Proceedings Paper:
Oba, T and Atwell, ES (2003) Using the HTK speech recogniser to anlayse prosody in a corpus of German spoken learner's English. In: Archer, D, Rayson, P, Wilson, A and McEnery, T, (eds.) UCREL Technical Paper number 16. Special issue. Proceedings of the Corpus Linguistics 2003 conference. Corpus Linguistics 2003 conference, 28-31 Mar 2003, Lancaster University, UK. Lancaster University , 591 - 598. ISBN 1862201315

Using the HTK speech recogniser to analyse prosody in a corpus of German spoken learners' English

Toshifumi Oba (tosh@comp.leeds.ac.uk)
and Eric Atwell (eric@comp.leeds.ac.uk)
School of Computing, University of Leeds, Leeds LS2 9JT, England

1. Introduction: speech research, prosody and learners' English

Intonation is important in human communication to help the listener to understand the meaning and attitude of the speaker (Brown, 1977; O'Connor, 1970). Language students and teachers see intonation as a part of the structure of the language (Tench, 1996). However, acquisition of proper intonation is difficult for non-native speakers and requires repeated practice (O'Connor, 1970). For example, the Interactive Spoken Language Education (ISLE) project found that intonation was the biggest problem for German learners of English, but the project did not tackle intonation (Atwell et al, 1999).

Intonation is also important for speech recognition to retrieve correct semantic and syntactic information and to successfully identify words (Rodman, 1999: Werner and Keller, 1994). It not only influences acoustic models, but can also contribute for language models; Taylor et al (1997; 1998) focused on the relationship of intonation and 'move types' of the dialogue to constrain the number of possible language models. However, intonation is generally limited to be that of native-speakers in speech recognition research.

Speech recognition research has been done for non-native speakers too. The works can be categorized into two groups: multilingual speech recognition; and use of speech recogniser for foreign language learning. Uebler (1998) investigated bilingual (Italian and German) and multi-dialectical speech recognition by assuming the two languages being one. Stemmer et al (2001) developed acoustic models of foreign words which appeared in German dialog, such as English words in film titles. In the FLUENCY project, Eskenazi (1996) used a speech recogniser to detect foreign speakers' pronunciation errors for second language training. This work also involved prosody looking at a correlation of pronunciation and prosody errors. However, little research has been undertaken on speech recognition targeted on non-native speakers' intonation: the majority of speech recognition research focuses on pronunciation of native speakers. While there have been relatively few studies on speech recognition dealing with prosody or non-native speakers, those for non-native speakers' prosody have received even less attention.

The use of Speech recognition has been investigated in Computer-Assisted Language Learning (CALL) such as in Eskenazi (1996) and Witt and Young (1997). However, the ISLE project reported that detection of errors and providing feedback to the learners were not well developed in existing language learning software (Atwell et al, 1999). For example, a well-known software, 'Tell Me More' of Auralog, improved the detection and the feedback for pronunciation practice by pointing out error phonemes and showing a 3D animation to visualize the 'model' articulation. However, its technology for intonation practice is still poor. Eskenazi (1999) mentions that a visual display is more effective than oral instructions for intonation practice and 'Tell Me More' displays a waveform and pitch curve, which traces the amplitude and frequency variations, respectively, of both user's voice and a 'model' utterance. However, it neither points out the placement of the intonation errors, nor provides suggestions to improve intonation, leaving the comparison tasks to the users. Investigation of Germans speakers' English prosody should help to improve these technologies, by pointing out their common weakness in English intonation, which should be included in exercises.

Table 1 shows a variety of speech recognition research and which fields are involved in this research. Research fields to be considered are pronunciation, prosody, native speaker, non-native speaker, monolingual, multilingual and the HTK. 'Y' represents a field which a research deals with: otherwise 'N' is given. More precisely, 'Y' is marked in each column as follows:

Non/Mul: when speech recogniser are for non-native speakers' utterances or for multi-languages;
German: when speech recogniser deals with German speakers' English;
Inton: when speech recogniser considers intonation features;
G/B: when the research deals with 'goodness' and 'badness' of intonation;
HTK: when the research exploits the HTK based-speech recogniser.

