

**Conclusion:** This Letter has described ROM-based algorithms and architectures for computing  $x \sin(\theta)$  and  $x \cos(\theta)$ . Some properties of sine and cosine functions are used to reduce the required table size. The proposed architecture can be pipelined to provide an  $x \sin(\theta)$  or  $x \cos(\theta)$  per addition or table look-up time. Simulation results show that acceptable precision can be achieved in a feasible ROM size. An application of the proposed method is also presented. The proposed coordinate rotator is pipelined to provide a throughput of 0.25 rotations per addition or table look-up time.

20th January 1992

H.-M. Jong, L.-G. Chen and T.-D. Chiueh (Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan, Republic of China)

#### References

- 1 HWANG, K. (Ed.): 'Computer arithmetic' (John Wiley & Sons, New York, 1979), pp. 201-206
- 2 TAYLOR, E. J., GILL, R., JOSEPH, J., and RADKE, J.: 'A 20 bit logarithmic number system processor', *IEEE Trans.*, 1988, C-37, pp. 190-200
- 3 CONSIDINE, V.: 'CORDIC trigonometric function generator for DSP'. Proc. IEEE ICASSP, 1989, pp. 2381-2384

## NEURAL NETWORK SYNTHESISER OF PAUSE DURATION FOR MANDARIN TEXT-TO-SPEECH

Shaw-Hwa Hwang and Sin-Horng Chen

*Indexing terms:* Speech synthesis; Signal processing, Neural networks

A neural network based approach of pause-duration synthesis for Mandarin text-to-speech is proposed. It uses an MLP to replace explicit synthesis rules for generating pause duration from input text. By properly training the MLP using a large set of utterances, phonological rules of producing pause duration are automatically learned. Experimental results confirmed that this is a promising approach.

**Introduction:** In Mandarin Chinese, each character is pronounced as a monosyllable. Pause duration between two successive monosyllables plays an important role in the naturalness of sentential speech; pause duration is thus important prosodic information in synthesising Mandarin text-to-speech. Traditionally, it is synthesised by a rule-based approach [1-2]. Phonological rules are invoked to imitate the human pronunciation process of generating pause duration from a given text. Although a rule-based approach is simple, the process of rule inference is tedious. Besides, because various linguistic features may interactively affect the pronunciation of pause duration, rules are usually incomplete.

In this Letter, a neural network based approach of pause-duration synthesis is proposed. The basic idea is similar to the NETalk used for assigning parameters of allophones to each English character according to the context [3]. In our method, a multilayer perceptron (MLP) is employed to replace explicit synthesis rules to generate pause duration according to the context. By properly training the MLP with a large training set using the error back-propagation algorithm, phonological rules are expected to be automatically deduced and implicitly memorised. The MLP can hence be taken as a mechanism of pause-duration synthesis. Intensive study of linguistics for rule inference is therefore unnecessary.

**Proposed system:** Fig. 1 shows the block diagram of the proposed system. It consists of two main parts. The first one is a text analysis in which some linguistic features representing the context are extracted. The second one is a single-output MLP serving as the mechanism of generating pause duration from these linguistic features. It is noted that the nonlinear operation in the output node of an MLP is removed for linearly generating multilevel values of pause duration.

720

Although linguistic features on various levels may affect the pronunciation of pause duration, in this study only some relevant linguistic features, listed below, are extracted from neighbouring context in the text analysis:

(1) The type of initial of the ensuing syllable: six broad types of initial listed below were used:

- (a) /m, n, l, r, 'null'/
- (b) /h, sh, shi/
- (c) /b, d, g/
- (d) /tz, j, ji/
- (e) /p, t, k/
- (f) /ts, ch, chi, f, s/

(2) The tone of the ensuing syllable: five lexical tones were used

(3) One positional parameter: is the ensuing syllable the ending syllable of a sentence?

(4) Two phrasal parameters:

- (a) does the processing location precede a polysyllabic phrase?
- (b) does the processing location lie within a polysyllabic phrase?

(5) Others: does there exist an intentional pause or breath?

