Ingeniería y Ciencia
ISSN: 1794-9165
ingciencia@eafit.edu.co
Universidad EAFIT
Colombia

Ballesteros L., Dora M.; Renza, Diego; Duvan Ortiz, Héctor
Función resumen perceptual para verificación de integridad en audio forense
Ingeniería y Ciencia, vol. 13, núm. 25, enero-junio, 2017, pp. 167-183
Universidad EAFIT
Medellín, Colombia

Ingeniería y Ciencia
ISSN:1794-9165 | ISSN-e: 2256-4314
ing. cienc., vol. 13, no. 25, pp. 167–183, enero-junio. 2017.
http://www.eafit.edu.co/ingciencia

Función resumen perceptual para verificación de integridad en audio forense

Dora M. Ballesteros L. [1], Diego Renza [2]y Héctor Duvan Ortiz [3]

Recepción: 01-02-2017 | Aceptación: 18-04-2017 | En línea: 08-05-2017

PACS:43.72.Uv

doi:10.17230/ingciencia.13.25.7

Resumen

En este trabajo se propone una función que permite calcular un código resumen a partir de los parámetros de una señal de voz. Esta función está basada en el ordenamiento de los coeficientes espectrales en un proceso de imitación entre el espectro de la señal de voz y el espectro de una señal de ruido gaussiano generada localmente. El método de función resumen está orientado a la verificación de integridad forense en señales de voz, con un enfoque perceptual, que implica que la función resumen no cambia si la señal sufre modificaciones que no alteran el contenido (como re-cuantización), pero que si cambia ante modificaciones como recorte y adición de ruido. Se realizaron diversas pruebas para verificar el enfoque perceptual del método resumen propuesto y se compararon los resultados frente a modificaciones utilizando métodos tradicionales.

Palabras clave: Función resumen; función perceptual; integridad; audio forense.

[1] Universidad Militar Nueva Granada, dora.ballesteros@unimilitar.edu.co, http://orcid.org/0000-0003-3864-818X, Bogotá D.C., Colombia.
[2] Universidad Militar Nueva Granada, diego.renza@unimilitar.edu.co, http://orcid.org/0000-0001-8073-3594, Bogotá D.C., Colombia.
[3] Universidad Militar Nueva Granada, hduvanortiz@gmail.com, http://orcid.org/0000-0001-7299-4700, Bogotá D.C., Colombia.

Perceptual Digest Function for Verifying Integrity in Audio Forensics

Abstract
In this work we propose a function that allows to calculate a summary code from the parameters of a voice signal. This function is based on ordering of spectral coefficients obtained by means of the application of the Fast Fourier Transform (FFT), using a locally generated reference function (Gaussian random noise). The proposed method is oriented to the verification of integrity in forensic voice signals. The proposed methodology has a perceptual approach, which implies that the resulting code is maintained, even when modifications are made, particularly those that do not affect the sensitive content of the signal, such as re-quantization processes.

Key words: Hash function; perceptual hash; integrity; audio forensics.

1 Introducción

A pesar de la constante innovación en el área de la seguridad de la información, la alteración de contenidos es actualmente uno de los grandes problemas que aqueja a diversas áreas, y que puede tener una repercusión negativa y directa sobre los actores que basan sus actividades en la difusión de estos. Esto ha dado origen a que partiendo de conocer que los contenidos pueden ser manipulados de forma directa o indirecta, se generen métodos que permitan detectar si la información ha sido cambiada, de qué manera se hizo y de qué magnitud fue la alteración [1]. En el área de contenidos multimedia, este problema se agudiza con el aumento en el número de soluciones digitales que permiten editar el contenido de una grabación de voz y la facilidad de utilizarlos en un computador o celular, lo que implica que cada día sea más difícil demostrar y verificar la integridad de registros de voz.

Uno de los campos en los que la verificación de integridad juega un papel importante es el de audio forense, área que corresponde al tratamiento de señales de voz con el objetivo de determinar si su contenido representa una evidencia que pueda ser utilizada en la resolución de casos, en su mayoría, de ámbito jurídico. Algunas de las alteraciones que se pueden encontrar a través de esta práctica corresponden por ejemplo a la adición de ruido para deteriorar la señal de voz o para dar un falso escenario, realizar recortes o

empalmes de conversaciones, o generar pérdidas de información debido a alteraciones de los parámetros base de la señal de voz (frecuencia, amplitud, formato, entre otros) [2].

