7th International Conference on DEVELOPMENT AND APPLICATION SYSTEMS
Suceava, Romania, May 27 – 29, 2004

KNOWLEDGE BASED TRANSCRIPTION OF HANDWRITTEN PITMAN'S SHORTHAND USING WORD FREQUENCY AND CONTEXT

Swe MYO HTWE[1], Colin HIGGINS[2], Graham LEEDHAM[3], Ma YANG[4]
*[1)-2)]The University of Nottingham
School of Computer Science and IT,
Wallaton Road, Nottingham, NG8 1BB, UK
[3)-4)]Nanyang Technological University,
School of Computer Engineering,
N4-#2C-77 Nanyang Avenue, Singapore 639798
[1)]smh@cs.nott.ac.uk, [2)]cah@cs.nott.ac.uk
[3)]asgleedham@ntu.edu.sg, [4)]mayang@pmail.ntu.edu.sg*

Abstract. *The paper proposes the computer transcription of handwritten Pitman shorthand as a mean of rapid text entry to handheld devices. Handwritten outlines are bound to be variation from writers to writers and it causes pattern recognition to be prone to errors, however these imperfections can be restored by the use of heuristic approach in the interpretation stage. The transcription accuracy can be improved by the combination of three factors: firstly, incorporating contextual knowledge as used by human readers; secondly, applying knowledge of the most frequent words of Pitman shorthand; and finally, adding knowledge of collocation. Statistical analysis of a Shorthand lexicon is presented and distribution of transcription accuracy based on accuracy of segmentation is discussed in the paper. Experiments using a phonetic Lexicon with 5000 entries show that the approach is efficient and produces a satisfactory transcription accuracy of 94%.*

Keywords: *Pitman shorthand, unigram approach, shorthand lexicon, most frequently used words*

Introduction
Motivation

Shorthand is a speech-recording medium practiced in real time English reporting community at a practical rate of about 120-180 words per minute. Computer assisted machine shorthand (Palantype and Stenotype) are widely used in present court reporting, but their major drawbacks of importability and additional space requirement of a keyboard limit their use and negate their use in a mobile environment. Handheld devices like Tablet PCs and Personal Digital Assistants (PDAs) are gradually taking a significant role in business practice, but their lack of high-speed text entry restricts their use such that most mobile rapid note-takers retain the traditional way of using a notepad and a pencil to record speech using shorthand or speedwriting. Therefore, innovation in the automated recognition and transcription of handwritten shorthand which is ideal for handheld computers has become a timely research topic.

Background

Evaluation of the potential of Pitman shorthand [1][2] as a means of rapid pen driven text entry to a computer has been reported since the 1980's. Research in the 1990's [3][4] emphasized the segmentation and recognition of Pitman outlines, however recent work [5][6] has concentrated more on backend transcription of shorthand primitives into English text.

As Pitman records speech phonetically, transcription of handwritten Pitman shorthand is related to techniques applied in keyboard driven shorthand machines. However Pitman requires the extra step of interpreting pattern primitives such as loops, strokes or hooks into a valid phonetic sequence beforehand. In initial work on shorthand machines [7], a simple phonetic code conversion algorithm was used to produce the most appropriate spelling for phonemes that the palantypist keyed. In later work, Newell et al [8] proposed a longest match transcription algorithm, which improved overall system performance, but still produced spelling and word boundary errors.

Detailed research in the automatic transcription of handwritten Pitman shorthand has been conducted by Leedham & Downton [9] and the transcription process has been categorised into two major sections: - *conversion of pattern primitives into phonetic representation using production rules and transliteration of phonetic strings into correct orthography English.* The most recent work by Nagabuhushan & Anami [6] proposed a dictionary supported transcription algorithm. Their work overviewed current transcription performance and concluded that further work is required in the homophones solution area.

