

# MÉTODO PARA LA EVALUACIÓN OBJETIVA DE LA INTELIGIBILIDAD EN SALAS DE PALABRA "IAIS": PLANTEAMIENTO (PARTE I)

REFERENCIA PACS: 43.71.Gv

Yebra Calleja, M<sup>a</sup> Soledad; González Etxabe, Iosu; Vera Guarinos, Jenaro  
Departamento de Física, Ingeniería de Sistemas y Teoría de Señal.  
Universidad de Alicante.  
Escuela Politécnica Superior de Alicante  
Edif.: Politécnica II  
Campus de San Vicente del Raspeig  
Apdo: 99.  
03080-Alicante  
Tlf: 965 90 - 9756 / 9751  
Fax: 965 90 9750  
Email: [jenaro@disc.ua.es](mailto:jenaro@disc.ua.es); [myebra@dfists.ua.es](mailto:myebra@dfists.ua.es)

## ABSTRACT

The 'Articulation Index' (**AI**) is an objective method for intelligibility evaluation in transmission address of dedicated speech electroacoustic systems. Extrapolating this method to no amplified natural speech case, binaural recordings in a daily teaching situation for different speakers in several classrooms, are carried out. Signal/Noise ratio like the difference of the L10 and L90 percentil is chosen now. Starting from these registrations, an **In Situ Articulation Index (IAIS)** settles down. This objective parameter takes account in a natural way of those subjective aspects that occur in realistic situations: speaker, enclosure and listeners.

## RESUMEN

El 'Índice de Articulación' (**AI**) es un método objetivo para la evaluación de la inteligibilidad en sistemas de transmisión de información electroacústicos. Se plantea extrapolar dicho método al caso de habla natural no amplificada. Se realizan grabaciones binaurales de distintos oradores en varias aulas en una situación de docencia cotidiana. Se elige como relación Señal/Ruido la diferencia de los percentiles L10 y L90. Se establece un Índice de Articulación a partir de estos registros **In Situ (IAIS)**. Dicho parámetro refleja de forma objetiva aquellos aspectos subjetivos que se dan de forma natural en una situación real: orador, recinto y oyentes.

## INTRODUCCIÓN

En este trabajo se va a plantear un método para la evaluación objetiva de la inteligibilidad de la palabra, basado en el conocido "Índice de Articulación" (**AI**), al que se le denomina por analogía "Índice de Articulación In Situ" (**IAIS**).

## OBJETIVOS

Establecer qué parámetros se pueden extraer de la actividad sonora que se produce en el transcurso de una clase normal, de modo que su valoración esté directamente ligada a la comprensión verbal. Estudiar el peso que le corresponde a cada una de las variables que intervienen en el proceso de comunicación oral: Tipo de orador, características acústicas de la sala, comportamiento del auditorio.

## METODOLOGÍA

El carácter práctico y realista que quiere mostrar el método **IAIS**, sólo se puede conseguir con medidas en tiempo real; es necesario trabajar con registros sonoros obtenidos 'in situ'. Por lo que se decide grabar en varias aulas con distintos oradores y oyentes, a partir de éstas se obtiene de forma excluyente el nivel del habla y del ruido. Éste procedimiento es uno de los pilares fundamentales de nuestro proceso.

Las manipulaciones posteriores se afrontan siguiendo las indicaciones establecidas por la norma que es el germen de nuestra idea: ANSI Standard S3.5-1969 para el cálculo del Índice de Articulación (AI).

Se usaron dos configuraciones para el sistema de registro de la señal:

Registro monoaural: Sonómetro (S-NA 27 de Rion Co.) como transductor y acondicionador de la señal, junto con un ordenador portátil como dispositivo de almacenamiento.

Registro binaural: Torso y cabeza artificial (HMS III.0 de Head acoustics GMBH); que digitaliza directamente la señal, lo que se aprovecha para conectarla a un DAT (DAP1 de TASCAM) que realiza la tarea del almacenamiento.



Fig. 1: Grabación in situ real

El primer paso es decidir los puntos en los que se van a realizar las grabaciones. De todos los posibles, siempre existirá uno fijo que se utiliza para caracterizar al orador y se toma como referencia: frente al orador y a una distancia de  $1\div 1,5$  m. Los demás puntos se distribuyen conforme a la geometría de cada aula: entre 8 y 13 puntos para aulas de  $124\div 180$  m<sup>2</sup>.

