

## RELACIÓN ENTRE LOS PORCENTAJES DE INTELIGIBILIDAD DE PRUEBAS DE ARTICULACIÓN Y VALORES DE STI

Sommerhoff Jorge <sup>1</sup>, Rosas Claudia <sup>2</sup>

(1) Instituto de Acústica, Universidad Austral de Chile, Valdivia, Chile.  
jsommerh@uach.cl

(2) Instituto de Lingüística y Literatura, Universidad Austral de Chile, Valdivia, Chile.  
claudiarosas@uach.cl

### Resumen

Un corpus de 1000 logatomos, de estructura consonante-vocal-consonante (CVC), agrupados en 10 listas fonéticamente balanceadas de 100 palabras cada una, correspondiente al idioma hispanoamericano, se utilizó para estudiar la relación entre los porcentajes de inteligibilidad de pruebas de articulación y valores de STI (Speech Transmission Index) medidos en condiciones acústicas deficientes o malas, es decir, de muy baja inteligibilidad. En la hipótesis que se estudia se desea determinar si para un mismo STI, medido en dos salas acústicamente diferentes pero que se caracterizan por su mala inteligibilidad ( $STI < 0,40$ ), las pruebas de articulación entregan porcentajes similares de inteligibilidad. En la primera sala, la mala inteligibilidad es preponderantemente producto de las distorsiones en el dominio del tiempo del canal de transmisión, y en la segunda sala, preponderantemente producto de las distorsiones en el dominio de la frecuencia del canal de transmisión.

Los resultados de las mediciones indican que el promedio del porcentaje de inteligibilidad de las pruebas de articulación realizados en la sala que presenta solo distorsiones en el dominio del tiempo, son mayores al promedio del porcentaje de inteligibilidad de las pruebas de articulación realizados en la sala que solo presenta distorsiones en el dominio de la frecuencia, ambas salas con un mismo STI. La brecha de esta diferencia se incrementa a medida que disminuye el STI.

**Palabras-clave:** inteligibilidad de la palabra, STI, articulación, logatomo, fonema.

### Abstract

A 1000 consonant-vowel-consonant structure logatoms corpus (CVC-structure), grouped in 10 phonetically equally balanced lists of 100 words each, was used to study the relationship between the articulation intelligibility percentages and STI (Speech Transmission Index) values measured in poor or bad acoustic conditions, ie with very low intelligibility. In the hypothesis under study it wants to be known if whether for a same STI, acoustically measured in two different rooms which are characterized by poor intelligibility ( $STI < 0.40$ ), the articulation tests deliver similar percentages of intelligibility. In the first room, poor intelligibility is predominantly due to the distortions in the time domain of the transmission channel, and in the second room, predominantly the result of distortions in the frequency domain of the transmission channel.

The results of the measurements show that the average percentage of intelligibility of the articulation tests performed in the room with only distortions in the time domain, is greater than the average percentage of intelligibility tests performed in the room that presents only distortions in the frequency domain, both rooms presenting the same STI. The difference increases with a decreasing STI.

**Keywords:** Speech Intelligibility, STI, articulation, logatom, phoneme.

**PACS no. 43.71.+m**

## 1 Introducción

Una buena inteligibilidad del habla es el objetivo principal del diseño acústico en salas donde la comunicación de voz es de suma importancia, no sólo en las aulas, salas de conferencias o auditorios, sino también en lugares como aeropuertos o centros de trabajo donde los mensajes importantes o instrucciones de emergencia deben ser transmitidos.

En términos generales, se pueden utilizar dos métodos diferentes para medir inteligibilidad de la voz; la evaluación subjetiva basada en la utilización de locutores y auditores, y la evaluación objetiva basada en los parámetros físicos del canal de transmisión. Sin embargo, el resultado de cualquier medida objetiva debe correlacionarse con el resultado de la medida subjetiva, ya que el objetivo final de la evaluación es cuantificar la cantidad de elementos del habla que son reconocidos correctamente por un oyente.

