

Evaluación de técnicas de reducción de ruido en habla

Torres, M.E.; Rufiner, H.L.*; Aronson, L.**; Martínez, C.E.***;
Milone, D.H.*; Tomassi, D.R.*****

Resumen

El presente trabajo evalúa la inteligibilidad y la calidad de señales de voz luego de ser procesadas por un conjunto de técnicas de reducción de ruido. La inteligibilidad se mide en porcentaje de palabras repetidas correctamente en una prueba subjetiva de reconocimiento, y se discuten las sustituciones fonéticas más frecuentes en términos de matrices de confusión. La calidad de las señales obtenidas se evalúa en forma subjetiva y también objetiva a partir de un conjunto de medidas seleccionadas al efecto. En este trabajo se presenta la evaluación de algunas de las técnicas clásicas más utilizadas, tales como Sustracción Espectral, Filtrado de Wiener y Ephraim-Malah. Además se muestran los resultados de la evaluación preliminar de técnicas más recientes, como las basadas en transformada ondita. Se presenta y discute el desempeño relativo de cada algoritmo considerado.

Palabras clave: bioingeniería, prótesis auditivas, algoritmos de reducción de ruido, transformada ondita, inteligibilidad

Artículo derivado de un proyecto conjunto de la Agencia Nacional de Promoción Científica y Tecnológica -ANPCyT y la Universidad Nacional de Entre Ríos -UNER; Directores: Dra. M.E. Torres y Dr. H.L. Rufiner. Facultad de Ingeniería -FI-, UNER (Oro Verde, Argentina); recibido en noviembre 2009; admitido en diciembre 2009.

Autores: *Consejo Nacional de Investigaciones Científicas y Técnicas -CONICET-; FI, UNER (Oro Verde, Argentina) y Facultad de Ingeniería en Ciencias Hídricas -FICH-, Universidad Nacional del Litoral -UNL (Santa Fe, Argentina). **FI, UNER. ***FI, UNER y FICH, UNL. ****CONICET y FICH, UNL. E-mail: metorres@santafe-conicet.gov.ar

Evaluation of noise reduction techniques in speech signals

Torres, M.E.; Rufiner, H.L.*; Aronson, L.**; Martínez, C.E.***;
Milone, D.H.*; Tomassi, D.R.*****

Abstract

The present work evaluates the intelligibility and the quality of speech signals after being processed using a group of noise reduction techniques. The intelligibility is measured in percentage of correctly recognized words in a subjective test, and the most frequent phonetic substitutions are discussed in terms of confusion matrices. The quality of the obtained signals is evaluated also in an objective way starting from a group of measures selected ad-hoc. In this work the evaluation is presented for some of the most commonly employed classic techniques, such as Spectral Subtraction, Wiener and Ephraim-Malah filtering. The results of the preliminary evaluation of more recent techniques are also shown, as the ones based in the Wavelet transform. The relative performance of each considered algorithm is presented and discussed.

Key words: bioengineering, auditory prostheses, noise reduction algorithms, wavelet transform, intelligibility

Paper derived from a research Project funded by Agencia Nacional de Promoción Científica y Tecnológica -ANPCyT- and Universidad Nacional de Entre Ríos -UNER-; Directed by M.E. Torres (PhD) and H.L. Rufiner (PhD); Facultad de Ingeniería (FI), UNER (Oro Verde, Argentina); submitted in November 2009; admitted in December 2009.

Authors: *Consejo Nacional de Investigaciones Científicas y Técnicas -CONICET-; FI, UNER (Oro Verde, Argentina) and Facultad de Ingeniería en Ciencias Hídricas -FICH-, Universidad Nacional del Litoral -UNL- (Santa Fe, Argentina). **FI, UNER. ***FI, UNER and FICH, UNL. ****CONICET and FICH, UNL. E-mail: metorres@santafe-conicet.gov.ar

I. Introducción

El desarrollo de algoritmos de reducción de ruido para señales de voz se ha visto notablemente intensificado en los últimos años. La evaluación de esos métodos, sin embargo, pocas veces obedece a un criterio uniforme entre los distintos autores. Esta falta de uniformidad se manifiesta no sólo en la diversidad de medidas y métodos de evaluación utilizados, sino también en las distintas técnicas de referencia sobre las cuales se comparan los nuevos resultados. Por otra parte, pocas veces se tienen en cuenta los posibles efectos de las características particulares del idioma y dialectos en la validación del desempeño tanto de estrategias de procesamiento como de los métodos para su evaluación. Cuando los nuevos métodos de procesamiento introducidos están destinados a asistir a pacientes con alguna discapacidad auditiva, la situación es aun más desfavorable, ya que muchas veces las medidas disponibles sólo han sido validadas con sujetos normo-oyentes. Advirtiendo esta situación, el presente trabajo representa un primer paso hacia la construcción de una base de referencia de algoritmos de reducción de ruido en habla, basada en una evaluación comparativa de su desempeño en el marco del idioma español. Para ello, se diseñó una batería de señales de voz contaminadas con distintos tipos e intensidades de ruido (a partir de diferentes relaciones señal-ruido o SNR), la que fue procesada con un conjunto reducido de algoritmos de supresión de ruido clásicos y basados en transformada ondita. Las señales obtenidas con cada técnica fueron luego evaluadas en términos de calidad e inteligibilidad del habla. En la bibliografía aparecen trabajos recientes similares con medidas objetivas y subjetivas pero en idioma inglés [22,23,28]. La influencia del idioma en los resultados subjetivos de este tipo de técnicas no se ha cuantificado aún. Sin embargo, es de esperar que existan diferencias significativas debido a las variaciones de importancia relativa de los fonemas en cada idioma y a la forma disímil en que el ruido afecta su inteligibilidad. Una versión resumida de este material, que sólo contempla los resultados obtenidos con las técnicas clásicas, fue publicada en [36]. En este nuevo trabajo se presentan también, en forma resumida, los resultados de una compara-