Es sabido que en situaciones reales, el habla va acompañada siempre de ruido de fondo ajeno (ambiente) y reverberación (que a su vez forma parte del mensaje y en gran medida se adiciona al ruido) siendo éste uno de los mayores problemas para la inteligibilidad de la palabra. Por otro lado, atendiendo a la estructura temporal se observó que en los registros sonoros se podía reconocer lo que es habla y ruido por su huella energética (véase Fig. 2). De manera genérica el ruido de fondo se muestra como un murmullo casi continuo, sin apenas interrupciones. En cambio el habla, con mayor nivel habitualmente, aparece de manera interrumpida y con picos continuos de muy corta duración.

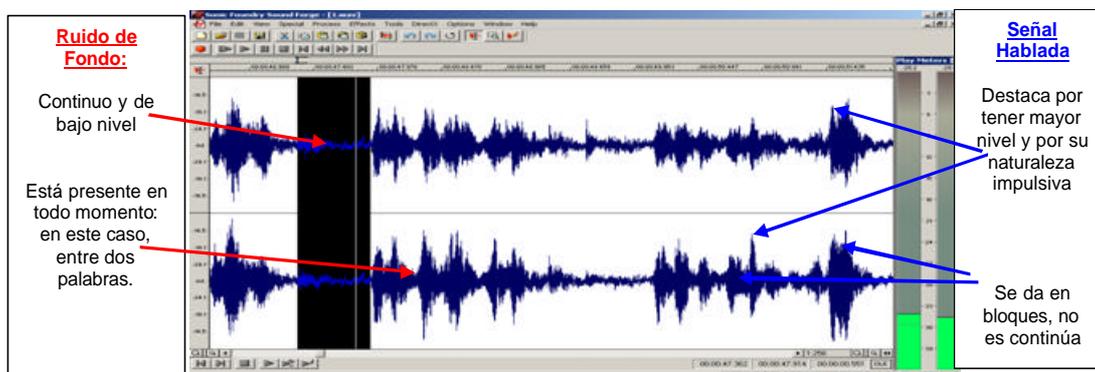


Fig. 2: Historia temporal de una grabación 'in situ'

El paso siguiente será determinar la duración mínima de los registros. La falta de uniformidad del habla natural y el hecho de que el discurso transcurre en ambientes relativamente reverberantes y ruidosos, hace que la elección de este parámetro de grabación condicione fuertemente el proceso de extracción que se elija para determinar qué es ruido y qué es nivel de voz.

En la bibliografía es usual encontrar que el espectro promedio, valor “rms”, del habla de un orador se puede determinar a partir de aproximadamente un minuto de grabación, e incluso menos. Como las condiciones de nuestras experiencias no parecían estar comprendidas en los trabajos consultados, se decide analizar el histograma de Niveles de Presión sonora instantánea de los registros ‘in situ’ para varios periodos de tiempo:

Los resultados estadísticos de la distribución de Niveles de Presión Sonora para una duración de un minuto aproximadamente se pueden ver en la figura (Fig. 3).

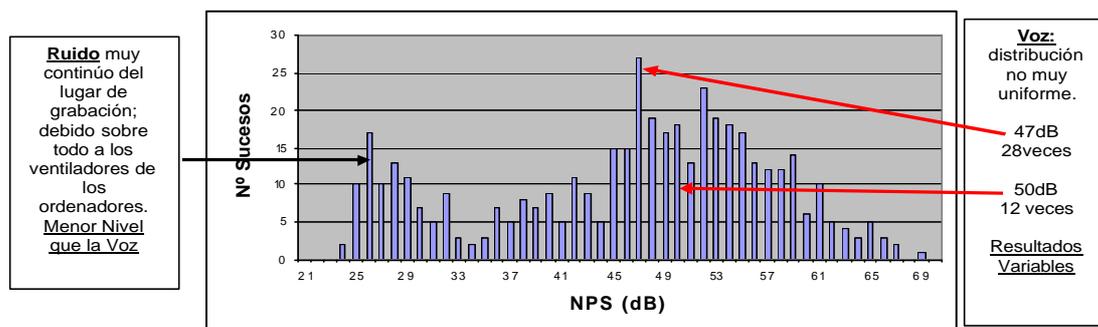


Fig. 3: Histograma para t= 57" de los Niveles de Presión Sonora para 2 KHz

Es evidente que para esta situación la distribución estadística no tiene un aspecto uniforme, lo que denota un tiempo de grabación insuficiente, sobre todo debido a que el ruido no es estacionario en situaciones reales como la nuestra.

