A Novel Intonation Model for Synthesis of Speaking Styles

Andy Tams – Computer Science, University of Essex
Mark Tatham – Language and Linguistics, University of Essex

Reproduced from *WebSLS* (the website student journal of the International Speech Communication Association – fully refereed e-journal) February 2002.

ABSTRACT

In this paper we describe the adaptation and extension of the Fujisaki model for English intonation of speaking styles. This is for a prosodic model of a radio news broadcaster. Intonational phonological type models are unable to capture the fine threaded realisation differences of individual read aloud styles, and a quantitative approach is required. The model is used to analyse a corpus of read aloud speech styles, concentrating on real rather than laboratory speech. Progress towards a novel synthesis model, drawing on modern theories of speech production is described.

INTRODUCTION

Applications for a reading style include spoken newspapers for the blind, weather information over the telephone, and auditory presentation of instructions for complex hand free tasks amongst many others. Each of these requires a distinct specific reading mode because of differing requirements. Examples are the rapid fast scanning of a text in a talking newspaper, or to increase comprehension and intelligibility in a slow and careful reading style. This paper is concerned with replicating the prosodics of a radio news broadcaster, since correct intonation and rhythm can lead to high acceptability of the synthetic speech (Carlson, 1992).

One problem is that speaking styles are not clearly defined in the literature. The majority of the research concentrates on two main types of data (Llisteri, 1992): spontaneous speech from more or less unprepared situations, and the speech read from a previously prepared text. These basic categories are further elaborated with the addition of loosely defined labels such as ‘connected’, and ‘professional’ across a continuum of environments and material. Spontaneous speech is gathered under laboratory conditions. Researchers use variations of an interview technique, described by authors as directed or semi-directed. The subject answers questions about their everyday life or a similar topic, often in the form of a monologue with a minimum of intervention from the interviewer. This speech is then compared with read speech, with the same speaker reading an edited transcript of the earlier spontaneous speech. Subsequent analysis shows prosodic differences, such as lower f0 range, speaking rate changes, and vowel reduction (Koopmans Van-Beinum, 1992).

Higuchi *et al.* (1997) used the Fujisaki model of intonation (Fujisaki, 1992) to implement speaking styles for text-to-speech synthesis (hence TTS). They measured parameters of the model for four speaking styles, and derived rules for style conversion. Abe (1997) uses a statistical approach over a wider range of prosodic parameters. Both approaches were only partially successful, as a result of the use of global characteristics such as f0 range.

Our starting point is the definition of speaking style from Eskenazi (1992):

> ‘we define style to be the expression of information about the dialect and socio-economic background of the speaker, information about the manner in which he is expressing himself (formal, casual, reading, etc.) and information on the image he has of the speakers(s) he is addressing (slowing down for the hard of hearing, or foreigners, etc.). Style may overlap, but does not encompass the range of a speaker’s emotion or attitude.’

This definition is chosen because it describes speaking style over several dimensions, and it discriminates between speaking style and emotional correlates. Llisteri (1992) states that two major perspectives have played a role in the definition of speaking styles. Phonetic and phonological techniques, studying the acoustic and linguistic characteristics of language are one approach. Another is the sociolinguistic and psychological perspective related to the use of language in a variety of contexts and situations. Eskenazi (1993) states that more attention should be paid to the sociolinguistic approach, calling for a data driven approach. This was extended in Tams *et al.* (1995) arguing that a definition of speaking styles must include an application component. The problem with the data driven approach is that the data is ambiguous, with no clear division between styles, so a satisfactory definition of speaking style must consider the context, situation, audience, and aims of the communication - the environment of the speaker, as the primary factor.

[*footnote:* Eskenazi (1993) defines 'data' driven as including both a sociolinguistic component plus the phonetic and phonological behaviour reported in the literature.]

