

Estudio de la correlación entre STI y test de inteligibilidad subjetivo*

Study of the correlation between STI and subjective intelligibility test

Jorge Sommerhoff¹, Claudia Rosas²

¹Universidad Austral de Chile, Facultad de Ciencias de la Ingeniería. Instituto de Acústica, Valdivia, Chile. Correo electrónico: jsommerh@uach.cl

²Universidad Austral de Chile, Facultad de Filosofía y Humanidades, Instituto de Lingüística y Literatura, Valdivia, Chile. Correo electrónico: claudiarosas@uach.cl

La inteligibilidad de la palabra se puede medir por métodos subjetivos y objetivos. Sin embargo, los métodos objetivos deben tener la propiedad de poder predecir la respuesta subjetiva de acuerdo al corpus del test, a las condiciones acústicas del canal de transmisión e idioma. En este trabajo se ha determinado la correlación del método objetivo STI con la respuesta subjetiva de un corpus de logatomos de estructura CVC, confeccionado sin barreras lingüísticas para el idioma español, con el fin de cuantificar la inteligibilidad en salas destinadas a locución normal (p.e. salas de clases). Se ha determinado que en malas condiciones acústicas, ya sea por una elevada reverberación o alto ruido de fondo, un mismo valor de STI tiene dos valores distintos de correlación con la respuesta subjetiva, resultado que da una luz de alerta en la aplicación del STI en malas condiciones acústicas y que requiere mayor investigación.

Palabras clave: inteligibilidad, STI, métodos subjetivos y objetivos.

The speech intelligibility can be measured by subjective and objective methods. However, objective methods should have the property of being able to predict subjective response according to the test corpus, the acoustic conditions of the transmission channel and language. In this study, it has been established the correlation of STI objective method with the subjective response of a corpus made of CVC structure logatoms, built without language barriers for Spanish language, with the aim of quantifying the intelligibility of rooms designed for normal speech (i.e. classrooms). It has been determined that in poor acoustic conditions, due to a high reverberation or to high noise background, the same STI value has two different values of correlation with the subjective response, a result that gives a warning light in using STI in poor acoustic conditions and that needs further study.

Key words: intelligibility, STI, subjective and objective methods.

* Este artículo forma parte de los resultados parciales del proyecto de investigación Fondecyt Regular N° 1090249.

1. INTRODUCCIÓN

La inteligibilidad de la palabra es una medida de la eficacia en la comprensión de la voz hablada, que, de una forma u otra, cuantifica el porcentaje del mensaje comprendido correctamente.

Existen métodos subjetivos y objetivos de medición de la inteligibilidad de la palabra. Los métodos objetivos facilitan los procedimientos de medición por sobre los subjetivos, pero sus resultados deben finalmente concordar y poder expresarse en términos de los tests subjetivos. Luego, las curvas de correlación entre mediciones subjetivas y objetivas de inteligibilidad son importantes para determinar la utilidad y aplicación de los métodos objetivos en distintos lenguajes y condiciones acústicas de las salas.

Uno de los modelos de predicción objetivo más ampliamente utilizado es el Índice de Transmisión de la Voz (Speech Transmission Index - STI) desarrollado por Steeneken y Houtgast (1980). El STI se basa en la generación y análisis de una señal de prueba artificial modulada en amplitud que reemplaza la señal de la voz. Este método supone que la inteligibilidad del habla se mantiene con la acústica de la sala solo si la modulación de la señal artificial se transmite sin alteración desde la fuente hasta la posición del auditor. Como se expresa en la ecuación 1, tanto el tiempo de reverberación T como la relación señal ruido S/R pueden reducir esta modulación. La reducción de modulación en el punto de recepción está asociada con la pérdida de inteligibilidad y su resultado se expresa en un índice cuyo rango va de 0 a 1. La técnica STI está estandarizada por la norma IEC 60268-16 (2003).

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

donde F corresponde a una de las catorce frecuencias de modulación del método.

La Figura 1 muestra curvas de igual STI para distintas combinaciones de tiempo de reverberación y relaciones de señal/ruido. Esto significa que en una sala con bajo tiempo de reverberación y mala relación S/R el valor STI medido puede ser el mismo que el medido en una sala con alta reverberación y buena relación S/R .

Una vez que el STI se ha calculado o medido, su valor se puede convertir a una de las muchas mediciones de inteligibilidad de la palabra, pero en cada caso se debe conocer la curva de correlación que lo relaciona al corpus empleado. Por ejemplo, la Figura 2 muestra función de regresión polinomial de tercer orden que se obtuvo con los porcentajes de inteligibilidad utilizando la lista de palabras fonéticamente balanceadas de ANSI [4] y las correspondientes mediciones de STI, curva análoga a la informada por Steeneken y Houtgast (1989).

Un aspecto importante cuando se trata de una prueba de inteligibilidad con voz hablada es el corpus utilizado como material de lectura. Los resultados de las investigaciones realizadas por Hearcom (2005) describen y seleccionan el material disponible para realizar pruebas de inteligibilidad en siete idiomas: danés, holandés, inglés (británico), francés, alemán, polaco y sueco. Los corpus de palabras consideradas en esta publicación están basados en oraciones, palabras monosilábicas (consonante-vocal-consonante - CVC), o tripletas de números. Para el idioma español hay al menos dos corpus de 100 palabras cada uno (Fuchs y Osuna 1965; Miñana 1969) y fueron estudiados por Sommerhoff y Rosas (2007).

