



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Parallelization of Speech Compression Algorithm Based on Human Auditory System on Multicore System

¹Teddy Surya Gunawan, ²Mira Kartiwi and ¹Othman O. Khalifa

¹Department of Electrical and Computer Engineering, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia

²Department of Information Systems, International Islamic University Malaysia, Jalan Gombak, 53100 Kuala Lumpur, Malaysia

Abstract: Human auditory system has been successfully employed in speech compression to reduce bit rate requirement. Previous researches on speech compression stated that the use of simultaneous masking and/or temporal masking reduced the bit rate requirement while maintained perceptual quality. However, the benefit of using auditory masking in speech coding was outweighed by the amount of computation required to calculate masking threshold. Nevertheless, the current advances in microprocessor technology shows that to overcome heat generation and power cap, a multicore processor integrates two or more independent cores into a single package become popular. The objective of this research is to develop and implement a novel parallel speech compression algorithm based on human auditory system on a multicore system. To achieve a scalable parallel speech coding algorithm, Single Program Multiple Data (SPMD) programming model was used, in which a single program was written for all cores. Matlab parallel computing toolbox was used in the implementation. Finally, the performance of the developed parallel algorithm was evaluated using Perceptual Evaluation of Speech Quality (PESQ) and parallel execution time. Results show that the average PESQ score and pulse reduction for all 30 files when both auditory masking models were used is around 4.00 (transparent quality) and 41.12%, respectively. Moreover, the maximum speed achievable on our parallel experiment was around 2.45 if four cores were fully utilized.

Key words: Human auditory system, simultaneous masking, temporal masking, speech compression, PESQ, SPMD, parallel execution time

INTRODUCTION

Nowadays, major development of computer architecture is the implementation of highly integrated chips. This trend has been promoting the development of processors that can perform the functions associated with the entire system. Monolithic processor is very expensive to develop, requires high power consumption but with limited return on investment. Therefore, it is apparently much more effective to build modular processors with multicores rather than building single core higher clock frequency. Further development of multicores by Intel and AMD, not only combines multiple CPU (Central Processing Unit) but also integrates both CPU and GPU (Graphics Processing Unit) to achieve high performance computing with low power requirement.

The most commercially significant multi-core processors are those used in personal computers, primarily from Intel and AMD (Borkar *et al.*, 2006; Geer, 2005) and game consoles, e.g., the eight-core Cell processor in the PS3 and the three-core Xenon processor

in the Xbox 360 (Gorder, 2007; Petrini *et al.*, 2007). Several design avenues have been explored both in academia, such as the Raw multiprocessor and TRIPS and in the industrial world, with notable examples being the AMD Opteron, IBM Power5, Sun Niagara, Intel Montecito and many others (Herlihy and Shavit, 2008; Wang *et al.*, 2009). Other manufacturer such as NVidia focuses on the modern Graphic Processing Unit (GPU) which is optimized for executing a large number of threads in parallel (Rinaldi *et al.*, 2011). In the software part, a new parallel programming paradigm has been proposed by Shah and Cherkoot (2011), in which GUI (Graphical User Interface) based is used for synchronization scheme. These advances in hardware and software part means that parallel programming will become more popular in the future.

Multicore has been utilized to speedup various digital signal processing algorithms that require high computational time. Multicore and cluster system has been used for MPEG4 and H.264 encoding and decoding of a video signal (Berekovic *et al.*, 2002; Gunawan, 2001;

Lu *et al.*, 2008; Nishihara *et al.*, 2008). Chen *et al.* (2008) have been implemented parallel Hough transform in multicore systems to enhance the computing performance of feature extraction for image analysis and computer vision. Petrini *et al.* (2007) and Williams *et al.* (2008), have implemented parallel graphic 3D and lattice Boltzmann simulation. Hemalatha and Vivekanandan (2008) proposed parallel implementation of k-means algorithm which is normally used for signal classification. Parallel processing also has been used in increasing the performance of database access (Kadry and Smaili, 2008) particle swarm optimization algorithm (Hernane *et al.*, 2010), visualization of large medical images (Kofahi and Qureshi, 2005). These researches showed that in the future multicore and parallel programming will become more utilized in everyday life.

In terms of speech processing algorithms, multicore system also have been utilized to enhance the performance of parallel speech understanding system (Chung *et al.*, 1993) and speech recognition (Ou *et al.*, 2008). Other speech processing algorithms, such as speech coding (Gunawan *et al.*, 2005, 2011), speech enhancement and speech recognition (Frikha and Hamida, 2007; Ilyas *et al.*, 2010) also require high amount of processing time. Recent advances in multi-core system make it a natural choice and viable option for solving high computation requirements of the speech compression algorithm using simultaneous and temporal masking models. Although many researches have been done on multicore system, however, none has been focused on other parallel speech algorithms such as speech coding exploiting human auditory system. Therefore, the objective of this paper is to develop novel parallel speech coding algorithm and to implement and evaluate on a multicore system.