By more closely examining the notion of an environment, this can be modelled as a determiner on possible strategies and actions of the speaker. Speakers use different strategies and mechanisms to achieve the same goal (a recognised speech style). The environment dictates constraints on the speaker, and offers a level of organisation not purely linguistic (c.f. Ladd: 1996 which categorises some speech phenomena as non linguistic) which effects prosodic realisation. The interaction of the constraints and processes must be understood to give an explanation of the mechanisms and performance of the speaking style. Cross comparison of speaking styles, characterising differences in terms of abstract high level phonological parameters - describes the *what*, without considering the dynamic constraint satisfaction that takes place during the speech production process - it does not account for the *why*.

The intonation model used is an adaptation of the Fujisaki (1992, 1995) model. The most important requirement is that it can capture fine grained distinctions between styles. This model captures f0 curves very accurately. We show that these can be matched to linguistic categories, and the model is extended for English. This draws on a specialised corpus for read aloud speaking styles, described in the next section.

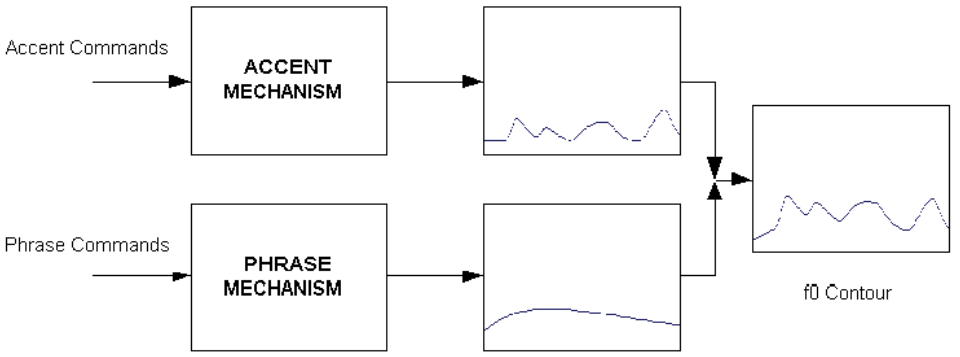

Figure 1:The Fujisaki Model - Accent and phrase components are summed to form the f0 contour.

The following section is the analysis of intonation, and the penultimate section describes a model for TTS. A standard intonational phonological type model is introduced for the low level Fujisaki approach, and this is then extended in a novel way. Finally, preliminary evaluation of the model is reported.

THE CORPUS

The RadioNews corpus described in this paper consists of British English radio news broadcasts and professionally read radio news data. This corpus was inspired by and based on the Boston University Radio News corpus, devised by Ostendorf *et al.* (1995). In their corpus they have also included extensive phonetic alignments and ToBI labelling. This approach is not followed here, because with the growth in automatic labelling techniques, it is no longer a difficult and time consuming task to add a novel or additional annotation. Therefore, no standardised prosodic annotation is included as part of the corpus proper. This corpus contains five hours of speech from radio news broadcasts from three UK stations: BBC Radio 4 (R4), BBC Radio 1 (R1), and Classic FM (CFM). Between the stations there are marked differences in the form and contents of the broadcasts. It is divided into two sections of material, radio news and lab speech.]

The radio news is primarily of nine data sets, recorded from the actual broadcasts (7-12 per data set). Eight of these are for an individual speaker (4 male, 4 female), with half from R4 and two each for CFM and R1. The data sets were designed to allow extensive coverage of a small number of speakers. The lab speech contains 22 recordings in 4 different read aloud styles (neutral, radio, advertisement, and bored), with examples for each station, by a professional speaker. The speaker has experience of radio news broadcasting and laryngograph data is available for these recordings. Additionally the speaker read a page from a novel, serving as a 'control' example of non-news speech. The lab news is designed principally for speaking styles and variability research and is approximately one and a half hours in length.

All broadcasts are annotated with an orthographic transcription, Part-Of-Speech tags, and phonetic alignment. Other annotations will include syllable markings and hand marking of pitch periods, but so far this has only been completed for data set DS-1 (7 recorded broadcasts, approximately 2 hours of speech, speaker BP).

