Evolution of Adult Male Oral Tract Shapes for Close and Open Vowels

David M. Howard
Intelligent Systems Research Group
Department of Electronics
University of York
Heslington, York, YO10 5DD
+44 1904 432405
dh@ohm.york.ac.uk

Andy M. Tyrrell
Intelligent Systems Research Group
Department of Electronics
University of York
Heslington, York, YO10 5DD
amt@ohm.york.ac.uk

Crispin Cooper
Intelligent Systems Research Group
Department of Electronics
University of York
Heslington, York, YO10 5DD

ABSTRACT
In this paper, we describe an experiment to evolve oral tract (mouth) shapes for a set of vowels for two adult males. Target vowels were recorded in an acoustically anechoic room along with the output from an electrolaryngograph, which monitors vocal fold vibrations electrically via two electrodes placed externally on the neck of the speaker at larynx level. Physical modelling digital waveguide synthesis using a two dimensional virtual oral tract was employed in the experiments. A population of 50 randomly shaped oral tracts were set up at the start of the evolution procedure, and the fitness of each was tested with respect to each of the target vowels. The resulting vowels are compared spectrally alongside the evolved oral tract shapes. The results for open vowels were close to the targets whereas those for close vowels were not. It is suggested that this is because of the associated narrow constriction to which special attention needs to be paid in the future.

Categories and Subject Descriptors
J.3 [**Life and Medical Sciences**]: Health, Medical information systems; I.2.6 [**Artificial Intelligence**]: Learning, Parameter learning

General Terms
Algorithms, Measurement, Experimentation, Human Factors.

Keywords
Evolution, bio-inspired computing, oral tract shape, vowel synthesis.

1. INTRODUCTION
Electronic voice synthesis has a number of applications today, and it is perfectly possible for it to produce a highly *intelligible* speech output. However, its output is rarely, if ever, mistaken for the voice of a human; that is, it rarely sounds *natural.* Sorting out and understanding which elements of human speech production contribute to our perception of naturalness is a key goal of today's speech research.

In this work, physical modelling synthesis techniques are applied to speech synthesis. Inspiration is taken from the experience and success gained with the successful use of physical modeling synthesis in electronic music synthesis[1], where outputs are often described as being *natural sounding*, or *organic.* This is in comparison with the outputs obtained from more traditional electronic musical instruments that make use of techniques such as additive, subtractive, wavetable or sampling synthesis [e.g. 2], which are often described as being *cold* or *lifeless* by players and audience alike, and listeners find that the more tradition music synthesis methods become less interesting with extended exposure [3].

Physical modelling synthesis has the added attraction for performing musicians that it makes use of virtual structures that are directly related to those that make up physical musical instruments such as strings (one dimensional - 1D), membranes (2D) and blocks (3D) [1,3]. The construction of virtual instruments is therefore a very intuitive high-level process that does not require any practical knowledge of physics or acoustics. The synthesis system is controlled by parameters that are also highly intuitive because they relate directly to how humans interact with real acoustic instruments, for example, by striking, plucking or bowing [4]. Tactile haptic feedback has also been employed to give the player a physical sense of actually playing an acoustic vibrating instrument [5].

If physical modeling synthesis can enhance the naturalness of sound in the musical domain, it seems very plausible to apply it in the voice synthesis arena.

The main method used for speech synthesis is formant synthesis, which is based on the source-filter model of Fant [6]. Whilst Holmes has shown that formant synthesis can produce an output which is indistinguishable from the natural original [7], this was only achieved following a painstaking synthesis by analysis approach over many months. For this work, detailed comparisons were made between time-frequency-amplitude spectrograms [8] of the original and synthesized versions to enable changes to be made to the synthesis parameters to increase the acoustic similarity between them. Although this does demonstrate that similarity can be obtained, the application of the source-filter model is compromised because spectral changes that are in reality due to the source are compensated for by changes made to settings of the filter. This offers no gain in terms of a better

GECCO'07, July 7–11, 2007, London, England, United Kingdom.

understanding of what is needed to improve naturalness in voice synthesis. In addition, it has more recently been demonstrated that there are important non-linear interactions between the source and the filter that the original source/filter model does not take into account, and it is likely that these make an essential contribution to the naturalness of the perceived output [9].

Thus a move from a formant synthesis technique to a new method, physical modeling, seems to be therefore appropriate. Effects such as non-linear source-tract interaction will emerge as a consequence of the application of the physical modelling process itself. Physical modelling requires that the physical attributes of the system can be specified appropriately. These are usually gained from the outputs obtained from Functional Magnetic Resonance Imaging (fMRI) of the vocal tract [10].