En la gráfica siguiente (Fig. 4) se puede observar que al aumentar el tiempo de registro hasta los cinco minutos aproximadamente, su densidad espectral es mucho más uniforme: presentando una doble ‘campana’ lo que demuestra la existencia de palabra y ruido simultáneamente pero claramente discernibles, por lo que se optó por esta duración como la idónea.

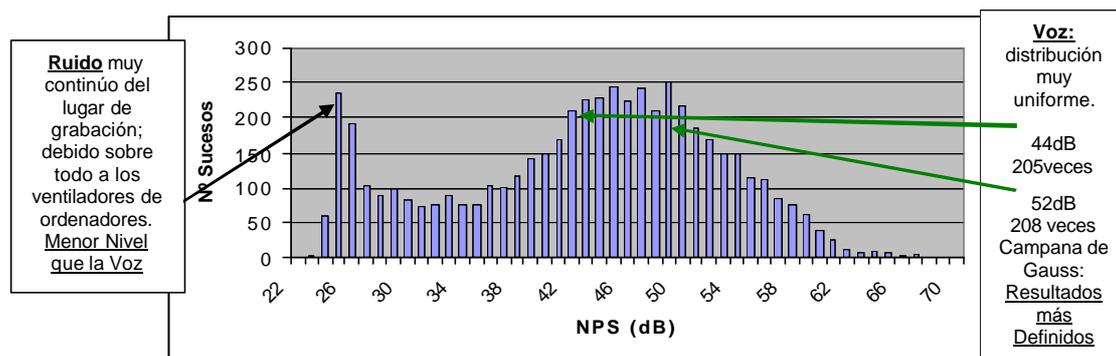


Fig. 4: Histograma para t= 5' de los Niveles de Presión Sonora para 2 KHz

Uno de los fundamentos y “novedades” que aporta el método **IAIS**, es la obtención del nivel del espectro del habla y del nivel del espectro del ruido por métodos estadísticos idénticos a los que habitualmente usamos en acústica medioambiental (Índices percentiles: Ln).

El factor L10 o nivel superado tan sólo durante el 10% de de la grabación, se considerará como el nivel de voz máximo presente durante el discurso. Por otro lado, con el factor L90 se extraerán los valores superados durante el 90% de duración del fragmento de audio, que se puede considerar sin género de dudas como ruido en estado puro. Los

resultados obtenidos con la elección clásica de L90 y L10 parecen seguir pautas lógicas y concuerdan con lo que se podía intuir al escuchar los fragmentos grabados.

Las grabaciones se han realizado durante clases de estudios de Ciencias, donde se realizan ejercicios numéricos y se utiliza la pizarra para aclarar conceptos. Estas acciones: escribir, borrar la pizarra, resolver problemas, poner transparencias etc..., hacen que el discurso del profesor no sea ni mucho menos continuo, al contrario de como ocurre en las conversaciones cotidianas. Dicha interrupciones, silencios largos no usuales en el habla conversacional, se han denominado “falsos silencios”, que aunque no crean una disminución ni aumento de la inteligibilidad del habla en sí, estadísticamente hablando “falsean” los resultados, ya que en esos momentos no hay nivel de habla y desvirtúan el valor del ruido de fondo estimado por el L90.

Para eliminar este problema y llegar a tener una grabación con habla fluida conversacional típica, la solución es obvia: eliminar los falsos silencios. Pero, ¿Qué se considera falso silencio y qué no? Se decidió afrontar el problema de una manera subjetiva basada en la experiencia y tras muchas horas de escucha virtual, sopesando y contrastando varios aspectos, se vio que la dependencia principal se podía atribuir al Tipo de clase impartida y al Ritmo del orador.

Estudiando estos dos aspectos, se puede concluir que a la hora de recortar los falsos silencios se debe considerar como tal, a todo aquel superior a 2 segundos como norma general, y 1.5 segundos para situaciones de ritmo muy lenta: clase de problemas.

