v) Analyse the tones of a clear native speaker of Chinese to verify the software before moving on to a trainee.

(vi) Build the tools into an application which produces written Chinese (probably in pin-yin), invites the user to speak the words or phrase, and then analyses the result. The correct pronunciation could then be given as sound output.

(vii) Introduce the option of producing Chinese words and inviting the user to identify the tones via a keyboard input.

(viii) Develop a suitable 'graphical user interface' with convenient feedback from the analysis.

(ix) Subjectively test the software on a user and, by seeking the advice of a professional teacher, give an assessment of its potential for further development.

BMGC08-4: Digitising speech for storage and transmission

```
Project: BMGC08-4.
Area: Digital media processing.
Supervisor: Barry Cheetham
Difficulty grading: SH/INF=F, BM=F, CM=F.
```

Max number of students who can do this project: 1

Wired telephone systems generally restrict speech to a frequency range of about 300-3.4kHz. This band-limited speech may be digitized at 64kb/s by sampling it at 8000 samples per second and representing each sample by an 8-bit number, with non-uniform (A or mu law) quantization steps. Some loss of quality is incurred by this digitization process, and the first aim of this project is to implement the process and apply it to stored speech segments to investigate the loss in quality. A comparison may be made with higher quality (CD quality) sound sampled at 44.1 kHz. For many applications, such as mobile telephony, 64 kb/s is too high a bit-rate. The project will then proceed to program and investigate a range of techniques capable of digitizing speech at lower bit-rates. The evaluation of speech quality by a technique similar to 'MOS' scoring will be investigated.

For this project, understanding some of the basic principles of speech and sound perception will be required. Attending Comp30291 'Digital Media Processing' in the first semester and learning MATLAB would be an advantage.

The main ‘speech processing’ objectives are as follows:

(i) Given a number of data files, each containing about 3 minutes of speech sampled at 44.1 kHz, 16 kHz and 8 kHz with 16 bits/sample uniform quantization, the first objective is to develop a program which reads these files in short sections to display and analyze them. The important features of speech signals should be documented. A wave-analyzer will be useful here.

(ii) A simulation of the effect of a 300 Hz high-pass filter and 8 bit/sample A-law or mu-law quantization to produce speech which is indicative of telephone quality digitized speech.

(ii) A fully documented program for demonstrating the effect on speech quality of reducing the bit-rate from 64 kb/s to 32 kb/s and then to 16 kb/s simply by reducing the number of quantization levels per sample from 64 to 16 then 4 . Uniform, non-uniform and code-book quantization should be considered.

(iii) An implementation and demonstration of linear and/or adaptive delta modulation for digitizing speech at 16kb/s.

(iv) An implementation of a simple linear predictive coder for digitizing speech at 2.4 kb/s

(v) Improvement to the simple linear predictive coder or an implementation of a different coding technique for comparison. An adventurous student could do some research here, for example by investigating a ‘variable bandwidth speech coder’ which increases the bandwidth for unvoiced speech

(vi) An investigation of the use of spectral analysis and 'objective MOS type' scoring to measure speech quality.

BMGC08-5: Digitising music for storage and transmission

```
Project: BMGC08-4.
Area: Digital media processing.
Supervisor: Barry Cheetham
Difficulty grading: SH/INF=F, BM=F, CM=F.
```
Max number of students who can do this project: 1

The project will consider the principles of MP3 compression as applied to music and speech. A 'demonstration' MP3 encoder and decoder could be developed to help people to better understand the theory of MP3 sound compression.

For this project, understanding some of the basic principles of speech and sound perception will be required. Attending Comp30291 'Digital Media Processing' in the first semester and learning MATLAB would be an advantage.

The objectives are to develop a 'demonstration' MP3 encoder in MATLAB and investigate its performance.

BMGC08-6: Automatic hearing tester, equalizer and hearing loss simulator

```
Project: BMGC08-6.
Area: Digital media processing.
Supervisor: Barry Cheetham
Difficulty grading: SH/INF=F, BM=F, CM=F.
```
Max number of students who can do this project: 1

The first aim is to design, build and test a unit for testing the hearing of a person who fears his or her hearing may have been damaged. The unit will generate a sequence of tones of increasing frequency. The intensity of each tone will be very low initially, and will be gradually increased until the patient indicates that he or she has heard the sound, by pressing a button. The unit will record the threshold intensity of audibility for each frequency and hence will allow a frequency-response graph to be drawn for the hearing of the person being tested. The unit will be implemented as a software package for a personal computer equipped with a good quality sound card. It should be calibrated against equipment currently in use, and an instruction manual should be produced.

Once the hearing tester has been implemented, an 'equaliser' may by designed to allow a hearing impaired person to listen to music from a CD with the gain-response modified to equalise frequency dependent losses of sensitivity as measured by the tester. Also a "hearing impairment simulation" filter may be designed to allow a person with normal hearing to experience the type of hearing loss measured by the hearing tester. A music teacher needed such a facility recently to help with the task of trying to set a fair musical examination for a severely deaf, but highly talented young musician.

The project will require high-level language programming, graphics and the design of a convenient graphical user interface. A visit to a local hospital routinely producing audiograms will, if possible, be arranged. Attending Comp30291 'Digital Media Processing' in the first semester and learning MATLAB may be an advantage.

The objectives are as follows:

(i) A working prototype system capable of testing the hearing of a subject and plotting an "audiogram" on the PC screen
(ii) Careful optimisation of the software, especially with regard to the signal generation. Switching the output signal on and off must be done gradually rather than suddenly to avoid transients being heard.
(iii) An 'equaliser' allowing a hearing impaired person to listen to music from a CD with the gain-response modified to equalise frequency dependent losses of sensitivity.
(iv) A 'hearing impairment simulation' filter to allow a person with normal hearing to experience the type of hearing loss measured by the hearing tester.
(v) A convenient graphical user interface for the hearing tester, equalizer and hearing loss simulator.