Table 1: Features of Various Speech Recognition Research

Research Reference	Non-N	German	Inton	G/P	HTK
(Taylor, 1998)	N	N	Y	N	Y
(Uebler, 1998)	Y	N	N	N	N
(Stemmer, 2001)	Y	Y	N	N	N
(Teixeira, 1996)	Y	Y	N	N	Y
(Hansen, 1995)	Y	Y	Y	Y	N
(Yan and Vaseghi, 2002)	N	N	Y	N	Y
(Jurafsky et al, 1994)	Y	Y	N	N	N
(Berkling et al,1998)	Y	N	N	N	Y
(Oba and Atwell, 2003)	Y	Y	Y	Y	Y

The HTK (Hidden Markov Model Toolkit) is a free and portable toolkit for building and manipulating Hidden Markov Models (HMMs) primarily for speech recognition research, although it has been widely used for other topics such as speech synthesis, character recognition and DNA sequencing (HTK3, 2000; Young et al, 2001). Taylor et al (1998) exploited the HTK-based speech recogniser to provide word hypotheses constrained by 'move types' correlated with intonation. However, direct analysis of prosody using the HTK has not been the focus of any past research.

The ISLE project exploited the IHAPI HMM-based speech recogniser to improve the performance of computer-based English learning systems, such as providing clear feedback by specifying error words and phones (Atwell et al, 1999; 2003). The project collected a corpus of audio recordings of 23 Italian, 23 German Spoken Learners' and 2 native speakers' English, in which subjects read aloud samples of English text and dialogue selected from typical second language learning exercises, such as pronunciation and stress placement training using minimal pairs and polysyllabic words. (Atwell et al, 2000b; 2003; Menzel et al, 2000). The audio files were time-aligned to graphemic and phonetic transcriptions, and speaker errors were annotated at the word- and the phone-level, to highlight pronunciation errors such as phone realisation problems and misplaced word stress assignments. We re-used the first three blocks of the corpus, Block A through C, which contained 82 sentences edited from 'The Ascent of Everest' (Hunt, 1996). As the rest of the corpus generally consisted of shorter sentences or just words without a uniform topic, the first three blocks were the best for prosodic analysis using a speech recogniser.

Our research analysed German spoken learners' English prosody re-using the ISLE speech corpus by using the HTK-based speech recogniser. There were three main stages to the research: prosodic annotation of the English text in the corpus, following a model devised for speech synthesis; native speakers' assessments of the intonation abilities of the 23 German speakers; and speech recognition experiments using the HTK.

2. Prosodic annotation

Prosodic annotation was done following the set of instructions or informal algorithm in (Knowles, 1996), to predict 'model' intonation patterns for written English text, to be passed to a speech synthesiser. The annotation was done to all 27 sentences of Block A of the ISLE corpus. All the tags were added by hand using Windows Excel. The process consisted of four steps: prosodic parsing to divide the text into blocks; assembling blocks to form a potential tone group; adapting lightening rules to merge the groups into an actual tone group; deciding the type of the tone group. The annotation was generally achieved by simple mappings of each step such as from grammatical tags (Atwell et al 2000a) and transition markers to

assembling of tone units. The marked-up prosody was compared with the native speakers' recordings and some of the patterns were modified.

The 27 sentences consisted of 429 words and were divided into 84 tone groups: 1 'low rise', 3 'high rise', 52 'fall-rise' and 28 'fall' patterns. The prosodic annotation produced many 'fall-rise' patterns as predicted in the instructions. Tone types of the first 10 sentences, which were used for evaluating German speakers' intonation abilities, were modified by comparing against 2 native speakers' recordings. The 10 sentences consisted 157 words retained 15 'fall-rise' and 10 'fall' patterns after canceling 1 'low rise', 2 'high rise' and 4 'fall-rise' patterns. This modification was undertaken as we tended to create more tone groups than found in actual native speakers' utterances. This is due to an ambiguity of the annotation instructions; when it was not clear from the instructions if the two successive blocks should be merged to form a single tone group, we left them as individual tone groups in most such cases. However, the instructions generally required simple mappings such as grammatical tags to the degrees of accentual types, which could be handled by non-linguists.

We hoped to use the modified prosodic patterns as a 'native speaker target', against which to compare learners' actual prosodic patterns, so we investigated automated methods to extract prosodic features from the learners' speech-files. Unfortunately, we were unable to automatically predict markup equivalent to the synthesiser cues, so could not directly compare the learners against this model.