ción preliminar entre varias técnicas que utilizan la transformada ondita [24]. No se ha reportado hasta el momento una comparación completa entre técnicas clásicas y basadas en onditas.

II. Técnicas de reducción de ruido

Para este trabajo se escogió un conjunto de estrategias de reducción de ruido que son frecuentemente utilizadas para comparar el desempeño de nuevas estrategias. A continuación se describen brevemente cada una de ellas.

A. Sustracción Espectral: Esta técnica intenta estimar la señal limpia a partir del espectro de la señal ruidosa, sustrayéndole una versión estimada del ruido que es generalmente actualizada durante los intervalos de ausencia de voz [6]. Existen varias modificaciones a la formulación básica [5,11,38]. La alternativa seguida aquí es la correspondiente a [5].

B. Filtro de Wiener: Este método intenta minimizar el error cuadrático medio entre la señal limpia y la estimada. Entre las aproximaciones más comunes se encuentran las motivadas por la sustracción espectral [38], los enfoques iterativos [11] y las que emplean estimadores recursivos de la relación señal - ruido. En este trabajo se empleó un estimador para la SNR del tipo desarrollado en [15].

C. Estimadores de Ephraim y Malah: Estos estimadores minimizan el error cuadrático medio para la amplitud o para el logaritmo de la amplitud del espectro de la señal limpia, suponiendo que tanto el ruido como la señal de interés son procesos aleatorios independientes que siguen una distribución gaussiana [15,16]. Además de las particularidades de la expresión del estimador, un aspecto importante de estas técnicas es la forma en que estiman la SNR en cada segmento de la señal [7].

D. Técnicas basadas en onditas: Existen dos paradigmas básicos de filtrado basados en transformada ondita: uno se basa en el análisis de singularidades de la señal [29]; el otro, aplica un umbral en el dominio ondita y elimina o reduce el valor de ciertos coeficientes [13,14].

En la primera técnica, los autores desarrollaron un algoritmo que remueve ruido blanco, discriminando éste de las singularidades de la señal a

través del análisis del “módulo máximo” de los coeficientes ondita [29]. Sin embargo, esta técnica no cobró mucha difusión a lo largo de los años, por lo cual no será considerada para nuestro estudio. Por su parte, la técnica de umbralamiento desarrollada por Donoho *et al.* [13,14] es la base de la mayoría de los algoritmos actuales y será explicada a continuación en su formulación original y sus variantes más importantes.

La estrategia de umbralamiento de los coeficientes ondita se basa en la observación de que, en muchas señales, la mayor parte de la energía se concentra en pocas bandas de la descomposición ondita. Los valores de los coeficientes en dichos niveles son relativamente grandes comparados con los del ruido, cuya energía suele distribuirse a su vez en un gran número de coeficientes ondita [33].

A pesar de que existen diversos algoritmos que proponen modificaciones a la técnica original de Donoho, las tres etapas siguientes se hallan siempre presentes [13]:

1. Aplicar la transformada ondita a la señal, obteniendo los coeficientes de la descomposición.
2. Aplicar algún tipo de umbralamiento (no-lineal).
3. Aplicar la transformada ondita inversa para recuperar señal limpiada.

Una de las principales diferencias radica en el tipo de umbralamiento utilizado y la forma de estimar el umbral óptimo. Sin embargo, hay otros aspectos que también permiten diferenciar los distintos algoritmos aplicados. Los dos tipos de umbralamientos más comunes son el duro y el suave. El umbralamiento duro elimina los coeficientes que se encuentran por debajo de cierto umbral elegido. El umbralamiento suave, además de lo anterior, modifica también los demás coeficientes. Existen además variantes que evitan las discontinuidades.

Para la estimación del umbral, en la mayoría de los algoritmos es indispensable estimar el nivel de ruido de la señal en estudio. Esto introduce diferentes alternativas.

Otra variante entre los algoritmos es el tipo de transformada utilizada, siendo la más común la transformada ondita diádica, pero se emplea también la transformada de paquetes de onditas. Finalmente, se suelen realizar procesamientos adicionales a la señal para mejorar los resultados obtenidos.

Partiendo del anterior resumen de características, se seleccionaron un total de 13 algoritmos diferentes para su evaluación: Gao [18], Lang [26], Seok [32], Soon [34], Bahoura [4], Chang [8], Cohen [9], Kim [25], Sheikhzadeh [33], Fu [17], Saeed [31], Lei [27], Akhaee [1], Arefeen [2]. Todos estos algoritmos se implementaron o se consiguió una versión implementada por parte de los autores. A partir de los resultados obtenidos en pruebas objetivas de calidad e inteligibilidad preliminares (con un conjunto reducido de datos) se seleccionaron sólo 6 de ellos para intervenir en las pruebas exhaustivas. Los algoritmos que mostraron mejor comportamiento en estas pruebas y resultaron seleccionados fueron: Akhaee, Arefeen, Bahoura, Chang, Lang y Lei.

