

## Efectos de la duración y del contenido lingüístico en la identificación perceptiva de hablantes

Effects of duration and linguistic content in perceptual identification of speakers

**Ana Lidia Munguía Duarte**

**Amalia Hoyos Arvizu**

Universidad de Sonora

Sonora, México

[amunguia@capomo.uson.mx](mailto:amunguia@capomo.uson.mx)

[ahoyos@capomo.uson.mx](mailto:ahoyos@capomo.uson.mx)



### Resumen

En este estudio realizamos un experimento de identificación de hablantes con el fin de probar si la hipótesis de que ciertos sonidos son más efectivos durante la identificación perceptiva de los hablantes se puede sostener en el caso de lenguas como el español. De acuerdo a Amino & Arai (2006, 2007, 2008a, 2008b, 2009) y Amino, *et al.* (2005, 2006, 2006b, 2006c, 2007) algunos sonidos son más efectivos durante la identificación perceptiva particularmente los nasales, los mismos autores también sugieren que el punto de articulación es determinante. Los experimentos realizados por estos autores utilizaron estímulos del japonés, y la mayoría de éstos fueron palabras monosilábicas con significado. En nuestro estudio, los estímulos consisten de palabras bisilábicas con contenido semántico, palabras monosilábicas y trisilábicas carentes de significado y oraciones que incluían a estos tres tipos de estructuras. Los resultados de nuestro experimento mostraron que en la identificación de hablantes del español las consonantes nasales son más efectivas que las orales; sin embargo, el punto de articulación resultó, porcentualmente, determinante en el grupo de las consonantes nasales, pero no en el de las orales. La duración de los estímulos resultó ser un factor determinante para la efectividad en la identificación de voces cuando comparamos estructuras CV contra oraciones pero no cuando contrastamos estructuras CVCV contra oraciones. En los casos donde los datos arrojaron una diferencia significativa, ésta no tuvo un nivel de confianza alto.

**Palabras clave:** Identificación de hablantes, nasales, orales, duración, velocidad de habla.

### Abstract

In this study we conducted a human speaker identification to test the hypothesis that certain sounds are more effective during perceptiva identification of speakers can be sustained in the case of languages like Spanish. According to Amino & Arai (2006, 2007, 2008, 2008b, 2009) and Amino, *et al.* (2005, 2006, 2006b, 2006c, 2007) some sounds are more effective during perceptual identification, they report that nasal consonants are more effective than oral ones; they also suggest that the articulation's place is crucial in the identification. Experiments by these authors used stimuli of Japanese, and most of these were monosyllabic words with meaning. In our study, the stimuli consist of disyllabic words with semantic content, and mono and tri-syllabics meaningless words, we also use sentences including the above words. The results showed that in the identification of Spanish speakers the nasal consonants were more effective than oral obstruent consonants, but the articulation's place was determinant for nasal consonants but not for orals. The length of the stimulus was a determining factor for effectiveness in identifying voices contrasting CV structures vs. sentences but not CVCV structures vs. sentences. In cases where of the data showed a significant difference, it did not have a high confidence level.

**Key words:** Speaker identification, nasals, oral, duration, speech rate.

## 1. INTRODUCCIÓN

Diariamente cada uno de nosotros vive la experiencia de identificar, auditivamente, a algún pariente, amigo, colega, conductor de radio y/o televisión, lo cual nos indica que los sonidos mismos conllevan algún tipo de información que nos permite esta identificación.

La investigación sobre la identificación y el reconocimiento de voces se ha abordado desde diferentes perspectivas, (1) la sociolingüística que estudia criterios sociales como edad, sexo, conocimiento previo del hablante, etc., (2) la fonética acústica que estudia, detalladamente, las características individuales de la producción del habla de cada sujeto, y recientemente, (3) desde la fonética perceptiva que estudia el proceso de descodificación de las características del contenido lingüístico de la información o mensaje emitido por el hablante.

Los estudios sobre identificación de hablantes se han concentrado en la investigación acústica y perceptiva; Battaner *et al.* (2003) presentan una revisión exhaustiva de estudios sobre identificación de hablantes y enlistan algunos de los parámetros que han sido considerados pertinentes para esta tarea. Muestran que algunos de estos parámetros son muy resistentes a variables externas experimentadas por el hablante, como el estrés, condiciones fisiológicas, etc., mientras que otros resultan por demás sensibles a estas variables. Entre los primeros destaca el estudio de LTAS (*Long Term Averaged Spectrum*), los valores medios así como la anchura de los formantes y los sonidos turbulentos; dentro del segundo grupo destacan los parámetros relacionados con el F0, el estudio de sonidos que se vean afectados por efectos coarticulatorios y en especial el estudio de las consonantes nasales. En este artículo, dentro del marco de la fonética perceptiva,

presentaremos los resultados de un experimento donde se midieron los efectos de la duración y el contenido lingüístico en la identificación de hablantes, estudiaremos en particular sonidos orales (plosivos) y sonidos nasales.

