

Codificación y comprensión de señales de voz con cuantización vectorial no determinística

Max Suell Dutra

Dr.-Ing. Mecânica,

Gerhard Mercator Universität Gh Duisburg

Professor Associado COPPE, Diretor Grupo de Pesquisa em Projetos de Maquinas e Robótica,

Universidade Federal do Rio de Janeiro

Rio de Janeiro, Brasil

max@mecanica.coppe.ufrj.br

César Hernando Valencia Niño

MSc. da Engenharia Mecânica,

Universidade Federal do Rio de Janeiro

Docente Tempo Completo,

Investigador Grupo AMSCP,

Universidad Santo Tomás de Aquino USTA

Bucaramanga, Colombia

cesar.valencia@ufrj.br

Rodolfo Sánchez García

Ing. de Telecomunicaciones - Universidad Santo Tomás

Acanto Telecomunicaciones y Redes S.A.

Cisco Certified Network Associated (CCNA)

rsanchez@acanto.cl

Resumen— En los sistemas de transmisión [1] actuales es de vital importancia garantizar la integridad de la información y a su vez, que ésta se encuentre protegida de cualquier ambiente externo que exponga la misma a fuentes no deseadas. Por esto la codificación se ha convertido en una de las herramientas más importantes para garantizar la seguridad y exclusividad de la información entre su origen y destino; frecuentemente encontramos que las señales que involucran información son de tamaños mayores y, por lo tanto, su procesamiento y transmisión es más complicado, por esto la alternativa de utilizar métodos de compresión de datos también es usada frecuentemente; en este artículo es utilizada la técnica de cuantización vectorial no determinística para codificar y comprimir la información. De la misma forma, es posible encontrar que al elevar el grado de compresión e implementar algoritmos de codificación se encuentran problemas que comprometen la calidad de la información, es posible que al querer recuperar los datos originales características propias del procesamiento de datos generan pérdidas en la inteligibilidad de la señal original, para esto son utilizadas técnicas diversas como el algoritmo LMS de filtrado adaptativo y así obtener un aprovechamiento significativo.

Palabras clave— Codificación, Compresión, Cuantización Vectorial, Ruido de Cuantización, Filtrado Adaptativo.

Abstract— In the present systems of transmission [1] is vitally important to guarantee the integrity of the informa-

tion and, that as well this one is protected of any external atmosphere that exposes the same to sources non-wished. It is codification has become one of the tools most important to guarantee the security and exclusive feature of the information between its origin and destiny; frequently we found that the signals that involve information are of sizes majors and therefore its processing and transmission is more complicated, by this the alternative to use methods of data compression also is used frequently, in this article the technique of non-deterministic vector quantization is used to codify and to compress the information. The same form, it is possible to find that when elevating the compression degree and implementing codification algorithms are problems that menace the quality of the information, is possible that when wanting to recover the own original data characteristic of the data processing they generate lost in the intelligibility of the original signal, for this they are used technical diverse like adaptive filtering algorithm LMS and thus to obtain a significant advantage.

Keywords— Codification, Compression, Vector Quantization, Quantization Noise, Adaptive Filtering.

I. INTRODUCCIÓN

En repetidas ocasiones la naturaleza de las señales de voz dificulta cualquier tipo de trabajo que sea requerido para reducir o mejorar las cualida-

des de este tipo de señales. Es así como características adicionales a los sistemas de comunicaciones tradicionales como la codificación son realmente difíciles de implementar e, inclusive, en algunos casos imposibles debido al alto costo computacional[2]; mediante la utilización de la cuantización vectorial es posible obtener codificaciones que incrementan el nivel de seguridad en la información y, además, es posible conseguir tasas de compresión[3]-[4] bastante altas. En el presente artículo es explicado el desarrollo completo de esta aplicación para señales de voz además de técnicas de filtrado para devolver la calidad a la señal de salida debido a problemas ocurridos durante el algoritmo de cuantización.

El desarrollo de los contenidos es el siguiente, en la parte II es presentada la teoría general de cuantización vectorial, para la parte III son expuestos los resultados obtenidos utilizando VQ¹ y finalmente en la parte IV son presentados los resultados del filtrado adaptativo utilizado en la señal de salida.

II. CUANTIZACIÓN VECTORIAL

A diferencia de la cuantización escalar que opera sobre datos aislados la cuantización vectorial[5]-[6] opera sobre bloques de datos, estos bloques son conocidos como vectores.