III. Material de prueba y Métodos

Para las pruebas se empleó la *Batería de Evaluación de Pacientes con Prótesis Auditiva* (BEPPA), desarrollada especialmente en forma conjunta por la Fundación Arauz y la Facultad de Ingeniería de la Universidad Nacional de Entre Ríos [3]. El corpus consta de un listado de consonantes en contexto vocálico, conjuntos de monosílabos y transiciones formánticas y conjuntos de frases de uso cotidiano. Todas las señales fueron grabadas en una cámara anecoica y contaminadas con ruido blanco (WHITE) y murmullo (BABBLE). Las grabaciones corresponden a dos sujetos argentinos nativos, uno de sexo masculino y otro de sexo femenino. Los archivos de ruido se tomaron de la base de datos NOISEX (37) y fueron adicionados computacionalmente en distintas SNRs (sin tener en cuenta el conocido efecto Lombard)¹.

Las pruebas efectuadas comprenden la evaluación de la inteligibilidad y de la calidad del habla obtenida con las distintas estrategias de procesamiento consideradas.

III.1. Evaluación de inteligibilidad

La evaluación de inteligibilidad se efectuó a través de pruebas subjetivas con sujetos normo-oyentes. La normalidad de la audición de los sujetos participantes fue analizada a través de audiometría de tonos puros. Las señales contaminadas con ruido blanco y murmullo a SNR

de -5, 0, y 5 dB y posteriormente procesadas por uno de los algoritmos de reducción de ruido, fueron presentadas en forma binaural a los participantes. El material presentado consistió en una secuencia de palabras de cada uno de los subconjuntos que constituyen la batería. La reproducción del material se efectuó a una intensidad de 65 dB SPL (Sound Pressure Level), y tuvo lugar dentro de una cámara anecoica.

Para determinar el desempeño de cada algoritmo se tomaron diez registros para cada tipo y condición de ruido. Se compararon las respuestas con las referencias de las elocuciones presentadas y se determinaron los índices de error de reconocimiento correspondientes a cada algoritmo en cada condición de ruido. Se construyeron también matrices de confusión de consonantes a fin de visualizar los fonemas más comprometidos con cada uno de los algoritmos. Estos arreglos fueron contruidos sólo en base a los registros en los cuales era identificable una sustitución, supresión o inserción de fonemas. Los casos en los cuales la respuesta dada por el oyente difirió notablemente de la referencia presentada no fueron considerados.

III.2. Evaluación de calidad

La evaluación de la calidad se realizó mediante una prueba subjetiva y una prueba objetiva. En la prueba subjetiva, los oyentes fueron consultados para calificar el habla procesada en cada condición de ruido, de acuerdo a cuatro dimensiones: claridad, ruido residual, confort, y criterio global de aceptación. En todos los casos se utilizó una escala de calificación decimal, siendo de 10 puntos la mejor calificación posible en todos los casos.

La evaluación objetiva de la calidad del habla se efectuó contemplando cuatro medidas:

- *Log Area Ratio (LAR)*: Es una medida basada en la diferencia entre los coeficientes de predicción lineal obtenidos para la señal limpia y para la señal procesada [19]. Su buena correlación con la aceptación general de los algoritmos en pruebas subjetivas tanto con normo-oyentes como con pacientes hipoacúsicos ha sido reportada por algunos autores [30].
- *Log-Likelihood Ratio (LLR)*: Esta medida es también conocida con

el nombre de *distancia de Itakura*, y al igual que la anterior, se basa en la diferencia entre los coeficientes de predicción lineal para la señal limpia y para la señal procesada [19], pero con una ponderación diferente de los espectros obtenidos.

- *SegSNR*: esta medida se basa en la promediación de la relación señal-ruido obtenida para cada segmento. Si bien ha mostrado no tener una gran correlación con los resultados subjetivos en muchos casos, es una de las medidas más empleadas en el reporte de resultados de nuevas técnicas.

- *Qc*: Esta medida brinda una estimación de la similitud de las representaciones internas de la señal limpia y de la procesada, basándose en un modelo psicoacústico del proceso periférico de audición [10, 20]. Algunos autores [30] han reportado su buena correlación con datos subjetivos asociados a pacientes hipoacúsicos.