En estudios recientes se ha reportado que el procesamiento del contenido fonológico y el de la identidad del hablante ocurren de manera independiente; sin embargo, autores como Nygaard (2005) y Goggin *et al* (1991) sugieren que hay evidencias de que existe una interacción entre ambos. Siguiendo esta línea, Amino & Arai (2006, 2007, 2008a, 2008b, 2009) y Amino *et al.* (2005, 2006a, 2006b, 2006c, 2007) han realizado algunos experimentos cuyos resultados les permiten argumentar que ciertos sonidos son más efectivos durante la identificación perceptiva de los hablantes; específicamente, estos autores aseguran para el caso particular del japonés, que los estímulos donde prevalecen los sonidos nasales son más efectivos para la identificación que aquellos donde únicamente ocurren sonidos orales. Además, aseguran que, al menos en el caso de las nasales, el punto de articulación también es relevante en la identificación de hablantes ya que, en sus experimentos, la identificación fue más efectiva cuando las nasales eran coronales y menos cuando eran labiales.

En otro estudio, Amino *et al.* (2007) evaluaron la relevancia de la estructura de la sílaba así como las transiciones de los formantes de las vocales durante la identificación perceptiva de hablantes. Concluyeron, de este estudio, que las estructuras silábicas con una consonante en posición de ataque resultan más efectivas en la identificación del hablante que aquellas que no tienen una consonante en dicha posición, y que en el caso de sílabas con consonantes en posición de coda, la efectividad de identificación aumentó cuando la consonante resultó ser una nasal. Con respecto al efecto de la transición de formantes, dicen los autores que las características de éstos no parecen ser determinantes para la identificación.

El propósito de este estudio es examinar si las propiedades fonético-fonológicas de los sonidos, así como la duración del contenido lingüístico resulta determinante en la identificación perceptiva entre hablantes del español mexicano. Aunque en un principio se planteó considerar la variable de conocimiento/contacto previo entre los sujetos del experimento y los participantes en la identificación, por cuestiones de espacio hemos incluido en este artículo únicamente los resultados de participantes que no conocían ni habían escuchado previamente a los informantes; es decir, en todos los casos, los oyentes –sujetos participantes en el experimento escuchaban por primera vez a los hablantes –sujetos que proporcionaron los estímulos.

## **2. IDENTIFICACIÓN PERCEPTIVA DE HABLANTES: ESTUDIOS PREVIOS**

La realización de experimentos que estudian las variables que influyen en la identificación perceptiva de hablantes no es nueva; sin embargo, desde que en muchos países se ha incrementado el uso de este tipo de información como evidencia judicial, también ha ido en aumento la investigación sobre los parámetros relevantes para la identificación de hablantes, y particularmente, aquella relativa a la relevancia del contenido lingüístico de los estímulos, así como el factor de conocimiento previo entre los sujetos para una mayor efectividad en la identificación de hablantes. Destacan los experimentos desarrollados en

lenguas como el inglés, el ruso y el japonés; en el caso de las dos primeras la mayoría de los experimentos se ha realizado con identificaciones automatizadas no con sujetos participantes, mientras que en el japonés, lengua en la cual se ha desarrollado el mayor número de experimentos, todos los estudios han sido con sujetos participantes. Tratándose del español, entre los estudios disponibles está el de Battaner *et al.* (2003) donde se estudia la variación intra y entre hablantes con la finalidad de aplicar tal conocimiento al reconocimiento automático de hablantes y a la práctica de la fonética judicial. Los autores revisan una serie de parámetros acústicos para determinar su relevancia en la individualidad de las voces y destacan una serie de parámetros que permiten identificar a los hablantes, asimismo, argumentan la poca validez para esta misma tarea de otros parámetros, como el estudio de las consonantes nasales. Por el contrario, mediante un experimento realizado con hablantes del inglés, autores como Glenn & Kleiner (1968) asumen que la estructura fisiológica –forma y tamaño- del aparato vocálico individualiza la voz de los hablantes, particularmente durante la producción de sonidos nasales ya que tanto la cavidad como los articuladores donde se producen dichas sonidos se mantiene relativamente fijo.