El índice de transmisión del habla (STI) es una métrica física objetiva que mide la calidad de las transmisiones de voz [1]. Esta técnica está estandarizado por la norma internacional IEC 60268-16 [2]. Como herramienta, es buen punto de referencia para comparar la calidad acústica de salas para el habla, pero según lo indicado por esta norma, es importante darse cuenta de que las bases del modelo STI siguen siendo una simplificación del proceso humano, por ello, debe conocerse su relación con los resultados estadísticos de los Tes. de inteligibilidad subjetivos, que utilizan frases, palabras o fonemas .

Los resultados de las pruebas subjetivas, llamados genéricamente Pruebas de Articulación, se interpretan como la probabilidad de transcribir correctamente los fonemas y sílabas modificados por un canal de transmisión, que en condiciones ideales alcanza un valor medio de 98,5% [3]. La prueba de la articulación no requiere que se transmita ningún significado, sin embargo, los resultados de inteligibilidad dependerá del tipo y la combinación de fonemas utilizados en el corpus de sílabas o frases [4].

En este trabajo se utilizó un corpus para el idioma hispano americano, de 1000 logatomos, agrupados en 10 listas fonéticamente balanceadas de 100 logatomos cada uno. La correlación de los resultados de las pruebas de articulación y los valores de STI medidos, se evaluó en diferentes condiciones de distorsión acústica y con un esfuerzo vocal normal del hablante masculino. Se realizaron mediciones y pruebas de articulación en dos salas con condiciones acústicas deficientes o malas, donde una sala tenía principalmente una alta distorsión de dominio de tiempo (habitación con elevado tiempo de reverberación y bajo ruido de fondo muy bajo), y la otra sala, principalmente con una alta distorsión en el dominio de la frecuencia (habitación con baja tiempo de reverberación y elevado ruido de fondo).

## 2 El STI y su relación con las pruebas de articulación

Los algoritmos originales de STI [1] y su relación con el índice de articulación [3] suponen que la predicción de la inteligibilidad del habla depende de una contribución ponderada de 7 bandas de octava. Como tal, el método STI se basa en modular en amplitud un ruido en cada una de estas bandas de octava, donde las frecuencias de modulación se eligen para que coincidan con las de habla natural. La pérdida de la modulación que se produce en el canal de transmisión representa la pérdida de modulación de la voz hablada, y por lo tanto, constituye una medida de la pérdida de la inteligibilidad [5].

La función de transferencia de modulación  $m(F)$  se la en la ecuación (1).

$$m(F) = \left[ 1 + \left( 2\pi F \frac{T}{13,8} \right)^2 \right]^{-1/2} \left[ 1 + 10^{(-S/N)/10} \right]^{-1} \quad \text{Esq.(1)}$$

donde,  $m(F)$  = factor de reducción de modulación.

$F$  = frecuencias de modulación en amplitud (0.63, 0.80, 1.0, 1.25, 1.6, 2.0, 2.5, 3.15, 4.0, 5.0, 6.25, 8.0, 10.0 y 12.5 Hz).

$S/N$  = relación señal-ruido en bandas de octava desde los 125 Hz a los 8 kHz.

$T$  = tiempo de reverberación en bandas de octava desde los 125 Hz a los 8 kHz.

Esta función describe en qué medida la modulación  $m$  se transfiere desde la fuente al receptor, donde un valor bajo de  $m$  indica una reducción significativa de inteligibilidad de la voz. Esta ecuación indica que  $m$  es dependiente de dos variables acústicas mensurables y presentes en los canales de transmisión, a saber, la relación señal a ruido  $S/N$ , que tiene en cuenta las distorsiones en el dominio de la frecuencia, y el tiempo de reverberación, que tiene en cuenta las distorsiones en el dominio del tiempo.