IV. Resultados

En las pruebas subjetivas de inteligibilidad, el porcentaje de palabras no reconocidas por los oyentes para cada algoritmo clásico se resume en la **Tabla I**. Como puede apreciarse, los resultados correspondientes a la regla de Ephraim y Malah (LogSTSA) son los mejores con ambos tipos de ruido. Comparando la influencia de ambos tipos de ruido puede notarse, además, que el desempeño de Wiener es similar en ruido blanco y murmullo, mientras que PSS y LogSTSA muestran desempeños mejores en ruido blanco, siendo las diferencias más sensibles cuanto más baja es la SNR. Tras realizar un análisis de las correspondientes matrices de confusión entre fonemas se pudo establecer que, para la contaminación con ruido murmullo, los algoritmos estudiados presentan dificultades principalmente para la discriminación de las consonantes oclusivas sordas /p/, /t/ y /k/; /g/, oclusiva sonora, y en menor medida para /b/ y /d/, oclusivas sonoras. En todos los casos se observó que también es importante la omisión de fonemas registrada en condiciones elevadas de ruido, comprometiendo principalmente las oclusivas y nasales. Por otra parte, es significativa la inserción de fonemas registrada con sustracción espectral. Pudo establecerse un eleva-

do porcentaje de reconocimiento de las consonantes fricativas en ruido de murmullo, observándose una disminución con ruido blanco, resultando especialmente significativas las confusiones del fonema fricativo /f/. Estas observaciones ponen en evidencia la influencia de las características espectrales del ruido en la inteligibilidad alcanzada. El número de elocuciones no repetidas por los oyentes por resultarles incomprensibles fue mayor en el caso de sustracción espectral, siendo menor para las reglas propuestas en [15,16].

Tabla I. Pruebas subjetivas: Porcentaje de palabras mal reconocidas (Clásicas)

Ruido	Blanco			Murmullo		
	- 5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB
Algoritmo						
PSS	15,70	5,95	1,90	23,60	10,70	4,36
Wiener	15,20	5,71	4,50	12,10	5,27	2,38
LogSTSA	7,10	3,70	1,05	13,10	4,76	2,38

Los resultados de las pruebas subjetivas de evaluación de calidad se muestran en la **Tabla II**. Como puede apreciarse, el algoritmo de Ephraim y Malah fue encontrado, en general, superior en todos los aspectos, seguido del filtro de Wiener y por último la técnica de sustracción espectral. Puede también apreciarse que Wiener y LogSTSA alcanzan calificaciones similares en condiciones de contaminación con murmullo con SNR no negativa.

Tabla II. Pruebas subjetivas: evaluación de calidad del habla (Clásicas)

SNR	Blanco			Murmullo			Blanco			Murmullo		
	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB
Algoritmo	APRECIACIÓN GENERAL						CLARIDAD					
PSS	4.10	5.40	6.70	3.10	4.90	6.60	3.70	5.40	6.20	3.40	4.70	6.90
Wiener	5.50	6.10	7.80	4.20	6.30	7.70	5.30	6.00	7.70	4.40	6.30	7.50
LogSTSA	6.20	7.00	8.50	5.00	6.50	7.80	6.10	6.90	8.50	5.00	6.10	7.90
	CONFORT						RUIDO RESIDUAL					
PSS	3.90	4.80	6.50	3.50	4.10	6.00	3.90	4.80	6.90	2.70	4.30	5.40
Wiener	5.20	6.40	6.90	5.00	6.10	7.30	3.90	5.30	7.60	3.80	6.10	7.00
LogSTSA	6.10	7.10	8.20	5.10	6.30	7.30	5.50	6.90	7.90	5.00	6.30	6.80

Para la evaluación objetiva con las medidas consideradas anteriormente, se incorporaron además señales contaminadas con ruido en relaciones de 10 y 15 dB. También se incluyó otra variante de los algoritmos de Ephraim-Malah (STSA) [15], y una técnica basada en una representación de la señal en cosenos discretos (DCT) (35). Los resultados se muestran en la **Figuras 1**, para contaminación con murmullo. La **Fig. 1.a** muestra que la medida LAR favorece a LogSTSA frente a PSS y a Wiener en todas las SNR consideradas (debe tenerse en cuenta que una menor distancia representa una mayor similitud entre la señal procesada y la señal limpia de referencia). Sin embargo, a diferencia de lo ocurrido en las pruebas subjetivas, PSS es mejor ponderada que WIENER. Además, la diferencia de desempeño entre LogSTSA y PSS permanece aproximadamente constante en todas las condiciones, mientras que el desempeño relativamente inferior atribuido a Wiener es más evidente a alta SNR. Nótese también que las técnicas STSA y DCT muestran el mejor desempeño para SNRs menores a 5 dB. Para LLR (**Fig. 1.b**), PSS resulta mejor que Wiener y que LogSTSA en el rango comprendido entre -5 y 5 dB. Aún más, en la mayor parte de este intervalo Wiener es superior a LogSTSA. Todo esto se contrapone con los resultados obtenidos en las pruebas subjetivas. Para el caso de SegSNR (**Fig. 1.c**), puede apreciarse que todas las técnicas obtienen una calificación similar, con excepción de Wiener, que obtiene una evaluación inferior. Si bien LogSTSA es también el que muestra mejor desempeño con esta medida, la calificación relativa entre PSS y Wiener se contrapone nuevamente a los resultados subjetivos. La similitud observada en la valoración de las distintas técnicas no permite determinar un ordenamiento en cuanto a sus desempeños. Con Qc (**Fig. 1.d**), por su parte, PSS es valorada pobremente, en tanto que el resto de las técnicas obtiene una calificación similar más alta. Entre ellas, Wiener muestra un mejor comportamiento que STSA, lo cual también se contrapone con los resultados subjetivos obtenidos con normo-oyentes.