La verificación de integridad es de gran importancia en el ámbito forense ya que permite corroborar que un contenido no ha cambiado, y a través de este tipo de métodos proporcionar una herramienta útil para el manejo de evidencias digitales. Los métodos más comunes que permiten verificar la integridad de contenidos corresponden al marcado frágil y al uso de funciones resumen [3]. El marcado frágil corresponde a la inserción de una determinada información, llamada firma, en un medio digital, de tal forma que al momento de realizarse algún tipo de manipulación sobre el contenido del medio digital, la marca se destruya; esto a su vez, permite corroborar la integridad de dicho contenido [4]. Por otra parte, las funciones resumen permiten generar un mensaje de salida basado en el contenido del medio digital; este proceso se realiza tanto en el medio original como en el medio que se cree que ha sido modificado. En cualquiera de los dos casos, al comparar los respectivos códigos (o firmas) es posible encontrar diferencias que puedan indicar alteraciones en el contenido [5].

1.1 Marcado frágil

Los métodos de marcado frágil propuestos hasta el momento han buscado alternativas que ofrezcan un alto nivel de imperceptibilidad de la firma y resistencia a manipulaciones de forma pero no de contenido. En este contexto, se pueden resaltar trabajos tales como un esquema de marcado en audio basado en descomposición de valores singulares y evolución diferencial utilizando cuantización dinámica, este método otorga gran robustez, imperceptibilidad y fundamenta su uso en la protección de derechos de autor [6]. También se encuentran trabajos en los cuales las investigaciones se enfocan en mejorar funciones ya existentes, como es el caso del planteamiento de un algoritmo genético para el marcado frágil de señales de audio que busca reducir la distorsión generada al aplicar métodos de bit menos significativo, lo que a su vez mejora la relación señal pico a ruido (PSNR) [4]. Sin embargo, una falencia presente en el uso del marcado frágil se debe a que la inserción de una firma puede considerarse como una alteración directa del contenido.

Por otra parte, los trabajos basados en funciones resumen centran su investigación en reducir costo computacional y a la vez evitar colisiones (repetición de códigos partiendo de contenidos diferentes). Entre los trabajos realizados a partir de la implementación de funciones resumen, podemos resaltar el diseño de un esquema de autenticación enfocado en señales de audio con compresión AAC (Advanced Audio Coding), que se destaca por el uso de este tipo de formato representativo de los archivos de audio digitales en Internet [7]; otro método destacado corresponde al diseño de una función resumen robusta que permite extraer algunos mili-segundos de un audio, y a partir de este fragmento se comparan sus características generales para verificar su integridad con respecto al audio completo [8]. Así mismo, recientemente se ha propuesto el diseño de una función resumen basada en el uso de matrices no negativas (NMF) y coeficientes obtenidos de la aplicación de la transformada discreta de coseno (DCT), esto permite realizar reducción de información, aumentar la robustez y disminuir el costo computacional [9].

1.2 Funciones resumen

El uso de una función resumen corresponde a la práctica por medio de la cual se representa un contenido C a través de una cadena de caracteres H. Aquí C representa la información de entrada que se quiere proteger, verificar y/o autenticar, mientras que H es la información de salida basada en las propiedades de la entrada C [10]. Los parámetros más relevantes para determinar el nivel de efectividad de una función resumen se pueden obtener a partir del cumplimiento de ciertas propiedades representativas de este tipo de métodos, teniendo en cuenta el fin para el cual fue diseñada la función. Entre estas propiedades se destacan las siguientes:

- Bajo costo: el cálculo de H debe requerir bajo costo computacional [11].
- Compresión: la cadena de caracteres de H al ser una representación de C, debe ser de longitud menor [12].
- Uniformidad: se debe minimizar el número de posibles colisiones, es decir, que dos entradas distintas generen la misma salida [12].

- Determinista: dos entradas C iguales, deben generar salidas H iguales. Esto implica que no haya ninguna modificación de contenido y que valores externos como los metadatos no influyan en la generación de H [12].

- Unidireccional: la función se debe diseñar de tal manera que, al tener un valor de H sea computacionalmente imposible revertir el proceso para encontrar C [13].

- Longitud: la cantidad de elementos presentes en H debe ser preestablecido [11].