However, fMRI data acquisition is hampered by a number of practical factors. The fMRI environment is very acoustically noisy which means that the acoustic feedback paths to the ears of the subject are compromised. Spoken or sung sounds produced whilst lying in the machine are rarely representative of the subject's natural output.. The subject has to lie supine in the machine which is not a normal speaking position from the point of view of posture, breathing or vocal tract soft tissue orientation due to gravitational forces acting perpendicularly to normal. It also takes a considerable time (in speech production terms) to acquire an fMRI image; and a subject has to hold an articulatory gesture still for some seconds. Voice data gathered in an fMRI experiment is therefore potentially compromised in terms of the degree to which it represents normal vocal behaviour, and an alternative method would be potentially beneficial.

The approach adopted in this paper for establishing oral tract areas for different vowels is novel. It is based on computing techniques that are inspired by the principles of evolution, and therefore offers a direct alternative to placing subjects in an fMRI machine or LPC analysis. Oral tract shapes are evolved and tested using the two dimensional physical modeling synthesis techniques of Mullen et al [11,12], who have demonstrated that a digital waveguide mesh (DWM) provides a highly successful method for vowel synthesis. Currently, only oral sounds can be synthesized because there is no nasal tract (nose), but this is all that is required for the synthesis of isolated vowels.

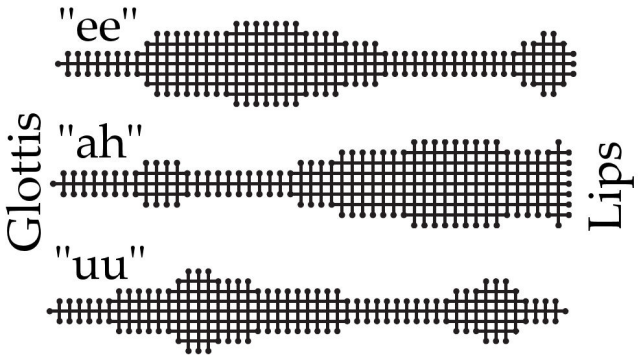

Figure 1: Two dimensional oral tract area waveguide mesh element representations of the vowels in *beat* ("ee"), *Bart* ("ah") and *booed* ("uu") derived from fMRI data for an adult male.

The vocal tract areas used are derived from fMRI images of the oral tract of n adult male [10]. Figure 1 shows two dimensional oral tract area waveguide mesh element representations for the vowels in *beat* ("ee"), *Bart* ("ah") and *booed* ("uu"). The differences in oral tract area shapes are clearly visible. Physical modeling synthesis applies an excitation at the glottis (larynx or voice box) end of the mesh for voiced vowels. This is the commonly used LF model [13] implemented as a wavetable oscillator to enable its f0 value to be readily altered to model pitch changes. The output is summed from the elements at the lip end of the mesh.

When synthesizing steady oral vowels as in this experiment, the mesh shape remains constant. For continuous speech synthesis however, the shape of the mesh has to be varied dynamically in order to enable the synthesis of sounds such as diphthongs (e.g. the words *eye* or *ear*) for which the oral tract area shape changes dynamically. To enable the oral tract area to be varied dynamically, Mullen et al [14] implemented an *impedance contour map* along the length of the DWM representation of the oral tract to enable the mesh shape to be changed without the need to remove or replace mesh elements during synthesis, which ensures that there are no audible discontinuities.

There is a method for calculating the shape of the oral tract using an extension of linear predictive coding (LPC) [15]. Rossiter et al. [16] implemented a real-time oral tract area display based on LPC analysis as part of their ALBERT system [17] for real-time visual feedback for training professional singers. More recently, this display has been incorporated into the WinSingad PC Windows-based real-time singing training system [18, 19]. The oral tract results derived using the LPC technique assume an all-pole vocal tract acoustic frequency response, which is not always true. It is not a physical model, and it can produce oral tract shapes that are not necessarily unique, since more than one oral tract tube shape can produce a particular sound [20].

The main aim of the work is to establish whether or not an evolutionary computation technique can successfully evolve oral tract shapes that can produce plausible sounding vowels when compared to the natural originals. If successful, it might offer improvements over LPC-based oral tract area estimation.