SPEECH COMPRESSION ALGORITHM

The auditory models utilization in speech and audio coding is by no means new and their applications include low bit rate speech coding (Black and Zeytinoglu, 1995) through to MPEG audio compression (Bradenburg *et al.*, 1994). Simultaneous masking is a frequency domain phenomenon in which a low-level signal can be rendered inaudible by a simultaneously occurring stronger signal if both signals are sufficiently close in frequency. Temporal masking is a time domain phenomenon in which two stimuli occur within a small interval of time (Zwicker and Zwicker, 1991). The discussion of basic speech compression algorithm in this research is based on (Gunawan *et al.*, 2005).

Gammatone analysis/synthesis filter bank: Figure 1 shows the serial algorithm of speech coding using

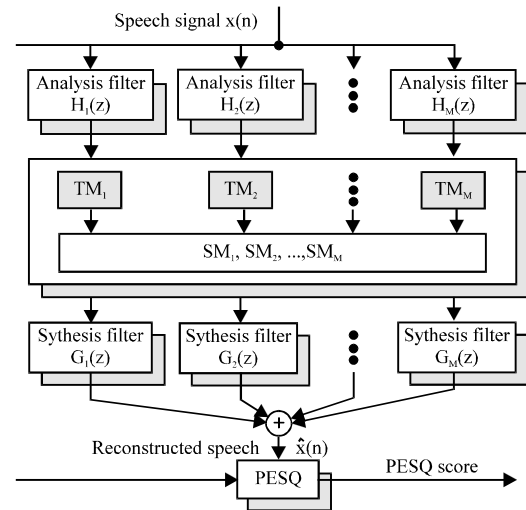


Fig. 1: Gammatone based speech compression algorithm (Gunawan *et al.*, 2005)

gammatone front end, in which parallelization of this serial algorithm will be the focus of this research.

Gammatone filters are implemented using FIR filters to achieve linear phase filters with identical delay in each critical band. To achieve linear phase filters, the synthesis filters $g_m(n)$ is designed to be the time reverse of its analysis filters $h_m(n)$. The analysis filter for each subband m is obtained using the following expression:

$$h_m(n) = a(nT)^{N-1} e^{-2\pi n \text{ERB}(f_c) nT} \cos(2\pi f_c nT + \phi) \quad (1)$$

where, f_c is the centre frequency for each subband m , T is the sampling period and N is the gammatone filter order ($N = 4$). The analysis filter bank output is followed by a half-wave rectifier to simulate the behaviour of the inner hair cells. Furthermore, a simple peak picking operation is implemented to simulate the nature of the neuron firing. This process results in a series of critical band pulse trains, where the pulses retain the amplitudes of the critical band signals from which they were derived. The simultaneous and/or temporal masking operation is then applied to the pulses, in order to remove the perceptually irrelevant peaks.

Auditory masking models: In this research, a simultaneous masking model similar to that used in MPEG (Black and Zeytinoglu, 1995) was utilized to calculate the masking threshold, SM , for each critical band. It has been shown that simultaneous masking removes redundant pulses in the structure shown in Fig. 1. While for temporal masking model, the model that was developed by (Najafzadeh *et al.*, 2003) was used as follows:

$$M_r(t, m) = a(b - \log_{10} \Delta t)(L(t, m) - c) \quad (2)$$

where, $M_f(t, m)$ is the amount of forward masking (dB) in the m th band, Δt is the time difference between the masker and the maskee in milliseconds, $L(t, m)$ is the masker level (dB) and a, b and c , are parameters that can be derived from psychoacoustic data. More details of masking models calculation can be found by Gunawan *et al.* (2005).

PARALLEL SPEECH CODING ALGORITHM

In this research, the hardware used was a computer system with AMD quad cores 2.5 GHz and 6 GBytes of memory. For the software part, it used Windows 7 64 bits and Matlab 2011a with Parallel Computing Toolbox and Distributed Computing Server. The available parallel programming models in Matlab Parallel Computing Toolbox are parallel-for (parfor), distributed array, Single Program Multiple Data (SPMD) and interactive parallel computation with pmode (MathWorks, 2010). In this paper, Single Program Multiple Data (SPMD) parallel programming models was chosen as the parallel program written for multicores system can be extended easily into a cluster system with many computers to further speed up

the computation. Moreover, SPMD allows the fine tuning of computational and communication parts so that the effective parallelization will be achieved.