From the orthographic transcription a phonetic transcription was generated. This is input to a phonetic alignment process using the Mbrolign tool (Malfrere and Dutoit 1997), based on a dynamic time warping algorithm. This process is not error free, and has been hand corrected for DS-1. Part of speech tags are computed by the Festival TTS system (Black *et al.* 1988), which implements a probabilistic tagger. This uses the CUOVALD lexicon, selected because it is the most acceptable in terms the balance between coverage and accuracy.

F0 data includes hand marked pitch periods. The most important requirement of this process is to mark the periodic sections of the waveform consistently, ensured by placing the pitch mark at a zero crossing. Since this point can be found automatically, this gives the process the necessary rigour. Recordings were first pre-processed by adjusting their bias to zero. Regions of voiced speech were selected, with boundary regions decisions influenced by phonological voicing representations. The most consistent speech feature, such as the start of a periodic 'hump' or a pitch excursion is then annotated. For all files, both lab and broadcasts, a second f0 representation was also computed using the super resolution pitch detection (SRPD) algorithm developed by Medan *et al.* (1991) and implemented by Bagshaw *et al.* (1993). This produces good results, comparable to contours computed from the hand marked procedure above.

The structure of the corpus was dictated by the speaking styles coverage criteria, practical considerations in recording the broadcasts and broadcasters, length of the recording session possible with the professional speaker, and annotation and analysis time constraints. Within this structure there is a change in emphasis from phonetically balanced to content balance. Compared to the Boson corpus, variation is captured by careful selection of recordings for a single speaker. In terms of content, semantic labels are not included but the corpus has been designed to facilitate such studies by inclusion of comparable broadcasts (in terms of content, successive and parallel accounts). This enables the structure of the discourse (i.e. news stories) to be examined over time.

THE FUJISAKI MODEL

The basis for the intonation model is the source filter approach developed over two decades by Fujisaki (1992) for Japanese. This pays particular attention to the way the f0 contour is

generated, treating the f0 contour as a linear superposition of accent and phrase commands. The phrase command acts over the domain of the intonation phrase, shaped as an initial rise followed by a long fall to an asymptote line. This is generated by a phrase control mechanism, activated by a pulse command with varying magnitude. The accent command is a local peak on accented syllable, generated by the accent control mechanism. This is called by a binary step function, with duration and amplitude parameters. The model is described mathematically, using a linear 2nd order equation for ln f0:

$$\ln f0(t) = \ln f\min + \sum_{i=1}^{I} ApiGpi(t - Toi) + \sum_{j=1}^{J} Aaj\{Gaj(t - T1j) - Gaj(t - T2j)\} \quad (1)$$

$$Gpi(t) = \alpha_i^2 e^{-\alpha t} \quad \text{for } t \geq 0 \quad (2)$$

$$Gpi(t) = 0 \quad \text{for } t < 0 \quad (3)$$

$$Gaj(t) = Min\left[1 - (1 + \beta jt)e^{-\beta jt}, \gamma\right] \quad \text{for } t \geq 0 \quad (4)$$

$$Gaj(t) = 0 \quad \text{for } t < 0 \quad (5)$$

With the parameters:

f_{min} asymptotic value of f0 in the absence of accent commands
I number of phrase commands
J number of accent commands
Api magnitude of the ith phrase command
Toi onset of the ith phrase command
T1j onset of the jth accent command
T2j offset of the jth accent command
αi natural angular frequency of the phrase mechanism for the ith phrase command
βj natural angular frequency of the accent mechanism for the jth accent command
γ ceiling of the accent component

The model components are critically damped second order systems, with the two sets of parameters for the phrase and accent equations above. The model parameters, which include the angular frequency of the components above are generally regarded as constant for the utterance, though some variations of the model use smaller domains (Mobius 1997).