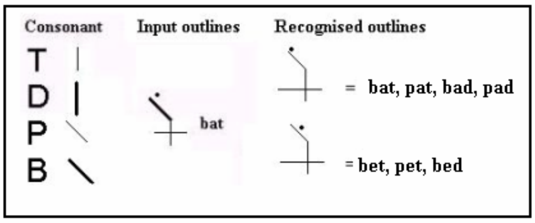


Figure 1. Illustration of construction of a Pitman shorthand outline

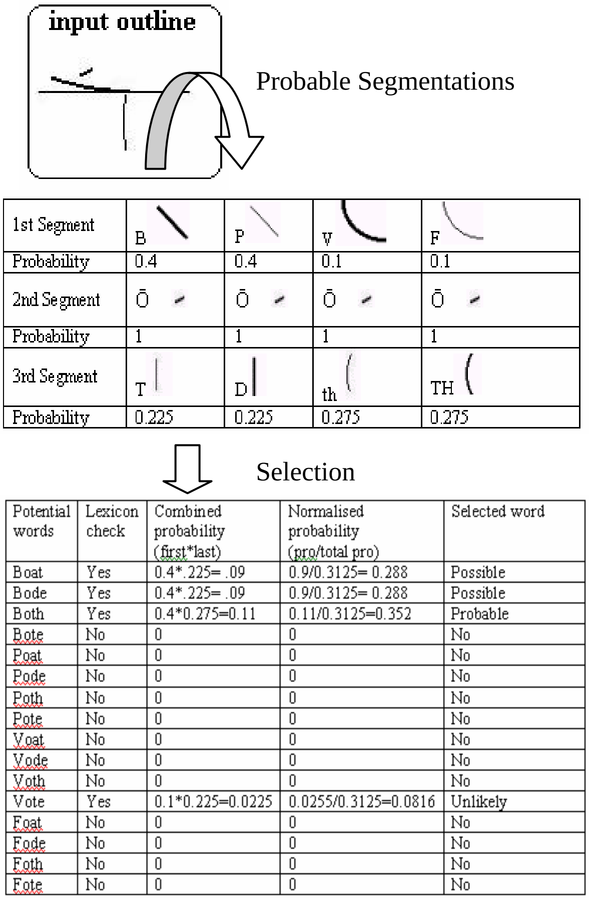
In general, homophones (outlines which are pronounced the same but have different spellings) are caused by two factors which start in the recognition stage. Firstly, most recognizers do not detect line thickness, whereas Pitman defines similar sounding consonants by the same stokes and differentiates between voiced and unvoiced sounds by the line thickness. Secondly, it is sometimes difficult to determine an accurate location for a vowel in the recognition stage even though the exact sound of a vowel is indicated by the writing position of a vowel symbol (i.e. beginning, middle or end of a consonant stroke) in Pitman shorthand. These two factors affect outline uniqueness and raise the occurrence of homophones. Figure 1 illustrates samples of Pitman shorthand outline and simulates the incidence of homophones.

In this paper, further approaches in the semantic transcription using the knowledge of word frequency and context is proposed. Due to parallel development of a recognizer system and the transcription engine, segmentation data was not available at the beginning of the work and it is assumed that input to the transcription system is a ranked list of Pitman basic features with interpreted phonetic values.

Use of Contextual Knowledge

Vocalised outlines are prone to homophones in Pitman shorthand. Transcription of a handwritten note is feasible often only by the original writer as an extensive use of contextual knowledge and sometimes memory is required [9]. Vocabulary is part of the contextual clues in reading shorthand script and the lexicon used varies depending on the level of English and the domain. Lexicon lookup methods with specific domain information are key to simulating the vocabulary knowledge in an automated system. However, depending on experience, Pitman writers omit vowels in long outlines, or sometimes just write down an essential one. Therefore, it is advisable to create different versions of shorthand dictionaries with and without vowels with the search based on the length of an outline. In general, there are 4.2 phonemes per word on average [10] and it in turn means the average length of Pitman shorthand is 4.2 segments per outline (spo). Our system uses a filter value of 3 spo to distinguish between short and long outlines. Any length less than 3 spo is taken as short and interpretation needs the full knowledge of vowel positions in an outline, and vice versa. On the whole, lexicon lookup methods are only practical in filtering non-valid words, and they are unable to deal with ambiguous candidates for a single outline. The slope of a stroke and position of an outline is vital in Pitman shorthand and a minor deviation can lead to different representations.