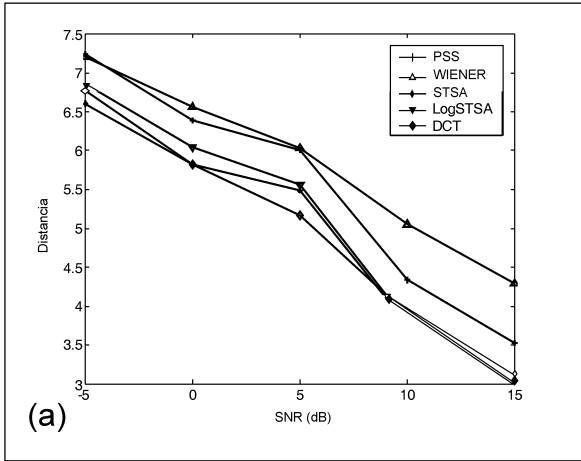


Figura 1 a: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos clásicos para el caso de habla femenina contaminada con ruido murmullo. Estimación de la calidad para distintas SNR utilizando LAR como medida.

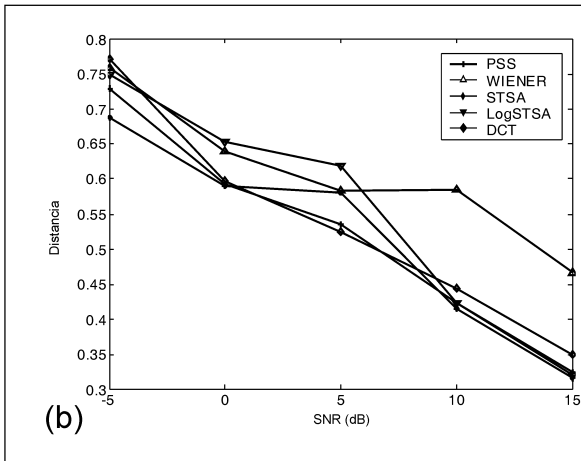


Figura 1 b: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos clásicos para el caso de habla femenina contaminada con ruido murmullo. Comparación usando LLR.

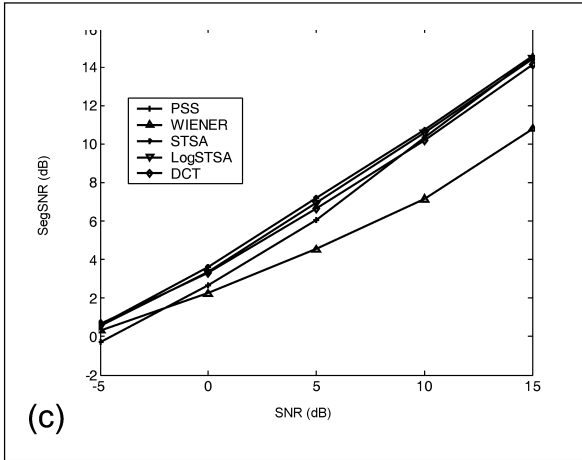


Figura 1 c: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos clásicos para el caso de habla femenina contaminada con ruido murmullo. Comparación por medio de SegSNR.

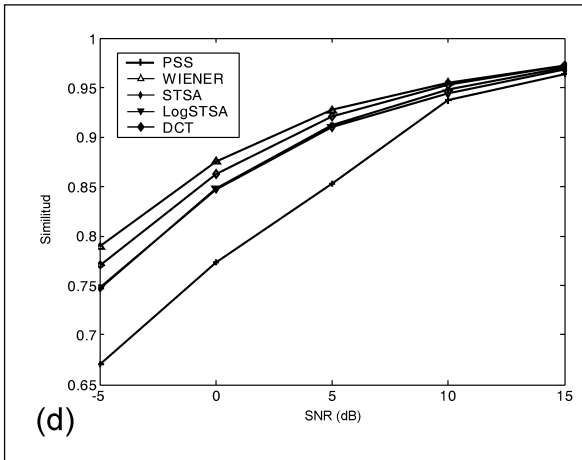


Figura 1 d: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos clásicos para el caso de habla femenina contaminada con ruido murmullo. Estimación de calidad empleando Qc.

Para el caso de onditas, el porcentaje de palabras mal reconocidas por los oyentes para cada algoritmo estudiado se resume en la **Tabla III**. En la **Tabla IV** se muestran los resultados de las pruebas subjetivas de calidad correspondientes para estos métodos, sugiriendo una relativa superioridad del método de Lang. Sin embargo, del análisis conjunto de ambas tablas no puede inferirse que alguno de los métodos sea significativamente mejor que los demás. Por otra parte, debe destacarse que los resultados obtenidos para los métodos clásicos y los basados en onditas no admiten comparación directa debido a que fueron realizados en distintas poblaciones de sujetos. En tal sentido, sería necesario considerar los posibles efectos de la variabilidad de juicio de los oyentes y el número y experiencia de éstos en pruebas de este tipo, lo que puede haber influenciado los resultados de las pruebas subjetivas en ambos grupos. Sin embargo, la marcada diferencia de desempeño puesta en relieve en estos resultados sugiere una clara ventaja de los métodos clásicos. Esto parece concordar con resultados indirectamente sugeridos por Hu y Loizeau, quienes en [21] proponen una técnica de limpieza basada en onditas y recientemente en [22] y [23] realizan comparaciones de diversos métodos no incluyendo a los basados en onditas.