En el cuadro 1 sintetizamos los estudios relacionados con las características del contenido lingüístico de los estímulos, pero no con la parte fisiológica. La segunda columna del cuadro contiene información sobre el número y sexo de los hablantes, es decir, las personas que contribuyeron en la obtención de los datos lingüísticos, los estímulos; la tercera contiene el número de participantes: los sujetos que escucharon e identificaron a los hablantes; en la cuarta columna se enlistan las características de los estímulos y la lengua sobre la cual se realizó el experimento; y, finalmente, en la última columna se resumen los resultados obtenidos de dichos experimentos. En este cuadro destacan tres aspectos: 1) la gran cantidad de estudios realizados con voces de hablantes de sexo masculino; 2) las características de la estructura de la palabra en el japonés -tendencia a estructuras monosilábicas-, y 3) el conocimiento previo entre sujetos hablantes y sujetos participantes - en ocho de los diez experimentos citados había una relación de contacto previo entre ellos.

Cuadro 1. Estudios previos sobre identificación de voces

	Nº DE HABLANTES	Nº DE PARTICIPANTES	MATERIALES (LENGUA)	SONIDOS MÁS EFECTIVOS
Matsui <i>et al.</i> (1993)	8, M	11, conocidos	Sílabas CVC extraídas (Japonés)	Depende de los hablantes
Kitamura & Akagi (1995)	5, M	8, conocidos	Vocales aisladas (Japonés)	/ú/
Amino (2003)	3, F	14, conocidos	Vocales/ monosílabos aislados (Japonés)	/ú/ y nasales
Amino (2004)	3, M y 2, F	18, conocidos	Monosílabos extraídos (Japonés)	Nasales y consonantes coronales sonoras
Amino <i>et al.</i> (2005)	10, M	5, conocidos	Monosílabos extraídos (Japonés)	Nasales
Amino <i>et al.</i> (2006)	8, M	2, M y 8, F conocidos	monosílabos [V, VV VN, CV, CV,	Sílabas con ataque silábico, sílabas pesadas.

			CVV, CVN] (Japonés)	Nasales, Alveolares
Amino & Arai (2006)	4, M	16, desconocidos	Monosílabos extraídos (Japonés)	conocimiento previo del hablante; nasales coronales
Amino <i>et al.</i> (2007)	10, M	5 M, conocidos	Monosílabos CV extraídos. (Japonés)	Nasales, sonoros>> sordos
Amino & Arai (2007)	8, M	2 M y 6 F conocidos	Monosílabos [V, VV VN, CV, CV, CVV, CVN] (Japonés)	Sílabas con ataque silábico, sílabas pesadas con codas nasales. Nasales, Alveolares
Amino & Arai (2008)	4 M	15, desconocidos	Monosílabos	Nasales, coronales y dorsales. La diferencia en los rasgos fonológicos es más relevante que la duración del estímulo

Una característica común en las investigaciones anteriores es que los estímulos que contienen sonidos nasales son mucho más efectivos durante las pruebas de identificación de voces. Además, una tendencia fuerte a que las consonantes coronales y la longitud de los estímulos también sean factores determinantes en la identificación de hablantes.

### 3. METODOLOGÍA

#### 3.1. Hablantes

El tamaño y composición del conjunto de sujetos en los experimentos de identificación llevados a cabo con hablantes desconocidos es un factor determinante para la efectividad de la identificación: cuanto más pequeña sea la muestra mayor efectividad habrá. En atención a lo anterior, los datos para esta investigación se obtuvieron de cuatro sujetos masculinos cuya edad oscila en un rango de 22 a 28 años, nacidos en Hermosillo y con residencia prolongada y continua en la ciudad, hijos de padres originarios del mismo lugar. En el cuadro 2 se muestran los datos generales de los hablantes.

Cuadro 2. Datos generales de los hablantes.

Identificación del hablante	Edad	Altura	Peso	F0 (Hz) Media	F0 (Hz) DS	Número sil./sg.	Número seg./sg.
H1	24	1.75	65	118	10.4	6.9	15
H2	22	1.84	73	130	20.2	6.1	13.6
H3	28	1.76	95	152	22.4	5.8	13
H4	23	1.86	68	132	14.7	7.7	17

En el cuadro 2 observamos que el hablante 4 es el más veloz en la producción de sílabas y segmentos por segundos; mientras que el más lento es el hablante 3. Aunque esta variable no ha sido considerada en los estudios previos nos pareció importante incluirla en nuestro experimento para observar si constituye una variable influyente.

### 3.2. Participantes

Se seleccionaron 10 participantes que identificaran a cuatro hablantes a quienes nunca antes habían escuchado. Al igual que los colaboradores hablantes, los sujetos participantes también son hablantes nativos de español, con una residencia de al menos 10 años en la ciudad de Hermosillo; la edad promedio de los participantes, todos ellos de sexo femenino, es de 22 a 32 años, aproximadamente, y se cuidó que ninguno de los participantes tuviera algún problema de audición.