Figura 1
Curvas de contorno de igual STI para diferentes relaciones S/R y tiempos de reverberación T

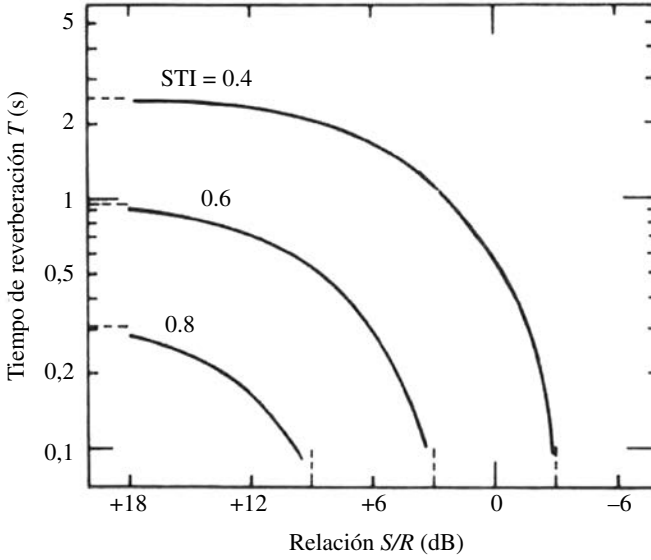
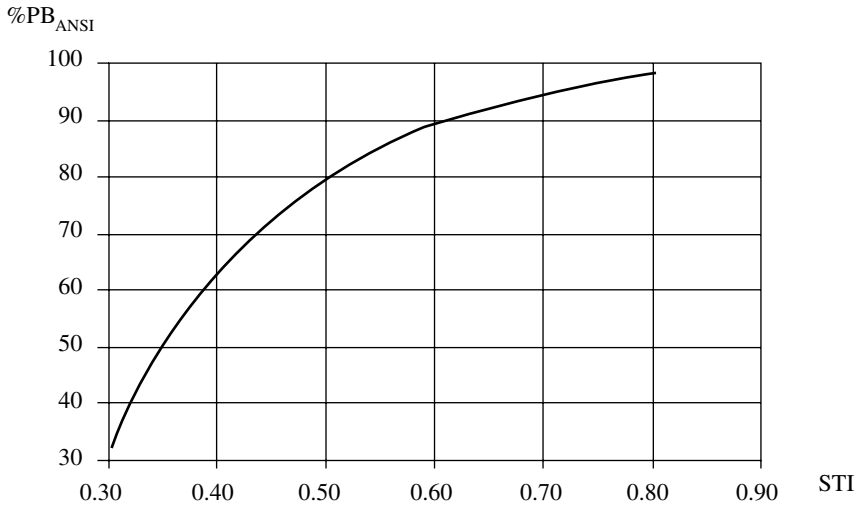


Figura 2
Función de regresión polinomial de tercer orden de 50 puntajes de porcentaje de inteligibilidad utilizando las listas de palabras fonéticamente balanceadas de ANSI y las correspondientes mediciones de STI



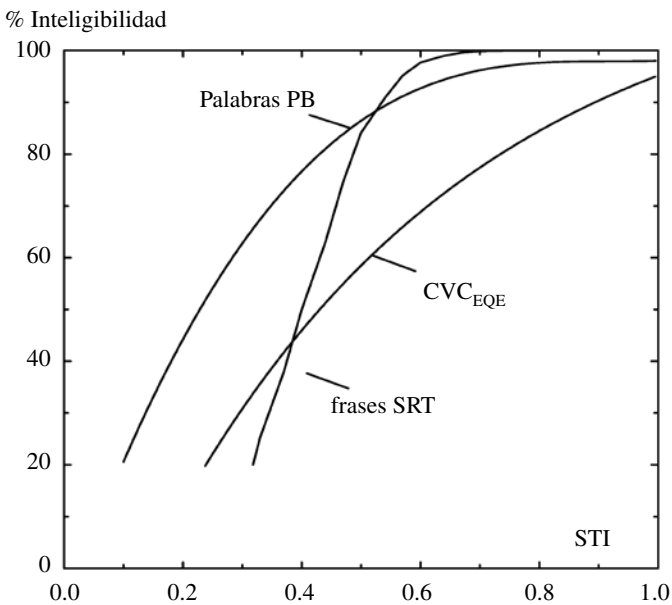
$$\%PB_{ANSI} = 788,27 STI^3 - 1643,9 STI^2 + 1179,3STI - 196,3 \quad (2)$$

La Figura 3 muestra las relaciones entre el STI y el porcentaje de inteligibilidad de las listas de palabras fonéticamente balanceadas PB de la conocida lista de Harvard (palabras en inglés), listas de frases utilizadas en los tests de umbrales de recepción del habla (Speech Reception Threshold SRT), y listas de palabras sin sentido de estructura CVC con igual distribución de fonemas holandeses utilizada por Steeneken y Houtgast (1980). De estas curvas se desprende que la prueba subjetiva más útil para calificar la calidad de una sala para la voz hablada (p.e. sala de clases) corresponde al test de palabras monosilábicas de composición CVC, dado que el porcentaje de inteligibilidad cambia en cada categoría de calidad, permitiendo diferenciar salas “excelentes” de salas con categoría “buenas”. De acuerdo con la norma ISO 9921:2002 (2003), los intervalos de clasificación de la inteligibilidad se dan en la Tabla 1.