Es posible establecer correlaciones temporales entre los bloques o vectores y así disminuir el error de cuantización. Optimizando el número de centroides utilizados, tanto el codificador como el decodificador contienen un conjunto de vectores M-dimensionales llamado diccionario o libro de códigos. Los vectores que componen este diccionario son llamados vectores código[7], de tal forma que éstos son seleccionados para que sean representativos de las secuencias de muestras que emite la fuente. Por último, el diccionario asigna a cada uno de estos vectores código un índice, cuando la señal de entrada llega al codificador éste busca dentro del diccionario el vector código más parecido al

vector de entrada y transmite para el decodificador el índice asignado previamente a este vector código, el decodificador genera este mismo vector código basándose en el índice recibido y entrega como salida la secuencia de componentes.

$$F = \text{Log}_2(K) \quad (1)$$

Para un diccionario de K vectores código es necesario según (1), un determinado número F de bits para representar el número total de índices, por lo tanto, es posible determinar la longitud media conociendo tanto F como la longitud de cada uno de los vectores código, tal como está en (2).

$$r = F/L \quad (2)$$

A. Condición de Proximidad

En un diccionario C para asignar un vector código a una entrada es utilizado el criterio de proximidad[8]-[9], en el que dado un vector X de entrada el vector código[10] más cercano K_j es el que verifica (3).

$$\|X - K_j\|^2 \leq \|X - K_i\|^2, \forall K_i \in C \quad (3)$$

B. Evaluación del Desempeño

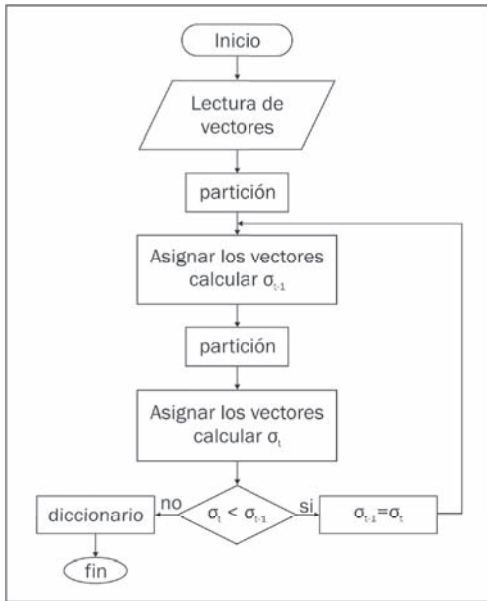
Para determinar el desempeño[11] obtenido durante la codificación basta con encontrar la distorsión promedio entre la entrada y la reproducción de salida mediante (4).

$$\bar{d} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n d(X_i, \hat{X}_i) \quad (4)$$

En la Fig. 1, es presentado el diagrama de flujo utilizado para conseguir la codificación y compresión de la información, basándose en un proceso iterativo que busca obtener la distorsión más baja posible.

¹ Vector Quantization.

FIG. 1. DIAGRAMA DE FLUJO UTILIZADO PARA OBTENER LA MÍNIMA DISTORSIÓN



Fuente: Autores del proyecto

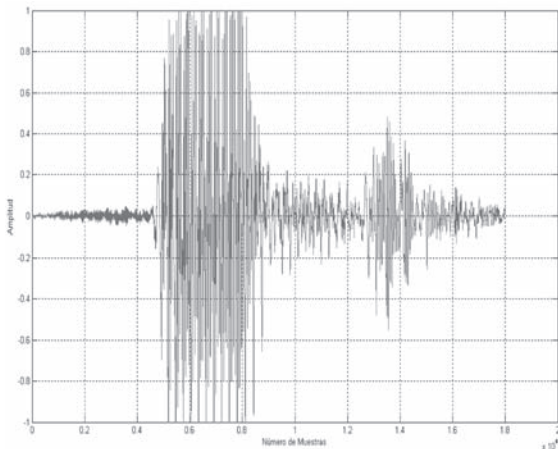
III. RESULTADOS OBTENIDOS UTILIZANDO VQ

Para la realización del proceso descrito anteriormente fue tomada una señal de voz y digitalizada con el fin de aplicar la técnica de cuantización vectorial y conseguir resultados concluyentes tanto para codificación como para compresión.

A. Adquisición de la Señal

La señal adquirida representa una palabra compuesta por dos sílabas, el proceso fue realizado utilizando una $f_s=44.1\text{KHz}$ y $k=16$, la señal obtenida se presenta en la Fig. 2.