### 3.3. Estímulos

La grabación de los estímulos se realizó con una grabadora Marantz PMD y un micrófono ShureSM48 en un ambiente relativamente quieto.

Se pidió a los sujetos hablantes que leyeran una lista de palabras y pseudopalabras, una serie de oraciones que contenía la lista previa de palabras y un pequeño texto; se les solicitó que las repitieran tres veces y que su voz fuera lo más natural posible. Dado que el propósito central del experimento es evaluar la influencia de las propiedades articulatorias, así como la duración de los sonidos, los estímulos grabados incluyen monosílabos sin significado, bisílabos con significado, trisílabos sin significado y frases que contuvieran los segmentos bajo estudio. En todos los casos el núcleo vocálico fue controlado: siempre se usó la vocal baja [a], con el fin de que el experimento fuera lo más simple posible para los participantes.

Los datos fueron grabados exclusivamente para este experimento, todos los estímulos contienen cualquiera de las siguientes 6 consonantes, las 3 orales menos sonoras /p, t, k/ y 3 consonantes nasales, /m, n, ɲ/. Algunos ejemplos del tipo de estructuras presentadas a los participantes en el experimento se muestran en (1):

(1) monosílabos	[pa, ta, ka, ma, na, ɲa]
bisílabos con significado	[papa, tata, kaka, mama, nana, maɲa]
trisílabos carentes de significado	[apapa, atata, akaka, amama, anana, aɲaɲa]
oraciones	[pepe pide papas, mi mamá me ama] [me gusta más el grupo apapa, me gusta más el grupo amama]

Los datos utilizados para medir la velocidad de producción de los hablantes se obtuvieron de una muestra de un minuto, tomada de la grabación de la lectura de un texto narrativo. La grabación del texto narrativo se realizó en una fecha posterior a la grabación de los estímulos.

### 3.4. Procedimiento

Al igual que para la grabación de los estímulos, el experimento se realizó en un espacio donde se evitó, hasta donde fue posible, el ruido. Los participantes escucharon los estímulos usando una computadora personal y audífonos Sony MDR-V100; las tareas se realizaron con el programa Praat utilizando el formato de experimentos de opción múltiple obligada, MFC.

Antes de la sesión del experimento definitivo se realizó una sesión de familiarización; en esta fase, los participantes escucharon un enunciado común producido por cada uno de los informantes, [le'ʃe la 'kulpá al te'lefono ke so'naβa]. Los participantes escucharon estas expresiones tantas veces como quisieron, hasta que estuvieran seguros de que podrían reconocer a los hablantes. Para evitar cualquier tipo de confusión, los hablantes fueron presentados a los participantes con un número de identificación del 1 al 4 y las expresiones fueron presentadas en ese mismo orden.

Después de la sesión de familiarización, se realizó una sesión de práctica; ésta consistió en la presentación, aleatoria, de ocho enunciados, dos por cada hablante. Se cuidó que estos enunciados contuvieran el menor número posible de consonantes oclusivas menos sonoras y de consonantes nasales, véase ejemplos en (2).

- (2) mayo de dosmilonce  
se solicita doble

Esta sesión se repitió hasta que el participante logró identificar a los hablantes con un promedio de 90% de certeza. Los sujetos participantes invirtieron aproximadamente 15 minutos y escucharon entre 5 y 15 veces las expresiones.

Una vez terminada la sesión de familiarización se pasó a la realización del experimento sin tomar ningún tipo de receso entre estas dos tareas. El experimento consistió de un número total de 124 estímulos, correspondientes a 6 sílabas del tipo CV, 9 palabras de la forma CVCV, 6 pseudopalabras de la forma VCVCV y 10 enunciados para cada uno de los cuatro hablantes. Al igual que en la sesión de familiarización, los estímulos también fueron presentados de manera aleatoria. Durante la realización del experimento no se permitió a los participantes escuchar las expresiones de referencia; es decir, no podían volver a escuchar las oraciones presentadas en la sesión de familiarización. Se les permitió un receso después de 62 intentos.

## 4. RESULTADOS

Siguiendo la metodología de Amino & Arai (2007), los resultados del experimento fueron evaluados a partir del porcentaje de identificación correcta; se consideró el 25% de identificación como el nivel mínimo de oportunidad para ser considerado dentro del experimento, cualquier porcentaje por debajo del 25% fue descartado para este análisis.