2. RECORDING THE TARGET VOWELS

In order to evolve oral tract shapes, target vowels and some form of input signal are required for each speaker and each vowel. It is not possible in practice to gain direct access to the glottal pressure wave during the production of a vowel, and some other means is needed to record an input signal. Here, we are dealing with voiced vowels, that is, vowels in which the vibrating vocal folds in the larynx provide the acoustic excitation to the oral tract. The electrolaryngograph [21] enables vocal fold vibration to be monitored non-invasively, and it provides an output waveform (Lx) that relates directly to the nature of the vibrating vocal folds. Although the Lx waveform is not a representation of either the glottal flow or glottal pressure waveform, it is an audio waveform that is directly related to the source of voiced sounds: vocal fold vibration.

The vowels were recorded in the acoustically semi-anechoic chamber belonging to the Department of Electronics at the University of York, UK. Two male and two female adult subjects

were recorded producing the eight vowels in the words: *boot*, *beat*, *bet*, *Bart*, *bat*, *but*, *Bert*, and *bought*. Each vowel was produced thrice, using a rising, falling and flat pitch contour. The Lx waveform from the electrolaryngograph was recorded simultaneously with the speech pressure waveform from a Sennheiser MKH-20 omni-directional microphone and an RME quad microphone amplifier. These two audio channels were recorded in stereo using an Edirol R4 hard disk recorder in PCM *.wav* format at 16bits resolution and a sampling rate of 44.1kHz. The *.wav* data were transferred digitally to a PC computer for processing.

The microphone recordings of the vowels provided the targets during the evolution process and the simultaneously recorded Lx waveforms provided the associated inputs for the physical modelling synthesis process. It should be noted that Lx waveform is not really appropriate as an input waveform for speech sound production, since it is not directly related to the glottal pressure waveform. However, it does contain many of the natural characteristics of the excitation for voiced speech, and it is very easy to obtain compared with any methodology that is available for measuring glottal pressure, such as inverse filtering. If it works as an input waveform in the evolution process and a satisfactory output waveform is produced, then it will indicate additionally something of the power of the evolution technique in terms of how it is able to take account of make of such differences.

3. EVOLVING ORAL TRACT SHAPES

Bio-inspired computing enables techniques such as genetic evolution to be employed as a computational tool in a number of application areas that involve design and optimization [23-25]. A genetic evolution computational technique requires that there is some way of testing a result from a particular member of a generation, and this is usually achieved by means of deriving an output based on the application of an appropriate input.

In the case of evolving oral tract shapes, an input and target output waveform are required, which in this case are the Lx waveform and vowel speech pressure waveform respectively, and the input is applied to a physical modeling synthesis model, the output from which can be compared to the natural target waveform. Further details in relation to the methodology can be found in [22] and its application for sung sounds in [26].

In order that oral tract shapes can be evolved, a definition in the form of a genome is required, Figure 2 shows how the genome of the oral tract is defined for an arbitrary oral tract shape. The genome itself indicates the number of mesh elements on either side of the midline at that point. One element must remain at each point to enable the acoustic pressure to propagate along the tract – we are not dealing with complete constrictions of the oral tract in this work. The example shown in figure 1 therefore has the unique genome: 1210202101021121. Since this work is only involved with oral vowels, the tract can never be fully constricted, so there is: (1) a minimum of one element at every position along the mesh, and (2) an odd number of elements across the mesh at all positions.

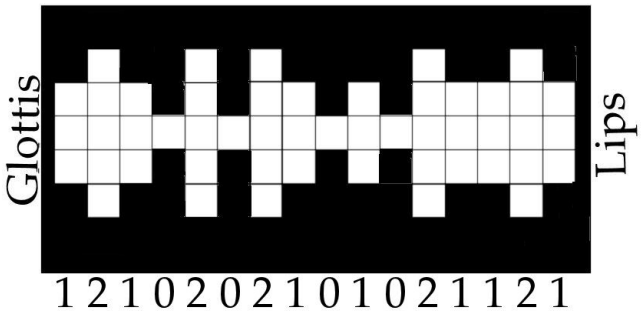

Figure 2: Oral tract genome in which each digit indicates how many elements there are either side of the midline, around which the mesh is symmetric. In this example, the genome is: 1210202101021121.

The oral tract has a length of 16 mesh elements and a minimum and maximum width of 1 and 5 elements respectively as illustrated in figure 2. The initial population of genotypes or individuals is established by setting up 50 oral tract shapes with randomly shaped oral tract waveguide meshes. The Lx waveform is applied at the glottis end and the output monitored at the lip end and compared with the target natural original in order to evaluate the genotypes for their fitness as a solution. The fitness evaluation is based on comparing the difference between the amplitude frequency spectrum of a 50 ms window taken during the steady-state portion of the target vowel and the output from the mesh.

