

DIFERENCIAS ENTRE LOCUTORES EN CONSONANTES NASALES DE UN CORPUS DE LA POLICÍA DE INVESTIGACIONES DE CHILE

Felipe Farías Álvarez ^a, Claudia Rosas Aguilar ^b

(a) Instituto de Acústica, Universidad Austral de Chile, Casilla 567, Valdivia, Chile, felipefarias@vtr.net
(b) Instituto de Lingüística y Literatura, Universidad Austral de Chile, Casilla 567, Valdivia, Chile, claudiarosas@uach.cl

Resumen

Las consonantes nasales son los únicos sonidos del habla en que se produce un cierre de los órganos articulatorios bucales y una apertura de la cavidad nasal para el paso del aire, lo cual les otorga características particulares respecto de los otros sonidos del habla.

En este estudio se realiza un análisis de formantes de las consonantes nasales en un corpus de voces de casos cerrados, registrado por la Policía de Investigaciones de Chile entre los años 2006 y 2009. Se comparan las muestras asociadas a un locutor entre ellas y entre diferentes hablantes, considerando además la posición de las nasales. Los resultados obtenidos en las comparaciones muestran mayores variaciones entre diferentes locutores que entre muestras del mismo hablante, en prácticamente todos los parámetros considerados, lo cual da soporte a la característica locutor-dependiente de las consonantes nasales, recomendable para el reconocimiento forense de locutores.

Palabras-clave: Nasales, Acústica forense, Reconocimiento de locutores.

Abstract

Nasal consonants are the only sounds in human speech which produce a closure of the oral articulatory organs and an opening of the nasal cavity for passing the air, which gives them unique characteristics compared to other speech sounds.

In this study, a formant analysis of nasal consonants is made using a corpus of voices of closed cases, recorded by the Investigative Police of Chile between 2006 and 2009. The samples of the same speaker and between speakers are compared, also considering the position of the nasals. The results obtained in the comparisons show greater variations between different speakers than the samples of the same speaker, in practically all of the parameters considered, which supports the speaker-dependent feature of nasal consonants, recommended for forensic speaker recognition.

Keywords: Nasals, Forensic acoustics, Speaker recognition.

PACS no. 43.72.Uv, 43.70.Kv

1 Introducción

El reconocimiento o comparación forense de locutores se ha perfilado en las últimas décadas como una prueba válida y relevante en los casos policiales. En Chile, la Reforma Procesal Penal y su correspondiente implementación atrajo la atención de las voces grabadas como posibles pruebas en la determinación de inocencia o culpabilidad de los sujetos llevados a juicio oral y, con ello, el desarrollo de los análisis periciales [8].

Considerando esta nueva situación, en Chile han surgido diversas iniciativas que apuntan a abordar la problemática de la prueba de reconocimiento de voz. Una de estas es la de la Policía de Investigaciones de Chile, la cual creó en el año 2003 la sección de Sonido en el Laboratorio de Criminalística, que entre otras funciones, tiene la identificación y análisis de voz. El año 2006, la ahora llamada sección de Sonido y Audiovisual de este laboratorio, incorporó en sus peritajes el reconocimiento de hablantes con la herramienta biométrica Batvox [6].

En este contexto, surge desde fines de 2005 una base de datos de grabaciones utilizadas como pruebas, con diferentes condiciones, pero que involuntariamente se transforma en una muestra representativa de la calidad y las características lingüísticas de las voces y grabaciones que se encuentran o se podrían encontrar en la realidad chilena.

Asumiendo el desafío que implica la determinación de una población de referencia de voz para las comparaciones forenses, surge el proyecto "Voces en contextos periciales para el reconocimiento automático", del cual este trabajo es parte. Una primera etapa consiste en conocer las características de la base de datos existente y también establecer parámetros que pudieran ser útiles, y por tanto considerados en etapas futuras, siempre bajo la perspectiva forense del problema.

Se propone establecer la idoneidad de las consonantes nasales en el reconocimiento forense de locutores, principalmente debido a las características particulares de estas por sobre otros sonidos del habla.

2 Características de las nasales

Las nasales tienen la particularidad de ser los únicos sonidos del habla donde se cierra la cavidad oral y el paso del aire se produce por la cavidad nasal. Esta propiedad les otorga características únicas, ya que la cavidad nasal es muy compleja en forma y tamaño, lo cual hace que las nasales sean muy particulares en cada locutor [4].

Una de estas características es que se introducen ceros a la función de transferencia, por lo tanto, no se puede modelar el espectro de las nasales como una función de sólo polos [7]. La función de transferencia se hace más compleja, y también más particular para cada hablante.