En lo que sigue de esta sección, presentaremos los resultados de las tres hipótesis de este estudio: que las consonantes nasales permiten identificar con mayor facilidad a los

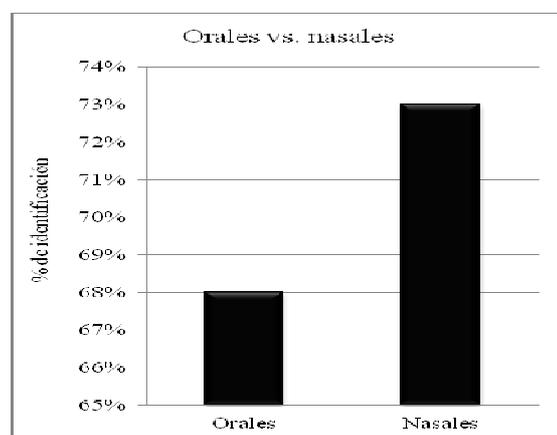
hablantes, que los segmentos corales son más efectivos que los no corales, y que la longitud del contenido lingüístico también participa en la identificación del hablante.

#### 4.1. Hipótesis 1

Los estímulos que contienen contenido nasal son más efectivos para la identificación de hablantes que aquellos que contienen únicamente contenido oral (Amino, *et al.* 2005, 2006, 2006b, 2006c, 2007; Amino & Arai 2006, 2007, 2008, 2008b, 2009).

En la figura 1 observamos que el porcentaje de identificación es más alto para los estímulos con consonantes nasales que para aquellos que contenían únicamente consonantes orales. Los resultados de este experimento coinciden con la tendencia de que los sonidos nasales son más efectivos que los obstruyentes para la identificación de hablantes.

Figura 1. Resultados de identificación por grupos de consonantes.



Las diferencias encontradas en las medias 32.3% para las consonantes producidas en la cavidad nasal, y 27% para aquellas producidas en la cavidad oral son significativas en un análisis de varianza ( $F(1,10) = 9.51, p < 0.0115$ ).

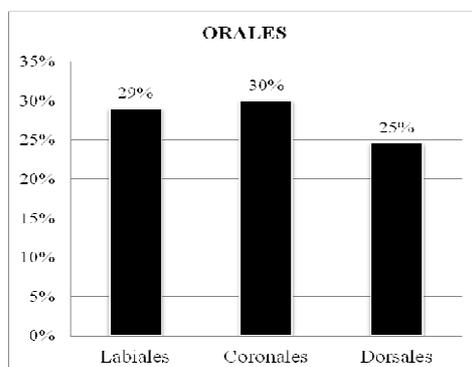
#### 4.2. Hipótesis 2

Los estímulos que contienen nasales corales son más efectivos que los que contienen nasales labiales para la identificación de hablantes. (Amino, *et al.* 2005, 2006, 2006b, 2006c, 2007).

Con respecto a la hipótesis anterior, en el caso de las consonantes orales, los resultados indican que el punto de articulación no resultó ser determinante en la efectividad de identificación de voces, la media de labiales y corales es muy similar, en tanto que la media para las dorsales fue relativamente menor. Las diferencias encontradas en las medias: 29% para labiales, 30% para corales y 29% para dorsales no resultó ser significativa en un análisis de varianza ( $F(2,6) = 2.73$ ).

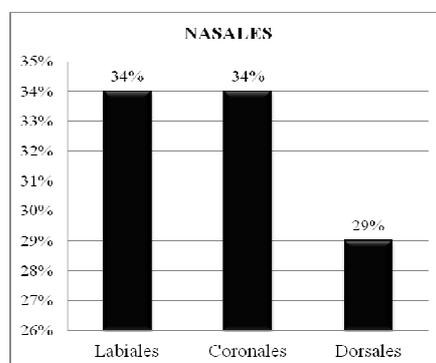
En la figura 2 se observa una tendencia, mínima, a identificar con mayor facilidad aquellos estímulos que contenían consonantes coronales, con un 30% de efectividad, comparado al 29% y 25% de las labiales y las dorsales, respectivamente.

Figura 2. Identificación de obstruyentes bajo el criterio de las características articulatorias.



En el caso de las nasales, como se observa en la figura 3, hubo una mayor identificación de estímulos cuando éstos contenían o una nasal labial o una coronal. En estos casos hubo una diferencia entre medias de 34% para labiales y coronales, mientras que para las consonantes dorsales fue de 29%. Un análisis de varianza revela que los resultados son significativos ( $F(2,6) = 5.37, p = .046$ ) aunque con un nivel de confiabilidad muy bajo.

Figura 3. Identificación de nasales bajo el criterio de las características articulatorias.