The ten (20%) genotypes that have the closest spectral match are deemed to be the fittest are then copied to the next generation, where they are used as the basis for offspring creation. The remaining forty (80%) genotypes are discarded and thereby excluded from the next generation. Mutation and crossover operators are defined which operate on one or two genotypes respectively to create new members of the new generation from the retained ten fittest genotypes. The process is iterated until the population converges to a solution, which is established by fitness results that remain stable when compared to the target. In this case, the evolution process was run over 50 generations, and it was repeated twice with a new random set of starter genotypes each time, giving a total of three runs for each vowel.

The target vowels selected for the experiment were those uttered with flat intonation contour, since then the section selected for fitness evaluation would have one less degree of change (pitch) associated with it. A comparison was carried out with vowels spoken on a rising intonation contour and the results were very similar. Since this was a new application for evolutionary computation techniques, various forms of modifications were tried to attempt to improve the overall results. This is a form of human intervention which was based on informal listening to the final outputs to decide whether or not they would pass phonetically as the target vowel. The results of this human intervention by informal experimenter listening are given in Table 1 using a three point scale as follows to indicate the extent to which the resulting vowels from the three evolution runs pass phonetically as the target vowels: 1: all three pass; 2: one out of three pass; 3: none pass.

Table 1: Results for the two male speakers (M1, M2) for the eight vowels uttered on a flat intonation contour. Numeric data indicate how the results over three evolution runs were perceived in terms of passing phonetically as the target vowel during the human process (1: all three runs pass; 2: one out of three runs pass; 3: no runs pass). The effect of modifications (A) and (B) are also shown where there was an observed improvement (A+, B+) or detriment (A-, B-). (Modification B was only applied to vowels *boot*, *beat*, and *bought*.)

Vowel	M1		M2	
boot	3		3	A-,B-
beat	3		3	B+
bat	1		2	A+
but	1		2	A+
bet	2	A-	2	A-
Bert	1		3	A+
Bart	3	A-	1	A-
bought	1	B-	3	B-

Two modifications (denoted as A and B) were implemented in an attempt to improve on these results.

Modification A was originally implemented as a way of speeding up the fitness evaluation process by halving the portion of the target signal that was used (25ms rather than 50ms). It turned out that some of the evolved vowels for one subject (M2) were improved (shown by A+ entries in table 1). The improvement was not universal though. There were none for subject M1, and some of the resulting vowels were worsened (shown as A- entries in table 1).

Modification B was inspired by the observation that the results for those vowels that are produced with a narrow articulation of the tongue with the palate (particularly *beat* and *boot*, but also *bought*), were rated "3" for both subjects (see table 1). It was hypothesized that this was due to the low spatial resolution of the oral tract model. A four-fold increase in mesh resolution was implemented, increasing the search space by approximately one million, but a good solution was not then not achievable and initial evolution runs failed to produce any scores of "1" or "2". Then the width of the mesh at the ten predefined points along its length was limited to 1, 5, 11 or 17 mesh nodes, corresponding to 2.8mm, 13.8mm, 30.3mm and 46.8mm respectively. This was only employed for *boot*, *beat*, and *bought*, and the results are shown in table 1 as "B+" or "B-" where a difference was observed. One vowel (*beat*) for subject M2 was improved, and the vowel in *bought* was worsened for both subjects.

4. RESULTS

Results are presented in two forms: oral tract area shapes and spectral comparisons between the evolved vowels and the targets. To guide the comparisons, the results of a listening test for the evolved vowels for the two male subjects are listed in table 2. In this listening test [27], twelve subjects were asked to decide whether each of the evolved vowels would pass phonetically as its original target, thereby providing an indication as to how close the vowels were in perceptual terms. The results are averaged across all 12 listeners, and they can be compared to the oral tract shape and spectral data presented below.

The listening test results confirm the scores given during the human intervention process (see table 1), with high listening test results appearing for vowels which gain a human intervener score of "1" or "2", and low scores for those vowels with a "3". It is hypothesized that evolved vowels with high scores should exhibit spectra that are close to those for their target originals.

Oral tract areas resulting from the evolution process followed by long-term average spectra for the evolved and target vowels are presented in the next sections.

Table 2: Average responses (%) from 12 listeners who were asked whether each evolved vowel would pass phonetically as the target original for the two male speakers (M1, M2).

Vowel	M1 (%)	M2 (%)
boot	0.0	7.7
beat	23.1	7.7
bat	100.0	100.0
but	100.0	76.9
bet	61.5	92.3
Bert	100.0	53.8
Bart	76.9	92.3
bought	100.0	15.4

.