Tabla 1
 Clasificación de la inteligibilidad de acuerdo a la norma ISO 9921 para resultados de pruebas de inteligibilidad subjetiva utilizando palabras sin sentido de estructura CVC, como también para pruebas de inteligibilidad objetiva con el método STI

Categoría	Excelente	Buena	Regular	Deficiente	Mala
CVC	> 81%	70% a 81%	53% a 70%	31% a 53%	< 31%
STI	> 0,75	0,60 a 0,75	0,45 a 0,60	0,30 a 0,45	< 0,30

Figura 3
 Relaciones entre el STI y el porcentaje de inteligibilidad de las listas de palabras fonéticamente balanceadas PB de la conocida lista de Harvard, listas de frases utilizadas en los tests de umbrales de recepción del habla SRT y listas de palabras sin sentido de estructura CVC con igual distribución de fonemas holandeses utilizada por Steeneken y Houtgast



Rosas y Sommerhoff (2008) estudiaron el comportamiento de un corpus de 1200 logatomos en español de estructura CVC, en listas fonéticamente balanceadas y correlacionadas con el STI. Se concluyó la importancia de las consonantes en el porcentaje de error, especialmente de la consonante posterior a la vocal (implosiva). También, en este trabajo se infieren distintas curvas de correlación obtenidas por la manipulación del corpus de 1200 logatomos (exclusión sucesiva de los logatomos más difíciles de escuchar). Independientemente de las características acústicas de la sala en que se cuantifica la inteligibilidad, estas experiencias indican que la variación de los resultados es dependiente de los fonemas que constituyen la estructura CVC utilizada, como también producto de barreras lingüísticas por la utilización de combinaciones CVC perceptivamente confusas (p.e. ñich, o yep/llep).

La Tabla 2 muestra las estructuras posibles del español con sus correspondientes porcentajes de frecuencia (Cfr. Quilis 1999).

Tabla 2
Estructuras silábicas del español ordenadas según frecuencia

Fonema	Fonema	Fonema	Fonema	Fonema	Frecuencia
	C	V			55,81%
	C	V	C		21,61%
		V			9,91%
		V	C		8,39%
C	C	V			3,14%
C	C	V	C		0,98%
		V	C	C	0,13%
	C	V	C	C	0,02%
C	C	V	C	C	0,01%
Total					100%

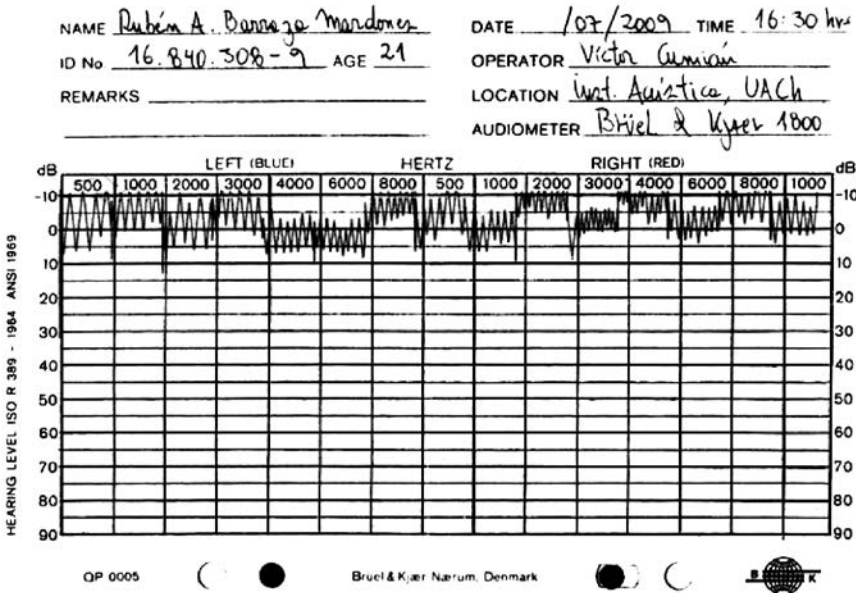
A partir de estas experiencias en esta investigación se estudia el comportamiento de un corpus de logatomos CVC que no contenga fonemas o combinación de fonemas que distorsionen los resultados de inteligibilidad producto de sus barreras lingüísticas o fonoaudiológicas que permita cuantificar las cualidades acústicas de una sala para la transmisión de la voz hablada.

2. MATERIALES Y MÉTODOS

2.1. AUDITORES

Para la investigación se utilizaron como auditores 10 jóvenes universitarios. Todos fueron instruidos de acuerdo a las recomendaciones de la norma ANSI S3.2 [4] y a todos se les aplicó una audiometría para establecer que sus condiciones auditivas fuesen normales. El trabajo realizado por ellos en las audiciones fue cancelado con recursos asignados al proyecto.

Figura 4
Ejemplo de una audiometría.