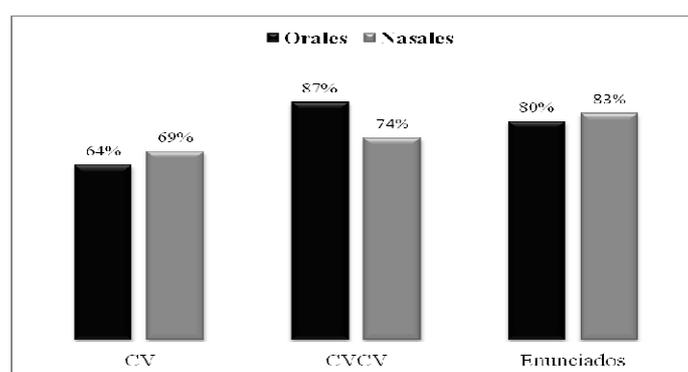
En este punto de la investigación, y de acuerdo a los resultados obtenidos a través del análisis de varianza, no podemos afirmar que la hipótesis de que las características articulatorias son determinantes en la identificación de voces sea válida. De acuerdo al análisis de varianza, ésta fue significativa únicamente en el caso de las nasales; y aún en este grupo es cuestionable puesto que no podemos afirmar-tajantemente- que la diferencia sea significativa entre los tres tipos consonantes.

### 4.3. Hipótesis 3

Los estímulos de mayor duración permiten mayor efectividad en el reconocimiento de hablantes. Bricker & Pruzansky (1976) y Amino & Arai (2008, 2008b) reportan que la

longitud del estímulo puede ser un factor determinante en la identificación de hablantes; sin embargo, en el experimento que realizamos esta variable no resultó ser estadísticamente significativa para la identificación de los sujetos hablantes, como podemos ver en la figura 4.

Figura 4. Porcentaje de identificación de acuerdo a la longitud del estímulo.



En efecto, a los participantes les resultó más fácil identificar cualquier forma que fuera mayor a un monosílabo; sin embargo, no hay diferencias realmente significativas en la identificación de formas de dos o tres sílabas respecto a las formas enunciativas. Es decir, los participantes identificaron al hablante en un porcentaje similar cuando éste produjo estructuras cortas, del tipo CV o CVCV. El análisis de varianza revela que cuando contrastamos estructuras del tipo CV contra estructuras de mayor longitud, como las correspondientes a enunciados, la hipótesis resulta significativa, cuando se trata de la producción de consonantes orales, ( $F(1,4) = 11.7$ ,  $p = .026$ ) pero no cuando se trata de las consonantes nasales ( $F(1,4) = 2.7$ ).

## 5. FACTORES RELATIVOS A LOS HABLANTES

En el cuadro 3 se muestra la matriz de confusión entre hablantes. Resalta en éste el alto porcentaje de identificación del hablante 3, con un 87% de identificación por parte de los participantes, en comparación al 53% que obtuvo el hablante 4. También es de notar el alto porcentaje de confusión en el hablante 1 y el hablante 4: cuando el estímulo correspondió a H1 los participantes lo confundieron un 22.5% con H4, y cuando el estímulo correspondió a H4 se le confundió también 22.5 con H1. Un total de 50% de confusiones entre estos dos hablantes.

Cuadro 3. Matriz de confusión entre los 4 hablantes; el % de respuesta percibida (columna) por el hablante real (filas).

	H1	H2	H3	H4
H1	60.3	10.6	6.5	<b>22.5</b>
H2	10.3	73.2	22.5	19
H3	3.2	2.2	<b>87</b>	5.5
H4	<b>22.5</b>	19	4.0	<b>53</b>

Amino & Arai (2008a, 2008b) consideran, dentro de los factores que pueden crear confusiones en la identificación, las frecuencias fundamentales de las vocales. Ellos pudieron medir las frecuencias fundamentales de las vocales ya que en su estudio contemplan como variable a seguir, para la identificación de hablantes, la diferencia entre vocales anteriores y vocales posteriores. Dado que en la muestra de datos de este experimento restringimos el núcleo vocálico de los estímulos a una única vocal, no tenemos suficiente información para evaluar si éste pudiera ser uno de los factores que inciden en el reconocimiento y/o confusión de voces durante la identificación. Hashimoto *et al.* (1998), reportan que el promedio de F0 de cada hablante puede también ser un factor a considerar.

En esta etapa de nuestra investigación destacamos una posible correlación entre velocidad de habla y efectividad de reconocimiento del hablante; sin embargo, por ahora únicamente lo planteamos como una hipótesis que tendremos que validar. En el cuadro 4 se muestra el número de sílabas y segmentos por segundo para cada hablante, así como el porcentaje de identificación por parte de los participantes en el experimento.

Cuadro 4. Correlación entre % de identificación y velocidad de habla de cada hablante.

