[[1]](#footnote-1)

Project Noise-Killer

Mikkelsen-Löth, Marcos, *Legajo 56645,* Romarís, Juan Manuel, *Legajo 57108,* De León, Nicolás, *Legajo 57232,* Martorell, Ariel, *Legajo 56209,* Vigón, Tomás, *Legajo 57327*

*Abstract*—Se presenta un algoritmo de supresión de ruido, permitiendo reducir los efectos espectrales del mismo sobre el espectro de voz. El método de resta espectral resulta computacionalmente eficiente y es independiente del tipo de procesador utilizado. Este método funciona requiriendo el mismo tiempo de cálculo que otros métodos más complejos de supresión, restando una estimación del espectro del ruido calculada en tiempos sin voz. Además de esto se le aplican otros procedimientos secundarios que incrementan la atenuación el ruido residual. Este algoritmo puede ser utilizado para reducir el nivel de ruido en pistas de audio grabadas o inclusive en tiempo real, siempre y cuando el ruido sea estacionario.

# Introducción

E

l ruido de fondo que se encuentra generalmente en grabaciones de cualquier tipo ya sea de música en vivo o en trasmisiones en vivo de cualquier tipo de evento puede bajar la calidad de los procesadores de voz digitales que se utilizan para aplicaciones de compresión de audio, reconocimiento y autentificación. Estos efectos pueden ser reducidos mediante el uso de micrófonos canceladores de ruido, modificaciones internas a los algoritmos procesadores de voz para compensar explícitamente la contaminación de la señal o reducción de ruido en el preprocesador.

El objetivo en este proyecto es crear una técnica que pueda suprimir el ruido implementando un algoritmo de forma eficiente. Para realizar dicho algoritmo se procedió a utilizar una estimación de la magnitud del espectro de la pista de audio ya sin ruido mediante la sustracción de la magnitud espectral del ruido a la pista de audio original. Esto se logra promediando obteniendo una estimación del ruido en momentos donde solo este se encuentra presente.

# Análisis del método de sustracción de ruido

## Marco teórico

El estimador de ruido utilizado se obtiene al realizar una estimación del espectro del ruido de la pista que posee dicho ruido. Esta información espectral se obtiene mediante la medición de dicho ruido en intervalos de la pista en donde solo se encuentra este.

Asumiendo que este ruido de fondo es estacionario en momentos donde se presencia audio útil y en los que no, su composición espectral tenderá a mantenerse constante. Justamente por esto es necesario destacar que si el ruido varía a otro tipo de ruido con otro estado estacionario debe existir un mínimo de tiempo (alrededor de 300 ms en el que el nuevo ruido se mantenga estacionario) para poder lograr estimar la nueva magnitud espectral del ruido.

Conocida la composición espectral de este ruido, se asume que es posible reducir este ruido solamente restando su aporte a la composición espectral del audio total teniendo en cuenta el audio útil/deseado.

## Modelo matemático

Para este procedimiento se asume que se posee una señal discreta *x(n)* que tiene como componentes la señal con el audio original, pasado por una ventana, *s(k)* y una señal de ruido pasada por la misma ventana *n(k)* de la siguiente manera:

cuya transformada de Fourier es la siguiente:

El filtro de sustracción espectral se calcula reemplazando el espectro del ruido con espectros ya medidos por lo que se reemplaza por el valor promedio del ruido que se obtuvo en las mediciones que se realizaron en el periodo de actividad donde no hay habla de la pista. En cuanto a la fase de es reemplazada con la fase de . Con estas substituciones obtenemos el estimador de sustracción espectral:

o

Donde:

Debido a que justamente se estima el ruido estacionario presente, existirá un error de estimación dado por:

cuyos efectos auditivos se pueden reducir con las especificaciones se mencionan a continuación.

## Promedio de la magnitud

Como el error espectral es la diferencia entre el ruido N y su media la utilización de un promedio local de la magnitud puede ayudar a una reducción del error. Entonces se reemplaza por donde:

es la iesima transformada de *x(k)* con el cual se obtiene la siguiente expresión:

donde

donde se puede observar que si se toman más valores para realizar el promedio se tiene que converge .

Como el habla es no-estacionaria esto conlleva al problema de que solo hay un tiempo limitado en el que se puede promediar. Aquí uno llega a una relación de compromiso donde promediando mayor cantidad de ventanas puede obtenerse una mejor reducción del ruido de fondo, pero haciendo que posiblemente se reduzca la inteligibilidad del audio (como el habla no es estacionaria, promediar ventanas consecutivas del mismo puede llegar a reducir considerablemente la inteligibilidad del mismo).