Numerical optimisation of an analysis by synthesis procedure can be used to find the magnitude and timing of the underlying processes. In the Fujisaki model this is a hill climbing search of parameters guided by linguistic constraints, but in this work we have also combined it with statistical models (see below). These constraints are primary and necessary, because the search would produce an arbitrary number of accent and command phrases (over-fitting) to produce an optimal mathematical approximation to the contour otherwise.

Declination can be explained as a negative impulse to reset the phrase component. This model has been linked into physiological and physical mechanisms of the laryngeal system, based on f0 transition data in singing. The model has been applied to several languages

including German (Mobius 1997, Mixdorff 1997), and preliminary studies for English (Fujisaki and Ohno 1995). The model has been used in this paper because it has physiological and physical justifications, is quantitative, synthesis is straightforward, and the mapping between the quasi-discrete inputs and continuous output of the accent and phrase mechanisms.

ANALYSIS

A data set of 200 utterances, representative of the radio broadcaster style, was selected from DS-1. Half of this set was selected on the basis of intonational behaviour and linguistic structure, the other half was randomly selected from the data set. This was used for adapting the Fujisaki model to English and data analysis. Adaptation is necessary because the model has difficulty handling all the accents of English (see below), with Japanese having fewer intonational contrasts.

For the analysis performed in this paper, all of the f0 contours are processed by five point median smoothing to remove segmental perturbations (the original contours are also retained). The analysis procedure consists of the following stages:

1. Processing the corpus annotations, extracting word boundaries.
2. A linguistic and statistical analysis to generate candidates for phrase and accent commands. A rudimentary pause is performed to mark clause boundaries, followed by accent candidates. This uses the Festival TTS system. This is followed by additional processing using rule based syntactic constraints, incorporating marked pauses into the analysis. Complex nominals are treated as a single unit.
3. The candidate accent and phrase commands are then input to an optimisation process. Alternatively, candidate phrase and accent commands can be generated by hand using adapted version of the SFS annotation tool and the rendering process.

The rendering process performs numerical optimisation, using two modules: render and search. Render is a trivial implementation of the equations (10 to (5), and the slow rise component described below. This takes as input files of accent and phrase commands, but no linguistic constraints are applied at this stage. This can then be combined with a search module, subject to linguistic and well formedness constraints (for example each accent command is associated with a word, words cannot more than one. Numerical optimisation is then performed for parameter values, using a hill climbing search of the parameter space.

Analysis shows that the magnitude of the phrase command is influenced by the length of the preceding phrase and syntactic cohesion of the boundary, with a maximum for utterance initial position. The location of the phrase command can be inferred from the portions of the f0 contour delimited by unaccented syllables (where f0 falls). By comparison to the other speaking styles data in the corpus, different speaking styles show variation in accent realisation and phrase commands. The variation of model parameters for different speaking styles is also being investigated, though in most formulations of the model they are kept constant. However we have found that range is an important correlate of speaking styles, hence this approach is not adopted in this work. The use of this quantitative model, with its fine grained control structure and different components allow the modelling of speaker strategies in the realisation of speaking styles.

THE SYNTHESIS MODEL

The number of prosodic phrases in an utterance is denoted by the number of positive phrase commands. The onset of a prosodic phrase is marked by the adjustment of the declination line. Boundaries are marked by pauses, a pitch accent on the last syllable, or a boundary tone. The boundary can be shifted by accentuation requirements. The Fujisaki model does not have a clearly defined phonology for English, so an appropriate framework has been devised. Problems exist for long rising sections of the f0 contour (Taylor, 1994), for example in interrogative statements, which are not very common in this corpus.

