European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

HIGH-SPEED CONTROL OF A SPEECH SYNTHESISER BY STENOTYPE KEYBOARD

J.L.Arnott *

ABSTRACT

Stenotype keyboards are the fastest proven means for encoding verbal information into machine-compatible form, hence they can be used as interfaces for real-time control of speech synthesisers. Systems have been developed for the British Palantype and American Stenograph keyboards, and these have been operated at real speech rates by skilled stenotypists. As well as being used with prepared texts, the systems were trialled successfully in spontaneous dialogue, the brisk speech rate allowing the dialogues to be fluent, articulate and relatively natural.

INTRODUCTION

The rate at which a human can convey information in synthetic speech form is limited by the communication rate of the man-machine interface. Human speech rates lie in the range 120-200 words/minute (wpm), but neither conventional typing (50-80 wpm), speed-typing (up to 100 wpm) nor written shorthand (90-120 wpm) can attain these word rates (ref 1). The fastest proven means of encoding verbal information into machine-compatible form is the stenotype keyboard, which can be operated, by a skilled stenotypist, at natural speech rates. This paper discusses the use of stenotype keyboards to control speech synthesisers at high word rates.

STENOTYPE KEYBOARDS

There are several stenotype keyboards, including orthographic and phonetic systems. The layouts of two English language phonetic keyboards, the British Palantype and the American Stenograph, are shown in Figure 1. Both keyboards have **a)** a left-hand group of initial consonants, **b)** a central group of vowel keys, and **c)** a right-hand group of final consonants. A skilled stenotypist would record spoken information on such a keyboard by **a)** segmenting each word into constituent syllables **b)** spelling phonetically each syllable, **c)** encoding this phonetic spelling for the keys available on the keyboard, and **d)** typing the syllable in one multi-key stroke of the keyboard. Certain characters which are not present on the keyboard (e.g. **B**, **D** and **W** (Palantype)) are stenotyped as multi-key codes (**P+**, **T+** and **MF** respectively). This method enables stenotypists to record verbal information at real speech rates. Any English

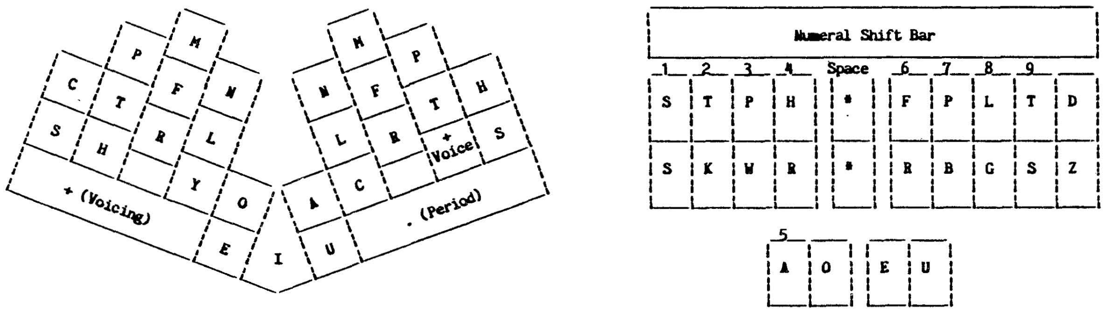

Figure 1: Layouts of Palantype (left) and Stenograph (right) Keyboards

* Microcomputer Centre, University of Dundee, Dundee DD1 4HN, Scotland.

10.21437/ECST.1987-116

word can be segmented, coded and stenotyped. The phonetic coding means that such a keyboard can be used as a control interface for a phonetic speech synthesiser (refs 2,3,4,5).

