

SUPRESIÓN DE RUIDO AUDIBLE UTILIZANDO ARRAYS MICROFÓNICOS

José Luis Sánchez Bote

Depto. de Ingeniería Audiovisual y Com. (DIAC)
Área de Tratamiento de Voz y Señales (ATVS)
Universidad Politécnica de Madrid
jbote@diac.upm.es

Joaquín González Rodríguez

Depto. de Ingeniería Audiovisual y Com. (DIAC)
Área de Tratamiento de Voz y Señales (ATVS)
Universidad Politécnica de Madrid
jgonzalez@diac.upm.es

Danilo Simón Zorita

Depto. de Ingeniería Audiovisual y Com. (DIAC)
Área de Tratamiento de Voz y Señales (ATVS)
Universidad Politécnica de Madrid
dsimon@diac.upm.es

Javier Ortega García

Depto. de Ingeniería Audiovisual y Com. (DIAC)
Área de Tratamiento de Voz y Señales (ATVS)
Universidad Politécnica de Madrid
jortega@diac.upm.es

ABSTRACT

The goal of the work presented in this paper is the implementation of a new method for speech enhancement using microphone arrays. This method gets noise reduction of speech signal using the masking properties of the human auditory system. Recent works with single channel speech processors, based in this method, have tested a good performance in noise reduction. This is achieved by filtering only the noise components that are over the masking threshold of the human auditory system. In this paper a multimicrophone recording system has been used to take advantage of its better performance in noise-free speech estimation, to improve the auditory noise suppressor of a single channel processor.

1. INTRODUCCIÓN

Las principales fuentes de degradación que se pueden encontrar en una señal de habla son el ruido y la reverberación. Para mejorar señal de habla con ruido y reverberación se deben atacar ambos problemas de manera diferente. Los métodos más recientes para mejorar voz contaminada con ruido se basan en la presunción de que la señal y el ruido se encuentran totalmente incorreladas. En esas condiciones, se pueden eliminar las componentes en frecuencia del ruido mediante el filtrado de su espectro a corto plazo. Para realizar esta tarea, en los últimos años se han venido usando dos métodos, la sustracción espectral [1] y el filtrado de Wiener [2]-[3]. No obstante, los trabajos más recientes apuntan a que se pueden introducir muchas mejoras en la limpieza de ruido si se consideran las características del sistema auditivo humano. Esta técnica, conocida como supresión de ruido audible (Audible Noise Suppression ó ANS) se ha aplicado con éxito en trabajos recientes [4]-[5] para señales de voz monocanales. El método ANS para la supresión de ruido se basa en el procesamiento del ruido por bandas críticas, de tal manera que se produzca una reducción del mismo tan sólo hasta el umbral en el que dicho ruido se manifiesta como audible. Dicho umbral es el llamado umbral de enmascaramiento y depende de la

frecuencia y de la cantidad de señal de voz libre de ruido que haya en una determinada banda crítica. Así el ruido residual inherente a todo procesamiento de limpieza de ruido puede ser “perceptualmente blanco”, es decir percibido como espectralmente blanco desde el punto de vista subjetivo. La novedad que introducimos en este trabajo es que se utiliza una señal de voz multicanal, procedente de un array microfónico para conseguir una mejor estimación de la señal de habla libre de ruido.