4.1 Oral tract areas

Figure 3 shows plots of the evolved oral tract areas for both subjects plotted in terms of mesh width genome value against distance from glottis to lips. The glottis and lips are at the left- and right-hand side of the figure respectively. The results for both subjects are plotted together for each vowel to enable direct comparison. It is not possible to offer a direct comparison with fMRI data, since the highly acoustical noisy levels associated with fMRI machines makes it impossible to make a useful audio recording.

What the plots do serve to provide is an indication of consistency between the results for the two speakers. It is reasonable to expect that the oral tract shapes for two male speakers should be essentially similar; the tongue will adopt a similar position for all speakers [8].

The oral tract shapes that exhibit the closest matches in terms of a general similarity in shape between the two male speakers are for the vowels in: *bat, but, bet, Bert,* and *Bart.* The other three (*boot, beat, bought*) exhibit greater differences. Of particular importance are the relative positions of constrictions in the tract, since these serve to move the formants around [28], and these results suggest that *boot, beat, bought* may have differences between their formants for these two speakers.

4.2 Spectral comparisons

In order to gain an impression of the success of this technique for some of the vowels, long-term average spectra (LTAS) are presented and considered. Since the vowels themselves were produced in isolation, the LTAS will be very similar in shape to the short-term spectrum which was used as the basis for the fitness function evaluation. In each case, the LTAS plotted is taken from the best evolution run, whether this be the original or modification A or B as indicated in table 1.

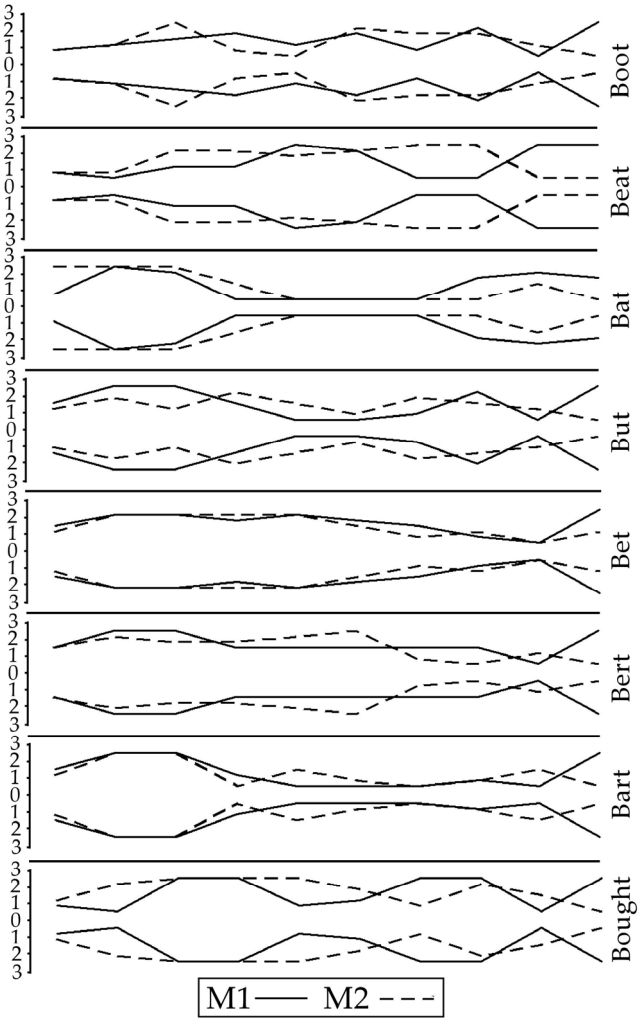

Figure 3: Evolved oral tract shapes for both male subjects (M1, M2) for the eight vowels (glottis on the left, lips on the right). The y-axis is calibrated in mesh elements.

Figures 4 and 5 show the LTAS for the subjects M1 and M2 respectively for each of the eight vowels. LTAS plots for the target original and evolved vowels are plotted on the same axes to enable direct comparison.

Observation of figure 4 for speaker M1 indicates that the spectral match is particularly good for the vowels in: *bat* and *but,* and the formants are very well matched for *bought.* These scored 100% in the listening test along with *Bert*, for which the overall match is not as clear but the lower formant peaks are well aligned.

The vowels with the narrowest constriction, *boot* and *beat*, are lacking spectral detail, and neither has evidence of the lowest formant peak (F1). This will make identifying them difficult, as evidenced by the listening test results (see table 2).