Figure 2 shows the flowchart of parallel speech compression algorithm. Initially, the program starts with initialization at each core. Then, a speech signal is partitioned using subband partition and distributed to N_{cores} configured in master-slave fashion. After that, each slave (or workers in Matlab) is filtered using gammatone filter. The peak picking operation is applied to each subband to obtain pulse train. Simultaneous and/or temporal masking threshold is then calculated and applied to these pulses, where pulses with amplitude below the masking threshold were removed. The critical band signals can then be reconstructed from the masked pulse trains by means of bandpass filtering and summing the outputs.

RESULTS AND DISCUSSION

Here, the performance of sequential code, in terms of subjective and objective quality of the enhanced speech, was evaluated. Furthermore, the performance of

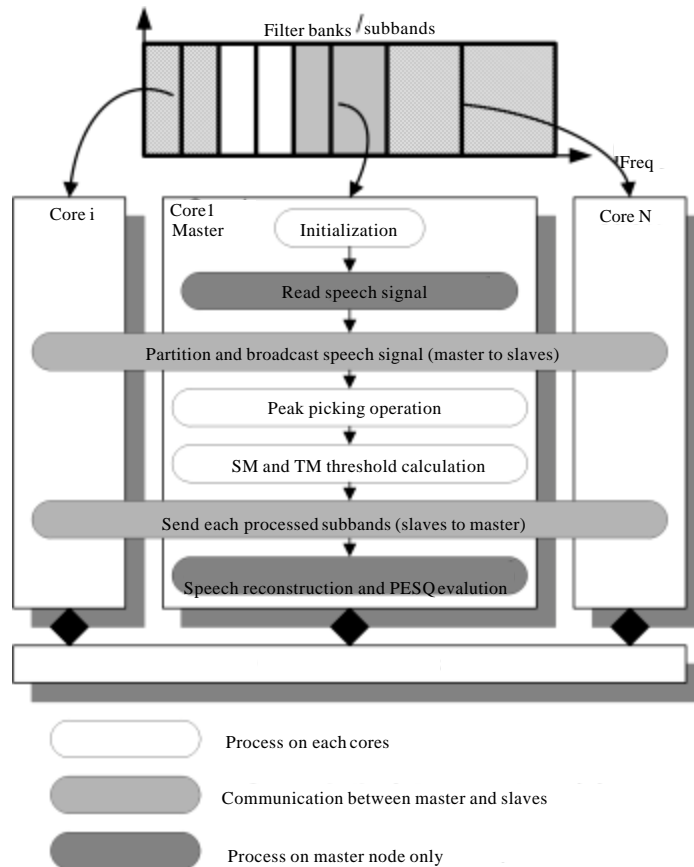


Fig. 2: Flowchart of parallel speech compression algorithm

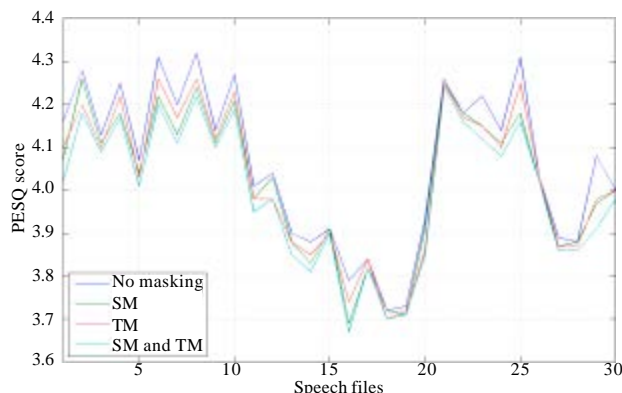


Fig. 3: PESQ score for various speech files

parallel code, in terms of speedup for various numbers of cores, was presented.

Datasets: To evaluate the performance of proposed algorithms, NOIZEUS datasets were used (Ma *et al.*, 2009). Thirty sentences from the IEEE sentence database were recorded in a sound-proof booth. The sentences were produced by three male and three female speakers and originally sampled at 25 kHz then downsampled to 8 kHz. The recorded signal number 1 to 10 and 21 to 25 were produced by male speakers, while the recorded signal 11 to 20 and 26 to 30 were produced by female speakers, respectively. The NOIZEUS datasets were used as it contains all phonemes in the American English language.

