

**MARCO DE TRABAJO TECNOLÓGICO PARA LA
REALIZACIÓN DE ESTUDIOS DE CARACTERIZACIÓN
ARTICULATORIA SOBRE IMÁGENES MRI**

**FRAMEWORK FOR THE DEVELOPMENT
OF ARTICULATORY CHARACTERIZATION
STUDIES OVER MRI IMAGES**

JOSE LUIS GARCÍA ARROYO
Universidad de Deusto
jlgarcia@deusto.es

BEGOÑA GARCÍA ZAPIRAIN
Universidad de Deusto
mbgarciazapi@deusto.es

IBON OLEAGORDIA RUIZ
Universidad de Deusto
ibruiz@deusto.es

AMAIA MÉNDEZ ZORRILLA
Universidad de Deusto
amaia.mendez@deusto.es

Artículo recibido el día: 6/03/2013

Artículo aceptado definitivamente el día: 30/07/2013

Estudios de Fonética Experimental, ISSN 1575-5533, XXII, 2013, pp. 367-404

RESUMEN

En este artículo se presenta un marco de trabajo tecnológico innovador diseñado y desarrollado por nuestro grupo de investigación para posibilitar la realización de estudios de caracterización articularia de los sonidos de una lengua a partir de medidas tomadas sobre secuencias de imágenes de cine-MRI. Como elemento fundamental se tiene la herramienta software de producción propia *DicomPas*, que permite realizar la toma de medidas de parámetros articulatorios sobre las secuencias de imágenes MRI y la ejecución de algoritmos *ad hoc* sobre dichas medidas, de cara al procesamiento de los datos, con vistas a la posterior extracción del conocimiento, en forma de generación de inferencias estadísticas o de inteligencia artificial. En estos momentos este marco de trabajo está siendo aplicado a la realización de diversos estudios en euskara y español de Euskadi, disponiéndose para ello de una base de datos con dos repositorios de imágenes tomadas en el plano medio sagital, correspondientes a 18 informantes diferentes.

Palabras clave: *caracterización articularia, cine-MRI, producción del habla, euskara, español.*

ABSTRACT

In this paper an innovative framework is presented, designed and developed by our research team to enable the accomplishment of research works concerning the articulatory characterization of the sounds of a language from measures taken over MRI image sequences. As fundamental element there is the *DicomPas* software tool, developed by our team, which allows to carry out the measures of articulatory parameters over the MRI image sequences and the execution of *ad hoc* algorithms over such measures, facing the data processing, with the view to the subsequent extraction of knowledge, in the form of the generation of statistical or artificial intelligence inferences. This framework is currently being applied to the achievement of diverse studies in Basque and Spanish of the Basque Country. To do so, a database with two repositories of images taken in the midsagittal plane, corresponding to 18 different informants, is available.

Keywords: *articulatory characterization, cine-MRI, speech production, Basque, Spanish.*

1. INTRODUCCIÓN

En los últimos años ha cobrado gran importancia en la investigación de la producción del habla el estudio de la caracterización articulatoria. Para ello, la técnica más usada es MRI (*Magnetic Resonance Imaging*) (Hornak, 1996), una tecnología de adquisición de imágenes que cumple simultáneamente las condiciones de generar imágenes de alta calidad en la captura de la información anatómica, tanto en el espacio bidimensional (2D) como tridimensional (3D), y ser no invasiva y segura para el informante. En este trabajo se presenta un marco de trabajo tecnológico para la realización de estudios de caracterización articulatoria sobre imágenes MRI.

Otras técnicas también utilizadas para la caracterización articulatoria son EMMA (*Electromagnetic Midsagittal Articulography*) (Romero, 2008), que permite obtener los datos cinemáticos relativos a diferentes puntos del aparato articulatorio en los que se han puesto sensores y EPG (*Electropalatography*) (Fernández Planas, 2008), que permite registrar los contactos de la lengua con el paladar. En ambos casos se obtiene información de interés, con muy buena resolución temporal. Sin embargo, existen ciertas desventajas. En el caso de EMMA, el proceso es invasivo y además la información está limitada a la trayectoria de los puntos elegidos; en el caso de EPG se trata de una técnica también invasiva y la información está limitada a los contactos de la lengua con el paladar, resultando además difícil en muchas ocasiones el ajuste del paladar artificial.

La ventaja de MRI sobre estas técnicas es ser una técnica no invasiva, que ofrece buen contraste entre regiones blandas, y que además proporciona información de todo el tracto vocal, incluida la cavidad laríngea y la nasal, de las cuales no se puede obtener información con EMMA y EPG. Como desventajas se tienen la imposibilidad de obtener información de los dientes debido a la ausencia en ellos de protones de hidrógeno, aunque existen técnicas indirectas para paliar este problema, como se puede ver en Takemoto *et al.* (2004), la obligación de estar tumbado durante el proceso de captura de imágenes, que a causa del efecto de la gravedad puede tener alguna influencia en la posición de la lengua y el volumen de la cavidad faríngea (Engwall, 2003a; Tiede *et al.*, 2000; Takemoto *et al.*, 2006), y la baja resolución temporal obtenida, además por supuesto de lo molesto que resulta para el informante. En cualquier caso, las tres tecnologías se complementan entre sí, aportando cada una información suplementaria a las demás; buenos ejemplos de estudios realizados con las tres técnicas se muestran en Engwall (2000); Engwall (2003b); Narayanan, Byrd y Kaun (1999).

Tal y como se explica en Narayanan *et al.* (2004), usando MRI se pueden realizar dos tipos de adquisiciones de imágenes: la estática, realizando una única captura, y la dinámica o de tiempo real, realizando la captura de una secuencia de imágenes durante el lapso de tiempo en el que se está produciendo el sonido, técnica conocida como cine-MRI.