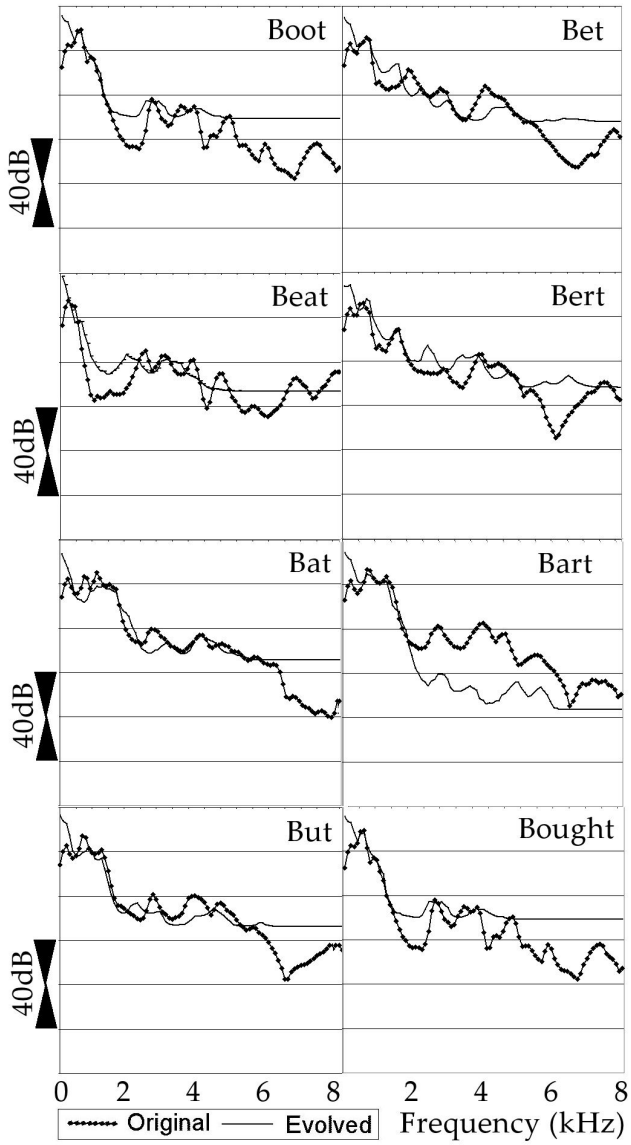

Figure 4: Long-term average spectra of the target original and evolved vowels for subject M1.

The vowel in *bet* exhibits peaks that are not apparent in the original, but the lowest two formants are quite well matched, so there will be some perceptual evidence of its phonetic origin as indicated in the listening test results. The remaining vowel in *Bart* is deficient in the overall matching of spectral trend by nearly 40 dB in places, which will somewhat hamper its phonetic identity, although its formant peaks are essentially well matched.

The results for speaker M2 are shown in figure 5. Here, the vowels in *bat, bet* and *Bart* scored well in the listening test (see

table 2), and each of these exhibits a good spectral match to the lower formant peaks, especially F1 and F2. The vowel in *but* did less well, and this is most likely due to the peak visible in the LTAS for the evolved version around 2kHz which is not apparent for the original. The other four vowels are deficient in their formant peaks, and they score poorly in the listening test. Of particular note is the fact that *boot* and *beat* show no evidence of a 1st formant peak (the same was true for M1), so once again the vowels with the narrowest constrictions are poorly matched to their targets.