Adicionalmente, en contenidos multimedia, es posible considerar dos condiciones adicionales: discriminación y robustez perceptual. La discriminación significa que al tener dos C totalmente diferentes, sus H correspondientes deben ser a su vez totalmente diferentes. Por su parte, en la robustez perceptual, al tener una modificación de C que no afecte su contenido sensible, la salida H del contenido original y el H del contenido modificado, no deben presentar diferencias [14]. Lo anterior implica que la clasificación y el uso de funciones resumen puede estar directamente ligado al tipo de contenido sobre el cual se calculará la cadena de caracteres. Por ejemplo, si hablamos de un contenido tal como una señal de voz, es posible hablar de un método perceptual y con cierto grado de discriminación; en este caso, el interés principal de la función resumen debe radicar en evaluar la preservación del contenido, más no en la forma de presentación del mismo. Esta práctica es comúnmente usada para la emisión de certificados, firmas digitales, generación y verificación de claves. Sus aplicaciones principales están orientadas a la protección y gestión de contenidos, debido a que permite distinguir algún tipo de modificación o manipulación de la información [8].

De acuerdo a lo anterior, las funciones resumen permiten verificar la integridad de un contenido de forma más sencilla debido a que la cadena de caracteres es mucho más pequeña que el contenido sobre el cual fue calculado. Además, al momento de verificar y comparar varios contenidos simultáneamente, el costo computacional es muy bajo. Estas cualidades hacen que las funciones resumen sean ampliamente usadas en los campos de criptografía, autenticación y análisis forense [15]. Entre los métodos más

representativos y difundidos, podemos resaltar los siguientes:

- SHA 512: pertenece a un conjunto de funciones resumen diseñadas por la Agencia de Seguridad Nacional (NSA) y el Instituto Nacional de Estándares y Tecnología (NIST) en el año 2001, que se diferencia de sus antecesores debido a que trabaja con un mayor número de iteraciones. Este algoritmo opera en ocho palabras de 64 bits, donde el contenido a ser codificado primero se divide en bloques de 1024 bits, que se procesan uno a la vez para posteriormente entregar una salida de 512 bits [16].

- MD5: Es un algoritmo de reducción de 128 bits diseñado por el profesor Ronald Rivest en 1991, que toma como entrada una cadena de texto y a su salida entrega un código de 128 bits. Este algoritmo se ejecuta en los siguientes pasos: se rellena el mensaje con bits de tal forma que a la longitud le falten 64 bits para ser múltiplo de 512, esta longitud se añade como dos palabras de 32 bits. Para realizar las operaciones se usan 4 buffers que corresponden a registros de 32 bits, procesando el mensaje en bloques de 16 bits y obteniendo una salida de 128 bits [17].

- TIGER: es una función de hash diseñada por Ross Anderson y Eli Biham en 1996 para procesadores de 64 bits, más rápido que SHA1 y MD5. Trabaja en bloques de 512 bits que provienen de usar 8 palabras de 64 bits, las que a su vez entregan 3 palabras de 64 bits. Esto genera una cadena final de 192 bits [18].

En este contexto, en el presente trabajo se propone una metodología que calcula una función resumen perceptual, diseñada para verificación de integridad forense de señales de voz digital. El algoritmo propuesto opera en el dominio de la frecuencia mediante la Transformada Rápida de Fourier (FFT), y está orientado a mantener el mismo código resultante al realizar modificaciones que no afectan el contenido sensible de la señal, tales como los procesos de recuantización.

2 Metodología

El método propuesto de generación de funciones resumen para señales de voz se basa en la habilidad de imitación de este tipo de señales a señales de ruido gaussiano, propuesto por los autores [19],[20],[21]. En la actual propuesta, la imitación se realiza en el dominio de la frecuencia en el bloque de ordenamiento.

La generación de la función resumen se realiza en segmentos de la señal con duración de 15 segundos, por lo cual si la señal es de longitud mayor, la misma se divide en segmentos de hasta 15 segundos.

A continuación se describe el esquema propuesto que permite obtener el código resumen de una señal de voz digital (Figura 1).

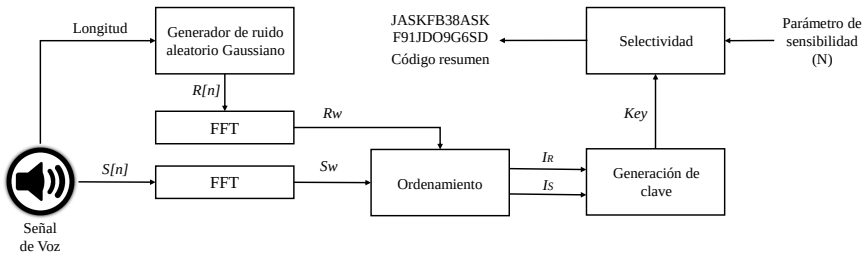

Figura 1: Método propuesto para la generación de la función resumen.