2.2. GRABACIÓN Y EMISIÓN DE LOS LOGATOMOS

Para mantener inalterable la emisión de los logatomos en las distintas pruebas y condiciones acústicas, la totalidad del corpus fue grabado y procesado para ser reproducido por una fuente de sonido que reemplaza la voz hablada.

Locutor. El locutor debe ser una persona con las competencias para dictar en forma fonéticamente correcta los logatomos. Lo anterior justificó que el corpus fuese leído por un miembro del equipo de investigación y familiarizada con el mismo.

Grabaciones. Las grabaciones se llevaron a cabo en la cámara anecoica del Instituto de Acústica de la Univ. Austral de Chile, con un grabador Tascam HD-P2 y un micrófono Audio-technica ATM10, con una distancia de 15 cm a 20 cm entre la locutora y el micrófono. El objetivo de utilizar como sala de grabación una sala anecoica es obtener un registro sonoro sin ruido y libre de reverberación. Las lecturas se realizaron a mediodía o media tarde, dado que muy temprano en la mañana las cuerdas vocales están flojas y muy tarde cansadas, las que pueden producir efectos glotales que no conviene que se mezclen con el logatomo.

Los logatomos grabados fueron normalizados, es decir, se igualaron al máximo el volumen promedio de cada uno de ellos, todo ello con el fin de producir una sonoridad semejante en su reproducción. También, en cada lista, los logatomos fueron grabados con un intervalo de silencio entre sí de dos segundos.

Como elemento de reproducción se utilizó un “Talkbox” NTI, elemento electroacústico que reemplaza la emisión de la voz humana, reproduciendo la señal con un nivel correspondiente a la voz normal (60 a 65 dB a un metro) y una directividad $Q = 2$.

2.3. CONFECCIÓN DEL CORPUS

El corpus se diseñó con el criterio de incluir todas las combinaciones CVC del español hispanoamericano que no distorsionen los resultados de inteligibilidad producto de sus barreras lingüísticas. Luego, está formado por logatomos del español hispanoamericano de estructura silábica CVC, que considera todas las posibilidades combinatorias de los fonemas, exceptuando las consonantes iniciales (CVC) y las consonantes finales (CVC) que a continuación se señalan y justifican:

I) *Consonantes iniciales*

- i) La **c** y **q** porque se articulan de la misma forma que la **k**.
- ii) La **h** porque no se articula.
- iii) La **x** porque se articula de la misma forma que la **j** (Jimena, Ximena).
- iv) La **v** porque se articula de la misma forma que la **b**.
- v) La **y** porque se articula de la misma forma que la **ll** (elle).
- vi) La **z** porque se articula de la misma forma que la **s**.

II) *Consonantes finales*

- i) En todos los casos donde existe neutralización [12] se ha elegido solo una de las consonantes que forma la oposición: **b** por **p**; **d** por **t**; **g** por **k**, **c** por **q**; **n** por **m** (ejemplos: captura → cabtura, Bagdad → Bagdat, Bagdad → Bakdad, cambio → canbio).
- ii) La **h** porque no se articula.
- iii) La **v** porque se articula de la misma forma que la **b**.
- iv) La **w** porque se articula de la misma forma que la **b** o **g** (wagneriano, washingtoniano).
- v) La **x** porque es la combinación de dos consonantes (fax o faks).
- vi) La **y** porque se articula de la misma forma que la **ll** (elle).
- vii) La **z** porque se articula de la misma forma que la **s**.
- viii) La **ñ** por su dificultad de pronunciarla como consonante final e inexistencia en esta ubicación en el lenguaje.

Con relación a la pronunciación de la consonante **r** se consideró utilizar la vibrante múltiple **rr** en la primera consonante (como por ejemplo en marrón) y la vibrante simple **r** en la segunda consonante (como por ejemplo en tomar).

Figura 5

Combinaciones de consonantes utilizadas en el corpus de logatomos con estructura CVC.

El cuadro a) muestra todas las consonantes que quedan después de consideraciones lingüísticas. El cuadro b) es semejante al cuadro a) excepto por la eliminación de la fila ñ y la columna ll, ambas retiradas por disminuir el puntaje de inteligibilidad en condiciones acústicas de cien por ciento de inteligibilidad

a)

		Consonante final																									
		b	c	ch	d	f	g	h	j	k	l	ll	m	n	ñ	p	q	r	s	t	v	w	x	y	z		
Consonante inicial	b																										
	c																										
	ch																										
	d																										
	f																										
	g																										
	h																										
	j																										
	k																										
	l																										
	ll																										
	m																										
	n																										
	ñ																										
	p																										
	q																										
	r																										
	s																										
	t																										
	v																										
	w																										
	x																										
	y																										
	z																										

b)

		Consonante final																									
		b	c	ch	d	f	g	h	j	k	l	ll	m	n	ñ	p	q	r	s	t	v	w	x	y	z		
Consonante inicial	b																										
	c																										
	ch																										
	d																										
	f																										
	g																										
	h																										
	j																										
	k																										
	l																										
	ll																										
	m																										
	n																										
	ñ																										
	p																										
	q																										
	r																										
	s																										
	t																										
	v																										
	w																										
	x																										
	y																										
	z																										

Como resultado de las consideraciones anteriores, la matriz de combinaciones de consonantes se muestra en la Figura 5 a), donde cada cuadro gris indica una combinación aceptada.