One of the interesting phenomena in reading English is that a sentence with spelling mistakes is usually comprehensible so long as the first and the last letters of each word are correct. For example, you may understand the following sentence even though it contains a number of typing errors - “Wornlgy seplled Egnlish words are sitll leiglbe as lnog as the frist and lsat ltteers are crroect.” A similar concept can be applied to shorthand transcription and ambiguous candidates of a single outline can be sorted upon the accuracy of the first and last consonant kernels. For instance, an input outline (Figure 2) is prone to ambiguous interpretation and the best match can be achieved by, say, choosing the interpretation with the highest combined accuracy of the first and last segments.



1st Segment	B	P	V	F
Probability	0.4	0.4	0.1	0.1
2nd Segment	Ō	Ō	Ō	Ō
Probability	1	1	1	1
3rd Segment	T	D	th	TH
Probability	0.225	0.225	0.275	0.275

Potential words	Lexicon check	Combined probability (first*last)	Normalised probability (pro/total pro)	Selected word
Boat	Yes	0.4*.225= .09	0.9/0.3125= 0.288	Possible
Bode	Yes	0.4*.225= .09	0.9/0.3125= 0.288	Possible
Both	Yes	0.4*0.275=0.11	0.11/0.3125=0.352	Probable
Bote	No	0	0	No
Poat	No	0	0	No
Pode	No	0	0	No
Poth	No	0	0	No
Pote	No	0	0	No
Voat	No	0	0	No
Vode	No	0	0	No
Voth	No	0	0	No
Vote	Yes	0.1*0.225=0.0225	0.0255/0.3125=0.0816	Unlikely
Foat	No	0	0	No
Fode	No	0	0	No
Foth	No	0	0	No
Fote	No	0	0	No

Figure 2: Transcription based on the combined accuracy of first and last segments

However, the above algorithm is possible only if the script is clearly written at the beginning and end of its outline. In normal English writing, people usually write clearly at the beginning of a word, but the script deteriorates and becomes ambiguous as it gets closer to the end, as in Figure 3.

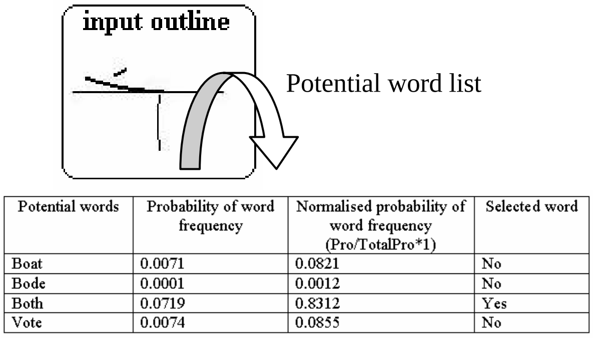
Figure 3: Illustration of normal handwritten script partially corrupted at the end

Our survey was done on 10 samples of shorthand notes handwritten by professional Pitman writers, and the study found that a majority of shorthand outlines are written clearly not only at the beginning but also at the end. Therefore, the above algorithm is believed to be effective enough to filter potential words from a wide range of instances.

Another way of reading ambiguous or highly distorted shorthand script is by spotting the most obvious segment of an outline and tracing the rest via this anchor segment. A similar technique can be replicated in the transcription system in such a way that a segment with the highest accuracy is considered as an anchor node and the search is based on local variables i.e., descendant primitives of the anchor segment. If the rule is applied to the example in Figure 2, the potential list will reduce to words starting with “B” or “P” i.e., “boat”, “both” and “bode” because stroke “B” or “P” is an anchor segment due to its high probability.

Use of most frequently used words

In order to cover multiple domains, our transcription system initially uses the first 5000 of most frequently used words and each word is tagged with its corresponding frequency value. If there are ambiguities in isolated outline during transcription, the one with the highest frequency value is chosen as a potential successor. This discrimination function may cause a transcription error, however additional factors like collation probability and contextual probability can be taken into account and the overall probability can be set as a final selection criteria. Figure 4 follows from Figure 2 and illustrates transcription based on word frequency. In practice, the probabilities illustrated in Figure 2 and Figure 4 will actually be combined together and then combined with sentence or phrase level collocation probabilities to give the best recognition results.