Una vez que se tiene al material sonoro grabado, durante el post proceso se calculan los factores estadísticos L10, L15 y L20 y los valores L80, L85 y L90 para cada banda de frecuencia. Aunque se prevé que los índices óptimos serán los de valor percentil de 10 y 90, calcularemos el resto para usarlos como referentes si fuera necesario.

ÍNDICE IAIS CANAL IZQUIERDO															
Frecuencia [Hz]	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
L10 (dB)	61,73	62,93	67,15	70,07	73,20	69,14	61,73	61,14	65,04	65,11	62,79	65,45	67,37	64,96	58,06
L15 (dB)	60,68	61,37	65,47	68,71	71,71	66,76	60,12	59,26	63,05	63,25	61,07	64,00	65,47	63,36	56,49
L20 (dB)	59,64	60,15	64,13	67,61	70,43	64,95	58,87	57,91	61,41	61,71	59,70	62,62	63,88	62,00	55,36
L80 (dB)	50,24	49,65	50,85	52,60	52,68	50,28	47,22	45,55	47,73	48,39	48,90	50,22	51,61	51,62	46,02
L85 (dB)	49,19	48,56	49,08	50,50	50,71	48,94	46,08	44,27	46,46	47,02	47,62	48,82	50,37	50,63	45,04
L90 (dB)	48,20	47,24	47,09	48,31	48,93	47,53	44,85	43,00	44,38	45,62	46,23	47,28	48,89	49,36	43,90
Peso	0,0004	0,001	0,001	0,0014	0,0014	0,002	0,002	0,0024	0,003	0,0037	0,0038	0,0034	0,0034	0,0024	0,0028
[L10 - L90] x Peso	0,0054	0,0157	0,0201	0,0305	0,0340	0,0432	0,0338	0,0435	0,0602	0,0721	0,0629	0,0618	0,0628	0,0374	0,0283
[L15 - L85] x Peso	0,0046	0,0128	0,0164	0,0255	0,0294	0,0356	0,0281	0,0360	0,0498	0,0600	0,0511	0,0516	0,0513	0,0306	0,0229
[L20 - L80] x Peso	0,0038	0,0105	0,0133	0,0210	0,0248	0,0294	0,0233	0,0297	0,0410	0,0493	0,0410	0,0422	0,0417	0,0249	0,0187
[L10 - L80] x Peso	0,0046	0,0133	0,0163	0,0245	0,0287	0,0377	0,0290	0,0374	0,0519	0,0619	0,0528	0,0518	0,0536	0,0320	0,0241
[L15 - L90] x Peso	0,0050	0,0141	0,0184	0,0286	0,0319	0,0395	0,0305	0,0390	0,0542	0,0652	0,0564	0,0568	0,0564	0,0336	0,0252
[L20 - L90] x Peso	0,0046	0,0129	0,0170	0,0270	0,0301	0,0348	0,0280	0,0358	0,0493	0,0595	0,0512	0,0522	0,0510	0,0303	0,0229
[L10 - L85] x Peso	0,0050	0,0144	0,0181	0,0274	0,0315	0,0404	0,0313	0,0405	0,0557	0,0669	0,0577	0,0565	0,0578	0,0344	0,0260
[L15 - L80] x Peso	0,0041	0,0117	0,0146	0,0226	0,0266	0,0330	0,0258	0,0329	0,0460	0,0550	0,0462	0,0468	0,0471	0,0282	0,0209
[L20 - L85] x Peso	0,0042	0,0116	0,0151	0,0240	0,0276	0,0320	0,0256	0,0327	0,0448	0,0543	0,0459	0,0469	0,0459	0,0273	0,0206
<b>AI 10-90 = 0,612    AI 15-85 = 0,506    AI 20-80 = 0,415    AI 10-80 = 0,520    AI 15-90 = 0,554</b>															
<b>AI 20-90 = 0,507    AI 10-85 = 0,564    AI 15-80 = 0,462    AI 20-85 = 0,459</b>															
<b>IAIS GLOBAL</b>															
<b>AI 10-90 = 0,652    AI 15-85 = 0,544    AI 20-80 = 0,453    AI 10-80 = 0,558    AI 15-90 = 0,594</b>															
<b>AI 20-90 = 0,547    AI 10-85 = 0,602    AI 15-80 = 0,499    AI 20-85 = 0,498</b>															

Lo más destacable es que los resultados correspondientes al nivel de habla (L10, L15 y L20), muestran un nivel de espectro de habla muy cercano al del espectro de voz “idealizado” sin esfuerzo vocal fuerte (70dBA). Por lo que con este método se puede distinguir perfectamente los niveles de habla y de ruido de fondo.