En una etapa inicial se han considerado los parámetros para caracterizar de mejor manera las particularidades de las consonantes nasales, en base a la bibliografía. Estos parámetros son función de transferencia, formantes, posición de la nasal y transiciones [4].

De estos 4 parámetros son los formantes los elementos centrales considerados inicialmente, principalmente porque se trata de un parámetro básico en reconocimiento de locutores, y que no ha sido abordado para las nasales en una base de datos con las características de la que aquí se presenta. Las transiciones han sido abordadas en *Speaker-dependent characteristics of the nasals* [2], aunque con un corpus completamente diferente. Dado que los formantes son el reflejo de la función de transferencia, esta última no ha sido considerada en esta investigación.

Finalmente, existen algunos estudios [5][7] que sugieren diferentes características de las nasales para diferentes posiciones de estas respecto de las vocales adyacentes (prevocálica, intervocálica y postvocálica).

Por lo anterior, se decide analizar formantes y separarlas de acuerdo a su posición, para luego analizar la variabilidad intra-locutor con la inter-locutor.

3 Base de datos

El corpus utilizado en este trabajo corresponde a archivos de audio de casos cerrados de la Policía de Investigaciones de Chile, registrados entre los años 2006 y 2009. Esta base de datos incluye los audios dubitados e indubitados que se han acumulado a lo largo de estos años, con diferentes calidades y duraciones. Se trata de aproximadamente 3500 archivos de audio.

Los archivos que conforman esta base de datos corresponden principalmente a escuchas telefónicas, grabados con equipos especializados para estos fines, modelo RT8100. Los archivos que se exportan, tienen un formato predeterminado no ajustable, que corresponde a un audio WAV a 8KHz de muestreo y 8 bits de cuantización.

No existe mayor información acerca de las características de formato, duración, calidad o algún otro alcance lingüístico que pudiera tener esta base de datos, por lo tanto, el trabajo acá descrito es la primera aproximación a la caracterización de los locutores presentes en este corpus.

3.1 Selección de archivos de la base de datos

Luego de una clasificación inicial realizada para reordenar los datos, se pueden observar algunas de las características del corpus. En base a esta primera clasificación, se decide considerar las grabaciones dubitadas (canal telefónico), con al menos 1 locutor varón adulto aparentemente chileno, con 20 segundos de duración o más y con una relación señal-ruido de 30 dB o más. Fueron tomadas 40 grabaciones al azar que cumplieran todas estas condiciones.

4 Metodología

Después de seleccionar los archivos de la base de datos completa, se llevó a cabo la segmentación de muestras a analizar.

El español tiene 3 fonemas nasales: bilabial /m/, lingualveolar /n/ y linguopalatal /ɲ/, cuyas grafías en español son m, n y ñ respectivamente, y funcionando en posición silábica explosiva, lo cual permite su oposición. En posición implosiva, se pierde el rasgo distintivo, neutralizándose la /m/ y /n/ y provocando el archifonema /N/, que sólo mantiene la característica nasal [1]. Si además de lo anterior consideramos la posición de la nasal, tenemos 9 posibilidades (3 fonemas en 3 posiciones) más una décima posibilidad que es cuando la nasal se encuentra en posición implosiva (archifonema /N/), por lo tanto, *a priori* tenemos 10 clasificaciones a utilizar.

Se utiliza el software PRAAT [3] para realizar la clasificación de los segmentos por cada archivo de audio, especificando no sólo consonante y posición, si no que también locutor en las grabaciones donde hubieran 2 voces de posible análisis. Se obtienen 1169 segmentos para ser analizados.

Después de la segmentación se extraen los formantes (f1 a f4) y f0 de manera automática, para los 1169 segmentos antes mencionados.

4.1 Comparación estadística por posición

Luego de obtener los datos para cada segmento, agrupados por archivo de audio y por locutor, se establece un mínimo de muestras aceptable para la comparación, considerando la poca cantidad de muestras para algunos locutores, se fija en 3 para poder tener comparaciones en todas las categorías encontradas.