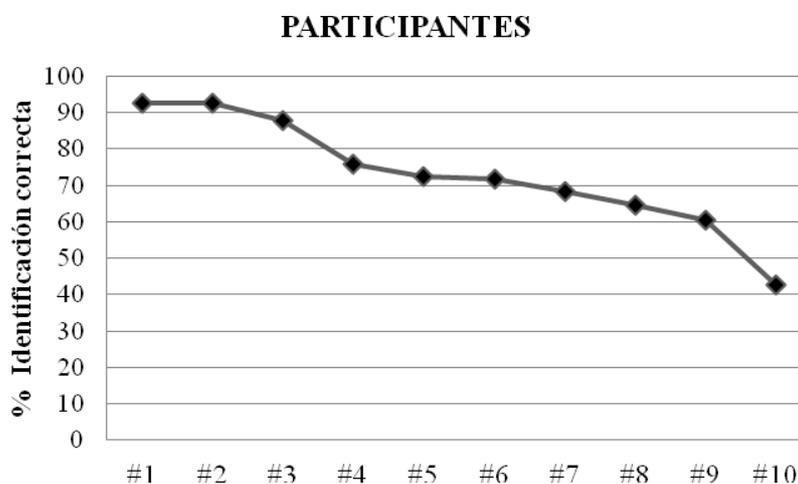
	Número sil./sg.	Número seg./sg.	% Identificación
H1	6.9	15	61.1 %
H2	6.10	13.6	73.2 %
H3	5.8	13	87 %
H4	7.7	17	53 %

En el cuadro 4 observamos que el hablante 4 quien produce un mayor número de sílabas y segmentos por segundo –casi 8 sílabas por segundo- obtuvo únicamente un 53% de identificaciones; en tanto que el hablante 3 quien produjo el menor número de sílabas y segmentos por segundo –aproximadamente 6 sílabas por segundo- fue el que obtuvo el porcentaje mayor, 87%, de identificaciones.

Además, los hablantes H1 y H4, quienes tuvieron el porcentaje más alto en la matriz de confusión, son los dos hablantes que muestran una mayor producción de segmentos por segundo.

Finalmente, en la figura 5 mostramos el porcentaje de identificación de los hablantes por parte de cada uno de los participantes. El máximo porcentaje de identificación fue de 92.7% y el mínimo de 42.7%, el promedio fue de 72.9%.

Figura 5. Actuación de los participantes en la identificación de hablantes.



## 6. CONCLUSIONES

El objetivo de este estudio fue corroborar, mediante un experimento de percepción, si existe una correlación entre los rasgos acústicos y la duración de los estímulos, con la identificación de los hablantes que los producen. De acuerdo a Amino & Arai (2007) y Amino, *et al.* (2005, 2006, 2006b, 2006c, 2007) algunos sonidos son más efectivos durante la identificación perceptiva, particularmente los nasales, lo mismo que el punto de articulación, también determinante. Los experimentos realizados por estos autores fueron realizados utilizando estímulos del japonés, y la mayoría de sus estímulos fueron palabras monosilábicas con significado.

En nuestro estudio no hemos diferenciado los resultados de estímulos sin significado -monosílabas y trisílabas- y elementos con significado -palabras y frases. Las conclusiones obtenidas a partir de los resultados de nuestro experimento son las siguientes:

- 1) Las consonantes nasales son más efectivas que las orales en la identificación de hablantes. Esta mayor efectividad es esperada dado que la cavidad de resonancia involucrada en las articulaciones nasales refleja información fisiológica del hablante (Glenn & Kleiner, 1968, Wolf, 1971).
- 2) En el grupo de las consonantes nasales, pero no en el de las orales, las coronales resultan ser más efectivas para la identificación de hablantes. Sin embargo, un análisis de varianza revela que el nivel de confianza para establecer esta significancia es muy bajo.
- 3) La estructura de los estímulos no resultó ser un factor determinante para la efectividad en la identificación de voces. Sin embargo, es de resaltar que cuando se

contrastaron estímulos monosilábicos y enunciados hubo una diferencia significativa cuando se trató de consonantes orales.

- 4) Hemos establecido de manera preliminar una correlación entre velocidad de habla e identificación de hablantes. Sin embargo, necesitamos corroborar si este factor es el que realmente está determinando la efectividad de la identificación o se suma a éste el promedio de F0 y las frecuencias fundamentales de cada hablante. En este estudio no es posible avanzar en este punto dado que los estímulos contienen únicamente la vocal baja [a].

Dados los anteriores resultados, en las siguientes fases de esta investigación habremos de considerar diversas variables tales como el incremento de los estímulos, la familiaridad o conocimiento previo entre sujetos –hablantes y participantes en el experimento; asimismo habremos de incluir entre los estímulos al menos dos vocales más.