Potential words	Probability of word frequency	Normalised probability of word frequency (Pro/TotalPro*1)	Selected word
Boat	0.0071	0.0821	No
Bode	0.0001	0.0012	No
Both	0.0719	0.8312	Yes
Vote	0.0074	0.0855	No

Figure 4: Illustration of transcription based on word frequency

On the other hand, there can be complete transcription failure due to outputs being unjustly discarded by a limited size lexicon. In normal text transcription, such a problem demands the serious action of either setting a larger dictionary or developing new data structures. For Pitman shorthand transcription we can benefit from the phonetic construction rules of an outline. Any non-dictionary outlines could be directly presented in the form of International Phonetic Alphabets (IPA) and users can change the phonetics into actual words. The beauty of this approach is it can achieve up to 100% correct transcription, but it is time consuming and perhaps, it will be not be favored by stenographers.

Error consideration

In practice, transcription efficiency depends on the performance of the segmentation and recognition phases. In textbook shorthand, there is an obvious distinction between thick and thin strokes, but this is not the practice of shorthand writers in speed recording. Therefore, writing pressure is neglected by the recognizer system and either voiced or unvoiced consonants are taken as the same consonants by the transcription system. The solution may be practical, but it raises an ambiguity rate by approximately 8% in the classification of the shorthand Lexicon.

Another limitation on transcription performance is imposed by the omission of vowels in an outline. Omitted positions are unpredictable as they vary widely from writers' experience or individual inclination. If the only remedy to this is by exclusion of vowel notations from the shorthand lexicon and matching up segmented features without vowel components, the new version of the lexicon is expected to have about 34% ambiguity (Figure 6).

Implementation and experimental results

The current goal of our experiment is to analyze the dynamic change of numbers of unique outlines with the growth of the shorthand lexicon. It also estimates a degree of transcription accuracy when ambiguous words are discriminated by word frequency. Input to the transcription system is simulated in the form of basic Pitman features with respective phonetic values and a dictionary of the 5000 most frequently used words reflects a Pitman shorthand lexicon.

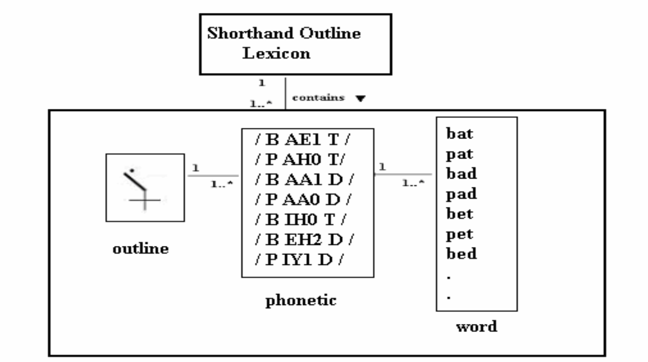


Figure 5: Illustration of the structure of a shorthand lexicon

The original lexicon is a phonetic dictionary with shorthand indexes built into it. Lexicon information is put into an object oriented data structure as illustrated in (Figure 5) and the whole lexicon is classified as a collection of lexicon objects. The purpose of our experiment is to monitor the growth of the lexicon object against the change of lexicon size to see if accuracy is related to size.