3. Human evaluation of intonation abilities and grouping of German speakers

Instead, we turned to expert human evaluation: a computational linguistics researcher and an English language-teaching (ELT) researcher subjectively assessed the intonation of the recorded utterances from German learners of English, by listening to the recorded utterances, comparing against the 'model' marked-up script, and counting perceived intonation errors. The judgments were used to partition the speakers: the speakers were divided into two groups by assigning the upper half of them, who made fewer intonation errors, to a 'good' intonation group; and the rest to a 'poor' intonation group. Three different groupings were done: two groupings based on each of the two evaluators; and the third based on agreement of the two evaluators. Speakers with exceptionally poor pronunciation as indicated by the ISLE corpus pronunciation markup were excluded in this grouping so that results of the following experiments would be independent from pronunciation ability.

The two evaluators, one computational linguistics researcher (Evaluator I) and one ELT researcher (Evaluator II), listened to the 10 utterances from each speaker, and compared their prosodic patterns against 'model' tone types from the annotation. If the evaluator perceived that all the tone types of each utterance were the same as the model patterns, the utterance was marked as 'correct'; otherwise 'error'. We separately counted the number of 'errors' for every speaker marked by each evaluator.

The agreement of two evaluators' judgments was at about 63 %. Evaluator II's judgment was stricter; this evaluator marked 109 errors out of 230 judgments (10 utterances from 23 speakers), while Evaluator I marked 78 errors. This was probably due to the difference of their judgments norms. Evaluator II mentioned that German speakers tended not to have a clear 'fall' in a 'fall-rise' pattern. This should result in marking more errors due to 'fall-rise'; however it can not be seen from the score sheet, as one judgment was made for one utterance.

As evaluators' judgments did not agree in some cases, three different groupings were done to 23 German speakers: Grouping I based on Evaluator I, Grouping II based on Evaluator II and Grouping III based on the agreement of the two evaluators. Before the groupings, 3 exceptionally 'poor' speakers were eliminated, so that the following HTK experiments should not be affected by pronunciation factors. In Grouping I and II, top 8 and bottom 8 speakers were categorized into 'good' and 'poor' intonation ability groups leaving 4 intermediate speakers each time. 5 'good' and 7 'poor' speakers were in the same groups in both groupings. In Grouping III, 7 'good' and 7 'poor' speakers were grouped by adding 2 speakers, who were categorized as 'good' by one evaluator and 'intermediate' by the other, into the 5 agreed 'good' speakers. Despite the high rate of disagreement from two evaluators, many speakers were categorized into

the same intonation ability groups. Therefore, it can be said that human evaluation was successful enough.

4. The HTK speech recognition experiments for analyzing prosody

Finally, the speech recognition experiments were undertaken using the HTK. Before starting the main experiments for prosodic analysis, we made two preparation experiments: for investigating best training models and several parameters for recognition tests; and for checking the influence of reducing training speakers to the recognition accuracy. In the main experiments, the HTK was used to train monophone and triphone Hidden Markov Models for a 'poor' intonation group and a 'good' intonation group separately. In each case, the training set excluded test speakers and each model was tested with the test speakers from both 'good' and 'poor' intonation group. This was achieved via cross-validation, repeating the experiment, taking out a different test-subset each time, and averaging the results. The whole process was repeated three times taking a different grouping. Results reveal that recognition accuracy becomes higher when models are trained with a group of same intonation ability as test speakers. Cross-merging experiments confirm that these results are generally consistent.

Before the main HTK speech recognition experiments for analysing German speakers' English prosody, the following HTK parameter settings were decided by preparatory HTK experiments:

- Monophone and 3-mixture word internal triphone HMMs would be trained.
- No language models would be used for recognition tests.
- Recognition accuracy was still reasonable, even when the number of training speakers were reduced from 17 to 12 and 6.
- WIP (Word Insertion Penalty) would be set to -60.0.

Three main experiments were undertaken using one of the three groupings: Experiment I through III taking Grouping I through III, respectively. The HMMs were trained with 6 'good' and 6 'poor' speakers separately, and each model was tested by the rest of speakers from both groups. This was repeated by taking different sets of test speakers each time.