Es de especial interés, para evaluar la inteligibilidad de las salas realizar prueba subjetivas con logatomos de estructura CVC, dado que la pendiente de la curva correlación con las mediciones objetivas STI permite una mejor discriminación de las diferentes cualidades acústicas de las salas.

La Figura 1 muestra la relación entre las mediciones de STI y los porcentajes de inteligibilidad de una prueba utilizando logatomos de estructura CVC realizada por Houtgast y Steeneken [6].

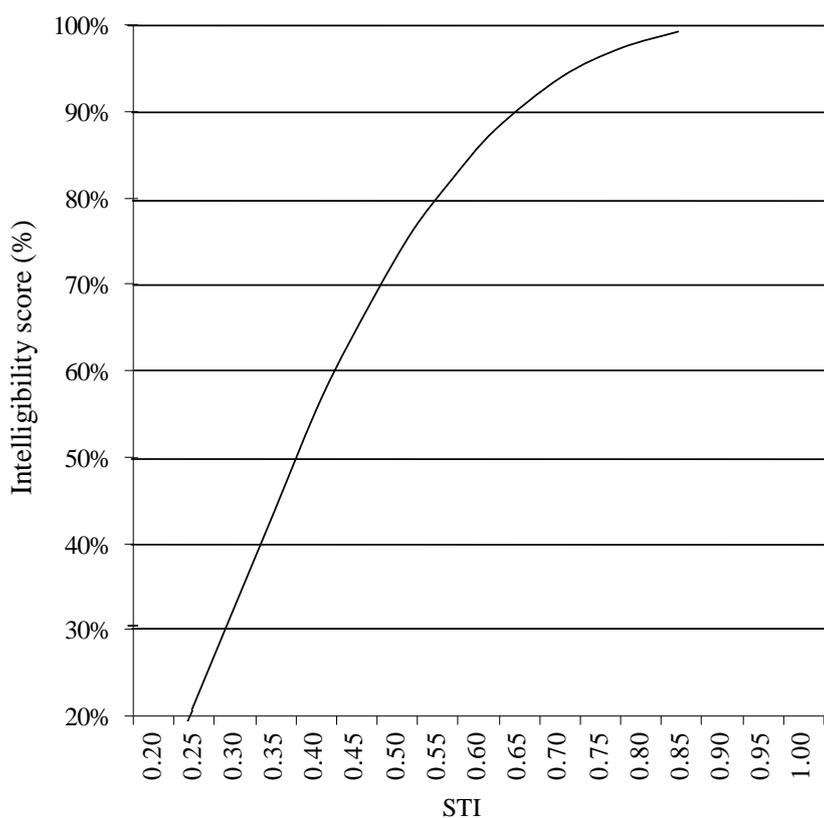


Figura 1. Relación entre las mediciones de STI y los porcentajes de inteligibilidad de una prueba que utiliza logatomos de estructura CVC

### 3 Materiales y métodos

#### 3.1 El corpus de logatomos.

La tabla 1 muestra la matriz consonantes utilizadas en el corpus [7], tanto de la primera consonante como de la segunda consonante.

Exceptuando la fila de la consonante 'ñ', la matriz está compuesta por 15 filas y 13 columnas, lo que da un total de 195 combinaciones. Si estas se combinan con las 5 vocales, da un total de 975 logatomos diferentes. Los 25 restantes se obtienen de la 'ñ' combinada con son las 5 columnas señalas en la tabla 2, y en una combinación con las 5 vocales.

Los logatomos están distribuidos en 10 listas balanceadas, de forma que la frecuencia de aparición de las distintas consonantes, tanto de la primera como de la segunda, son iguales en todas las listas. Lo mismos ocurre con las vocales. Estas listas han sido revisadas por los autores [7].