2.1 Generador de ruido aleatorio gaussiano

En este bloque se genera una señal de ruido gaussiano cuyo número de muestras es el mismo número de muestras de la señal de voz, y que adicionalmente comparte sus mismas estadísticas (promedio, desviación estándar). De acuerdo al estudio previo de señales de voz realizado por los autores en [19],[20],[21], se genera el ruido gaussiano con promedio igual a cero y desviación estándar igual a 0.19. Con estos valores, la distribución de probabilidad de la señal de voz y la señal de ruido gaussiano es similar y por consiguiente el proceso de imitación de espectro es viable.

Se aclara que la amplitud pico a pico de la señal de ruido gaussiano generado no afecta el proceso de imitación contemplado en el bloque de or-

denamiento. Es decir, se obtendrá el mismo resultado si la señal gaussianna es amplificada o atenuada.

2.2 Cálculo de componentes frecuenciales de la señal

El objetivo de este bloque es el de obtener el espectro tanto de la señal de entrada S_n, como de la señal de ruido generada R_n; este proceso se realiza por medio de la aplicación de la Transformada Rápida de Fourier. Aquí, el espectro de la señal de audio S_n se denomina como S_w; el espectro de la señal de ruido gaussiano R_n se denomina R_w. No se realiza ninguna manipulación previa a las señales antes del cálculo de su transformada. La cantidad de componentes de frecuencia de cada espectro, S_w y R_w, dependerá de la cantidad de muestras de las señales de entrada. En este caso el número de puntos que se toma para el cálculo de la transformada rápida de Fourier es la potencia de 2 inmediatamente superior al número de muestras de la señal.

Es importante resaltar que las señales de entrada al bloque no sufren ningún tipo de manipulación previa al cálculo de la Transformada de Fourier.

La transformada de Fourier se calcula a través de la ecuación 1.

$$X(k) = \sum_{n=1}^{N} x(n) e^{\frac{-2\pi j(k-1)(n-1)}{N}} \quad , \quad 1 \leq k < N \tag{1}$$

Donde en forma general $x[n]$ es la señal en el dominio del tiempo y $X(k)$ es su espectro.

2.3 Ordenamiento

En este bloque, los coeficientes de la magnitud de FFT de cada señal (voz y ruido) se ordenan de forma ascendente. Así mismo, en cada caso se genera un vector I que guarda las posiciones iniciales de los componentes de frecuencia, tanto de S_w como de R_w. Para el caso de S_w el vector de las posiciones se denomina I_S; mientras que para R_w se denomina I_R, tal como se describe en la Ecuación 2.

$$\begin{array}{lll} S_w = & [4\ 8\ 12\ 2\ 10\ 6\ ] & entonces, \quad I_S = [3\ 5\ 2\ 6\ 1\ 4\ ] \\ R_w = & [10\ 8\ 3\ 5\ 2\ 12\ ] & entonces, \quad I_R = [6\ 1\ 2\ 4\ 3\ 5\ ] \end{array} \tag{2}$$

El método de ordenamiento de los coeficientes espectrales utilizando como referencia una señal de ruido aleatorio gaussiano, está basado en el método propuesto en [21]. La diferencia principal de nuestra propuesta radica en que el ordenamiento se realiza en el dominio de la frecuencia y no en el dominio del tiempo, de tal forma que la imitación se realiza entre el espectro de cada señal.

2.4 Generación de clave

A partir de los vectores I_S e I_R, se genera una clave de ordenamiento de espectro, dada por la Ecuación 3.

$$Key(I_R) = I_S \tag{3}$$

Donde Key es la clave de ordenamiento. La anterior ecuación equivale a un proceso de indexación entre I_R e I_S, esto significa que el elemento ubicado en la posición I_R de la clave, corresponderá al valor de la posición I_R en el vector I_S. De forma general, esta clave, contiene números enteros en el rango de 1 hasta M, donde M es la longitud del vector S_w (ó R_w).