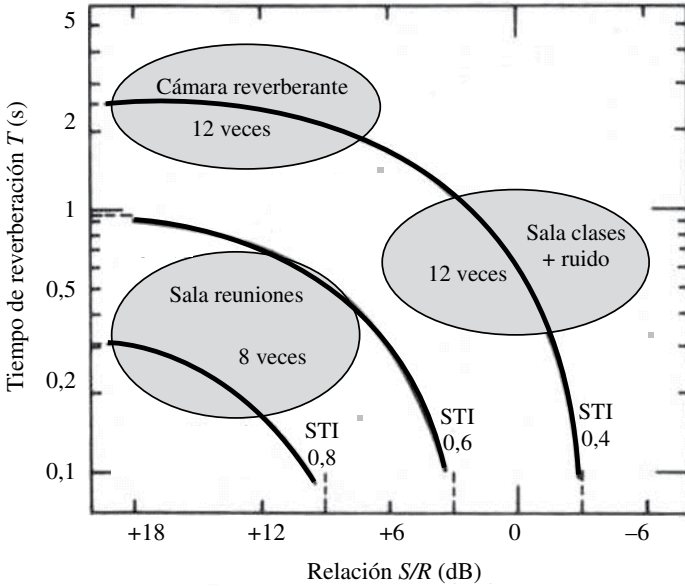
Esta matriz de combinaciones se puso a prueba en condiciones acústicas de cien por ciento inteligibilidad, reproduciendo la lista grabada a los auditores a través de audífonos. Dado que los resultados bajo estas condiciones fueron en general algo inferiores al cien por ciento, se detectó que la causa de ello era producto de las consonantes **ñ** en posición inicial y **ll** en posición final que en forma natural producían confusión al ser escuchadas en una estructura CVC. Por este motivo, para tener un corpus de respuesta subjetiva cercana al cien por ciento en condiciones acústicas ideales (o casi ideales), se optó por eliminar esta fila y columna quedando como matriz final la Figura 5 b). En su aplicación se espera que los errores en su escritura sean producto solo de las deficiencias acústicas de la sala.

El total de palabras que se conforman con las combinaciones de la matriz b) y las cinco vocales es de 750 logatomos. Con ellos se confeccionaron tres listas distintas de 750 logatomos, cada una compuestas de 15 sublistas de 50 logatomos.

2.4. CONDICIONES ACÚSTICAS PARA LAS MEDICIONES SUBJETIVAS Y MEDICIÓN DE STI

Para las mediciones se requerían salas con características acústicas que cubrieran todo el rango de calidad de acuerdo a su inteligibilidad, es decir, desde condiciones de sala “mala” a sala “excelente”. De acuerdo a la ecuación 1, las condiciones de inteligibilidad “mala” se pueden conseguir en salas con mucha reverberación, en salas con baja relación señal ruido, o con ambas condiciones. Para determinar si el resultado de la prueba subjetiva era estadísticamente el mismo en condiciones acústicas “regulares” a “malas”, se decidió separar la correlación entre STI y medición subjetiva en condiciones altamente reverberantes y correlación entre STI y medición subjetiva en condiciones de baja relación *S/R*. Para cumplir con estos requerimientos se utilizaron tres salas: una sala de reuniones para medir STI en condiciones acústicas “buenas” a “excelentes”, una cámara reverberante (200 m³) para medir STI en condiciones “regulares” a “malas” producto solo de un alto tiempo de reverberación, una sala de clases a la cual se le adicionó ruido rosa para medir STI en condiciones “regulares” a “malas” producto solo de una baja relación *S/R*. La Figura 6 ilustra las condiciones acústicas de las salas donde también se indica el número de veces que se dictaron los 750 logatomos en ellas.

Figura 6
Curvas de igual STI donde el área de las elipses superpuestas representa las condiciones acústicas de relación señal-ruido S/R y tiempos de reverberación T de las salas donde se realizaron las mediciones



2.5. AUDICIONES Y MEDICIÓN DE STI

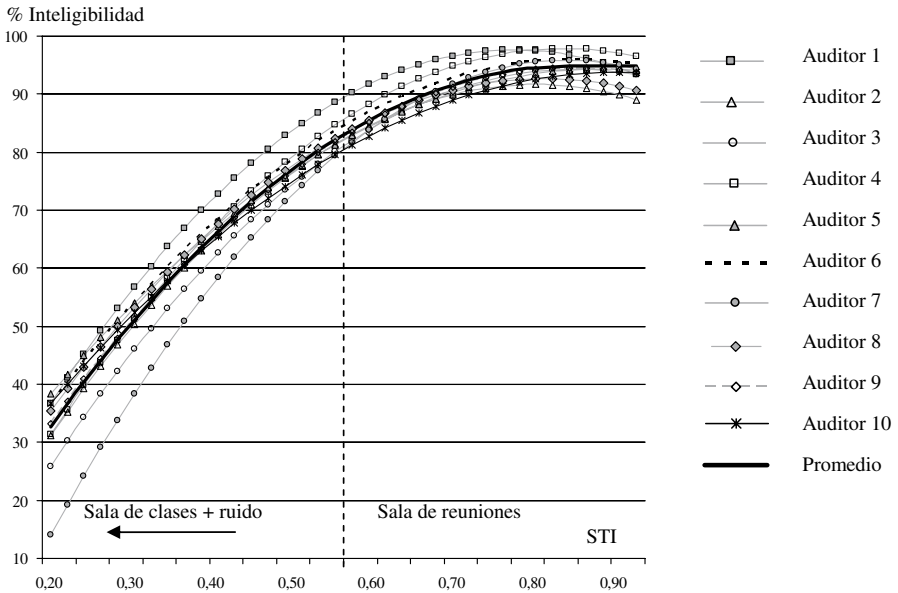
Cada auditor escuchó los 750 CVC logatomos 32 veces; 12 veces en cámara reverberante, 12 veces en sala de clases con ruido y 8 veces en sala de reuniones. Los 750 logatomos se ordenaron en sublistas de 50 logatomos, las que se dictaron con un descanso entre ellas. En cada asiento se midió la inteligibilidad con un medidor de STI de marca NTI, valor que se asociaba a cada una de las hojas de la prueba de audición.