Tabla III. Pruebas subjetivas: Porcentaje de palabras mal reconocidas (onditas)

Ruido Algoritmo	Blanco			Murmullo		
	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB
Akhaee	43,59	26,39	20,83	36,11	32,94	7,87
Arefeen	37,96	43,59	22,92	45,30	32,54	27,35
Bahoura	51,85	25,00	27,45	48,02	26,98	23,41
Chang	45,30	32,54	14,58	79,06	21,60	2,38
Lei	59,83	38,49	46,08	62,82	31,94	11,46
Lang	38,49	31,20	17,90	36,75	16,32	20,08

Tabla IV. Pruebas subjetivas: evaluación de calidad del habla (onditas)

SNR	Blanco			Murmullo			Blanco			Murmullo		
	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB	-5 dB	0 dB	5 dB
Algoritmo	APRECIACIÓN GENERAL						CLARIDAD					
Akhaee	4,05	4,75	4,70	3,55	4,70	5,80	3,95	4,50	4,90	3,65	4,45	5,90
Arefeen	3,95	3,90	5,00	3,15	4,50	4,95	4,10	4,10	5,15	3,15	4,50	5,00
Bahoura	4,00	4,50	5,15	3,55	4,45	5,15	4,05	4,65	5,65	3,30	4,85	5,05
Chang	4,35	4,60	5,35	2,90	4,75	6,00	4,20	4,70	5,90	2,60	4,65	5,90
Lei	3,75	3,60	3,85	3,00	4,20	5,15	3,65	3,70	3,75	2,90	4,05	5,15
Lang	4,25	4,15	6,05	4,20	4,85	5,75	4,25	4,15	5,05	4,55	4,95	5,50
	CONFORT						RUIDO RESIDUAL					
Akhaee	4,00	4,40	4,80	3,40	4,30	5,60	4,25	4,45	4,95	3,00	4,80	5,50
Arefeen	3,45	3,75	4,55	3,30	4,45	4,70	3,45	4,20	4,95	2,65	4,20	4,70
Bahoura	3,85	4,40	4,90	3,10	4,45	4,90	4,25	4,85	5,15	2,80	4,55	5,05
Chang	3,85	4,15	5,35	2,95	4,45	5,70	4,00	4,86	5,40	3,40	5,20	6,05
Lei	3,60	3,35	3,35	3,10	4,20	4,85	4,50	4,40	4,30	3,25	4,45	4,90
Lang	4,05	4,30	4,80	4,00	4,85	5,15	5,05	5,05	5,80	3,50	5,15	5,50

Los resultados objetivos para onditas pueden observarse en la **Figuras 2**. Un análisis global de los mismos indicaría una superioridad de los algoritmos propuestos por Akhaee y por Lang. De la comparación entre la **Fig/s. 1** y la **Fig/s. 2** se desprende que las pruebas objetivas muestran también un mejor desempeño general para los algoritmos clásicos.

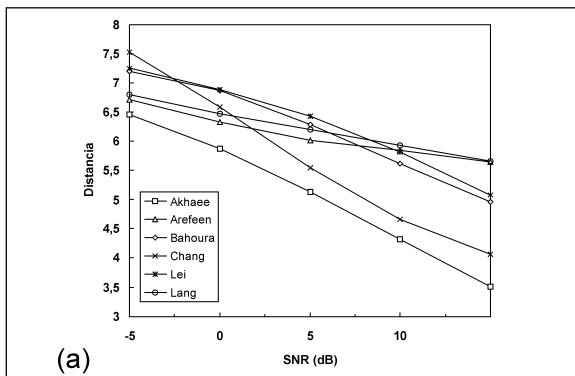


Figura 2 a: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos basados en onditas. Estimación de la calidad para distintas SNR utilizando LAR como medida.

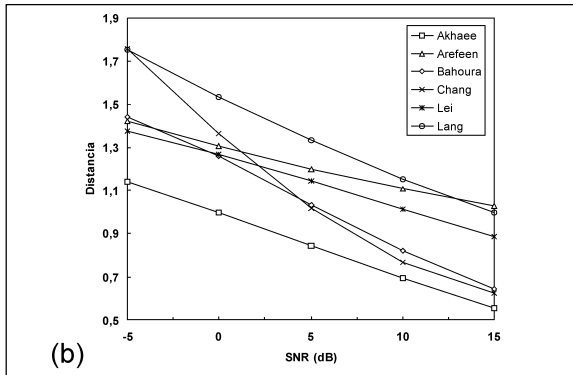


Figura 2 b: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos basados en ondas. Comparación usando LLR.

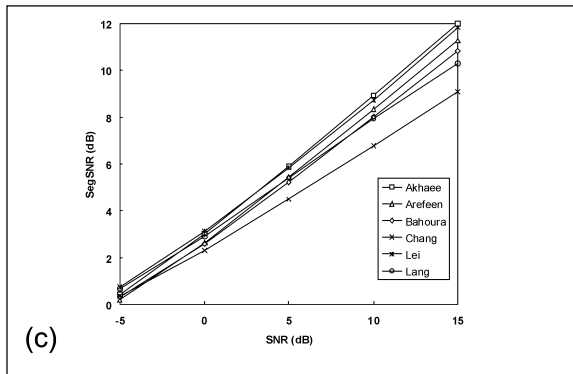


Figura 2 c: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos basados en ondas. Comparación por medio de SegSNR.