Además de esto, en el ejemplo mostrado se evidencia que los niveles de ruido de fondo definidos por muchas bibliografías como “ideales”, cercanos a 30-35dBA, se rebasan en mucho en una situación real como ésta (sólo había 45 alumnos ? 40% del aforo), con un nivel global de ruido (L80, L85 y L90) cercano a los 49 dBA.

El método propuesto **IAIS'**, como ya se ha apuntado anteriormente, se basa en la norma "ANSI S3.5-1969", donde se dicta el procedimiento para el cálculo del Índice de Articulación (**AI**). Se ha introducido una variación en la determinación de la relación Señal/Ruido con el propósito de su aplicación en medidas de la inteligibilidad "in situ" durante el transcurso de las pláticas en aulas de características usuales. Se utiliza como relación Señal/Ruido la diferencia de niveles percentiles L10 – L90.

Se sabe que todo el espectro audible no influye por igual sobre la inteligibilidad, esto se debe a dos causas interrelacionadas entre si inevitablemente, como son: el funcionamiento del mecanismo de audición y la producción del habla en los humanos. Con lo que se puede comprender que un enmascaramiento en frecuencias medias sea más influyente y, justifica la necesidad de utilizar ponderaciones (pesos) a la hora de calcular Índice de Articulación, tanto para el método **AI**, como para el **IAIS**.

F(Hz)	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
Peso	0,0004	0,0010	0,0010	0,0014	0,0014	0,0020	0,0020	0,0024	0,0030	0,0037	0,0038	0,0034	0,0034	0,0024	0,0020

TABLA: 1. Ponderación en frecuencias para la relación S/R (L10-L90)

La rutina algebraica para hallar el valor final de factor de inteligibilidad IAIS es:

Se resta el valor del nivel del habla L10 al nivel de ruido de fondo L90, obteniendo lo que sería la relación señal a ruido para cada banda de frecuencias, a este valor se le multiplica el peso correspondiente a la banda en la que está:  $(L10 - L90)_f \times \text{Peso}_f$ . Se añade el valor obtenido en la hoja de cálculo, para una vez hecho esto con todas, se suman obteniendo así el valor de IAIS, por ejemplo:

$$\sum_{200}^{5000} (L10 - L90)_f \times \text{Peso}_f = \text{IAIS}_{10-90} = 0,612 \quad (\text{ec.1})$$

El valor obtenido hasta ahora es el referido a un canal, en el ejemplo, el izquierdo. Se hace lo mismo para el canal derecho.

La inteligibilidad se interioriza como una sola sensación, puesto que la aportación de los dos oídos proporciona una única percepción. Se tuvieron ciertas dudas sobre la manera de lograr un valor único de IAIS partiendo de los resultados individuales y distintos de ambos oídos. No se conoce de manera clara la aportación que tiene cada uno ante las diversas situaciones, ni se tiene un modelo matemático de la percepción sonora binaural.

Se pensó en realizar la media aritmética entre los dos valores IAIS, pero lo que ocurre es que si a un oído llega menos nivel que al otro, la sensación global calculada baja drásticamente y esto no corresponde a la realidad. Por ello, se ha considerado que un comportamiento logarítmico quizás se ajuste más a la realidad. Por lo que la ecuación utilizada es

$$\text{IAIS}_{\text{global}} = 0,1 \cdot \log(10^{(\text{IAIS}_D \cdot 10)} + 10^{(\text{IAIS}_I \cdot 10)}) \quad (\text{ec. 2})$$

Con la que se consigue una adecuación más natural de la escala en los resultados para el valor global. Al representar el mapa de factores IAIS medidos en un aula se puede apreciar, ved la siguiente ilustración, lo dicho en párrafo precedente para las dos hipótesis barajadas.