Subjective and objective evaluation: We evaluated the quality of reconstructed speech with no masking, simultaneous masking, temporal masking and combined simultaneous and temporal masking. It was found that all configuration achieved transparent quality (defined here as $PESQ > 3.50$) as shown in Fig. 3. The pulse reduction (bit rate reduction) achieved from 30 speech files was from 0% (no masking) to 55.8% (combined masking threshold). The average PESQ score and pulse reduction for all 30 files when both auditory masking models were used is around 4.00 (transparent quality) and 41.12%, respectively. Note that, the transparent quality was confirmed by the informal listening test. Therefore, the combination of temporal masking and simultaneous masking produces an optimum quality (PESQ) and pulse saving (bit rate reduction). Nevertheless, the calculation of both masking models will further increase the computational time, in which parallel implementation will be the solution.

Parallel performance: The performance of parallel implementation of the proposed algorithm will be

Table 1: Parallel execution time and speedup for various numbers of cores

Number of cores	Parallel execution time (sec)	Speedup
1	0.93	1
2	0.52	1.79
3	0.41	2.27
4	0.38	2.45

evaluated in terms of processing time and speed up curve. A multicore system with quad cores was used for evaluation. The number of cores used was varied between one (serial program) until maximum four (parallel program). Table 1 shows the elapsed processing time when the number of cores is varied from one to four and averaged across 30 files. Note that, the experiment is repeated three times for each speech file and the parallel execution time is averaged. It is evident from the table that the higher number the cores, the faster the processing time.

Table 1 also shows the speedup for various numbers of cores. Ideally, the speed up increases linearly with higher number of cores. However, the ideal speed up can only be achieved if the communication time between master and slave is negligible and the data is equally distributed. It is found that the optimum number of cores for speech coding is two cores. Therefore, the speed up will be less than ideal if the number of cores is further increased. This could be due to the time to coordinate and communicate the data is higher than the actual computational time.

CONCLUSIONS

In this study, a parallel implementation of speech coding algorithm based on human auditory system has been presented. The average PESQ score and pulse reduction for all 30 files when both auditory masking models were used is around 4.00 (transparent quality) and 41.12%, respectively. Therefore, the combination of temporal masking and simultaneous masking produces an

optimum quality (PESQ) and pulse reduction. The speech compression algorithm was implemented using SPMD paradigm on a multicore system with maximum of four cores. Results show that the higher number of cores, the lower the processing time. However, it was found that the maximum speed up achievable on our experiment was around 2.45 if four cores were fully utilized. Further research into more efficient parallelization scheme will be conducted.

ACKNOWLEDGMENTS

The authors gratefully acknowledged that this research has been supported by IIUM research grant EDW B10-391.