Para el ejemplo dado anteriormente, el vector Key tendrá la forma dada en la Ecuación 4.

$$Key = [5\ 2\ 1\ 6\ 4\ 3\ ] \tag{4}$$

2.5 Selectividad

A continuación se seleccionan los primeros N valores del vector Key. Estos valores se representan con seis dígitos por dato, asignando ceros a la izquierda si es necesario. Por ejemplo, si el valor es 15 se escribe como 000015. La función resumen corresponde a la concatenación de los N valores de Key con formato de 6 dígitos por valor. Para N igual a 50, la función resumen

contiene 300 dígitos. Se enfatiza que independiente de la longitud de la señal de voz de entrada, la función resumen es de longitud fija, ajustable por el usuario. La cantidad de dígitos de la función resumen siempre será igual a N multiplicado por 6. La única condición del método es que la ventana de tiempo de la señal de voz sea máximo de 15 segundos, para señales con frecuencia de muestreo de 8 kHz.

3 Pruebas y resultados

Para la validación del método propuesto, se realizaron pruebas con el fin de determinar el cumplimiento de los parámetros que caracterizan las funciones resumen. Se utilizaron 300 señales de voz con una duración de 15 segundos, una frecuencia de muestreo de 8 kHz y una resolución de 16 bits/muestra. Estas señales corresponden a 50 locutores, de sexo femenino y masculino, en idioma español, provenientes de una base de datos propia con derechos de autor. Es importante aclarar que aunque solo se probaron señales de 8 kHz, el método propuesto es genérico y se puede extrapolar a señales de voz con mayor frecuencia de muestreo.

En cada una de las pruebas, se evalúa que la salida corresponde a la longitud establecida de 300 caracteres alfanuméricos, que no se presenten colisiones y que el algoritmo propuesto sea un método determinista.

Las pruebas listadas en la Tabla 1 fueron orientadas a la obtención de resultados para manipulaciones que sólo modifican la presentación del contenido (Recuantización), y manipulaciones que modifican el contenido sensible (Recorte y Adición de ruido). En cuanto a la recuantización, hay que recordar que la representación de señales de voz digitales precisa tanto de una frecuencia de muestreo, como de la cuantización de la amplitud de la señal. De esta forma, es posible pasar de un valor de cuantización a otro sin alterar el contenido sensible de la señal digital, ya que los parámetros que son cambiados modifican la presentación de la información pero no su contenido. Por su parte el recorte implica extraer, mover o eliminar fragmentos de audio a través de herramientas digitales. Este proceso corresponde a una alteración directa de contenido. En cuanto al ruido, la adición de cualquier clase de ruido corresponde a una alteración de contenido, esto se debe a que el ruido se mezcla con la señal original y altera la información inicial.

Tabla 1: Modificaciones realizadas a las señales de voz de prueba.

Tipo de modificación	Descripción de la modificación	Total de pruebas
1	Recuantización 16 bits a 24 bits	300
2	Recuantización 16 bits a 32 bits	300
3	Recorte de la señal	300
4	Adición de ruido	300

A partir de las modificaciones realizadas, los resultados se presentan en relación a la cantidad de variaciones encontradas en las cadenas de caracteres obtenidas y su porcentaje con respecto a la longitud de la salida. Las variaciones corresponden a la cantidad de cadenas de caracteres en las que se registraron cambios, el promedio se obtiene a partir del total de variaciones y del total de pruebas realizadas (300 pruebas). Mientras que el valor máximo y el valor mínimo representan la cantidad de caracteres que variaron en las cadenas.

En cada una de las modificaciones anteriormente citadas se realizó el proceso que se describe a continuación.

3.1 Recuantización de 16 bits a 24 bits

En esta prueba se aplicó recuantización sobre las señales de voz para obtener una resolución de 24 bits/muestra, se almacenó la señal de voz resultante y posteriormente se calculó el código resumen correspondiente para cada señal utilizando el método propuesto. Con la señal original y la señal manipulada, se calcularon las funciones resumen en cada caso y se compararon dígito a dígito. En la Tabla 2, fila 1, se presenta el consolidado de las 300 pruebas realizadas, es decir, 300 comparaciones entre el código resumen del audio con manipulación respecto al código resumen del audio original.

3.2 Recuantización de 16 bits a 32 bits

Se realizó el mismo proceso indicado en la prueba anterior, pero en este caso para obtener una resolución de 32 bits/muestra. En la Tabla 2, fila

Tabla 2: Consolidado de variaciones entre el código resumen de la señal manipulada utilizando el método propuesto respecto al código resumen de la señal original.