From the above discussions, a total of 15 binary contextual features were used. We note that the last feature is used to compensate for the effect of unusual pauses that occasionally occurred in the training utterances.

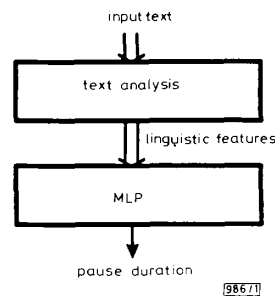


Fig. 1 Block diagram of proposed system

**Simulations:** The performance of the proposed approach was examined by simulations. Two sets of sentential utterances spoken by a female announcer were recorded from TV news. The first database comprising 2278 monosyllables was employed to train the MLP and the second database comprising 584 syllables was used for outside testing. All utterances are natural and fluent. They were manually segmented into syllable periods for extracting pause durations. Contextual features were also manually extracted.

Table 1 lists the average mismatch error between original and synthesised pause durations. Average mismatch errors of

Table 1 AVERAGE MISMATCH ERROR BETWEEN ORIGINAL AND SYNTHESISED PAUSE DURATIONS

	Mean	Absolute variation
Statistics of pause duration	7.33 frames	8-10 frames
	Average mismatch error	
Inside test	1.84 frames	
Outside test	1.88 frames	

1 frame = 3.75 ms.

1.84 frames (6.9 ms) and 1.88 frames (7.1 ms) were achieved for the inside and the outside tests, respectively. Comparing these results with the statistics of the training database with mean 7.33 frames and average absolute variation 8.10 frames, the system performed quite well. By closely analysing experimental results, we found that a very high hit rate, 95.83%, of correct determination as to whether a pause occurred or not was achieved for the inside test and 87.14% for the outside test.

**Conclusions:** A novel neural network based approach for synthesising pause duration for Mandarin text-to-speech has been presented. It simply uses an MLP to replace explicit synthesis rules as the mechanism of pause duration synthesis. Experimental results showed that, by this approach, reasonably good pause information can be efficiently generated using only some simple contextual features. It is thus a promising approach.

3rd February 1992

Shaw-Hwa Hwang and Sin-Horng Chen (Department of Communication Engineering and Center for Telecommunications Research, National Chiao Tung University, Taiwan, Republic of China)

#### References

- 1 LEE, L. S., TSENG, C. Y., and OUH-YOUNG, M.: 'The synthesis rules in a Chinese text-to-speech system', *IEEE Trans.*, 1989, ASSP-37, pp. 1309-1319
- 2 ZHANG, J.: 'Acoustic parameters and phonological rules of a text-to-speech system for Chinese'. Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, Japan, 1986, pp. 2023-2026
- 3 SEJNOWSKI, T. J., and ROSENBERG, C. J.: 'NETalk: a parallel network that learns to read aloud'. TR. JHU/EECS-86/01, The Johns Hopkins University Electrical Engineering and Computer Science, 1986

### LiNbO<sub>3</sub> WAVEGUIDE SHG DEVICE WITH FERROELECTRIC-DOMAIN-INVERTED GRATING FORMED BY ELECTRON-BEAM SCANNING

M. Fujimura, T. Suhara and H. Nishihara

*Indexing terms:* Integrated optics, Nonlinear optics, Optical waveguides, Harmonic generation, Electron beam lithography

A ferroelectric-domain-inverted grating was fabricated by electron beam scanning in LiNbO<sub>3</sub>. A waveguide second harmonic generation (SHG) device with the grating was fabricated and demonstrated for the first time. The experiments were performed using a CW-Nd:YAG laser, and normalised SHG conversion efficiency of 50%/W was obtained.

**Introduction:** The LiNbO<sub>3</sub> waveguide quasiphase-matching (QPM) second harmonic generation (SHG) device with ferroelectric-domain-inverted grating is one of the most promising devices for a compact short-wavelength coherent light source [1-5]. Formation of the domain-inverted grating is an important process in the fabrication of such devices. Inversion techniques such as Ti indiffusion into the +z face of LiNbO<sub>3</sub> [1, 2], Li<sub>2</sub>O outdiffusion from the +z face [3], and SiO<sub>2</sub> cladding on the +z face and heat treatment [4] have been used to form the domain-inverted gratings in waveguide SHG devices and waveguide SHG experiments have been performed.