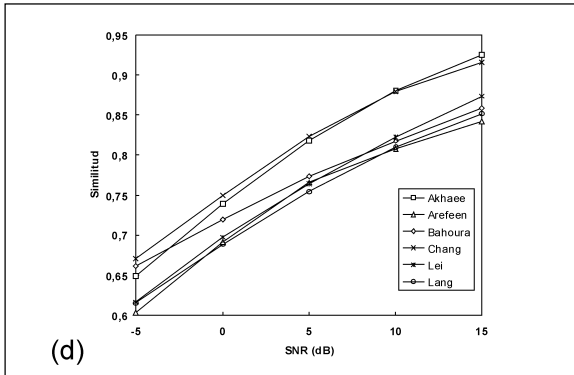


Figura 2 d: Resultados de la evaluación objetiva de la calidad del habla procesada con los algoritmos basados en onditas. Estimación de calidad empleando Qc. En todos los casos los resultados mostrados corresponden a habla contaminada con ruido murmullo.

V. Conclusiones

En este trabajo se han analizado e implementado un total de 18 técnicas de limpieza de ruido, 5 basadas en métodos clásicos y 13 no convencionales que utilizan la transformada ondita. Los mismos fueron evaluados mediante pruebas objetivas y subjetivas en individuos normo-oyentes utilizando una batería en idioma español nativo especialmente diseñada al efecto. El análisis de los resultados obtenidos pone de relieve la complejidad que entraña la evaluación exhaustiva de los distintos métodos de limpieza en habla destinados a su utilización en prótesis auditivas. Si bien los resultados generales obtenidos en nuestro caso para el idioma español parecerían estar en consonancia con los reportados por otros autores en idioma inglés, aún no permiten concluir cual o cuales serían las mejores técnicas a incluir en prótesis auditivas. En particular, restan pruebas con pacientes hipoacúsicos que demandaran un estudio con un protocolo específico.

Por otro lado, las contraposiciones observadas entre las mediciones subjetivas y objetivas de calidad sugieren la necesidad de explorar nuevas técnicas de medición. En trabajos recientes del grupo se ha encontrado una correlación muy importante entre las medidas subjetivas de calidad y la medida PESQ (*perceptual evaluation of speech quality*) para algoritmos de separación ciega de fuentes [12].

Notas

1. Sin tener en cuenta el conocido efecto Lombard.

Referencias bibliográficas

1. AKHAEI M., AMERI A., MARVASTI F. Speech enhancement by adaptive noise cancellation in the wavelet domain. En: *Fifth International Conference on Information, Communications and Signal Processing*. 2005, Dic.6-9; Bangkok, Thailand. p. 719-723.
2. AREFEEN Z., KAMRUL H., REZWAN K. Efficient hard and soft thresholding for wavelet speech enhancement. En: *XI European Signal Processing Conference (Eusipco)*, Vol.1; Toulouse, France; 2002. p. 507-510.
3. ARONSON L., MILONE D., MARTÍNEZ C., ESTIENNE P., TOMASSI D., RUFINER H., TORRES M. Batería para la evaluación del reconocimiento del habla en pacientes con prótesis auditiva. *Revista de la Federación Argentina de Sociedades de Otorrinolaringología*. 2007, 14 (1):17-24.
4. BAHOURA M., ROUAT J. Wavelet speech enhancement based on the Teager Energy Operator. *IEEE Signal Processing Letters*. 2001; 8 (1): 10-12.
5. BEROUTI M., SCHWARTZ R., MAKHOUL J. Enhancement of Speech Corrupted by Acoustic Noise En: *Proc. of the International Conference on Acoustics, Speech and Signal Processing*; 1979. p. 208-211.
6. BOLL S. F. Suppression of Acoustic Noise in Speech Using Spectral Subtraction. *IEEE Trans. On Acoustics, Speech and Signal Processing*. 1979; 27(2): 113-120.
7. CAPPÉ O. Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor. *EEE Trans. On Speech and Audio Processing*. 1994; 2: 345-349.
8. CHANG S., KWON Y., YANG S., KIM I. Speech enhancement for non-stationary noise environment by Adaptive Wavelet Packet. *Acoustics, Speech, and Signal Processing*. 2002; 1: 561-564.
9. COHEN I. Enhancement of speech using Bark-Scaled Wavelet Packet Decomposition. En: *Proceedings Eurospeech*, Aalborg, Denmark; 2001. p. 1933-1936.
10. DAU T., PUSCHEL D.; KOHLRAUSCH A. A Quantitative Model of the Effective Signal Processing in the Auditory System. Part I: Model Structure. *Journal of the Acoustical Society of America*. 1996; 99 (6): 3615-3622.
11. DELLER J., PROAKIS J., HANSEN J. *Discrete-Time Processing of Speech Signals*. New York: Prentice Hall; 1993.
12. DI PERSIA L., MILONE D., RUFINER H.L., YANAGIDA M. Perceptual evaluation of blind source separation for robust speech recognition. *Signal Processing*. 2008; 88 (10): 2578-2583.
13. DONOHO D. De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*. 1995; 41 (3): 613-627.
14. DONOHO D., JOHNSTONE J. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*. 1994; 81 (3): 425-455.
15. EPHRAIM Y., MALAH D. Speech Enhancement Using a Minimum Mean-Square Error Short Time Spectral Amplitude Estimator. *IEEE Trans. On Acoustics, Speech and Signal Processing*. 1984; 32: 1109-1121.
16. EPHRAIM Y., MALAH D. Speech Enhancement Using a Minimum Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Trans. On Acoustics, Speech and Signal Processing*. 1985; 33: 443-445.
17. FU Q.; WAN ERICA. A novel speech enhancement system based on wavelet denoising. *Center of Spoken Language Understanding, OGI School of Science and Engineering at OHSU*; (2003) <<http://speech.bme.ogi.edu/publications/ps/fu03.pdf>> [Agosto 2008].
18. GAO H. BRUCE A.G. Waveshrink with Semisoft Shrinkage. En: *Technical Report StatSci Division of MathSoft*. 1995, 39 <citeseer.comp.nus.edu.sg/438854.html> [21 Agosto 2008]
19. HANSEN JHL, PELLOM B. An Effective Quality Evaluation Protocol for Speech Enhancement Algorithms. En: *Proc. of the International Conference on Spoken Language Processing*. 1998; 7. p. 2819-2822.