Tipo de modif.	Número de variaciones	% de variaciones	Promedio	Valor máximo	Valor mínimo
1	12	4	0.16	5	0
2	0	0	0	0	0
3	300	100	263.58	275	253
4	300	100	263.35	275	243

2, se presenta el consolidado de las 300 pruebas para la manipulación de recuantización de 16 a 32 bits.

3.3 Recorte de la señal

En este ataque se eliminó un fragmento de cada señal de voz y se almacenaron las señales resultantes. Tomando como entrada las señales de voz modificadas, se calcularon las cadenas de caracteres por medio de la función resumen propuesta, los resultados se presentan en la Tabla 2, fila 3.

3.4 Adición de ruido

Finalmente se aplicó ruido aditivo a cada señal de audio, manteniendo una SNR de 20 dB. El tipo de ruido adicionado corresponde a aleatorio generado in situ. Luego, se aplicó la función resumen propuesta a las señales modificadas, obteniendo los resultados presentados en la Tabla 2, fila 4.

Con el fin de realizar un análisis perceptual del método propuesto en comparación con métodos representativos, se calcularon los códigos resumen aplicando los mismos tipos de modificaciones presentados en la Tabla 1, en este caso procesados mediante funciones resumen disponibles al público como lo son: SHA 512, Tiger 192 y MD5. El objetivo aquí es comparar los resultados obtenidos y presentar las diferencias resultantes respecto al método propuesto. Estos resultados fueron compilados y organizados en las Tablas 3, 4 y 5 de forma similar a como fueron presentados los correspondientes resultados del método propuesto (Tabla 2). Para efectos

comparativos, es importante tener en cuenta la longitud de la cadena de caracteres que genera cada uno de los métodos, es decir, la salida de la función resumen que se menciona en cada una de las tablas correspondientes. Para una comparación general de la variación obtenida en cada método, se recomienda evaluar el porcentaje de variación respectivo para cada modificación.

Tabla 3: Consolidado de variaciones entre el código resumen de la señal manipulada utilizando el método MD5 respecto al código resumen de la señal original.

Tipo de modif.	**Número de variaciones**	**% de variaciones**	**Promedio**	**Valor máximo**	**Valor mínimo**
1	300	100	30.06	32	28
2	300	100	29.92	32	28
3	300	100	30.11	32	28
4	300	100	29.96	32	28

Tabla 4: Consolidado de variaciones entre el código resumen de la señal manipulada utilizando el método TIGER 192 respecto al código resumen de la señal original.

Tipo de modif.	**Número de variaciones**	**% de variaciones**	**Promedio**	**Valor máximo**	**Valor mínimo**
1	300	100	44.5	47	42
2	300	100	44.45	47	42
3	300	100	44.66	47	42
4	300	100	44.42	47	42

Tabla 5: Consolidado de variaciones entre el código resumen de la señal manipulada utilizando el método SHA 512 respecto al código resumen de la señal original.

Tipo de modif.	**Número de variaciones**	**% de variaciones**	**Promedio**	**Valor máximo**	**Valor mínimo**
1	300	100	118.73	122	115
2	300	100	118.28	122	115
3	300	100	118.32	122	115
4	300	100	118.73	122	115