## Rectificación de media onda

Para cada frecuencia donde la magnitud contenido espectral de la señal con ruido es menor a la magnitud promedio del ruido estimado su salida es cambiada a 0. Esta modifiación es facil de implementar con una rectificación de media, haciendo que nuestro estimador se convierte en el siguiente:

donde

Con esta rectificación se puede sesgar para abajo la magnitud del espectro de cada frecuencia por un factor que es determinado por el sesgo del ruido a dicha frecuencia. El valor de los sesgos puede variar de frecuencia a frecuencia a su vez como en las distintas ventanas. La ventaja de este método es que el piso de ruido es reducido por un factor de , aunque presenta una desventaja en la situación en la que la suma de estos valores a una frecuencia dada es menor a la media lo que concluye en una incorrecta eliminación de información de la pista lo que podría reducir el grado de inteligibilidad.

## Atenuación adicional en períodos sin voz

El contenido de energía de en relación a permite tener un indicar preciso de si durante una ventana de análisis hay voz presente. Si no hay voz presente consistirá del ruido residual que queda después de hacer la rectificación de media onda. Para decidir si hay voz o no en una ventana analizada se define un umbral llamado T, donde empíricamente se llegó a un valor de -12dB. Este umbral se lo utiliza de la siguiente manera:

Si durante la ventana el valor obtenido es menor que el umbral T, se decide que no hubo voz en la misma. En caso de que suceda esto hay distintos caminos para tratar el audio antes de la resíntesis: no hacer nada, atenuar por un factor fijo, o fijar la salida en cero. Se encontró que tener una señal atenuada en momentos sin habla resultaba en audios de mejor calidad comparado a fijar la salida en cero. Entonces la salida espectral estimada tendrá la forma:

Donde .

# Implementación del algoritmo

Dado el desarrollo realizado en la sección anterior se puede proceder a realizar un algoritmo de análisis y síntesis. A continuación, se procederá a explicar cómo se implementó el algoritmo y que se tuvo en cuenta a la hora de realizar dicho algoritmo, cuyo diagrama en bloques puede ser observado en la siguiente figura.

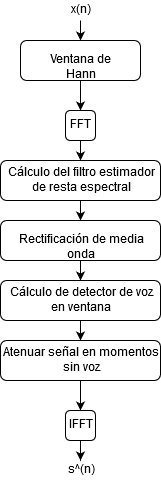


Fig. 1. Diagrama en bloques del algoritmo implementado.

## Obtención de a señal ventaneada

La pista de audio luego de ser muestreada es segmentada y multiplicada por una ventana de Hanning, luego las distintas ventanas son superpuestas con un factor overlap del 50%. Se trabajó con audios muestreados a 16kHz. El largo de las ventanas se eligió para que sean aproximadamente el doble de largo que la duración del máximo tono esperado para poder así obtener una mejor resolución en términos de frecuencia.

## Análisis en frecuencia

A cada ventana se le aplica una FFT con un tamaño igual al de la ventana utilizada.

## Promedio de la magnitud

Como se describió anteriormente la varianza del estimador del espectro del ruido se puede reducir promediando con la mayor cantidad de ventanas consecutivas posibles.

Sin embargo, como la voz es no-estacionaria hay un límite máximo para el número de ventanas que se pueden tomar para hacer un promedio local. Esta cantidad de ventanas depende cuantas pueden entrar en el intervalo de tiempo en el que consideramos a la voz como estacionaria.

Se presenta una relación de compromiso donde incrementar la cantidad de ventanas promediadas mejora la estimación espectral del ruido, pero reduce la inteligibilidad de la voz.

Según evidencia experimental de trabajos previos se puede notar que una buena relación entre la reducción de la varianza y la resolución en el tiempo si se toman 3 promedios (con ventanas de 256 puntos y frecuencia de sampleo 8kHz).

## Estimación del sesgo

Para poder obtener el estimador se debe realizar la resta entre la magnitud espectral de la media del ruido y con la magnitud del espectro total . Por lo tanto:

o

donde L es el ancho de la ventana utilizado en la FFT.

Luego de realizar dicha resta todos los valores que resultan ser negativos son puestos en 0, estos resultados se deben a que en esas frecuencias la suma entre la voz y entre el error local es menor a la del ruido esperado.