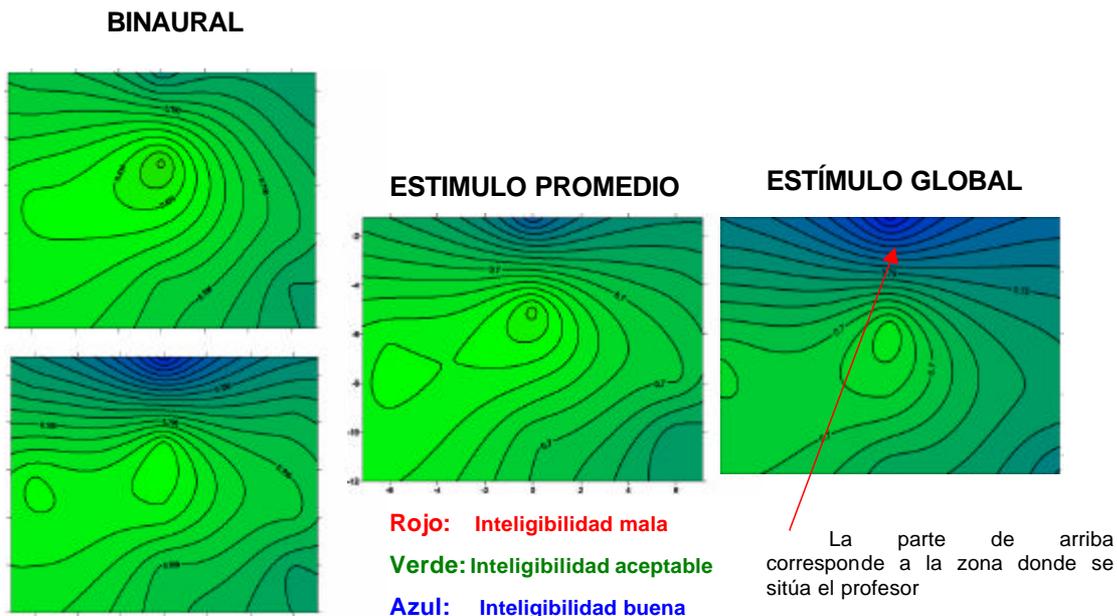


Fig. 5: Mapa de factores IAIS de inteligibilidad de un aula

En la imagen de la izquierda se muestra el mapa de IAIS del oído izquierdo y derecho consecutivamente. Esta profesora se situaba en la zona derecha del aula; por eso es mejor el IAIS del oído derecho, sirve también para ver la direccionalidad relativa del habla.

En la imagen central se muestra el estímulo que se obtendría si nuestra recepción siguiera la pauta de un promedio aritmético. Comparándola con la del oído derecho, la inteligibilidad es peor (contiene menos azules en la zona delantera del aula); se puede decir que al resultado del conjunto le falta naturalidad.

En la imagen derecha se puede observar que la aportación de cada oído refuerza el estímulo global de inteligibilidad obtenido; este gráfico se ha obtenido utilizando los criterios antes expuestos que serán los que se utilizarán a partir de ahora, para el cálculo del **IAIS**.

## CONCLUSIONES

Con el método planteado (**IAIS**) es posible obtener de manera sencilla y clara un valor de inteligibilidad objetivo, real e “in situ” que tiene en cuenta todos los parámetros que pensamos realmente influyentes, de forma conjunta con todas y cada una de sus características particulares: recinto, orador, público presente y entorno sonoro.

## BIBLIOGRAFÍA

**ANSI S3.5-1969**, *Methods for Calculation of the Articulation Index*. American National Standards Institute. (1969). New York: Acoustical Society of America.

**Gonzalez Etxabe, I.** ‘Estudio y Análisis de un método alternativo “in situ” para la valoración de la inteligibilidad objetiva en recintos’, P.F.C. de ITTSI. Universidad de Alicante (2005)

**Dunn, K.; White, S. D.** *Statistical Measurements on Conversational Speech*. J. Acoust. Soc. Am., **39**, 278-288 (1939).

**Nabtlek, A. K.; Robinson, P. K.** *Monoaural and binaural speech perception in reverberation for listeners of various ages*, J. Acoust. Soc. Am. **71**, 1242-1248 (1982).