In all of the three experiments, recognition accuracy was generally higher except 'poor' test speakers against monophone models when training and test speakers' intonation abilities were the same. Experiment II showed the most significant improvement, whose grouping was based on Evaluator II, who had a stricter norm on 'fall-rise' patterns. Average improvement of recognition accuracies by the agreement in triphone cases (with monophone cases) were:

- Experiment I: 10.13 % (13.20 %) among 'good' test speakers
  6.76 % (-2.85 %) among 'poor' test speakers

- Experiment II: 19.20 % (15.43 %) among 'good' test speakers
  15.50 % (1.67 %) among 'poor' test speakers

- Experiment III: 17.11 % (16.41 %) among 'good' test speakers
  13.14 % (-0.48 %) among 'poor' test speakers

The improvement was lower from 'poor' test speakers. This was because of one speaker, SESS0189, who was categorized into a 'poor' intonation group by both evaluators, but always had much higher recognition accuracy against models trained by 'good' intonation speakers. This must be because this speaker had different intonation error types from the other 'poor' speakers, while the rest of 'poor' speakers created similar intonation errors.

Although exceptionally 'poor' pronunciation speakers were excluded from the groupings, the following two support experiments gave supporting evidence that the above results were obtained by intonation abilities. These two experiments were done taking Grouping II, which showed the most significant improvement in previous experiments.

We counted correctly recognised 'keywords' for tone types. White (2002) found that the locus of accentual lengthening was shown to be the word, with the greatest lengthening tending to be at word edges. We called the word containing the last accented syllable of each tone group a 'keyword'. Improvement of recognition accuracy among the 'keywords', especially for 'fall-rise' patterns, was higher than that among all the words. This result showed that trained models were clearly distinguished by prosodic features, and 'poor' intonation speakers tended to show the difficulties at 'fall-rise' patterns as perceived by Evaluator II.

The other experiment was done taking two 'worst' and 'best' pronunciation speakers from 'good' and 'poor' intonation groups, as the former group tended to have slightly better scores on 'pronunciation' abilities. This result also showed the improvement when training and test speakers' intonation abilities agreed. This confirmed the result is not relevant to pronunciation factors.

Overall, we can conclude that the HTK was able to train clearly different HMMs according to training speakers' intonation abilities. We found that it was better to use models trained by speakers with the same intonation ability as the test speakers in order to achieve higher recognition accuracy, and that German speakers who showed 'poor' English intonation abilities, generally had similar errors.

5. Conclusions

Our research focussed on analysing non-native speakers' prosody using the HTK speech recogniser. As mentioned in the Introduction, the main focus of speech recognition researchers is generally on pronunciation factors, or if prosody is taken into account, they tend to deal with native speakers', that is, this research was unique, and important; we showed that a test German speaker should choose a model trained by other German speakers with the same intonation abilities as the test speaker, in order to obtain higher recognition accuracy. Therefore, it is worth considering intonation abilities for speech recognisers for non-native speakers.

There has not been previous research which used the HTK speech recogniser to directly analyse prosody. Our work proved that the HTK was able to deal with prosodic factors. The HTK trained two clearly different HMMs: those trained with similar prosodic patterns to native speakers' 'model' patterns; and those trained with common intonation error patterns among German speakers. Speech recognition researchers can deal with prosody using the HTK.

Our research suggests that foreign language learning software should be able to detect learners' intonation abilities unlike any existing educational software. The learning tool should contain different models separately trained by 'good' and 'poor' intonation speakers. By comparing recognition accuracies of 'keywords' for prosody against the two models, it should be possible to detect the accuracy of the learner's intonation and to point out intonation patterns where the learner especially showed weakness.

Fox (1984) and Grabe (1998) compared English and German intonations, and revealed that German language rarely had a similar 'fall-rise' pattern to that of English. One of our experiments implied that German speakers with 'poor' English intonation tended to have errors at 'fall-rise' patterns. German speakers of learning English require intensive practice of the 'fall-rise' pattern.

This research successfully showed that the agreement of training and test speakers' intonation abilities, 'good' or 'poor', brought about higher recognition accuracy. The intonation abilities were judged at only 'fall-rise' and 'fall' patterns; however, there are also other tone types, such as 'rise' and 'level'. This suggests that further investigations are required:

> Whether the same grouping would be given when all the tone types were taken into account in human evaluation of German speakers' English intonation abilities;
>
> If not, whether the different grouping would also shows the improvement of recognition accuracy by the agreement of intonation abilities.

We also need to consider the diversity of intonation errors. In the HTK experiments, one 'poor' intonation speaker showed the opposite result; recognition accuracy was better when models trained by 'good' intonation speakers were tested against this speaker. This was probably because this speaker had different types of intonation errors from the rest of 'poor' intonation speakers. Speech recognition should deal with this kind of exceptional speaker when it takes account of intonation abilities.