20. HANSEN M, KOLLMEIER B. Continuous Assessment of Time-Varying Speech Quality. *Journal of the Acoustical Society of America*. 1999; 106 (5): 2888-2899.
21. HU Y, LOIZOU PC. Speech enhancement based on wavelet thresholding the multitaper spectrum. *IEEE Transactions on Speech and Audio Processing*. 2004; 12 (1): 59-67.
22. HU Y, LOIZOU PC. Subjective comparison and evaluation of speech enhancement algorithms. *Speech Commun*. 2007; 49 (7-8): 588-601.
23. HU Y, LOIZOU PC. Evaluation of Objective Quality Measures for Speech Enhancement. *IEEE Trans on Audio, Speech, and Language Processing*. 2008; 16 (1): 229-238.
24. INGARAMO J.P., RISSO P.G. Técnicas de reducción de ruido en señales del habla basadas en la transformada ondita. [Tesis de grado, carrera de Bioingeniería] Facultad de Ingeniería, UNER; 2008 (inédito).
25. KIM I., YANG S., KWON Y. Speech enhancement using adaptive wavelet shrinkage. En: *Proceedings of International Conference on Industrial Electronics*; Pusan, Korea; 2001; Vol.1. p. 501-504.
26. LANG M., GUO H., ODEGARD J.E., BURRUS C.S., WELLS R.O. Jr. Noise reduction using an Undecimated Discrete Wavelet Transform. *IEEE Signal Processing Letters*. 1996; 3: 10-13.
27. LEI S., TUNG Y. Speech enhancement for nonstationary noises by wavelet packet transform and adaptive noise estimation. En: *Proceedings of 2005 International Symposium on Intelligent Signal Processing and Communication Systems*; Hong Kong, 2005; 28. p. 41-44.
28. LOIZOU P.C. *Speech Enhancement: Theory and Practice*. Boca Raton: CRC Press; 2007.
29. MALLAT S., WEN LIANG H. Singularity detection and processing with Wavelets. *IEEE Trans. Inform. Theory*. 1992; 38 (2): 617-643.
30. MARZINZIK M., KOLLMEIER B. Predicting the Subjective Quality of Noise Reduction Algorithms for Hearing Aids. *Acta Acustica*. 2003; 89: 521-529.
31. SAEED A., MOHAMMAD T., ROOHOLLAH R. Wavelet based speech enhancement using a new thresholding algorithm. En: *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*. Hong Kong, Oct.20-22, 2004. p. 238-241.
32. SEOK J., BAE K. Speech enhancement with reduction of noise components in the wavelet domain. *IEEE Int'l Conf on Acoustics, Speech, and Signal Processing*. 1997; 2: 1323-1326.
33. SHEIKHZADEH H., ABUTALEBI H. An improved wavelet-based speech enhancement system. En: *Proc. 7th European Conf. on Speech Communication and Technology*. Aalborg, Denmark, Sep. 2001; p. 1855-1858.
34. SOON I., KOH S., YEO C. Wavelet for speech denoising. *Speech and Image Technologies for Computing and Telecommunications*. 1997; 2: 479-482.
35. SOON I., KOH S., YEO C. Noisy speech enhancement using discrete cosine transform. *Speech Communication*. 1998; 24: 249-257.
36. TOMASSI D. R., ARONSON L., MARTÍNEZ C. E., MILONE D. H., TORRES M. E., RUFINER H. L. Evaluación de técnicas clásicas de reducción de ruido en señales de voz. *Revista de la Sociedad Argentina de Bioingeniería*. 2005; 11 (2): 19-28.
37. VARGA A., STEENEKEN H. Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems. *Speech Communication*. 1993; 12 (3): 247-251.
38. VASEGHI S.V. *Advanced Digital Signal Processing and Noise Reduction*, 2nd Ed. New York: John Wiley & Sons; 2000.