2. MEJORA DE SEÑAL MULTICANAL USANDO LAS PROPIEDADES DEL SISTEMA AUDITIVO HUMANO

La mayoría de los sistemas utilizados hasta ahora para la reducción de ruido de una señal de habla se basan en un filtrado de la voz ruidosa teniendo en cuenta la relación señal a ruido estimada ó SNR. En este trabajo se han usado los umbrales de enmascaramiento [6] del sistema auditivo humano para controlar la cantidad de reducción de ruido que se necesita en cada trama de señal de habla. El umbral de enmascaramiento es un nivel definido para cada una de las bandas críticas auditivas [6] e indica una cota superior por debajo de la cual el ruido queda enmascarado por la señal. De este modo, si el umbral de enmascaramiento es alto, no será necesaria mucha supresión de ruido y viceversa. Esto tiene el efecto de que la respuesta subjetiva del procesador es mucho mejor. Así por ejemplo, si se escucha una señal de voz procesada mediante sustracción espectral no lineal basada en SNR y procesada mediante el método auditivo o ANS, la última sonará mejor, aunque ambas tengan la misma evaluación objetiva de mejora en la SNR resultante. Tan sólo es necesario encontrar un filtro que realice una supresión de ruido teniendo en cuenta el nivel relativo del ruido en cada banda crítica. En este trabajo se ha utilizado el método descrito en [5].

2.1. Estimación de voz sin ruido usando arrays microfónicos

Los umbrales de enmascaramiento han de ser calculados utilizando la señal de voz libre de ruido y por lo tanto es necesario estimar esta última. En este trabajo se a utilizado un filtrado de Wiener con modificación de coherencia (Modified

Wiener method ó MW) como se describe en [2]. Este método se ha aplicado a un array de 15 micrófonos. La ganancia del filtro de Wiener es como sigue:

$$H_{MW}(\omega) = \begin{cases} \frac{\langle G_{xi\ yj}(\omega) \rangle - \langle G_{ni\ nj}(\omega) \rangle}{G_{xx}(\omega)} & \text{si } C(\omega) > CT \\ C(\omega)^\alpha & \text{si } C(\omega) < CT \end{cases} \quad (1)$$

con

$$C(\omega) = \frac{G_{x0}(\omega)}{\sqrt{G_{xx}(\omega)G_{00}(\omega)}} \quad (2)$$

donde $C(\omega)$ es la coherencia entre canales, $\langle G_{xi\ yj}(\omega) \rangle$ y $\langle G_{ni\ nj}(\omega) \rangle$ los espectros cruzados promedio sobre todos los canales, el segundo cuando se consideran tramas de ruido, es decir sin actividad de voz, $G_{xx}(\omega)$ es la estimación del autoespectro de la señal de voz ruidosa, obtenida mediante la conformación de haz de los canales microfónicos en tres subbandas de frecuencia, y α y CT (Coherence Threshold ó umbral de coherencia) son parámetros fijos. En la expresión (2) $G_{x0}(\omega)$ y $G_{00}(\omega)$ son respectivamente el espectro cruzado entra la señal conformada y el canal central del array o canal de referencia y el autoespectro de dicho canal de referencia. La expresión (1) se puede usar para obtener un estimador de señal de habla libre de ruido como sigue:

$$\hat{S}_{MW}(\omega) = H_{MW}(\omega) \cdot Y(\omega) \quad (3)$$

donde $\hat{Q}_{MW}(\omega)$ es la estimación de voz limpia procedente del filtrado de Wiener e $Y(\omega)$ es la señal de voz ruidosa. El resultado puede usarse directamente como señal limpia a la salida del sistema o, como ya se ha dicho, a modo de estimador de señal limpia para la implementación del método ANS, según se describe a continuación.

2.2. Mejora de voz utilizando umbrales de enmascaramiento

La señal limpia de salida por el método ANS es:

$$\hat{S}_{ANS}(\omega) = H_{ANS}(\omega) \cdot Y(\omega) \quad (4)$$

con $H_{ANS}(\omega)$ el filtro auditivo de mejora que viene dado [5] por,

$$H_{ANS}(\omega) = \frac{Y^v(\omega)}{a^v(\omega) + Y^v(\omega)} \quad (5)$$

siendo v un parámetro independiente de la frecuencia y $a(\omega)$ otro parámetro relacionado con el umbral de enmascaramiento. El umbral de enmascaramiento $T(\omega)$ se obtiene a partir de la señal limpia $\hat{Q}_{MW}(\omega)$ mediante el método descrito en [6]. El parámetro $a(\omega)$ se obtiene por:

$$a(\omega) = [N(\omega) + T(\omega)] \cdot \left(\frac{N(\omega)}{T(\omega)} \right)^{1/v} \quad (6)$$

con $N(\omega)$ el autoespectro de ruido que se puede estimar en tramas sin actividad de voz.