In this research, only German speakers were considered. It is worth investigating whether the same results would be obtained from other nationalities, and the possibility of adjusting the idea into multilingual speech recognition, in which there should be diversity even within the same intonation group because of influences from different mother languages.

A significant challenge is to use these results in real language-teaching systems. A lesson from the ISLE project is that theoretical results and practical of these results are quite different achievements!

We believe the analysis contributes to speech recognition research and foreign language learning technology. Our HTK experiments found better training models for German speakers with 'good' and 'poor' intonation speakers separately. Its results should help to find an effective way to train English speech recognition systems for German speakers with various English intonation abilities.

References

Atwell, E., Dementriou, G., Hughes, J., Schiffrin, A., Souter, C., and Wilcock, S. 2000a A comparative evaluation of modern English corpus grammatical annotation schemes. In: *International Computer Archive of Modern and Medieval English (ICAME) Journal*, vol.24, pp.7-23.

Atwell, E., Herron, D., Howarth, P., Morton, R. and Wick, H. 1999 *Pronunciation Training: requirements and solutions*, Project Report, Interactive Spoken Language Education Project LE4-8353, Deliverable D1.4. Cambridge: Entropic.

Atwell, E., Howarth, P., and Souter, C. 2003 The ISLE Corpus: Italian and German Spoken Learners' English In: *International Computer Archive of Modern and Medieval English (ICAME) Journal*, vol.27.

Atwell, E., Howarth, P., Souter, C., Baldo, P., Bisiani, R., Pezzotta, D., Bonaventura, P., Menzel, W., Herron, D., Morton, R., and Schmidt, J. 2000b User-Guided System Development in Interactive Spoken Language Education. In: *Natural Language Engineering journal*, *Special Issue on Best Practice in Spoken Language Dialogue Systems Engineering,* vol.6 (3-4), pp.229-241.

BEEP dictionary 1996 [Online]. Available from World Wide Web: <http://www-svr.eng.cam.ac.uk/comp.speech/Section1/Lexical/beep.html>.

Berkling, K., Zissman, M., Vonwiller, J., and Cleirigh, C. 1998 Improving Accent Identification through Knowledge of English Syllable Structure In: *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-1998)*, vol.2, 30 November-4 December 1998, Sydney. pp.89-92.

Brown, G. 1977 *Listening to Spoken English*. London: Longman.

Cruttenden, A. 1997 *Intonation*, 2nd edition. Cambridge: Cambridge University Press.

Eskenazi, M. 1996 Detection of foreign speakers' pronunciation errors for second language training In: *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-1996),* vol.3, 3-6 October 1996, Philadelphia. pp.1465-1468.

Eskenazi, M. 1999 Using Automatic Speech Processing for Foreign Language Pronunciation Tutoring: Some Issues and a Prototype In: *Language & Technology (LLT) Journal,* vol.2 (2), pp.62-76.

Evermann, G. 2002 *HTK History* [Online]. Available from World Wide Web: <http://htk.eng.cam.ac.uk/history.shtml>.

Fox, A. 1984 *German Intonation*. Oxford: Clarendon Press.

Grabe, E. 1998 *Comparative Intonational Phonology: English and German*, PhD thesis. Max-Planck-Institute for Psycholinguistic and University of Nijmegen.

Hansen, J.H.L. and Arslan, L.M. 1995 Foreign Accent Classification Using Source Generator Based Prosodic Features In: *Proceedings of the 1995 International Conference on Acoustic, Speech, and Signal Processing (ICASSP-1995)*, vol.1, 9-12 May 1995, Detroit. pp.836-839.

HTK 2000 [Online]. Available from World Wide Web: <http://htk.eng.cam.ac.uk/index.shtml>.

Hunt, J. 1996 *The Ascent of Everest*. Stuttgart: Ernst Klett Verlag, English Readers Series.

Interactive Spoken Language Education Non-Native Speech Data 1999 [CD-ROM]. Cambridge: Entropic.

Jurafsky, D and Martin, J.H. 2000 *Speech and Language Processing*. New Jersey: Pearson Higher Education.

Jurafsky, D., Wooters, C., Tajchman, G., Segel, J., Stolcke, A., Folser, E., and Morgan, N. 1994 The Berkeley Restaurant Project In: *Proceedings of the 3nd International Conference on Spoken Language rocessing (ICSLP-1994)*, September 1994, Yokohama. pp.2139-2142.