Tabla 1 – Matriz de consonantes utilizadas en la estructura CVC

		Segunda consonante del logatomo CVC																									
		b	c	ch	d	f	g	h	j	k	l	ll	m	n	ñ	p	q	r	s	t	v	w	x	y	z		
Primera consonante del logatomo CVC	b																										
	c																										
	ch																										
	d																										
	f																										
	g																										
	h																										
	j																										
	k																										
	l																										
	ll																										
	m																										
	n																										
	ñ																										
	p																										
	q																										
	r																										
	s																										
	t																										
	v																										
	w																										
	x																										
	y																										
	z																										

### 3.2 Grabación de los logatomos.

Todos los sonidos fueron articuladas de acuerdo con la norma que rige la interpretación fonética del español americano.

Todos los logatomos fueron leídos previa lectura de una frase portadora (10 frases diferentes). Las frases empleadas tenían por lo menos 4 sílabas y fueron las siguientes: "Escriban ahora", "El Siguiete es"; "Continuamos con"; "El Próximo es"; "Seguimos con"; "A continuación", "Avanzamos con"; "Escriban lo siguiente"; "El penúltimo es", "El último es". La frase portadora es importante para asegurar que todos los logatomos se pronuncian con la misma inflexión, como también, por el efecto de enmascaramiento que la reverberación de la frase de portadora produce en la primera consonante del logatomo. Los logatomos y la frase portadora se grabaron con el fin de mantener la misma voz del hablante en todas las pruebas. La voz grabada corresponde a la de un sujeto masculino. El conjunto de logatomos y frases portadoras fueron grabadas en una cámara anecoica, usando una grabadora digital Tascam HD-P2 y un micrófono condensador Audio-Technica ATM10. La razón de usar una cámara anecoica es obtener una grabación libre de ruidos y de reverberación. En la etapa de post-producción, se revisó la pronunciación y la inflexión de los logatomos, y los logatomos y frases portadoras grabados se normalizaron para producir el mismo volumen durante la reproducción, y un intervalo de silencio de 2,5 s entre cada frase (frase = frase de transición + logatomo), tiempo considerado suficiente para que un oyente anote el logatomo escuchado.

### 3.3 Grabación de los logatomos.

Como fuente de reproducción se utilizó Talkbox NTI. Este sistema electroacústico reemplaza la voz de un orador, y reproduce la voz del hablante con un volumen equivalente al esfuerzo de voz normal de un hablante masculino (60 dBA a 1 m de la boca) [2] y con una directividad  $Q = 2$ . Este sistema también se utiliza para reproducir la señal de prueba STI, en un nivel de presión sonora calibrada de 60 dBA a 1 m. En cuanto a la fuente de ruido, se utilizó un generador de ruido rosa AP600 Cesva conectado a un altavoz omnidireccional Cesva BP012.

La medición de la señal de STI se realizó por medio del instrumento NTI AL1-Acoustilyzer conectado a micrófono omnidireccional M201 de tipo I. El medidor de STI y el micrófono fueron calibrados antes de cada sesión de medición.

El montaje general se muestra en la figura 2.

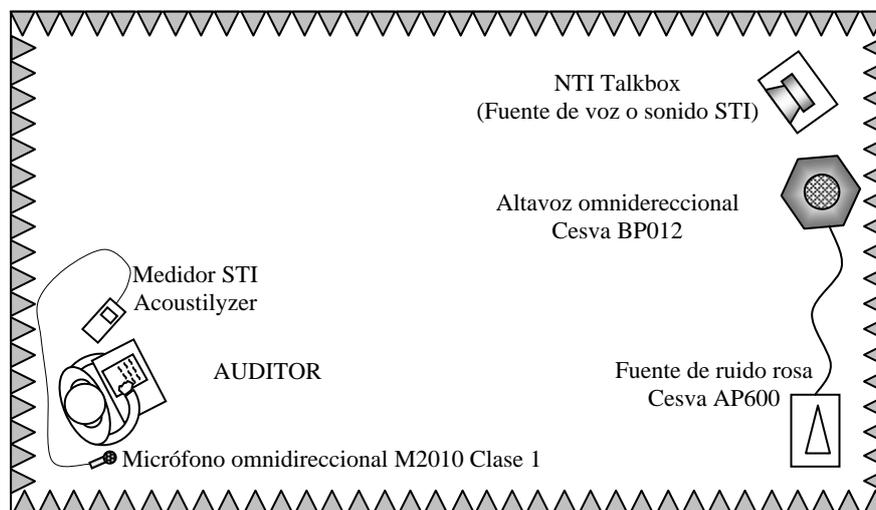


Figura 2. Montaje de instrumentos utilizados en la emisión de los logatomos y su medición.