Dado que los datos que se tienen para el análisis no cumplen con los requisitos de homogeneidad de varianza y normalidad que requieren los métodos paramétricos, se lleva a cabo la comparación con el método de Kruskal-Wallis. Se tienen 8 grupos de datos, cuyos resultados de comparación son los siguientes:

Tabla 1. Prueba de Kruskal-Wallis para nasales por posición (f1, f2 y f3).

	f1	f2	f3
Implosiva (/N/)	H(18) = 44,690, p = ,000	H(18) = 52,884, p = ,000	H(18) = 51,641, p = ,000
m prevocálica	H(11) = 25,509, p = ,008	H(11) = 34,498, p = ,000	H(11) = 32,579, p = ,001
m intervocálica	H(35) = 140,065, p = ,000	H(35) = 134,968, p = ,000	H(35) = 164,689, p = ,000
m postvocálica	H(4) = 9,520, p = ,049	H(4) = 9,952, p = ,041	H(4) = 14,372, p = ,006
n prevocálica	H(9) = 19,344, p = ,022	H(9) = 32,888, p = ,000	H(9) = 36,666, p = ,000
n intervocálica	H(32) = 124,463, p = ,000	H(32) = 87,869, p = ,000	H(32) = 126,249, p = ,000
n postvocálica	H(33) = 99,034, p = ,000	H(33) = 127,170, p = ,000	H(33) = 111,887, p = ,000
ñ	H(3) = 10,754, p = ,013	H(3) = 8,248, p = ,041	H(3) = 5,777, p = ,123

Tabla 2. Prueba de Kruskal-Wallis para nasales por posición (f4 y f0).

	f4	f0
Implosiva (/N/)	H(18) = 39,334, p = ,003	H(18) = 36,734, p = ,006
m prevocálica	H(11) = 28,770, p = ,002	H(11) = 36,932, p = ,000
m intervocálica	H(35) = 139,326, p = ,000	H(35) = 174,867, p = ,000
m postvocálica	H(4) = 2,603, p = ,626	H(4) = 12,834, p = ,012
n prevocálica	H(9) = 34,120, p = ,000	H(9) = 27,403, p = ,001
n intervocálica	H(32) = 101,683, p = ,000	H(32) = 120,380, p = ,000
n postvocálica	H(33) = 107,096, p = ,000	H(33) = 95,396, p = ,000
ñ (*)	H(3) = 8,021, p = ,046	H(3) = 9,676, p = ,022

(*) La ñ fue encontrada sólo en posición intervocálica.

4.2 Comparación sin posición

Después de haber realizado las comparaciones anteriores, donde se considera posición, se lleva a cabo una nueva comparación por consonante, pero esta vez sin considerar la posición de esta. Dado que sólo la m y la n se presentan en distintas posiciones, estas nuevas comparaciones son sólo de estas 2 consonantes. Los resultados obtenidos son los siguientes:

Tabla 3. Prueba de Kruskal-Wallis para nasales sin considerar posición (f1, f2 y f3).

	f1	f2	f3
m	H(36) = 177,026, p = ,000	H(36) = 170,841, p = ,000	H(36) = 200,507, p = ,000
n	H(39) = 202,494, p = ,000	H(39) = 232,793, p = ,000	H(39) = 252,105, p = ,000

Tabla 4. Prueba de Kruskal-Wallis para nasales sin considerar posición (f4 y f0).

	f4	f0
m	H(36) = 165,150, p = ,000	H(36) = 219,492, p = ,000
n	H(39) = 242,075, p = ,000	H(39) = 223,390, p = ,000

4.3 Medias por grupos comparados

A modo de caracterización del corpus utilizado, se obtienen los valores promedios de formantes y f0 de los grupos analizados:

Tabla 5 - Medias de formantes y f0 para grupos por consonante y posición.

Grupo	Nº de muestras	f1 (Hz)	f2 (Hz)	f3 (Hz)	f4 (Hz)	f0 (Hz)
/N/	111	382,75	1245,06	2182,36	2925,5	170,02
M pre	77	375,55	1226,99	2131,68	2915,52	149,86
M inter	335	442,2	1155,87	2093,82	2896,19	178,63
M post	43	411,44	1135,4	2167,9	2938,6	180,48
N pre	72	383,93	1301,83	2231,11	2979,32	149,07
N inter	261	417,61	1323,15	2129,51	2836,08	171,38
N post	243	391,49	1240,82	2179,9	2944,61	174,13
ñ	27	383,22	1268,88	2157,19	2972,85	164

Tabla 6 - Medias de formantes y f_0 para grupos sin considerar posición.

Grupo	N° de muestras	f1 (Hz)	f2 (Hz)	f3 (Hz)	f4 (Hz)	f_0 (Hz)
m	455	428,02	1165,97	2107,23	2903,47	173,92
n	576	402,38	1285,75	2163,47	2899,77	169,78
ñ	27	383,22	1268,88	2157,19	2972,85	164

Cabe reiterar que los resultados corresponden a muestras de grabaciones en canal telefónico, de varones adultos y aparentemente chilenos.