3. RESULTADOS

La Figura 7 muestra las curvas de correlación entre el promedio del % de Inteligibilidad de la respuesta subjetiva de cada auditor con relación al STI medido en cada puesto. El rango del STI en condiciones acústicas “regulares” a “malas” en la sala corresponde a las mediciones realizadas en la sala de clases con ruido adicionado. La Figura 8 muestra el conjunto de pares ordenados (STI, % Inteligibilidad) correspondientes a los resultados de las mediciones todos los auditores bajo las mismas condiciones acústicas de la Figura 7. La curva de correlación de estos puntos está dada por el polinomio de segundo orden de la ecuación 2, la cual es igual a la curva promedio de la Figura 7. La desviación estándar es de 4,43%.

Figura 7
 Curvas de correlación entre el promedio del % de Inteligibilidad de la respuesta subjetiva de cada auditor con relación al STI donde el rango del STI en condiciones acústicas “regulares” a “malas” corresponde a las mediciones realizadas en la sala de clases con ruido adicionado

$$\% \text{ Inteligibilidad} = -160,97 \text{ STI}^2 + 264,37 \text{ STI} - 13,72 \quad (2)$$



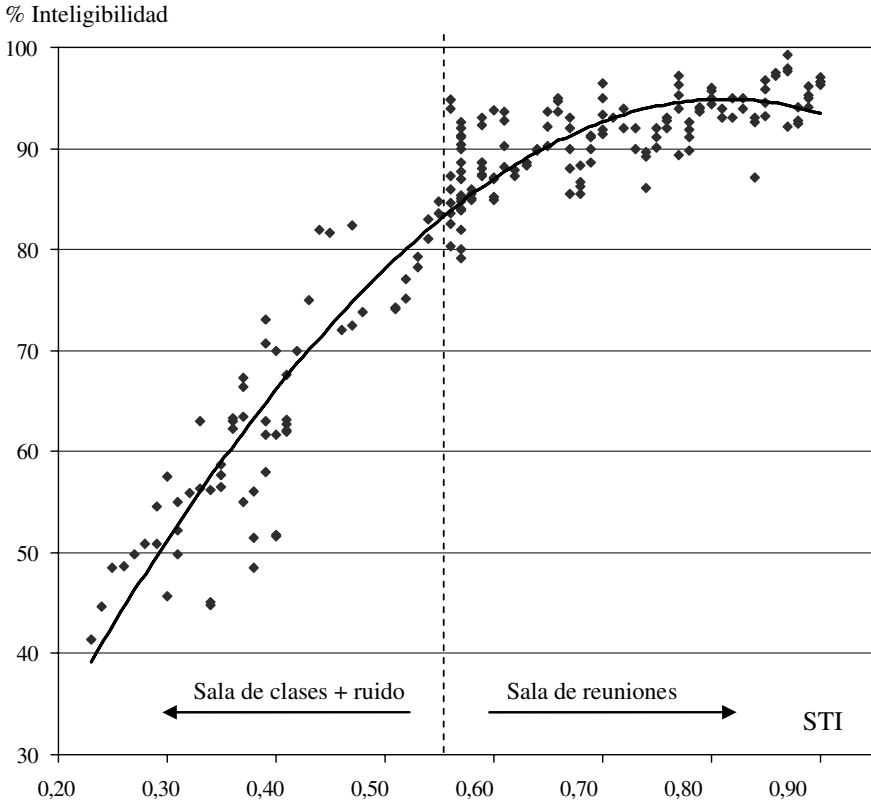
La Figura 9 muestra las curvas de correlación entre el promedio del % de Inteligibilidad de la respuesta subjetiva de cada auditor con relación al STI medido en cada puesto. El rango del STI en condiciones acústicas “regulares” a “malas” en la sala corresponde a las mediciones realizadas en la cámara de reverberación. La Figura 10 muestra el conjunto de pares ordenados (STI, % Inteligibilidad) correspondientes a los resultados de las mediciones todos los auditores bajo las mismas condiciones acústicas de la Figura 9. La curva de correlación de estos puntos está dada por el polinomio de segundo orden de la ecuación 3, la cual es igual a la curva promedio de la Figura 9. La desviación estándar es de 5,68%.

$$\% \text{ Inteligibilidad} = -64,67 \text{ STI}^2 + 131,08 \text{ STI} + 30,56 \quad (3)$$

Comparando la distribución de puntos de la Figura 8 con la Figura 10, se observa que para un mismo STI medido en condiciones reverberantes la audición es mejor que el medido en condiciones de ruido, existiendo eso sí mayor dispersión entre sujetos en la medición subjetiva en condiciones reverberantes que en condiciones de ruido.