FIG. 2. SEÑAL OBTENIDA

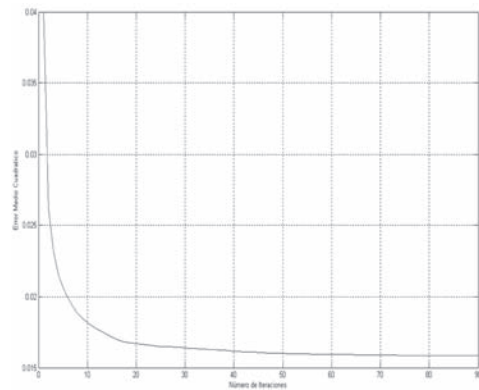


Fuente: Autores del proyecto

B. Proceso de Codificación y Compresión

Para realizar la cuantización vectorial es utilizado un modelo con $k=8$ bits, es decir, 256 niveles de cuantización, y un conjunto de entrenamiento no determinístico de (2,18001), cabe resaltar que anterior a este proceso la señal adquirida fue ajustada para tener un número total de 18001 muestras, el criterio de parada obedece a la inexistencia de variaciones de $1e-7$ en la distorsión después de cada iteración, en la Fig. 3. se presenta la curva de desempeño obtenida.

FIG. 3. CURVA DE DISTORSIÓN OBTENIDA DURANTE EL DISEÑO

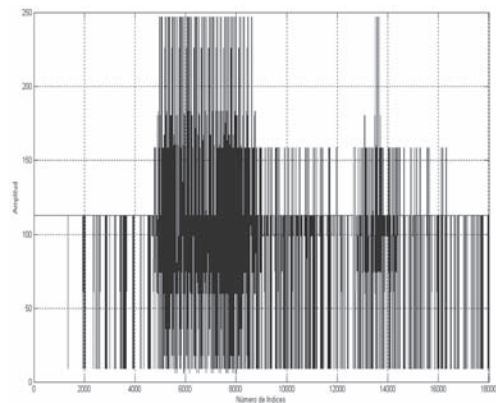


Fuente: Autores del proyecto

En la Figura 4 se presenta el conjunto total de índices transmitidos por el codificador, cada uno de ellos corresponde a cada vector código asignado por el diccionario.

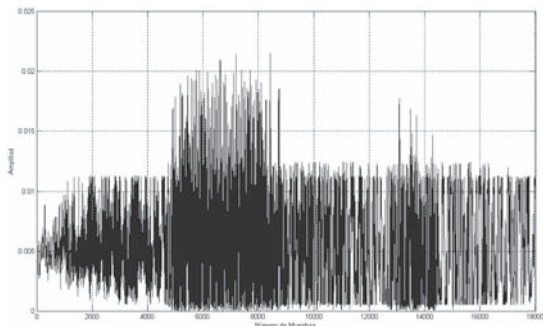
En la Figura 5 son exhibidos los rangos obtenidos de distorsión para cada uno de los índices transmitidos. Es apreciable que en las partes donde la amplitud de la señal original es mayor los rangos aumentaron hasta un 2.2% y los valores mínimos llegaron hasta 0.05%.