STENOTYPE-TO-SPEECH SYSTEM WITH NON-SIMULTANEOUS VOICING

A system was developed to enable a stenotypist to control a phonetic synthesiser on a phrase-by-phrase basis. The stenotypist would stenotype a complete phrase, terminate it, and the synthesiser would then synthesise the whole phrase. This system was termed a "non-simultaneous voicing", or "NSV", system because speech synthesis was not simultaneous with key activation. It had the specific advantage that a stenotypist could use a virtually standard keying technique. Apart from the need to maintain a rigorously phonetic coding regime, the stenotypist would perform as in a conventional (e.g. court reporting) task. (Stenotypists (Stenograph users in particular) can make use of non-phonetic codes and shortforms in conventional court reporting activity, a practice which must be avoided when using the speech output systems described here.)

The NSV system consisted of **a)** an electronic stenotype keyboard, **b)** a rule-based transliteration system to transform stenotype codes (e.g. P+ (=B) in Palantype) into phonetic codes, and **c)** a phonetic synthesiser. The transliteration system operated on similar principles to those used in text-to-speech processing (ref 6). The stenotypist could type at any speed, according to ability, but the synthesiser would set the rate of the synthetic speech.

USE OF NSV SYSTEM BY EXPERIENCED STENOTYPISTS

Experiments were devised to assess the feasibility of controlling a speech synthesiser from a stenotype keyboard. Professional stenotypists were commissioned to trial the stenotype-to-speech system. There were two experienced (full-speed court reporting) palantypists (Pal2 and Pal3), one experienced Stenograph user (Sten1), and one trainee palantypist (Pal1) with a maximum word rate of no more than 50 words/minute. In each experiment, a stenotypist was instructed to stenotype (in strictly phonetic code) selected texts (e.g. "The Rainbow"), these texts being presented in printed form. The synthetic speech was monitored, and the word rates were measured.

Steno-typist	**Mean Rate words/min**	**Peak Rate words/min**
Pal1	27.4	33.3
Pal2	47.3	94.3
Pal3	72.9	98.0
Sten1	52.4	54.7

Table 1: Measured Speech Rates from Stenotype NSV Speech System

Table 1 shows mean word rates achieved by the stenotypists in NSV mode, after brief familiarisation with the system. The highest rates were 50% (approx.) higher than those which a good Qwerty typist could achieve, although they were much lower than the maxima of which stenotypists are known to be capable. The rates were low because the stenotypists were unfamiliar with speech synthesis and strict phonetic coding, and because they were encouraged to attain coding accuracy rather than high speed. Speed and coding accuracy would improve with practice.

The fluency of the synthetic speech was limited somewhat by the phrase-by-phrase cycle of stenotyping-and-synthesis used by the NSV system. A system which effected simultaneous synthesis of speech was required, therefore, to give the stenotypist more flexible control over the speech synthesiser.

STENOTYPE-TO-SPEECH SYSTEM WITH SIMULTANEOUS VOICING

A "simultaneous voicing", or "SV", system was developed in order to give the stenotypist more immediate and spontaneous control over the speech synthesiser (somewhat similar to Voder operation (refs 7,8)). In the "SV" mode of operation, sounds are synthesised simultaneously with stenotype key depression, so that the synthesiser is "played", rather like a musical keyboard. Stenotypist and listener can hear the synthetic speech immediately upon key depression, so expression is more spontaneous and vital. It is also possible for the stenotypist to achieve a measure of prosodic control by varying the duration of speech sounds and syllables, simply by prolonging the depression of keys on the keyboard.

Steno-typist	Mean Rate words/min	Peak Rate words/min
Pal1	32.3	35.9
Pal2	88.9	108.5
Pal3	95.8	116.1
Sten1	50.8	56.0

Table 2: Measured Speech Rates from Stenotype SV Speech System

Table 2 shows speech rates achieved by the stenotypists in SV mode. The experienced palantypists were faster in SV than in NSV, approaching the lower regions of natural speech rates. The Stenograph user was less familiar with strict phonetic coding, however, and was therefore slower than the palantypists. Phonetic coding is possible on a Stenograph so a Stenograph user should be able to achieve at least low natural speech rates when good phonetic style has been acquired. The mean word rates of all these stenotypists would increase with further practice of synthesiser operation.