La Figura 11 muestra una comparación de las curvas de correlación de las Figuras 8 y 10 y de la curva CVC de la Figura 3. Ello nos indica que el corpus elaborado de

Figura 8
Conjunto de pares ordenados correspondientes a los resultados de las mediciones todos los auditores donde el rango del STI en condiciones acústicas “regulares” a “malas” corresponde a las mediciones realizadas en la sala de clases con ruido adicionado

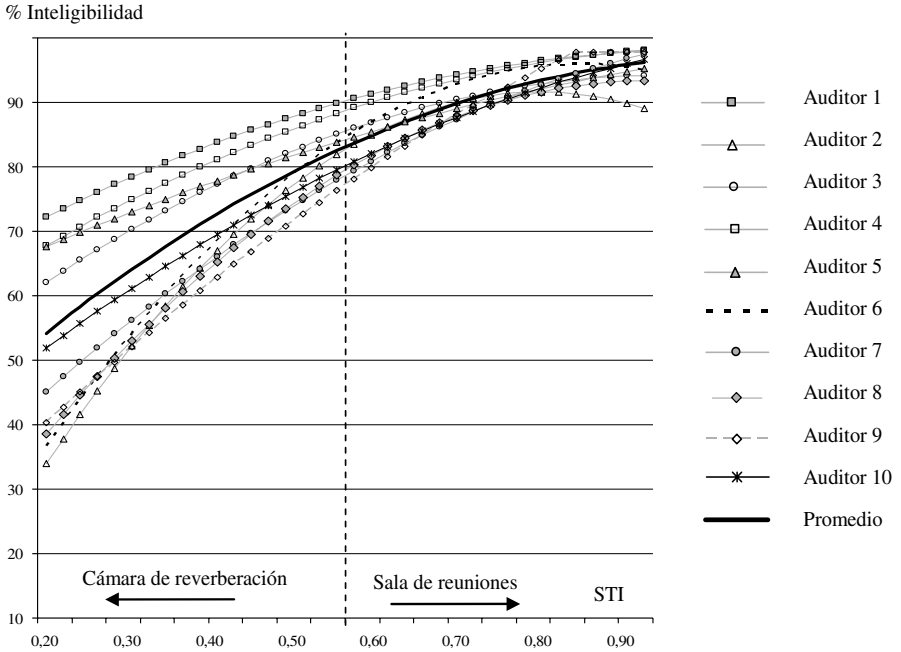


acuerdo a los criterios establecidos en este documento se escucha con mayor claridad que el corpus de Steeneken y Houtgast en condiciones acústicas adversas. También, para un mismo STI se ha encontrado una diferencia en la medición subjetiva en el rango de condiciones acústicas “regulares” a “malas” en condiciones reverberantes y condiciones con ruido, resultado que se investigará con mayor detención en el futuro.

4. CONCLUSIONES

Se construyó un corpus de 750 logatomos de estructura CVC para cuantificar subjetivamente la calidad acústica de una sala para transmitir la voz hablada normal. Los resultados de las mediciones de inteligibilidad realizadas con este corpus en

Figura 9
 Curvas de correlación entre el promedio del % de Inteligibilidad de la respuesta subjetiva de cada auditor con relación al STI donde el rango del STI en condiciones acústicas “regulares” a “malas” corresponde a las mediciones realizadas en la cámara de reverberación



condiciones acústicamente deficientes producto de la reverberación se separaron de los resultados de las mediciones de inteligibilidad realizadas en condiciones acústicamente deficientes producto de la baja relación señal/ruido. Las curvas que correlacionan estos dos conjuntos de datos a la medida objetiva STI difieren, lo que indicaría que la medida STI se debe utilizar con precaución al momento de evaluar salas con mucha reverberación o con mucho ruido. Estos resultados dan lugar a una revisión más profunda del comportamiento de la relación del STI y respuesta al test subjetivo en los rangos de baja inteligibilidad, como también, a la posibilidad de calificar las salas de acuerdo a la Tabla 1 con este corpus. La desviación estándar es mayor para el caso de la aplicación del corpus en condiciones reverberantes que en condiciones con baja relación señal/ruido.

Figura 10
Conjunto de pares ordenados correspondientes a los resultados de las mediciones todos los auditores donde el rango del STI en condiciones acústicas “regulares” a “malas” corresponden a las mediciones realizadas en la cámara de reverberación.