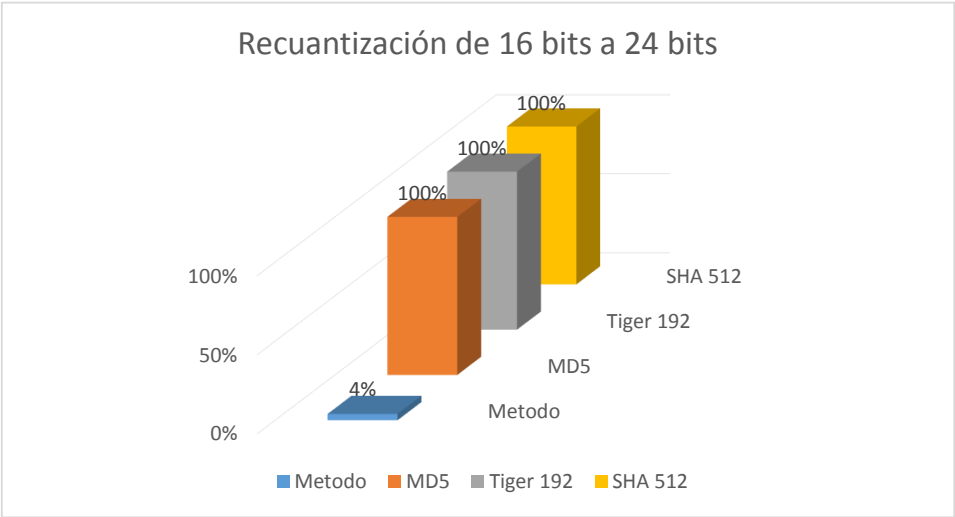

Figura 2: Diferencia (en %) entre el código resumen de la señal de voz original y el correspondiente a la señal manipulada con recuantización de 16 a 24 bits.

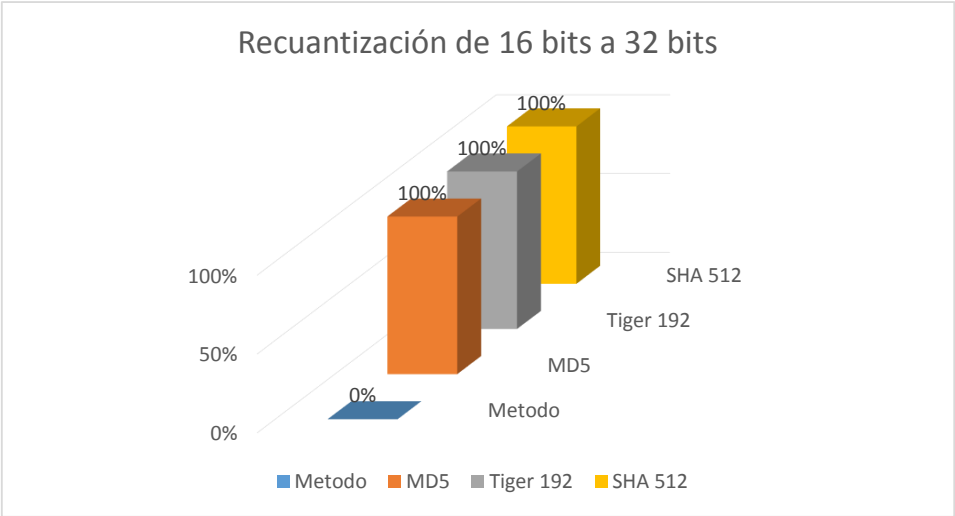

Figura 3: Diferencia (en %) entre el código resumen de la señal de voz original y el correspondiente a la señal manipulada con recuantización de 16 a 32 bits.

De manera general se puede observar en las Figuras 2 y 3, que el método propuesto da como resultado una respuesta óptima frente a las modificaciones realizadas a partir de recuantización, en comparación con los métodos MD5, Tiger 192 y SHA 512. De igual manera, y de acuerdo a las Tablas 2, 3, 4 y 5, es claro que se mantiene el grado de sensibilidad a alteraciones de contenido como lo son la adición de ruido y recorte, que corresponden a características presentes en las funciones resumen.

4 Conclusiones

En este trabajo se propone una función resumen diseñada para verificación de integridad forense en señales de voz digital. El algoritmo trabaja tomando como entrada una señal de voz que se divide en intervalos de hasta 15 segundos de duración y que entrega a su salida una cadena de 300 caracteres alfanuméricos. Sus principales fortalezas residen en el bajo costo computacional que requiere para su operación, adicionalmente tolera modificaciones de presentación de la información como lo es la recuantización, conservando las propiedades fundamentales de las funciones resumen.

Agradecimientos

Este trabajo fue financiado con recursos de la Vicerrectoría de Investigaciones de la Universidad Militar Nueva Granada bajo el proyecto IMP-ING-2136 de 2016.