### 3.4 Ambientes acústicos.

En las mediciones se utilizaron dos recintos con características acústicas que cubren la gama de calidad inteligibilidad, desde "mala" a "regular".

Para estas condiciones acústicas se realizaron evaluaciones en una sala con prácticamente solo distorsión el dominio del tiempo distorsión (habitación con tiempo de reverberación de alta y el ruido de fondo muy bajo), y en una sala con prácticamente solo distorsiones dominio de la frecuencia (habitación con muy tiempo de reverberación y el ruido de fondo alto). Las características de las dos salas son:

i) Sala de reverberación: Tiene un volumen de  $200 \text{ m}^3$  donde el tiempo de reverberación y ruido de fondo con el interior de una persona y los equipos de medición en ON, se muestran en las filas 2 y 3 de la Tabla 2.

Tabla 2. Tiempo de reverberación y ruido de fondo en cámara reverberante.

1	Frecuencia (Hz)	125	250	500	1k	2k	4k	8k
2	$T$ (s)	10.9	10.8	8.66	6.71	5.1	3.67	3.49
3	Ruido de fondo [dB]	17.7	18.1	18.5	19.4	20.2	20.8	20.7

El valor más bajo de STI que se obtuvo bajo estas condiciones fue de 0,27. Valores más altos de STI se lograron mediante la introducción en la habitación de placas y mantas fonoabsorbentes.

ii) Sala anecoica: Tiene un volumen de 90 m<sup>3</sup>, y el tiempo de reverberación y ruido de fondo con el interior de una persona y los equipos de medición en ON, se muestran en las filas 2 y 3 de la Tabla 3.

El nivel de presión del sonido STI y nivel de ruido rosa en el asiento del oyente para el valor más bajo medido de STI (STI = 0,20) se enumeran en las filas 5 y 6 de la Tabla 3. Se alcanzan valores más altos de STI reduciendo el nivel de presión sonora del ruido rosa en el interior de la habitación.

Tabla 3. Tiempo de reverberación y ruido de fondo en cámara anecoica.

1	Frecuencia (Hz)	125	250	500	1k	2k	4k	8k
2	Tiempo de reverberación $T$ (s)	0.10	0.09	0.07	0.06	0.06	0.05	0.04
3	Ruido de fondo [dB]	17.9	18.2	18.3	18.6	19.6	20.5	20.4
4	Nivel de presión sonora [dBA], de la señal STI en el punto de medición (4m de la fuente)	31.5	45.3	45.3	42.1	37	30	23
5	Nivel del ruido rosa [dBA] introducido a la sala para lograr un STI = 0.20	39.8	49.8	51.1	48.6	47.4	41.3	35.2
6	Relación señal ruido $S/N$ [dB]	-8.3	-4.5	-5.8	-6.5	-10.4	-11.3	-12.2

Los dos entornos acústicos se resumen en la Figura 3.