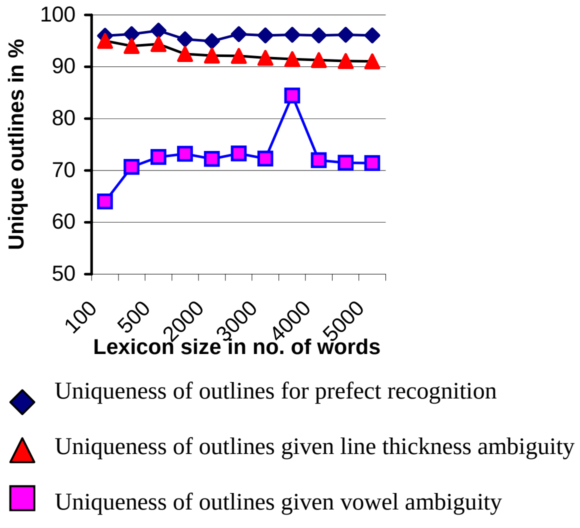


Figure 6: Analysis on shorthand Lexicon with the 5000 most frequently used words

The experimental results Figure 6 show that the growth rate is neither linear nor logarithmic but indicate that about 94% of the 5000 most common words have a unique shorthand notation. The maximum ambiguity is 5 potential words per outline and an average ambiguity is 3 potential words per shared outline. As seen in Figure 6, a fall occurs around a lexicon of 2000 words, which in turn means 5% of the most frequently used 2000 words have similar pronunciations. This indicates that real life transcription can expect at best 5% ambiguity rate for a vocabulary level of 2000 words. For larger lexicons the rate will be at best around 4%. The graph also shows there is a constant confusion rate after a Lexicon with 3000 words.

There is an exception to our experiment: - a group of about 90 commonly occurring words that are related to Pitman short-forms are constructed in the same way as vocalized outlines.

Discussion

Current findings are based on simulated data and further development needs be done on real data. A lexicon of the 5000 most common words is sufficient for a general area, but it is not enough for domain specific applications. Work needs to be done in optimizing dictionary lookup in the framework of multi-domain shorthand recognition, mostly towards developing new data structures for space restricted handheld devices. The current system has not complied complete rules of the Pitman system (e.g., rules of half length strokes, suffix, etc.,) and further work is required to analyze this area. The Graphical user interface (GUI) is another critical issue as a system that presents the user with choices for ambiguous words might be practical. Hence a closer study the operating systems of handheld devices and the design of the GUI needs to be done in the immediate future.

References

[1] Leedham C.G., Downton A.C., Brooks C.P. and Newwell A.F., (1984), *'On-line acquisition of Pitman's handwritten shorthand as a means of rapid data entry'*, Proc. 1st Int. Conf. On Human-Computer Interaction, London, UK, pp. 2.86-2.91

[2] Leedham C.G. and Downton A.C., (1986), *'On-line recognition of Pitman's shorthand: an evaluation of potential'*, Int. J. Man-Machine Studies, Vol.24, pp.375-393

[3] A. Nair A. and C.G. Leedham, (1992), *Evaluation of dynamic programming algorithms for the recognition of shortforms in Pitman's shorthand*, Pattern Recognition Letters, vol. 13, pp. 605-612.

[4] Y. Qiao and C.G. Leedham, (1993), *Segmentation and recognition of handwritten Pitman shorthand outlines using an interactive heuristic search*, Pattern Recognition, vol.26, No.3, pp.433-441

[5] P.Nagabhushan and Basavaraj.Anami, (2002) *"A knowledge-based approach for recognition of handwritten Pitman shorthand language strokes"*, Sadhana, Journal of Indian Academy of Sciences, Vol. 27, Part 5, pp. 685-698

[6] P.Nagabhushan and Basavaraj.Anami, (2002), *"Dictionary Supported Generation of English Text from Pitman Shorthand Scripted Phonetic Text"*, Language engineering conference, Hyderabad, India, pp.33

[7] Newell A.F., King J.A.F., (1977), *'Speech translation systems for the hearing impaired'*, Medical & Biological Engineering & Computing. : 15(5), p. 558-63

[8] Newell. A.F., Arnott.J.L., Dye R., Carins Y., (1991),*'A full-speed listening typewriter simulation'*, International Journal in Man-Machine studies: 35,p. 119-131,

[9] Leedham C.G., Downton A.C., (1990), *'Automatic recognition of Transcription of Pitman's Handwritten shorthand'*, In Plamondon R. and Leedham C.G. (Eds), Computer Processing of Handwriting, pp.235-269, World Scientific,

[10] Johannes C. Z., Conrad P., Arthur M.J., Mario B., *'Identical words are read differently in different languages'*, Psychological Science, 12, 379-384