It has been reported recently that electron beam (EB) scanning on the -z face induced the inversion without poling field at room temperature [6, 7]. This technique provided domain walls perpendicular to the surface and continued to the +z face. Bulk SHG experiments have been performed [7]. Such an inversion structure allows larger cross-sectional overlap between guided waves and the domain-inverted grating, which is an important requirement for waveguide SHG devices. Effective application of the inversion structure for implementing efficient waveguide SHG devices is expected. This Letter

reports a waveguide SHG device fabricated by EB scanning inversion for the first time.

**Domain inversion:** The +z face of LiNbO<sub>3</sub> of 0.5 mm thickness was coated with an Au film of 50 nm and EB was scanned on the -z face at room temperature to form domain inverted gratings of 6.4 μm period. The EB acceleration voltage, current, and diameter were 20 kV, 0.3 nA, and ~0.3 μm, respectively. The inversion pattern was observed after etching in a mixture of HF:HNO<sub>3</sub>, which etches only the -z face of LiNbO<sub>3</sub>.

Fig. 1 shows an SEM photograph of an etched top surface. Although the EB was scanned continuously along the grating line, the inversion occurred in segmented regions. The separation between two adjacent segments was ~1 μm, and it was independent of EB scanning speed that ranged from 0.02 to 0.1 mm/s. Because the electric field due to accumulated electrons may be related to the inversion pattern, EB scanning by another scanning mode may result in a different inversion pattern. EB scanning by a dotted-line-like mode was tried. The inversion pattern tended to be segmented similarly to the results for the continuous scanning mode. For large EB dot spacing, the separation of the domain segments corresponded to the EB spacing. However, for EB dot spacing less than 1 μm, the domain separation was 1 μm and independent of the EB dot spacing.

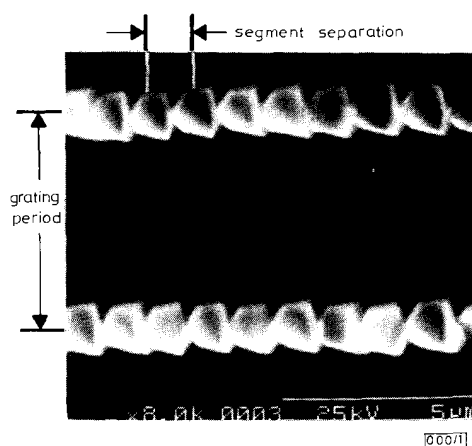


Fig. 1 SEM photograph of domain inverted grating fabricated by electron beam scanning (after etching)

It was found that excessive accumulation of electrons limited the area of the domain-inverted grating. Gratings up to a few square millimetres in area were successfully fabricated. However, for EB scanning over larger areas, domain inverted grating was obtained only in a part of the scanned area and no inversion occurred in the area scanned.

**Device fabrication:** A prototype waveguide QPM-SHG device was fabricated by EB scanning inversion. The device had a fanout domain-inverted grating and a channel waveguide array, as shown in Fig. 2, to compensate for residual phase mismatch [2]. The fanout grating was divided into three parts to obtain SHG interaction length of 3.3 mm against the grating area limitation mentioned before. The grating period ranged from 5.76 to 7.04 μm.

The domain-inverted grating was formed by dotted-line-like EB scanning with EB dot spacing 2.3 μm and scanning speed 0.03 mm/s. The channel waveguide array was fabricated by selective proton-exchange in pure benzoic acid and annealing [4].

**SHG experiments:** CW-Nd:YAG laser light of 1.06 μm wavelength was end-coupled through a ×20 lens and a TM-like fundamental guided-mode was excited in one of the channel waveguides. A guided-mode SH wave by first order QPM was obtained from the channel with the domain-inverted grating