REFERENCES

- Berekovic, M., H.J. Stolberg and P. Pirsch, 2002. Multicore system-on-chip architecture for MPEG-4 streaming video. *IEEE Trans. Circuits Sys. Video Technol.*, 12: 688-699.
- Black, M. and M. Zeytinoglu, 1995. Computationally efficient wavelet packet coding of wide-band stereo audio signals. *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, May 9-12, 1995, Detroit, USA., pp: 3075-3078.
- Borkar, S.Y., P. Dubey, K.C. Kahn, D.J. Kuck and H. Mulder *et al.*, 2006. Platform 2015: Intel processor and platform evolution for the next decade: Intel corporation. *Syst. Technol.*
- Bradenburg, K., G. Stoll, Y. Dehery, J. Johnston, L. kerkhof and E. Schroeder, 1994. The ISO/MPEG audio codec: A generic standard for coding of high quality digital audio. *J. Audio Eng. Soc.*, 42: 780-791.
- Chen, Y.K., W. Li, J. Li and T. Wang, 2008. Novel parallel Hough Transform on multi-core processors. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, March 31-April 4, 2008, Las Vegas, USA., pp: 1457-1460.
- Chung, S.H., R.F. DeMara and D.I. Moldovan, 1993. Pass: A parallel speech understanding system. *Proceedings of Proceedings of 9th Conference on Artificial Intelligence for Applications*, March 1-5, 1993, Orlando, USA., pp: 136-142.
- Frikha, M. and A.B. Hamida, 2007. Noise robuse isolated word recognition using speech feature enhancement techniques. *J. Applied Sci.*, 7: 3935-3942.
- Geer, D., 2005. Chip makers turn to multi-core processors. *Computer*, 38: 11-13.
- Gorder, P.F., 2007. Multicore processors for science and engineering. *Comput. Sci. Eng.*, 9: 3-7.
- Gunawan, T.S., 2001. Parallel Architectures and Algorithms for Motion Analysis. M.Eng. Thesis, Nanyang Technological University, Singapore.
- Gunawan, T.S., E. Ambikairajah and D. Sen, 2005. Speech and Audio Coding Using Temporal Masking. In: *Signal Processing for Telecommunications and Multimedia*, Wysocki, T.A., B. Honary and B.J. Wysocki (Eds.), Vol., 27, Springer, USA, ISBN: 9780387228471, pp: 31-42.
- Gunawan, T.S., O.O. Khalifa, A.A. Shafie and E. Ambikairajah, 2011. Speech compression using compressive sensing on a multicore system. *Proceedings of 4th International Conference on Mechatronics*, May 17-19, 2011, Kuala Lumpur, pp: 1-4.
- Hemalatha, M. and K. Vivekanandan, 2008. A semaphore based multiprocessing k-mean algorithm for massive biological data. *Asian J. Sci. Res.*, 1: 444-450.
- Herlihy, M. and N. Shavit, 2008. *The Art of Multiprocessor Programming*. Morgan Kaufmann Publishers USA..
- Hernane, S., Y. Hernane and M. Benyettou, 2010. An asynchronous parallel particle swarm optimization algorithm for a scheduling problem. *J. Applied Sci.*, 10: 664-669.
- Ilyas, M.Z., S.A. Samad, A. Hussain and K.A. Ishak, 2010. Improving speaker verification in noisy environments using adaptive filtering and hybrid classification technique. *Inform. Technol. J.*, 9: 107-115.
- Kadry, S. and K. Smaili, 2008. Massively parallel processing distributed database for business intelligence. *Inform. Technol. J.*, 7: 70-76.
- Kofahi, N.A. and K.U. Qureshi, 2005. An efficient approach of load balancing for parallel visualization of blood head vessel angiography on cluster of WS and PC. *J. Applied Sci.*, 5: 1056-1061.
- Lu, Q., J. Chen and Y. Yang, 2008. An MPEG4 simple profile decoder on a novel multicore architecture. *Proceedings of Proceedings of 2008 Congress on Image and Signal Processing*, May 27-30, 2008, Sanya, China, pp: 489-493.
- Ma, J., Y. Hu and O. Loizou, 2009. Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions. *J. Acoust. Soc. Am.*, 125: 3387-3405.
- MathWorks, 2010. *Parallel Computing Toolbox 5: User's Guide*. MathWorks, USA.
- Najafzadeh, H., H. Lahdidli, M. Lavoie and L. Thibault, 2003. Use of auditory temporal masking in the MPEG psychoacoustics model 2. *Proceedings of the 114th Convention*, March, 2003, Audio Engineering Society, USA.

- Nishihara, K., A. Hatabu and T. Moriyoshi, 2008. Parallelization of H.264 video decoder for embedded multicore processor. Proceedings of the International Conference on Multimedia and Expo 2008, June 23-April 26, 2008, Hannover, Germany, pp: 329-332.
- Ou, J., J. Cai and Q. Lin, 2008. Using SIMD technology to speed up likelihood computation in HMM-based speech recognition systems. Proceedings of International Conference on Audio, Language and Image Processing, July 7-9, 2008, Shanghai, China, pp: 123-127.
- Petrini, F., G. Fossum, J. Fernandez, A.L. Varbanescu, M. Kistler and M. Perrone, 2007. Multicore surprises: Lessons learned from optimizing Sweep3D on the Cell Broadband Engine. Proceedings of Proceedings of IEEE International Parallel and Distributed Processing Symposium, March 26-30, 2007, Long Beach, USA., pp: 1-10.
- Rinaldi, P.R., E.A. Dari, M.J. Venere and A. Clausse, 2011. Lattice-Boltzmann Navier-stokes simulation on graphic processing units. Asian J. Applied Sci., 4: 762-770.
- Shah, N. and V. Cheerkoot, 2011. A scheme based paradigm for concurrent programming. J. Software Eng., 5: 108-115.
- Wang, X., Z. Ji, C. Fu and M. Hu, 2009. A review of hardware transactional memory in multicore processors. Inform. Technol. J., 8: 965-970.
- Williams, S., J. Carter, L. Oliker, J. Shalf and K. Yelick, 2008. Lattice Boltzmann simulation optimization on leading multicore platforms. Proceedings of Proceedings of IEEE International Symposium on Parallel and Distributed Processing, April 14-18, 2008, Miami, USA., pp:1-14.
- Zwicker, E. and T. Zwicker, 1991. Audio engineering and psychoacoustics, matching signals to the final receiver, the human auditory system. J. Audio Eng. Soc., 39: 115-126.