This is solved for German by Mixdorff (1997) by introducing a slow rise component. This is like an accent component with onset T3 and offset T4, though with the time constant of the accent component, added as an extra term:

$$Gr(t) = 1 - (1 + \delta t)e^{-\delta t} \qquad \text{for } t \geq 0 \tag{6}$$

$$Gr(t) = 0 \qquad \text{for } t = 0 \tag{7}$$

This approach is adapted in this paper for English, but as a different gesture of the phrase command. This compensates the falling slope of the phrase command with an almost linear rise in ln f0. This change is to be consistent with the approach in the model for production described in the next section, and to solve theoretical problems with the phonology (noted by Ladd 1996). The intonation phonology introduces boundary tones, similar to Mixdorff, but with a different phrasal structure. This draws on the higher level Spruce symbolic representation (Morton *et al.* 1999). A sentence has a global slope set at L to H or H to L. This consists of one or more intonational phrases having local slopes at a lower level of abstraction, set at (L to H) or (H to L). Since this is a level that can be realised, the (L to H) is represented by modifying Gp with a Gr component (since the default is (H to L) by Gp). A phrase contains intonational words which may carry a pitch accent, realised by component Ga.

Figure 2 shows an example of an f0 contour synthesised using an implementation of the Fujisaki model for analysis purposes. This is not a numerically optimised curve, with accent and phrase commands selected according to linguistic constraints, and parameters quantified to discrete values.

Abe (1997) found that it is difficult to model the variations in intonation for read aloud speaking styles because the intonational phonology is often the same. However the prosodic realisation is different, so a model must be able to explain this. Blaauw (1992) stated that speaking styles differences may be the result of interacting parameters during speech production and this idea is used here. The aim is for a model of speaker strategies and constraints, introduced by the environment and the purpose of the communication.

Two requirements for this approach can be identified: a need for indeterminacy (a choice can be made between two or more processes to satisfy a constraint) and concurrency (to model interacting processes).

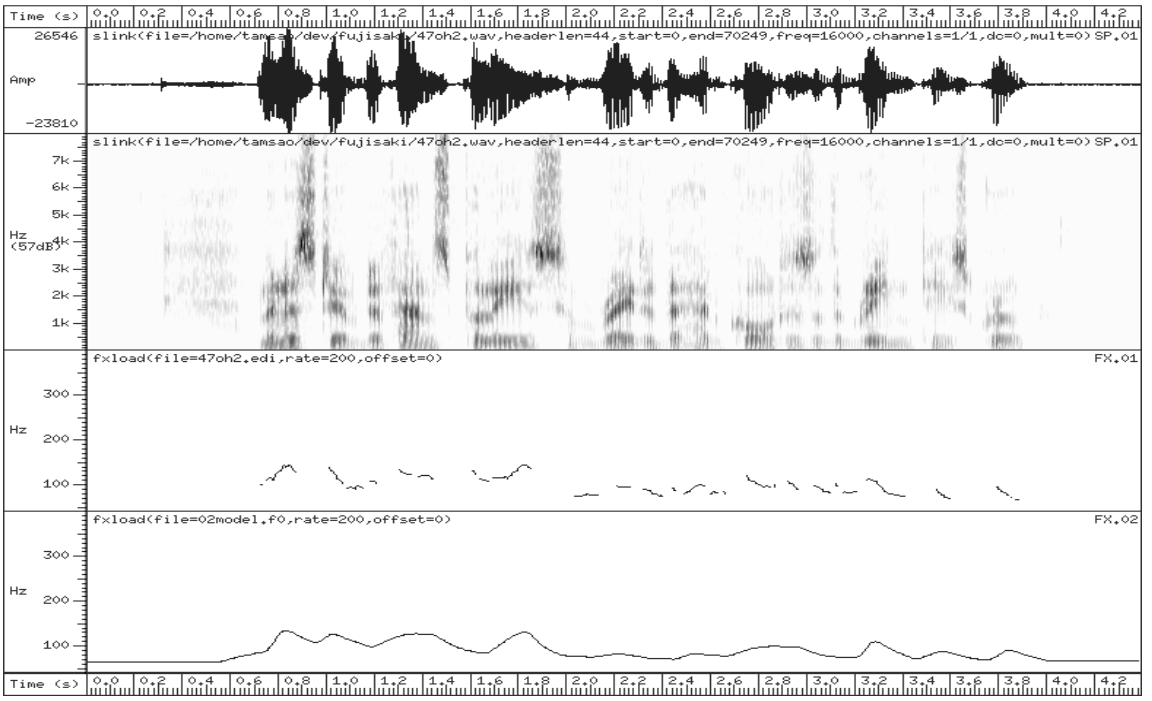

Fig. 2 Model Generated Contour - The utterance is 'ministry of defence police have raided the headquarters of the Greenpeace group', displayed using SFS developed by Mark Huckvale.