5 Análisis

Al realizar la prueba de Kruskal-Wallis para establecer si las diferencias inter-locutores son mayores a las intra-locutor, se obtiene un valor de significación menor o igual a 0,05 en casi todas las categorías antes descritas. Podemos establecer que las diferencias encontradas en la comparación son significativas cuando no se supera este valor (0,05), y por tanto, se rechaza la hipótesis nula de igualdad de medias.

Se observa que no se rechaza la hipótesis nula sólo en f4 de la m en posición postvocálica y en f3 de la ñ. En términos generales hay mayores diferencias entre muestras de distintos locutores que entre muestras del mismo locutor, exceptuando los 2 casos mencionados anteriormente.

En el caso de la ñ, se trata de la clasificación con menor cantidad de segmentos encontrados, los válidos que fueron utilizados en la comparación son solamente 16, por lo tanto, el encontrar un valor de significación mayor a 0,05 puede ser atribuible a un error producto de la poca cantidad de datos. Así mismo, la m en posición postvocálica es la segunda clasificación con menor cantidad de datos (21).

Luego al considerar las nasales sin posición, se observa que en ambos casos (m y n) la significación es del orden de 0,000, por lo tanto, las diferencias entre medias son altamente significativas. El resultado es similar al obtenido al considerar distintas categorías por posición, por lo tanto, al parecer la posición no afectaría significativamente sobre los valores de formantes.

6 Conclusiones

Al observar las comparaciones, se puede decir que en general, los resultados entregan gran variabilidad entre locutores, por sobre la variabilidad entre muestras del mismo locutor. Esto prueba la hipótesis con la que se trabaja, que es que las nasales tienen características altamente locutor-dependiente, por lo tanto, podrían ser relevantes en el reconocimiento forense de locutores.

Aunque no es posible establecer en este estudio cuánto más o menos locutor-dependiente es una nasal comparado a otro sonido del habla, ya que no se analizan otros sonidos, el hecho de que sean estadísticamente significativas las diferencias entre locutores otorga soporte a la hipótesis, y por lo tanto, ya es suficiente para expresar la idoneidad del uso de las nasales en comparaciones con fines forenses.

El uso de esta base de datos, de carácter inédito en Chile resulta ser, en términos generales, bastante provechoso principalmente porque se trata de casos reales, por lo tanto, las condiciones

técnicas debieran ser muy similares a las que se pudieran encontrar en el futuro en muestras de voz. Además no existen registros de experiencias previas que hayan hecho uso de una base de datos de este tipo para trabajos de investigación.

Por último, hay que destacar que los resultados de las tablas 5 y 6 debieran representar de manera fidedigna las características de formantes y f_0 de las nasales de un locutor chileno, en una grabación telefónica.

Agradecimientos

Este trabajo forma parte de la tesis de Magister "Análisis de los espectros nasales en un corpus de casos cerrados de la PDI", la cual a su vez es parte del proyecto FONDECYT 1110742, financiado por el Consejo de Investigación Científica y Tecnológica del Gobierno de Chile CONICYT.

Referencias

- [1] Albalá, M., Análisis y síntesis de las consonantes nasales, *RFE LXXII*, 1991.
- [2] Amino, K., Arai, T., Speaker-dependent characteristics of the nasals, *Forensic Science International* 185, 2009.
- [3] Boersma, Paul & Weenink, David (2010). Praat: doing phonetics by computer [Computer program]. *Version 5.1.25*, 2010. <http://www.praat.org/>
- [4] Farías, F. Características acústicas de las consonantes nasales, *Congreso Internacional de acústica y audio profesional INGEACUS 2011*, Valdivia, 9,10 y 11 de Noviembre de 2011.
- [5] Glass, J., Zue, V., Detection and recognition of nasal consonants in american english, *International Conference on Acoustics, Speech and Signal Processing*, Tokio, 1986.
- [6] Ochoa, F.; San Martín, C.; Carrillo, R. Identificación biométrica de locutores para el ámbito forense: Estado del arte, *VI Congreso Iberoamericano de Acústica - FIA 2008*, Buenos Aires, 5, 6 y 7 de noviembre de 2008.
- [7] Pruthi, T., Espy-Wilson, C., Acoustic parameters for automatic detection of nasal manner, *Speech Communication* 43, 2004.
- [8] Rosas, C.; Sommerhoff, J. La variación acústica del hablante: aplicaciones forenses, *VI Congreso Iberoamericano de Acústica - FIA 2008*, Buenos Aires, 5, 6 y 7 de noviembre de 2008.