Referencias

[1] N. Chen, H.-D. Xiao, and W. Wan, "Audio hash function based on non-negative matrix factorisation of mel-frequency cepstral coefficients," *IET Information Security*, vol. 5, no. 1, pp. 19–25, 2011. 168

[2] H. Malik, "Acoustic environment identification and its applications to audio forensics," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1827–1837, 2013. 169

[3] M. Ayoob and W. Adi, "Improving system reliability by joint usage of hash function bits and error correction coding," in *Emerging Security Technologies (EST), 2015 Sixth International Conference on.* IEEE, 2015, pp. 1–6. 169

[4] M. Zamani and A. B. A. Manaf, "Genetic algorithm for fragile audio watermarking," *Telecommunication Systems*, vol. 59, no. 3, pp. 291–304, 2015. 169

[5] J. Fridrich and M. Goljan, "Robust hash functions for digital watermarking," in *Information Technology: Coding and Computing, 2000. Proceedings. International Conference on.* IEEE, 2000, pp. 178–183. 169

[6] B. Lei, Y. Soon, and E.-L. Tan, "Robust svd-based audio watermarking scheme with differential evolution optimization," *IEEE transactions on audio, speech, and language processing*, vol. 21, no. 11, pp. 2368–2378, 2013. 169

[7] I. M. Maung, Y. Tew, and K. Wong, "Authentication for aac compressed audio using data hiding," in *Consumer Electronics-Taiwan (ICCE-TW), 2016 IEEE International Conference on.* IEEE, 2016, pp. 1–2. 170

[8] J. Haitsma, T. Kalker, and J. Oostveen, "Robust audio hashing for content identification," in *International Workshop on Content-Based Multimedia Indexing*, vol. 4. Citeseer, 2001, pp. 117–124. 170, 171

[9] J. Li, H. Wang, and Y. Jing, "Audio perceptual hashing based on nmf and mdct coefficients," *Chinese Journal of Electronics*, vol. 24, no. 3, pp. 579–588, 2015. 170

[10] M. S. Jain, M. V. Doshi, and M. T. Goyal, "Cryptanalytic jh and blake hash function for authentication and proposed work over blake-256 using c," *INTERNATIONAL JOURNAL OF COMPUTER TRENDS & TECHNOLOGY*, vol. 1, no. 4, pp. 1862–1866, 2013. 170

[11] R. Sobti and G. Geetha, "Cryptographic hash functions: a review," *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 2, pp. 461–479, 2012. 170, 171

[12] C. Qin, X. Chen, D. Ye, J. Wang, and X. Sun, "A novel image hashing scheme with perceptual robustness using block truncation coding," *Information Sciences*, vol. 361, pp. 84–99, 2016. 170, 171

[13] X. Lv and Z. J. Wang, "Perceptual image hashing based on shape contexts and local feature points," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1081–1093, 2012. 171

[14] A. Neelima and K. M. Singh, "A short survey on perceptual hash function," *ADBU Journal of Engineering Technology*, vol. 1, 2014. 171

[15] Z. Qiu-yu, R. Zhan-wei, X. Peng-fei, H. Yi-bo, and Y. Shuang, "Security analysis of speech perceptual hashing authentication algorithm," *International Journal of Security and Its Applications*, vol. 10, no. 1, pp. 103–118, 2016. 171

[16] H. E. Michail, A. Kotsiolis, A. Kakarountas, G. Athanasiou, and C. Goutis, "Hardware implementation of the totally self-checking sha-256 hash core," in *EUROCON 2015-International Conference on Computer as a Tool (EUROCON), IEEE.* IEEE, 2015, pp. 1–5. 172

[17] N. Chidambaram, P. Raj, K. Thenmozhi, and R. Amirtharajan, "Enhancing the security of customer data in cloud environments using a novel digital fingerprinting technique," *International Journal of Digital Multimedia Broadcasting*, vol. 2016, p. 1, 2016. 172

[18] D. Tomović, I. Ognjanović, and R. Šendelj, "Security challenges of integration of hash functions into cloud systems," in *Embedded Computing (MECO), 2015 4th Mediterranean Conference on.* IEEE, 2015, pp. 110–114. 172

[19] D. M. Ballesteros L., D. Renza, and S. Camacho, "An unconditionally secure speech scrambling scheme based on an imitation process to a gaussian noise signal," *J. Inf. Hiding Multimedia Sig. Process*, vol. 7, no. 2, pp. 233–242, 2016. 173

[20] D. M. Ballesteros L, D. Renza, and S. Camacho, "High scrambling degree in audio through imitation of an unintelligible signal," *Lecture Notes in Computer Science*, vol. 9703, pp. 251–259, 2016. 173

[21] D. M. Ballesteros L., D. Renza, and S. Camacho, "Security analysis of the speech scrambling method based on imitation of a super-gaussian signal," *J. Inf. Hiding Multimedia Sig. Process*, vol. 8, no. 1, pp. 156–167, 2017. 173, 175