A la luz de estos resultados también deberemos considerar si la velocidad de habla y la frecuencia fundamental son parámetros determinantes en la identificación de voces.

## 7. REFERENCIAS BIBLIOGRÁFICAS

Amino, Kanae. 2003. The Characteristics of the Japanese Phonemes in Speaker Identification. *Proceedings of Sophia University Linguistic Society*, 18. 32-43.

Amino, Kanae. 2004. Properties of the Japanese Phonemes in Aural Speaker Identification, *Technical Report IEICE*. 49-54.

Amino, Kanae & T. Arai. 2006. Speech Similarity in perceptual Speaker Identification. *Proceedings of 2006 meeting Journal of the Acoustical Society of Japan*. 273-274.

Amino, Kanae & T. Arai. 2007. "Effects of Stimulus Contents and Speaker Familiarity on perceptual Speaker Identification", *Acoustic Science Tech.*, vol.28, pp.128-130

Amino, Kanae & T. Arai. 2008a. Differential Effects of the Phonemes on Identification on Previously Unknow. <http://webistem.com/acoustics2008/acoustics2008/cd1/data/articles/001761.pdf> (8 de febrero de 2012).

Amino, Kanae & T. Arai. 2008b. Perception of Speaker Identity and its Relations to the Phonological Features. [http://www.splab.ee.sophia.ac.jp/papers/2008/2008\\_18.pdf](http://www.splab.ee.sophia.ac.jp/papers/2008/2008_18.pdf) (8 de febrero de 2012).

Amino, Kanae & T. Arai. 2009. Speaker-dependent characteristics of the nasals. *Forensic Science International*, 185. 21–28.

Amino, Kanae, T. Sugawara & T. Arai. 2005. Correspondences between the perception of the speaker individualities contained in speech sounds and their acoustic properties. *Proceedings of Interspeech*. 2025-2028

Amino, Kanae, T. Sugawara & Takayuki Arai. 2006a. Speaker similarities in human perception and their spectral properties. *Proceedings of Western Pacific Acoustics Conferences*. [http://splab.net/papers/2006/2006\\_10.pdf](http://splab.net/papers/2006/2006_10.pdf)

Amino, Kanae, T. Sugawara & T. Arai. 2006b. Idiosyncrasy of nasal sounds in human speaker identification and their acoustic properties. *Acoustic Science Tech*, 27.233-235.

Amino, Kanae, T. Sugawara & T. Arai. 2006c. Effects of the syllable structure on perceptual speaker identification. *IEICE Tech. Rep.*, 105.109-114.

Amino, Kanae, T. Arai, & T. Sugawara. 2007. Effects of the Phonological Content on Perceptual Speaker Identification. En C. Müller (ed.) *Speaker Classification II, LNAI 4441*. 83-92.

Battaner, E.; GIL, J; Marrero, V.; Llisterri, J.; Carbó, C.; Machuca, M.; de la Mota, C.; Ríos, A. 2003. VILE: Estudio acústico de la variación inter e intralocutor en español. *Actas del II Congreso de la Sociedad Española de Acústica Forense*. Barcelona: SEAF, Sociedad Española de Acústica Forense. 59-70. [http://liceu.uab.es/~joaquim/phonetics/VILE/VILE\\_SEAF03.pdf](http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_SEAF03.pdf).

Bricker, Peter & S. Pruzansky.1976. Speaker recognition. En N. Lass (ed.) *Experimental Phonetics*. London: Academic Press. 295-326.

Glenn, J. & N. Kleiner.1968. Speaker Identification Based on Nasal Phonation. *Acoustic Society of America*, 54:2. 368-372.

Goggin, J., C. Thompson, G. Strube, & L. Simental, L. 1991. The Role of Language Familiarity in Voice Identification. *Memory and Cognition*, 19. 448-458.

Hashimoto, M., S. Kitagawa & N. Higuchi.1998. Quantitative analyseis of acoustic features affecting speaker identification. *Journal of the Acoustical Society of Japan*, 54. 169-178.

Kitamura, T. & M. Akagi.1995. Speaker individualities in speech spectral envelopes. *Journal of the Acoustical Society of Japan*, 16:5. 283-289.

Matsui, T. I. Pollack, & S. Furui.1993. Perception of voice individuality using syllables in continuous speech. *Proceedings of the 1993 autumn meet. Acoustic Society of Japan*. 379-380.

Nygaard, L. 2005. Perceptual integration of linguistic and nonlinguistic properties of speech, En D. Pisoni and R. Remez (eds.) *The Handbook of Speech Perception*. Oxford: Blackwell Publishing. 390-413.

Wolf, J. 1971. Efficient Acoustic Parameters for Speaker Recognition. *Acoustic Society of America*, 51:6. 2044-2055.