This solves the problem of the intonational phonology approach being under specified for speaking styles, exploiting the concurrency implicit in the standard TTS model and models of speech production. The assumption of this approach is that the interaction of parameters is significant for speaking styles models.

A requirement for indeterminacy has been identified in analysis by Wichman and Knowles (1995). They argue that it is required because of a lack of agreement on boundaries by expert transcribers. This is not because of a lack of knowledge but a phenomenon that needs to be explicitly included in the theory. For production it is required to explain different strategies. In a study of pitch accent placement Ross *et al.* (1992) found that there is more than one possible accent structure for a story in an analysis of different speakers. They suggest a simple mechanism for allowing choice in accent prediction by classifying syllables into those requiring an accent, those that cannot, and those that can choose to take a pitch accent. This approach allows exactly that.

The model being developed is based on process algebra from computer science (Milner, 1992). Concurrent systems are represented as a collection of independent sequential processes communicating with each other in order to exchange information. A model consists of several communicating processes and at a suitable level of detail each process can be thought as sequential. An event is an observable activity at some level of abstraction. Levels of activity (a functional description) in a system are the events, and which events are included depends on the level of abstraction. Central to this model is the notion of a constraint.

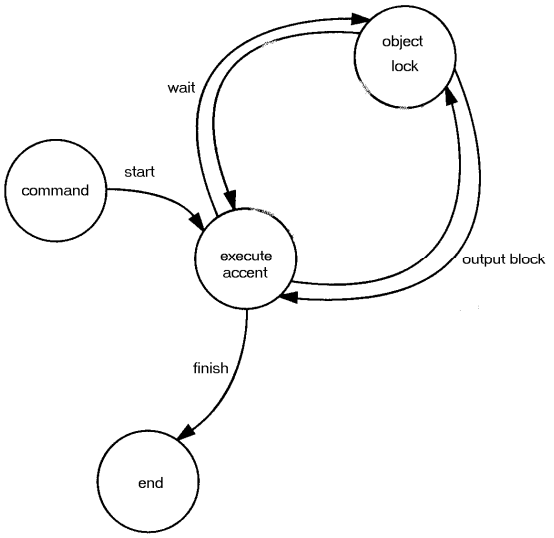

Fig. 3 Accent Component Realisation - This shows the realisation of the accent process, modelled as a thread. Execution can be blocked to interaction with other processes, allowing variability within timing relations.

In this framework communicating processes interact to satisfy constraints, by co-operation or competition. Since constraints are varied according to goals and the environment, this approach can be applied to a number of research areas. Examples include the relation between phrase length, the magnitude of the successive accent component, and boundary strength. Rhythmic constraints can also be formulated, for example slowing down speaking rate to give the hearer time to integrate important information.

One criticism of the Fujisaki model (Taylor 1992) is that it is sometimes difficult to give a linguistic justification for the placement of accent and phrase commands, in particular the slow rise effect. These can be viewed as a synchronisation restriction, where an accent or phrase command cannot be performed because of the execution of another generation process controlled at a higher level of abstraction. In particular this is noted in the application of Gr, which serves as a modifier on Gp. This explains the variation in phrase command onsets and difficulties in aligning these with linguistic boundaries. The Fujisaki model has been developed with each process operating to its own clock. Internal clocks can use synchronisation and entrainment mechanisms to realise timing relations and constraints. The proposed mechanism is shown in figure 3.

EVALUATION OF THE MODEL

Preliminary evaluation of the model using informal listening tests has been performed. The perceptual experiments aim to show the adequacy of the adaptation of the Fujisaki model to English for the radio news broadcaster speaking styles application.