FIG. 4. ÍNDICES TRANSMITIDOS



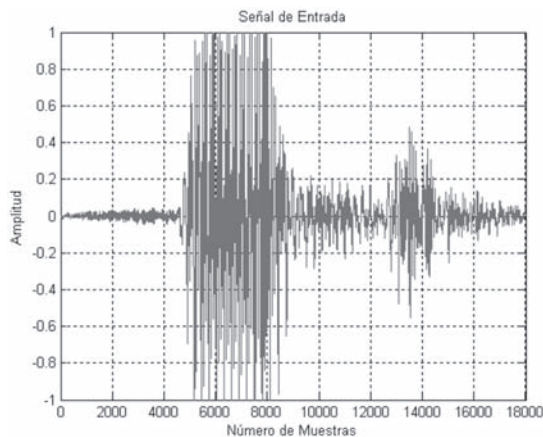
Fuente: Autores del proyecto

FIG. 5. RANGOS DE ERROR PARA LOS ÍNDICES TRANSMITIDOS



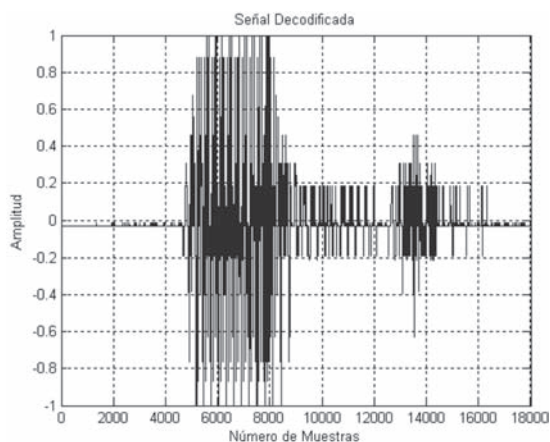
Fuente: Autores del proyecto

FIG. 6. SEÑAL DE ENTRADA.



Fuente: Autores del proyecto

FIG. 7. SEÑAL DE SALIDA



Fuente: Autores del proyecto

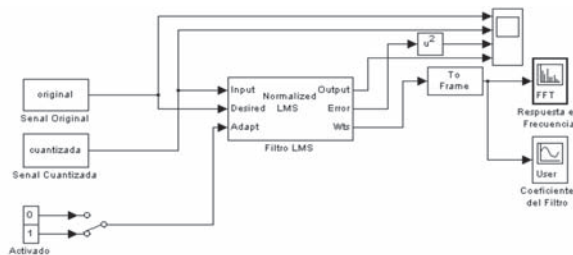
En las Figuras 6 y 7 son presentadas las señales de entrada y salida del modelo de cuantización vectorial, es notable que la similaridad garantiza la mayor parte de la información contenida; al realizar la reproducción audible de la señal de salida los resultados son buenos, habida cuenta

que es comprensible la información; sin embargo, algunas características ajenas a la señal original también son percibidas. Para esto en el próximo capítulo serán utilizadas técnicas de filtrado para lograr una mayor fidelidad de la señal. Como había sido anunciado, además de ser codificada la información también fue comprimida, el tamaño original era de 95.41Kbytes, después del proceso el tamaño de la señal de salida fue de 10.02Kbytes, consiguiendo utilizar solamente un 1/10 del tamaño original lo cual corresponde a un 10.50%.

IV. RESULTADOS OBTENIDOS UTILIZANDO FILTRADO ADAPTATIVO

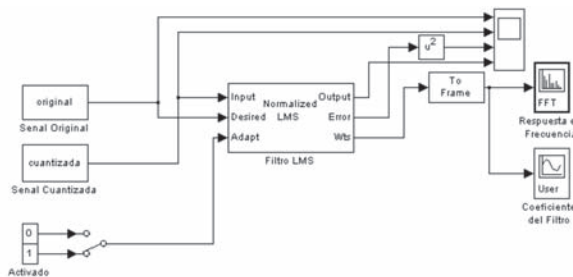
Para aumentar el grado de inteligibilidad de la señal decodificada, se hace necesario implementar algoritmos de tratamiento digital de señales para eliminar la información no deseada y recuperar características perdidas[12]. La técnica escogida es una de las más utilizadas y también la que tiene mejores resultados. Debido a esto los algoritmos para implementar este tipo de técnicas son de una complejidad mayor, sin embargo, fue necesario utilizar una segunda etapa en el filtrado para refinar la respuesta. Para esto fue utilizado un banco de filtros de 4 capas, las dos etapas del filtro son presentadas en las Fig. 8 y Fig. 9.

FIG. 8. DIAGRAMA DE BLOQUES DE LA PRIMERA ETAPA DEL FILTRO



Fuente: Autores del proyecto

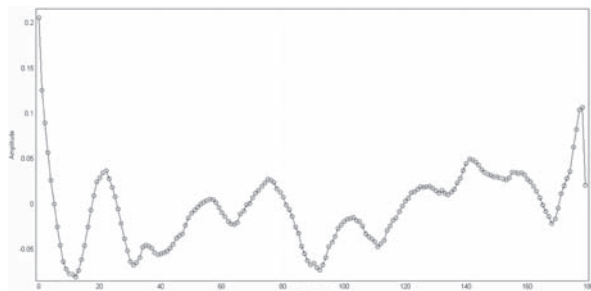
FIG. 9. DIAGRAMA DE BLOQUES DE LA SEGUNDA ETAPA DEL FILTRO.



Fuente: Autores del proyecto

Los resultados obtenidos permiten aumentar la recuperación de la inteligibilidad de la información, la Fig. 10 presenta el comportamiento de los coeficientes utilizados por el LMS[13].