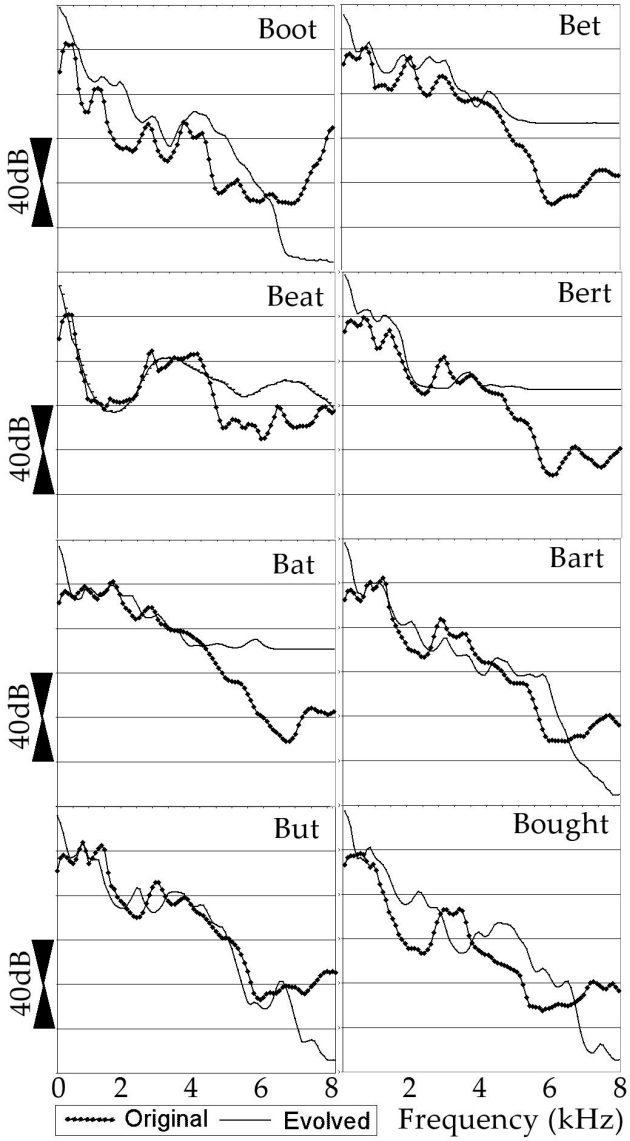

Figure 5: Long-term average spectra of the target original and evolved vowels for subject M2.

It is worth noting that some of the spectra lack detail in the high frequency region and this could well be due to the nature of the Lx waveform that has been used for their excitation. The Lx waveform relates to vocal fold closure and opening and it is not directly related to the glottal pressure waveform which is the true excitation waveform, as discussed above. Since it was recorded simultaneously with the target vowels themselves, each Lx waveform is unique, and therefore their spectra will be different.

5. CONCLUSIONS

This experiment has demonstrated that it is possible to evolve oral tract shapes using bio-inspired computation techniques, and that repeatable solutions can be achieved. Physical modelling provides an appropriate engine for the technique, and it has the advantage that it is directly related to the shape of the oral tract itself. The results indicate that there are issues with respect to evolving oral tract shapes for vowels that require a narrow oral tract constriction, particularly those in *boot* and *beat*.

Modifications were made in an attempt to improve the evolved results, and whilst there was evidence of improvements for some vowels, there are other vowels for which the results are worsened. The modifications employed therefore do not offer a universal solution in terms of improving the results, but it might be that some of their advantages could be taken advantage of in future implementations. It may be that multiple methods could be employed, leaving the fitness evaluation to select the closest result. There is plenty of scope for future work.

The technique itself has the potential to offer a non-invasive method for finding oral tract shapes that would obviate the use of either: (a) fMRI, which is acoustically noisy involving a supine position of the subject who has to hold a vocal tract posture for a number of seconds, or (b) LPC analysis, which has the potential to produce more than one solution for a given speech input that cannot as yet be constrained in a way appropriate to oral tract articulation.

The fact that solutions were evolved even though the excitation (Lx) was not fully equivalent to the natural glottal excitation during speech is, we believe, quite remarkable. The potential for further useful results being obtained using this technique, is moth promising.

6. ACKNOWLEDGMENTS

The authors thank the speakers and listeners for taking part in the experiments. This work is funded by a UK Engineering and Physical Sciences Research Council (EPSRC) postgraduate studentship, the Department of Electronics at the University of York, and the Future and Emerging Technologies programme (IST-FET) of the European Community, under grant IST-2000-28027 (POETIC). The information provided is the sole responsibility of the authors and does not reflect the Community's opinion. The Community is not responsible for any use that might be made of data appearing in this publication. The Swiss participants to the POETIC project are supported under grant 00.0529-1 by the Swiss government.

7. REFERENCES

[1] Pearson, M.D., and Howard, D.M. Recent developments with TAO physical modelling system, *Proceedings of the International Computer Music Conference*, **ICMC-96**, 1996:97-99.

[2] Dodge, C., and Jerse, T.A. *Computer music: synthesis, composition and performance*, New York: Schirmer Books, 1985.

[3] Howard, D.M. and Rimell, S. Real-time gesture-controlled physical modelling music synthesis with tactile feedback, EURASIP Journal of Applied Signal Processing (Special Issue on Model-based sound synthesis), 2004;7(15):1001-1006.

[4] Howard, D.M., Rimell, S., and Hunt, A.D. Force feedback gesture controlled physical modelling synthesis, *Proceedings of the Conference on New Musical Instruments for Musical Expression,* **NIME-03**, Montreal, 2003:95-98

[5] Howard, D.M., Rimell, S., Hunt, A.D., Kirk, P.R., and Tyrrell, A.M. Tactile feedback in the control of a physical modelling music synthesiser, In: *Proceedings of the 7th International Conference on Music Perception and Cognition*, Stevens, C., Burnham, D., McPherson, G., Schubert, E., and Renwick, J. (Eds.), Adelaide: Casual Publications, 2002:224-227.