The listening test methodology is based on an ABX procedure. The panel of three listeners were asked to select stimulus A or B as been closest to X in terms of its melodic characteristics. X used the original intonation of the recording, A and B were different intonation models.

The Festival TTS engine (Black *et al.* 1988) is used in the production of the stimuli. All stimuli had segmental durations extracted from the original recording, and used the same male voice. Three types of stimuli were produced:

1. A copy synthesis version, with f0 extracted from hand marking of pitch periods.

2. The f0 produced by the Festival TTS engine, using the ToBI model of intonation with contour target values predicted by linear regression (Black and Hunt 1996).
3. An f0 contour generated using the model described above.

For each utterance X was the copy synthesis version, A and B either the Festival or model versions. The presentation of the two stimulus types as A or B was randomised. Stimuli consisted of ten utterances from a broadcast of the portion of DS-1 set aside as test data.

Preliminary results indicate that listeners show a small preference for the model versions, thought his can be regarded as conclusive with such as small sample. Hence, this indicates the model is successful for this application. However the crude categorisation of the stimuli must be noted. Listeners were unable to consistently discriminate between the two types, and indicated that the two stimuli may sound different but deciding which was ‘closer’ is difficult. The results of these ongoing perceptual experiments will be published in a subsequent paper. Further extensions to the evaluation procedure will include objective measures, and opinion tests (to increase the categorisation choices).

CONCLUSION

This research is concerned with modelling speaking styles and to improve the naturalness of synthetic speech. A corpus of appropriate speaking styles data, concentrating on radio news broadcast recordings, real speech, and laboratory recordings has been collected. This has been annotated and analysed with a revised Fujisaki model of intonation, since this is the most appropriate for the defined criteria. The Fujisaki model uses two critically damped second order filters to generate f0 contours. The phrase component models long term effects such as declination (and its associated resets), the input parameter is a sequence of impulses. The accent component models pitch accents, and input to this component is a step function.

There is close agreement between the model and generated contours, suggesting some semblance of physiological and physical reality. It is not necessarily a universal model and additional laryngeal control is required for languages such as English. Further amendments are required, in particular the development of a phonological description to facilitate constraints on the phonetic parameters. A model for synthesis is being developed, modelling components as concurrent processes. This approach is novel and has implications for the architecture of TTS systems.

Informal listening tests have been performed to show that the qualitative approach, the foundation of the proposed model, offers improvements in speaking styles adaptation for naturalness of TTS (for applications such as automatic news reading). This uses a novel extension of the Fujisaki model, and further work is planned to give a quantitative evaluation of the predicted improvements in TTS quality. Future research questions of particular interest are its applicability to speaking styles apart from the limited domain of broadcasts, and the relation of individual speaker strategies to a specific speaking style. For the latter, care has been taken to the design of the corpus to facilitate such research.

ACKNOWLEDGEMENTS

This research was partially supported by BT Labs research contract ML605053. The laboratory recordings took place at BT Labs, thanks to Andrew Breen and Julian Page with engineering support by Paul Deans, Pete Jackson and Rod Ashworth. Special thanks to Eric Lewis for use of the cue points program for extracting pitch marks from Cool Edit.

REFERENCES

Abe (1997) 'Speaking Styles: Statistical Analysis and Synthesis by a text-to-Speech System' in *'Progress in Speech Synthesis'*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg, Springer-Verlag, pp. 495 –510.

Bagshaw, P. C., Hillier, S. M., and Jack, M. A. (1993) 'Enhanced pitch tracking and the processing of f0 contours for computer aided intonation teaching', Proceedings of Eurospeech 93, Vol. 2, pp. 1003-1006.

Blaauw, E. (1992) 'Phonetic Differences Between Read and Spontaneous Speech', *Proc. ICSLP*, Banff, Canada, pp. 755-758.

Black, A. and A. Hunt (1996) 'Generating f0 contours from ToBI labels using linear regression', ICSLP96, volume 3, Philadelphia, pp. 1385-1388.