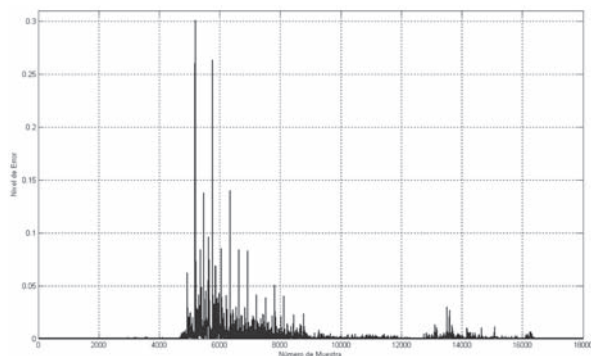
FIG. 10. COMPORTAMIENTO DE LOS ÍNDICES DEL LMS



Fuente: Autores del proyecto

En la Figura. 11 es presentado el nivel de error cuadrático[14] estimado entre la señal filtrada y la señal original. Nótese cómo el valor se hace más alto en las secciones de señal donde están presentes las sílabas de la palabra utilizada, esto es normal puesto que es en la parte donde más trabajo debió realizar el cuantizador y por supuesto donde más error existe.

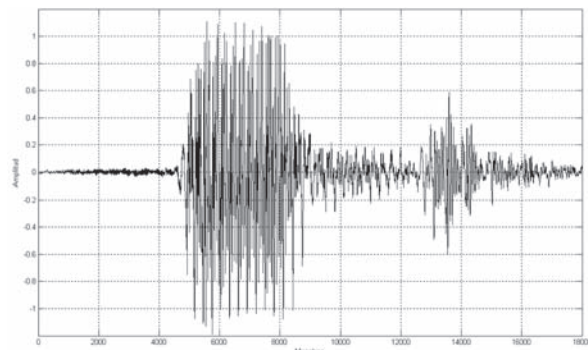
FIG. 11. NIVELES DE ERROR DEL LMS



Fuente: Autores del proyecto

En la Figura. 12 es expuesta la salida de la primera etapa de filtrado, esta entrega una respuesta más acorde con la señal original. Es posible ver cómo la señal entregada recupera aquellas terminaciones que caracterizan las propiedades de la voz, no como sucedía después de la cuantización donde la señal entregada era cortada abruptamente.

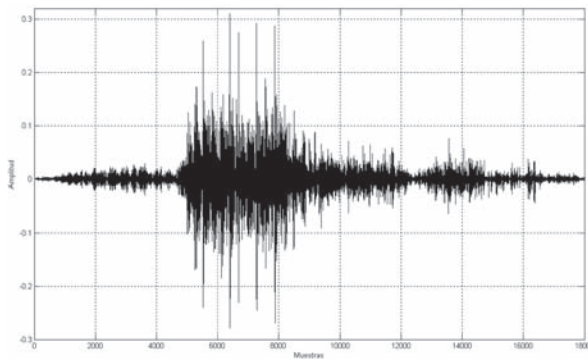
FIG. 12. SEÑAL DE SALIDA DEL LMS



Fuente: Autores del proyecto

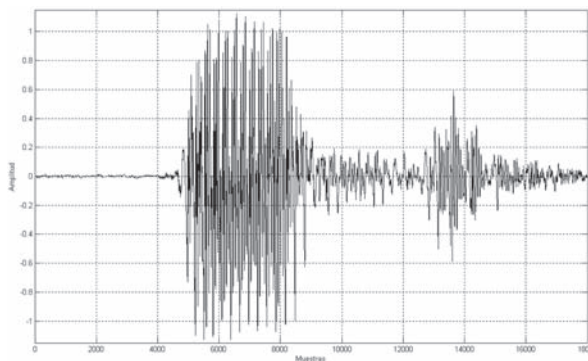
En la utilización del banco de filtros[15], fueron escogidas cuatro capas debido a que la respuesta con 3 era insuficiente y la respuesta con 5 era prácticamente la misma que con 4 capas, la única diferencia era en la complejidad del modelo. En la Fig. 13 es presentado el ruido identificado por el banco de filtros, los valores están dentro del mismo orden del nivel de error identificado por el LMS.