3. DESCRIPCIÓN DEL SISTEMA Y RESULTADOS

En la figura 1 se muestra el esquema del sistema de procesado aplicado a un array anidado de 15 micrófonos [2].

Para probar el sistema se ha utilizado la base de datos multimirófono procedente de la Universidad Carnegie Mellon [2] (subcorpus arr... de la tabla 1-a) y una base de datos propia a usando voz contaminada con ruido aleatorio (tabla 1-b). En esta

tabla GSNRA es la mejora en relación señal a ruido con ponderación A, GAI es la mejora en índice de articulación y GNMR es la mejora en relación ruido a enmascaradora.

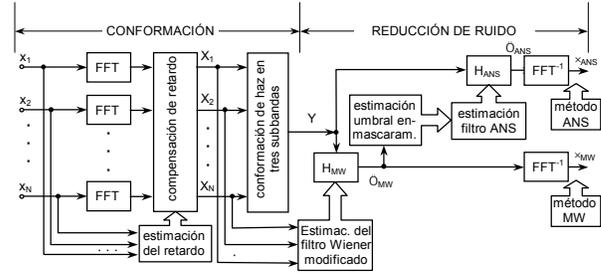


Figura 1. Esquema del procesador.

subcorpus	GSNRA(dB)		GAI		GNMR(dB)	
	mét. ANS	mét. MW	mét. ANS	mét. MW	mét. ANS	mét. MW
arr4A	2.0	0.9	0.07	0.020	6.7	3.1
arrC1A	1.0	0.4	0.03	-0.002	2.1	1.1
arrC3A	0.6	0.5	0.01	-0.002	1.8	1.3

(a)

SNR entr. (dB)	GSNRA(dB)		GAI		GNMR(dB)	
	mét. ANS	mét. MW	mét. ANS	mét. MW	mét. ANS	mét. MW
0	9.6	6.0	0.20	0.09	18.3	8.1
10	6.5	5.4	0.19	0.11	12.6	7.6

(b)

Tabla 1. Resultados de mejora de ruido. (a) Base de datos real. (b) Base de datos simulada.

4. CONCLUSIONES

Los resultados presentados en este trabajo obtenidos mediante la aplicación de un procesador auditivo multicanal han sido buenos, tanto desde el punto de vista objetivo (tabla 1) como desde el punto de vista subjetivo (estos resultados no se muestran en este trabajo), obteniéndose mejoras con respecto a los sistemas tradicionales basados en sustracción espectral o en procesado auditivo monocanal.

5. REFERENCIAS

- [1] Boll, S.F., "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. on Speech and Audio Processing*, vol. ASSP-27, pp.113-120, 1979.
- [2] Sánchez-Bote, J.L., González-Rodríguez, J., Ortega-García, J., "A new Approach to dereverberation and noise reduction with microphone arrays", *Proc.EUSIPCO*, pp.183-186, 2000.
- [3] González-Rodríguez J., Sánchez-Bote J.L. and Ortega-García, J., "Speech dereverberation and noise reduction with a combined microphone array approach", *Proc. ICASSP*, pp.1037-40, 2000.
- [4] Akbari, A., Le-Bouquin, R., Faucon, G., "Optimizing speech enhancement by exploiting masking properties of the human ear", *Proc. ICASSP*, pp.800-3, 1995.
- [5] Tsoukalas, D.E., Mourjopoulos, J.N., Kokkinakis, G., "Speech enhancement based on audible noise suppression", *IEEE Trans. on Speech and Audio Processing*, vol.5, no.6, pp.497-514, 1997.
- [6] Johnston, J.D., "Transform coding of audio signals using perceptual noise criteria", *IEEE Journal on Selected Areas in Comm.* vol.6, no.2, pp.314-23, 1988.