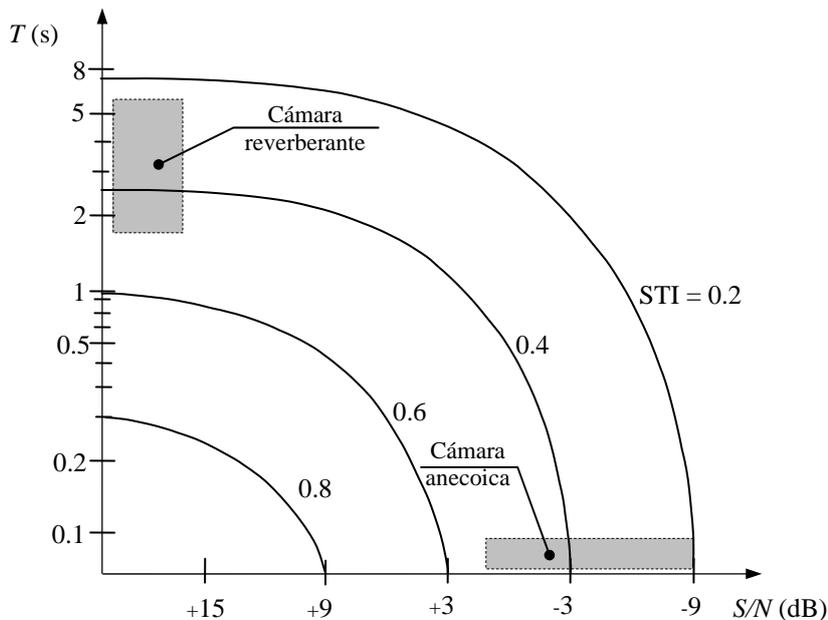


Figura 3. Familia de curvas STI en relación al tiempo de reverberación y relación señal-ruido.

En la figura 3, la familia de curvas STI están graficadas en función del tiempo de reverberación  $T(s)$  y de la relación señal-ruido  $S/N$ . Es importante destacar que para cualquier  $S/N > 15$  dB, las curvas de STI son prácticamente horizontales (especialmente con tiempo de reverberación alto), y que para cualquier  $T < 0,2$  s, las curvas de STI son prácticamente verticales (especialmente para relaciones de  $S/N$  bajos).

### 3.5 Auditores.

Diez oyentes (estudiantes universitarios), sometidos a una prueba audiométrica, fueron capacitados para responder a las palabras sin sentido como lo sugiere ANSI S3.2. [14]. Todos ellos obtuvieron puntuaciones iguales o superiores al 97% en condiciones acústicas ideales.

### 3.6 Sesiones de pruebas de articulación.

En cada sesión de pruebas de articulación, se dictaron tres listas consecutivas de 100 palabras cada una, con una pausa de aproximadamente 1 minuto entre cada lista. El tiempo total de cada sesión duró cerca de 25 minutos. Otra sesión se podía realizar después de hacer una pausa de 30 minutos.

## 4 Resultados y conclusiones.

Las curvas de correlación obtenidas en ambas salas y con distintas magnitudes de distorsión, pero siempre con valores de STI menores a 0,5, se muestran en la figura 4.