DIALOGUE USING AN SV STENOTYPE-TO-SPEECH SYSTEM

Dialogues were conducted in which each stenotypist "spoke" to a dialogue partner via the keyboard. The partner spoke with natural voice. The dialogues were unscripted and unprepared, and therefore spontaneous.

Steno-typist	Mean Rate words/min	Speaker Ratio	Listener Errors
Pal1	74.7	4.0	3
Pal2	125.1	2.5	2
Pal3	119.2	3.63	1
Sten1	60.4	2.14	1

Speaker Ratio =
Total Words Spoken/Total Words Steno'ed

Table 3: Dialogue in SV Mode: Word Rates

Table 3 shows that the mean word rates in the dialogues involving the two faster palantypists were within the lower range of natural speech rates. The natural speaker uttered more words than each stenotypist, mainly because the stenotypists were concise in their statements. The stenotypists were able to influence and contribute to the dialogue because they could control the synthesiser at high speed, and therefore interact fully with the natural speaker. The natural speaker (listener) made very few errors of comprehension, which indicated that the system functioned very effectively as a communication medium. The stenotypists were able to use unrestricted vocabulary - proper nouns such as place names and the names of musical composers were pronounced correctly and recognisably, a feature which compared well with the performance of most text-to-speech translation systems.

The SV system allowed the stenotypist to be more fluent and spontaneous than did the NSV system, although keying technique had to be more precise in the SV mode. This was justified, however, by the overall improvement in performance and flexibility, and the spontaneity of communication between the stenotypist and dialogue partner. A limitation of both the SV and the NSV systems was that the pitch of the synthetic speech could not be controlled from the keyboard. A fruitful aspect of future work would be to incorporate pitch control into the interface.

CONCLUSIONS ABOUT STENOTYPE-TO-SPEECH SYSTEMS

It can be concluded that a speech synthesiser can be controlled effectively from a stenotype keyboard, at speeds at least as fast as low natural speech rates, and that dialogue can be conducted using such a stenotype-to-speech system. Palantype and Stenograph keyboards are both suitable in this application, and word rates on both keyboards would increase with further familiarisation. The word rates which are possible with a stenotype-based system could not be attained by any other known man-machine interface method, indicating that stenotype keyboards are unique in this application. Stenotype-to-speech systems may have applications as speech prostheses for the manually-able speech impaired, and as training systems for conventional and sight-impaired stenotypists. They may also be of value in human factors experiments on the use of synthetic speech in the man-machine interface.

REFERENCES

1. A C Downton. "Simultaneous Transcription of Machine Shorthand for the Deaf", Ph.D. Thesis, Dept. of Electronics, University of Southampton, Southampton, U.K., 1982, p.7.

2. R D Steele and S L Wood, "An Efficient and Versatile Speech Prosthesis", Digest of Papers, Spring COMPCON 83, San Francisco, U.S.A., February/March 1983, pp.34-38.

3. R D Steele, "Stenotyped Input in Rehabilitation Applications", Proc. Sixth Annual Conf. on Rehab. Eng., San Diego, U.S.A., June 1983, pp.168-170.

4. G Stevens, D W Bell and J Bernstein, "Telephone Communication Between Deaf and Hearing Persons using Speech-to-Text and Text-to-Speech Conversion", Proc. Second Int Conf. on Rehab. Eng., Ottawa, 1984, pp.273-274.

5. J L Arnott and A F Newell, "Stenotype Shorthand and Speech Synthesis in Vocal Prosthesis for the Dexterous Speech Impaired", ibidem, pp.621-622.

6. H S Elovitz, R Johnson, A McHugh and J E Shore, "Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics", IEEE Trans. on ASSP, Vol.ASSP-24, No.6, Dec.1976, pp.446-459.

7. H Dudley, R R Reisz and S S A Watkins, "A Synthetic Speaker", Journal of the Franklin Institute, Vol.227, No.6, June 1939, pp.739-764.

8. H Dudley, "The Carrier Nature of Speech", Bell System Technical Journal, Vol.XIX, No.4, October 1940, pp.495-515.