## Atenuación adicional en períodos sin voz

La última mejora relacionada a la reducción de ruido es la supresión de la señal en períodos sin voz. Como fue mencionado antes se tiene que mantener un balance entre la magnitud y características del ruido mientras hay voz y mientras esta no está presente.

Un detector efectivo fue el propuesto previamente, planteando una relación de energía umbral entre la señal medida y la estimación del ruido. Este umbral tomó el valor de -12dB, obtenido empíricamente para que sólo los momentos sin voz fuesen atenuados.

## Síntesis

Luego de remover el sesgo y de aplicar la rectificación de media onda se reconstruye la señal a partir de la magnitud modificada correspondiente a la ventana central y utilizando la IFFT se obtiene la nueva pista de audio.

# Resultados

## Consideraciones

Para verificar la efectividad de este algoritmo se tuvieron en cuenta distintos factores. Primero se consideró un intervalo fijo de tiempo en el que solo se encontraba el ruido, dicho intervalo debía ser de por lo menos 300 ms. El tipo de ruido que se utilizó en las pruebas fue el ruido caracterizado comúnmente como ruido rosa.

## Pruebas con pistas de audios

Para probar la efectividad del algoritmo se utilizó una pista de audio muestreada a 16kHz. El ruido rosa se encuentra a un nivel constante en toda la pista y a su vez el contenido importante de dicha pista comienza alrededor del primer segundo del audio.

Luego de correr dicha pista por el algoritmo realizado se pudo notar una atenuación considerable del ruido. Analizando ambos espectrogramas, es decir uno antes y después de implementar el algoritmo y se los espectrogramas que se muestran en el apéndice.

Como se puede observar se nota una disminución general del ruido, pero a su vez se nota una pequeña distorsión en el área que corresponde al espectro del audio escuchado.

# Conclusión

Se puede concluir que el algoritmo implementado cumple con su función y logra disminuir el ruido, siempre y cuando este sea estacionario y no cambie su composición espectral por la duración del audio.

Apéndice

A continuación, se mostrarán los distintos espectrogramas de la señal, primero sin aplicar el algoritmo y después de aplicado el algoritmo:

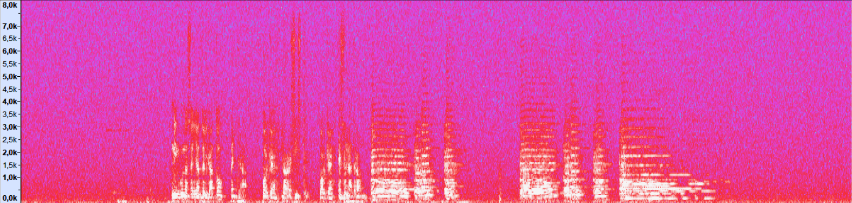


Fig. 2 Espectrograma del audio sin aplicar el algoritmo.

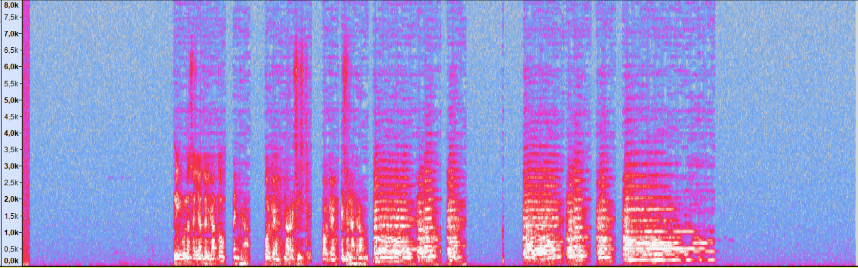


Fig. 3 Espectrograma del audio una vez aplicado el algoritmo.

Referencias

1. S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction” in *IEEE transactions on acoustics, speech, signal processing*, vol. ASSP-27, NO.2, April 1979, pp. 113-120.
2. J. Makhoul and J.Wolf, “Linear prediction and the spectral analysis of speech,” Bolt, Beranek, and Newman Inc., BBN rep. 2304, NTIS No. AD-749066, pp. 172-185
3. S.F. Boll and D. Pulsipher, “Noise suppression methods for robust speech processing,” Dep. Comput. Sci., Univ. Utah, Salt Lake City, Semi-Annu. Tech. Rep., Utec-Csc-77-202, pp. 50-54, Oct. 1977.

1. [↑](#footnote-ref-1)