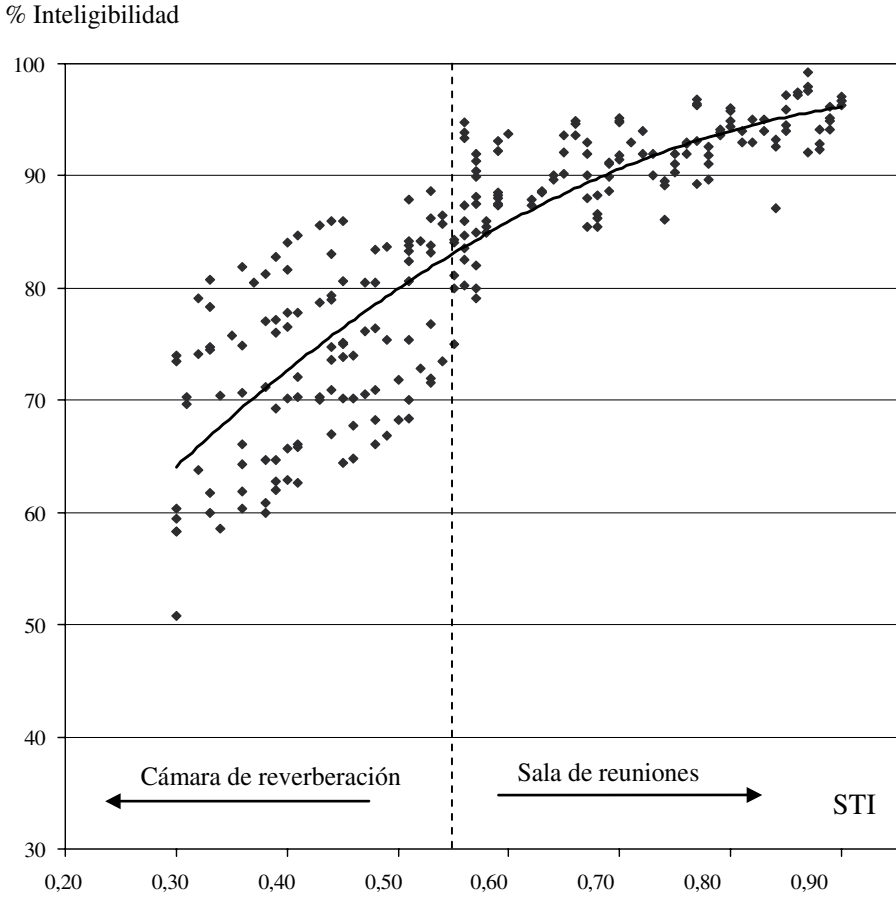
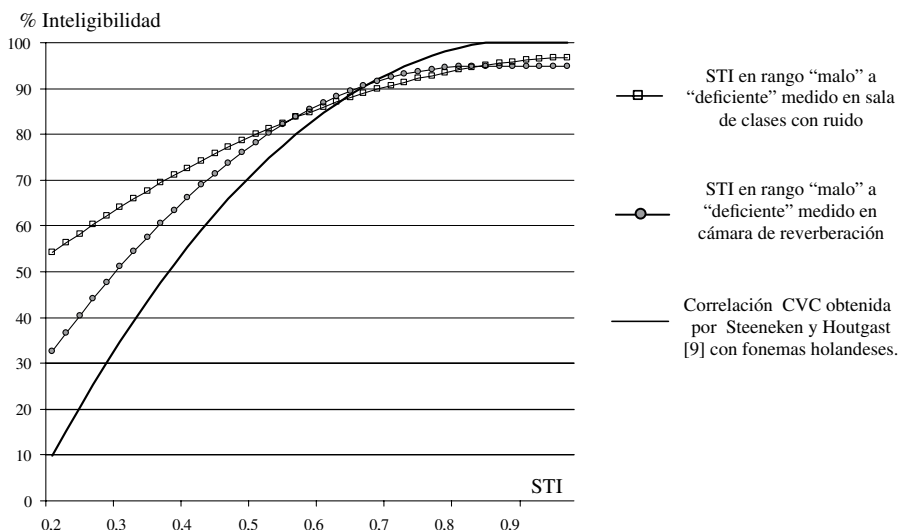


Figura 11
Comparación de las curvas de correlación de las Figuras 8 y 10,
y de la curva CVC de la Figura 3



OBRAS CITADAS

- Steeneken, H.J.M y T. Houtgast. 1980. "A physical method for measuring speech transmission quality". *J. Acoust. Soc. Amer* 67 N° 1.
- Sound system equipment – Part 16: Objective rating of speech intelligibility by speech transmission index. Geneva, Switzerland: International Electrotechnical Commission, International Standard IEC 60268-16; 2003.
- American National Standards Institute. 1989. "Method for Measuring the Intelligibility of Speech over Communication Systems", ANSI S3.2.
- Hearing in the Communication Society. 2005. Speech recognition tests for different languages, FP6-004171 HEARCOM.
- Fuchs, G. L. y J. Osuna . 1965. "Medición de inteligibilidad". *Memoria de las Primeras Jornadas Latinoamericanas de Acústica*. Universidad Nacional de Córdoba.
- Miñana, P. 1969. *Compendio práctico de acústica*. Barcelona: Labor.
- Sommerhoff, J. y C. Rosas. 2007. "Evaluación de la inteligibilidad del habla en español". *Estudios Filológicos* 42: 215-225.
- Steeneken, H.J.M. y T. Houtgast. 2002. "Validation of the STIr method with the revised model". *Speech Communication* 38: 413-425.
- International Standard Organization. "Ergonomics - Assessment of speech communication", ISO 9921:2003
- Rosas, C. y J. Sommerhoff. 2008. "Inteligibilidad acústica en español: una propuesta para su medición". *Estudios Filológicos* 43: 179-190.
- Quilis, Antonio. 1999. *Tratado de fonología y fonética españolas*. Madrid: Gredos.