Black, A. W., P. Taylor and R. Caley (1998) *'The Festival Speech Synthesis System'*, http://cstr.ed.ac.uk/.

Carlson, R. (1992) 'Synthesis: Modelling variability and constraints, *Speech Communication*, 11, North Holland, pp. 159-166.

Eskenazi, M. (1992) 'Changing Speech Styles: Strategies in Read Speech and Careful Spontaneous Speech', *Proc. ICSLP*, Banff, Canada, pp. 755-758.

Eskenazi, M. (1993) 'Trends in Speaking Styles Research', *Proc. Eurospeech'93*, Volume 1, Berlin, Germany, pp. 501-512.

Fujisaki, H. (1992) 'Modelling the Process of Fundamental Frequency Contour Generation', in *'Speech Perception, Production and Linguistic Structure'*, edited by Y. Tohkura, E. Vatikiotis-Bateson, Y. Sagisasaka, IOS Press, pp. 313 -328.

Fujisaki, H. and Ohno, S. (1995), 'Analysis and Modelling of Fundamental Frequency Contours of English Utterances', *Proc. Eurospeech'95*, Vol. 2, Madrid, Spain,

pp. 985 - 988.

Higuchi, N. Hirai, T. Y. Sagisaka (1997) 'Effects of Speaking Style on Parameters of Fundamental Frequency Contour' in *'Progress in Speech Synthesis'*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg, Springer-Verlag, pp. 417-428.

Koopmans-Van Beinum F. J. (1992) 'The role of focus words in natural and in synthetic speech: Acoustic aspects', *Speech Communication*, 11, pp. 439-452.

Ladd, R. (1996), *'Intonational phonology'*, Cambridge University Press.

Llisterri, J. (1992), 'Speaking styles in speech research', *'ELSNET/ESCA/SALT Workshop on Integrating Speech and Natural Language'*, Dublin, Ireland.

Malfrere, F., and T. Dutoit (1997) 'High Quality Speech Synthesis for Phonetic Speech Segmentation', *Proc. Eurospeech'97*, pp. 2631-2634.

Medan, Y., E. Yair and D. Chazan (1991) 'Super resolution pitch determination of speech signals', IEEE Trans. Signal Processing, Vol. 39, pp. 40-48.

Mixdorff, H. (1997) *'Modelling Patterns of German – Model-based Quantitative Analysis and Synthesis of F0 contours'*, unpublished PhD thesis, Technische Universitat Dresden.

Mobius (1997) 'Synthesizing German Intonation Contours' in *'Progress in Speech Synthesis'*, edited by J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg, Springer-Verlag, pp. 401 –415.

Morton, K., Tatham, M.and Lewis, L. (1999) 'A New Intonation Model for Text-to Speech Synthesis', in 'Proc. of the International Congress of Phonetic Sciences', San Francisco.

Ostendorf, M., P. J. Price, and S. Shattuck-Hufnagel (1995) 'The Boston University Radio News Corpus', Boston University Technical Report, ECS-95-001 March 1995, University Of Boston.

Ross, K., Ostendorf, M. and Shattuck-Hufnagel, S. (1992), 'Factors Affecting Pitch Accent Placement', *Proc. ICSLP*, Banff, Canada, pp. 365-368.

Tams, A., Tatham, M. and Page, J. H. (1995), 'Describing Speech Styles Using Prosody: A Pilot Study', *Proc. Eurospeech'95*, Vol. 3, Madrid, Spain,

pp. 2081-2084.

Taylor, P. (1992) *'A Phonetic Model of English Intonation'*, PhD thesis, University of Edinburgh, published by the Indiana Linguistics Club.

Taylor, P. (1994), 'The rise/fall/connection model of intonation', *Speech Communication*, 15, p. 169 - 186.

Wichman, A. and Knowles, G. (1995), 'How determinable are intonation units?', *Proc. ICPhS 95*, Volume 2, Stockholm, Sweden, pp. 223 - 225.