FIG. 13. RUIDO IDENTIFICADO Y RETIRADO POR EL BANCO DE FILTROS



Fuente: Autores del proyecto

FIG. 14. SEÑAL DE SALIDA FINAL



Fuente: Autores del proyecto

CONCLUSIONES

Los resultados obtenidos demuestran claramente que utilizando la técnica de cuantización vectorial es posible realizar codificación de una forma segura y además alcanzar altas tasas de compresión para datos de voz. Durante varias prácticas realizadas fue posible constatar que el aumento de niveles o número de bits genera un costo computacional innecesario, puesto que los valores de distorsión alcanzados no justifican el aumento. Para este tema es importante encontrar técnicas alternativas que optimicen el algoritmo tradicional manteniendo o aumentando las tasas de compresión utilizando cada vez más un número de bits menor y que consiga establecer valores admisibles para las iteraciones.

La utilización de técnicas de filtrado para obtener la recuperación de la inteligibilidad de la información decodificada es de vital importancia, por ello, es sugerida la utilización de estas técnicas siempre que sean usados métodos de cuantización para compresión y codificación.

Para trabajos futuros será importante implementar estos sistemas con sistemas bajo características de exigencia mayor, como por ejemplo, transmisiones continuas de datos y poder conseguir resultados on-line para aplicaciones de mayor exigencia.

REFERENCIAS

- [1] C. H. Valencia, M. S. Dutra, "Caracterización de la Relación Dinámica entre Intervalos de Silencio y Tamaño de Bits para el Canal de Sincronización de una Red Satelital utilizando Lógica Fuzzy", IEEE-Andescon 2008.
- [2] Y. Akaiwa, T. Nomura, and S. Minami, "An integrated voice and data radio access system" Vehicular Technology Conference, 1992 IEEE 42nd, Vol. 1, pp. 255-258.
- [3] G. ChuanYiong and Z. ShaoRen, "Dynamic compression and storage of voice information in the distributed voice information database" Communication Technology ICCT 1996, Vol. 1, pp. 127-200.
- [4] A. Benyassine, E. Sholomot, H.-Y. Su, D. Massaloux, C. Lamblin and J.-P. Petit, "ITU-T Recommendation G.729 AnnexB: a silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications" Communication Magazine IEEE 1997, Vol. 35, pp. 64-73.
- [5] M. Abe, S. Nakamura, K. Shikano and H. Kuwubara, "Voice conversion through vector quantization" International Acoustics, Speech and Signal Processing, 1998, Vol. 1, pp. 655-658.
- [6] A. Gersho and R. M. Gray, "Vector Quantization and Signal Compression", Ed. Springer, 1992, 760p, ISBN: 0792391810.
- [7] T. Lookabaugh, E. A. Riskin, P. A. Chou and R. M. Gray, "Variable rate vector quantization for speech, image and video compression" IEEE Transactions on Communications, 1993, Vol. 41, pp. 186-199.
- [8] H. R. S. Mohammadi and W. H. Holmes, "Low cost vector quantization methods for spectral coding in low rate speech coders" International Acoustics, Speech and Signal Processing, 1995, Vol. 1, pp. 720-723.
- [9] C. J. Kuo and L. Chang-Shyan, "Viterbi-based algorithm for side-match vector quantization over noisy channels" IEEE Transactions on Communications, 1996, Vol. 44, pp. 1455-1465.
- [10] C. Moreno and F. Labeau, "Variable Frame Size for Vector Quantization" IEEE International Conference on Acoustics, Speech and Signal Processing, 2006, Vol. 3, pp. III-III.
- [11] A. Mouchtaris, Y. Agiomyrqiannakis and Y. Stylianou, "Conditional Vector Quantization for Voice Conversion" IEEE International Conference on Acoustic, Speech and Signal Processing, 2007, Vol. 4, pp. IV-505-IV-508.
- [12] W. A. Sethares, J. A. Bucklew, "Local Stability of the Median LMS Filter", IEEE Transactions on Signal Processing, 1994, Vol. 42, No 11, pp. 2901-2905.
- [13] S. H. Leung, C. C. Chu, "Adaptive LMS Filter with Lattice Prefilter", Electronics Letters, 1997, Vol. 33, No 1, pp. 34-35.
- [14] H. J. Butterweck, "A Wave Theory of Long Adaptive Filters", IEEE Transactions on Circuits and Systems – Fundamental Theory and Applications, Vol. 48, No 6, pp. 739-747.
- [15] G. Saxena, S. Ganesan, M. Das, "Real Time Implementation of Adaptive Noise Cancellation", IEEE International Conference on Electro/Information Technology, pp. 431-436.