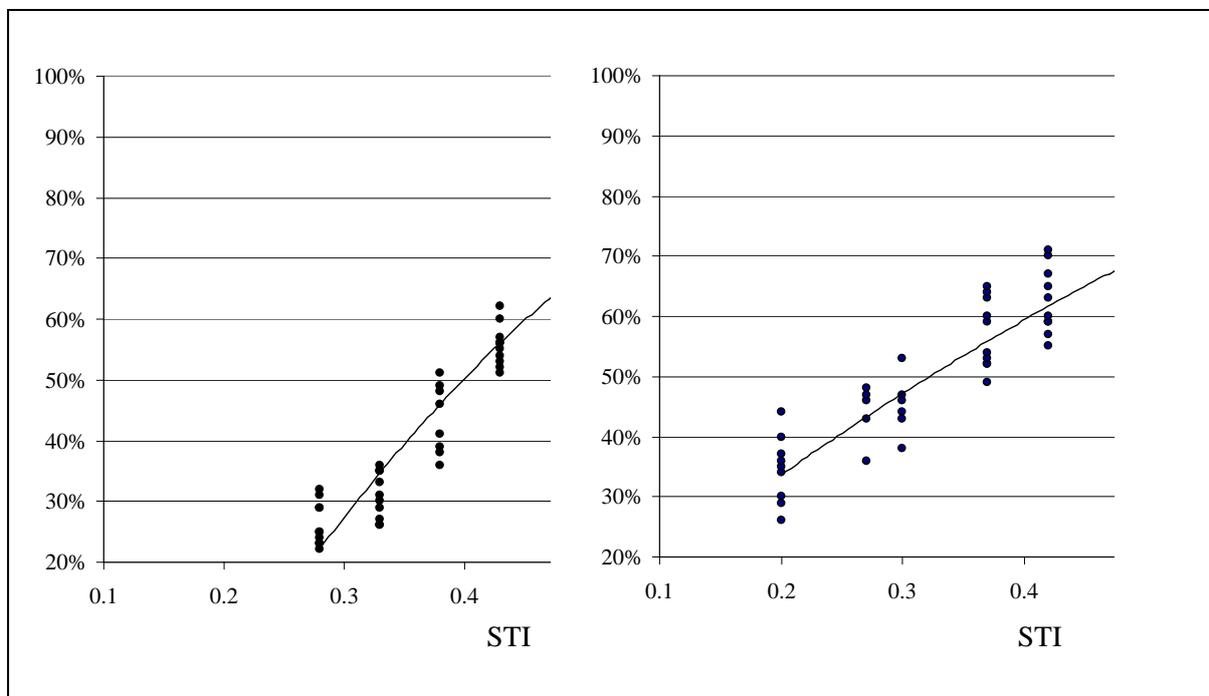


Figura 4. Resultados de las correlaciones entre pruebas de articulación y mediciones de STI en: a) sala con solo distorsión en el dominio del tiempo b) sala con solo distorsiones en el dominio de la frecuencia.

Se observa que en la sala reverberante la inteligibilidad es peor que en un sala con mucho ruido de fondo, a pesar de ambas tener un mismo STI. Por ejemplo, con un STI de 0,3 en la sala con elevada reverberación el porcentaje de inteligibilidad da en promedio 0,28% y en la sala con alto ruido de fondo, un promedio de 46%. La brecha de ambas curvas va disminuyendo a medida que el STI aumenta. Para un STI de 0,45, la curva de la sala reverberante es de 64% y para la sala con elevado ruido de fondo es de 67%.

Estos resultados inducen a pensar que la hipótesis planteada es admitida, lo que establece un camino de investigación para validarla completamente.

### **Agradecimientos**

Se agradece al Consejo de Investigación Científica y Tecnológica del Gobierno de Chile que, a través de su programa Fondecyt, financia el proyecto bajo el cual se inscribe el presente trabajo.

### **Referencias**

- [1] Steeneken, H.J.M.; Houtgast, T. A physical method for measuring speech transmission quality. *Journal of the Acoustical Society of America*, 67(1), 1980, pp. 318-326.
- [2] IEC 60268-16. Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index. *International Electrotechnical Commission*, Edition 4.0, 2011.
- [3] Fletcher, H. Speech and Hearing in Communication. *Acoustical Society of America*, Edición 1995.
- [4] Sommerhoff J., Rosas C. The evaluation of the intelligibility of speech in Spanish. *Estudios Filológicos* 42, 2007, pp. 215-225.
- [5] Steeneken, H.; Verhave, J.; McManus, S.; Jacob, K. Development of an Accurate, Handheld, Simple-to-use Meter for the Prediction of Speech Intelligibility, *Presented at ReproduceSound17*, Stratford-on-Avon, November 16th, 2001.
- [6] Houtgast, T.; Steeneken, H.J.M. The modulation transfer function in room acoustics. *Brüel & Kjaer Technical Review* N° 3, 1985, pp. 3-12.
- [7] Sommerhoff, J.; Rosas, C. Logatom corpus for the assessment of the intelligibility in Spanish speaking environments and its relation with STI measurements. *Applied Acoustics* 73, 2012, pp. 1190-1200.