[6] Fant, G. *The acoustic theory of speech production*, The Hague: Mouton, 1960.

[7] Holmes, J. Synthesis of natural-sounding speech using a formant synthesiser, In: *Frontiers of speech communications research*, Lindblom, B. and Oman, S. (Eds.), London: Academic Press, 1979:275-285.

[8] Howard, D.M. Practical voice measurement, In: *The voice clinic handbook*, Harris, T., Harris, S., Rubin, J.S., and Howard, D.M. (Eds.), London: Whurr Publishing Company, 1998.

[9] Titze, I.R. Theory of glottal airflow and source-filter interaction in speaking and singing, *Acta Acoustica united with Acustica*, 2004;90(4):641-648.

[10] Story, B.H., Titze, I.R., and Hoffman, E.A. Vocal tract area functions from magnetic resonance imaging. *Journal of the Acoustical Society of America*, 1996;100(1):537–554.

[11] Mullen, J., Howard, D.M. and Murphy, D.T. Digital waveguide mesh modelling of the vocal tract acoustics. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, **WASPAA**, 2003:119-122.

[12] Mullen, J., Howard, D.M., and Murphy, D.T. Waveguide Physical Modelling of Vocal Tract Acoustics: Improved Formant Bandwidth Control from Increased Model Dimensionality , *IEEE Transactions on Speech and Audio Processing*, 2006;14(3):964-971.

[13] Fant, G., Liljencrants, J., and Lin, Q.G. A four-parameter model of glottal flow, *Speech transmission Laboratories QPSR,* 1985;4:1-13.

[14] Mullen, J., Howard, D.M., and Murphy, D.T. Real-Time Dynamic Articulations in the 2D Waveguide Mesh Vocal Tract Model, *IEEE Transactions on Speech and Audio Processing*, 2007;**15**(2):577-585.

[15] Markel, J.D., and Gray, A.H. *Linear prediction of speech*, Berlin: Springer-Verlag, 1976.

[16] Rossiter, D.P., Howard, D.M., Downes, M. A real-time LPC-based vocal tract area display for voice development. *Journal of Voice*, 1995;**8**(4):314-319.

[17] Rossiter, D., and Howard, D.M. ALBERT: A real-time visual feedback computer tool for professional vocal development, *Journal of Voice*, 1996;**10**(4):321-336.

[18] Howard, D.M., Welch, G.F., Brereton, J., Himonides, E., DeCosta, M., Williams, J. and Howard, A.W. WinSingad: A real-time display for the singing studio, *Logopedics Phoniatrics Vocology*, 2004;**29**(3):135-144.

[19] Howard, D.M., Brereton, J., Welch, G.F., Himonides, E., DeCosta, M., William, J., and Howard, A.W. Are real-time displays of benefit in the singing studio? An exploratory study, *Journal of Voice,* 2007:**21**(1):20-34.

[20] Schroeter, J., and Sondhi, M.M. Techniques for estimating vocal-tract shapes from the speech signal, *IEEE Trans on Speech and Audio Processing*, 1994;2(1)II:133-150.

[21] Abberton, E.R.M., Howard, D.M., and Fourcin, A.J. Laryngographic assessment of normal voice: A tutorial, *Clinical Linguistics and Phonetics*, 1989;3(3):281-296.

[22] Cooper, C., Howard, D.M., Tyrrell, A.M., and Murphy, D. Singing Synthesis with an Evolved Waveguide Mesh Model, *IEEE Transactions on Speech and Audio Processing*, 2006;14(4):1454-1461.

[23] Koza, J., Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, 1992.

[24] Holland, J.H. *Adaption in Natural and Artificial Systems*. University of Michigan Press, 1975.

[25] Lones, M.A. and Tyrrell, A. M. Modelling biological evolvability: Implicit context and variation filtering in enzyme genetic programming, *BioSystems*, 2004;76(1–3):229–238,

[26] Cooper, C., Howard, D.M., and Tyrrell, A.M. Using GAs to create a waveguide model of the oral vocal tract, *Lecture notes in computer science*, Raidl, G.R. (Ed), 2004; 3005: 280-288.

[27] Howard, D.M., Tyrrell, A.M., Murphy, D.T., Cooper, C., and Mullen, J. (2007) Bio-inspired evolutionary oral tract shape modelling for physical modelling vocal synthesis, *Journal of Voice*, In Press.