Knowles, G. 1996 From text structure to prosodic structure. In: Knowles, G., Wichman, A. and Alderson, P., (editors). *Working with Speech*. Harlow: Addison Wesley Longman Limited. pp. 146-167.

Menzel, W., Atwell, E., Bonaventura, P., Herron, D., Howarth, P., Morton, R., and Souter, C. 2000 The ISLE Corpus of non-native spoken English. In: Gavrilidou, M., Carayannis, G., Markantionatou, S., Piperidis, S., Stainhaouer, G. (editors). *Proceedings of 2nd International Conference on Language Resources and Evaluation (LREC 2000)*, vol.2, 31 May-2 June 2000, Athens. pp.957-964.

Morton, R. 1999 *Recognition of Learner Speech*, Project Report, Interactive Spoken Language Education Project LE4-8353, Deliverable D3.3. Cambridge: Entropic.

Oba, T. and Atwell, E. 2003 Using the HTK Speech Recogniser to analyse prosody in a Corpus of German Spoken Learners' English. In: *Proceedings of the 2003 International Conference on Corpus Linguistics (CL- 2003)*, 28-31 March 2003, Lancaster.

O'Connor, J.D. and Arnold, G.F. 1970 *Intonation of Colloquial English*, 7th edition. London: Longman.

Rodman, R. D. 1999 Computer Speech Technology. Norwood: Artech House Inc.

Souter, C., Howarth, P., and Atwell, E. 1999 *Speech Data Collection and Annotation*, Project Report, Interactive Spoken Language Education Project LE4-8353, Deliverable D3.1. Cambridge: Entropic.

Stemmer, G., Nöth, E., and Niemann, H. 2001 Acoustic Modeling of Foreign Words in a German Speech Recognition System In: Dalsgaard, P., Lindberg, B., and Benner, H., (editors). *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech-2001)*, vol.4, 3-7 September 2001, Aalborg. pp.2745-2748.

Taylor, P., King, S., Isard, S. and Wright, H. 1998 Intonation and Dialog Context as Constraints for Speech Recognition. In: *Language and Speech*, vol.41 (3-4), pp.493-512.

Taylor, P., King, S., Isard, S., Wright, H., and Kowtko, J. 1997 Using Intonation to Constrain Language Models in Speech Recognition. In: *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech-1997)*, vol.5, 22-25 September 1997, Rhodes. pp.2763-2766.

Teixeira, C., Trancoso, I., and Sarralheiro, A. 1996 Accent Identification. In: *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP-1996),* vol.3, 3-6 October 1996, Philadelphia. pp.577-580.

Tench, P. 1996 *The Intonation Systems of English*. London: Cassell.

Thambiratnam, D. 2001 *[HTK-Users] Problem with HERest* [Online].
Available from World Wide Web:
<http://htk.eng.cam.ac.uk/pipermail/htk-users/2001-August/001145.html>.

Uebler, U., Schüßler, M., and Niemann, H. 1998 Bilingual and Dialectal Adaptation and Retraining In: *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-1998)*, vol.15, 30 November-4 December 1998, Sydney. pp.1815-1818.

Werner, S and Keller, E. 1994 Prosodic Aspects of Speech. In Keller, E. (editor) *Fundamentals of Speech Synthesis and Speech Recognition*. Chichester: John Wiley & Sons Ltd. pp.23-40.

White, L. 2000 *English speech timing: a domain and locus approach*, PhD thesis. University of Edinburgh.

Witt, S. and Young, S. 1997 Language Learning Based on Non-Native Speech Recognition In: *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech-1997)*, vol.2, 22-25 September 1997, Rhodes. pp.633-636.

Woodland, P. 2000 *HTK History* [Online]. Available from World Wide Web:
<http://htk.eng.cam.ac.uk/history.shtml>.

Yan, Q. and Vaseghi, S. 2002 A Comparative Analysis of UK and US English Accents in Recognition and Synthesis In: *Proceedings of the 2002 International Conference on Acoustic, Speech, and Signal Processing (ICASSP-2002)*, vol.1, 13-17 May 2002, Florida. pp.413-416.

Young, S., Everman, G., Kershaw, D., Moore, G., Odell, J., Ollason, D., Valtchev, V., and Woodland P. 2001 *The HTK Book 3.1*. Cambridge: Entropic.