Las adquisiciones estáticas de imágenes MRI se realizan si se desea llevar a cabo únicamente el estudio articulatorio de la morfología del tracto vocal, sin tener en cuenta su aspecto temporal, y permiten conseguir una alta calidad 3D. No obstante, tienen algunas desventajas; para conseguir esa buena calidad el informante se ve obligado a estar bastante tiempo sosteniendo la misma posición articulatoria (varios minutos en el estado actual de la tecnología MRI), lo cual puede reforzar el efecto de la gravedad en la posición de la lengua y el volumen de la cavidad faríngea, efectos ya comentados. Asimismo, el estar en un ambiente ruidoso y una posición articulatoria sostenida, puede derivar en un estado articulatorio alejado de lo normal (Engwall, 2003a; Takemoto *et al.*, 2006), y además, al carecerse de diferentes ejemplares con la información dinámica, si no se captura la articulación en la buena posición, es posible que la prueba no resulte del todo correcta.

Las adquisiciones dinámicas de imágenes, con la técnica cine-MRI, son necesarias si lo que se desea es realizar el estudio articulatorio en movimiento. Además, permiten en muchas ocasiones mejorar la descripción de la morfología del tracto vocal, ya que al disponerse de varios ejemplares sobre el mismo volumen, se puede centrar mejor el proceso con respecto a las estáticas. Como desventaja se tiene que, de esta manera, no se puede conseguir una calidad tan alta como en las estáticas, de hecho en el estado actual de la tecnología MRI éstas han de ser forzosamente de tipo 2D, o aproximaciones a 3D de baja calidad. La adquisición dinámica de imágenes está fuertemente restringida por las propias limitaciones técnicas de los dispositivos de captura, no obstante cada vez se están obteniendo mejores resultados, gracias a las mejoras tecnológicas, que posibilitan un menor tiempo entre imagen e imagen. Algunos de los trabajos más relevantes que se están realizando para la aceleración de este proceso se tienen en: Donoho (2006); Narayanan *et al.* (2004); Lustig, Donoho y Pauly (2007); Kim; Narayanan y Nayak (2009a) y Kim, Narayanan y Nayak (2009b).

Los estudios de caracterización articulatoria pueden ser realizados bien sobre 2D, la mayoría de las veces sobre el plano medio sagital, o bien sobre 3D. En Fitch y Giedd (1999) y Story, Titze y Hoffman (1996), se presentan dos excelentes estudios clásicos de 2D sobre el plano medio sagital y 3D, que presentan una orientación metodo-lógica que ha servido de guía para muchas de las investigaciones posteriores en ambos enfoques.

Las medidas sobre 3D son más exactas y completas, ya que el espacio real es tridimensional. No obstante, tienen algunas desventajas. En primer lugar, como ya se ha comentado anteriormente, la adquisición de imágenes es más costosa, ya que la generación de volúmenes 3D exige la captura de múltiples imágenes 2D, tema especialmente crítico si se pretende tomar múltiples imágenes en un lapso pequeño de tiempo. En segundo lugar, la toma de medidas en 3D es mucho más compleja.

Por ello, para la realización de muchas investigaciones, los fonetistas prefieren la toma de medidas sobre imágenes 2D (Beautemps *et al.*, 1996), normalmente sobre el plano medio sagital, que además de ser más sencillas se pueden considerar buenas estimaciones de las medidas reales en 3D, tal y como se ha confirmado en diferentes estudios (Badin *et al.*, 1998; Badin *et al.*, 2002), en los cuales se demuestra que para la mayoría de los problemas se cumple que los parámetros articulatorios del modelo 2D son buenos predictores de las formas 3D descritas. Además, la mayor parte de los valores de medidas tomadas en los últimos años a nivel mundial, en diferentes idiomas y para diferentes sonidos, están realizadas sobre imágenes 2D, especialmente del plano medio sagital, lo que permite contextualizar los experimentos realizados en un marco más robusto. En cualquier caso, no hay que perder de vista que el espacio real es 3D. De hecho, existen algunos aspectos de la caracterización articuladora que no son capturados en 2D y que se encuentran realmente en 3D, como se comenta en Badin y Serrurier (2006), lo cual también puede paliarse en cierta manera realizando diferentes capturas de 2D en los planos sagital, coronal o axial, para dar una perspectiva más completa, como se realiza en varios trabajos.

Varios de los estudios realizados hasta la fecha en diferentes idiomas utilizando la tecnología MRI, se tienen en: inglés (Baer, 1991; Narayanan, Alwan y Haker, 1997; Alwan, Narayanan y Haker, 1997; Narayanan *et al.*, 2004; Zhou, 2009; Zhou *et al.*, 2010), tamil (Narayanan, Byrd y Kaun, 1999; Narayanan *et al.*, 2004; Karthikeswaran y Dinakar, 2011), japonés (Tiede, Masaki y Vatiokis-Bateson, 2000; Takemoto *et al.*, 2004; Takemoto *et al.*, 2006), coreano (Yang, 1999), italiano (Romano y Badin, 2009), francés (Beautemps, Badin y Bailly, 1996; Badin *et al.*, 1998; Badin *et al.*, 2002; Serrurier y Badin, 2005, 2008), alemán (Hoole y Mooshammer, 2002), portugués (Martins *et al.*, 2008) y sueco (Engwall y Badin, 1999).

Aparte de los trabajos realizados por nuestro grupo de investigación, en lo que respecta al español, el único trabajo de caracterización articuladora sobre MRI realizado hasta la fecha, al menos del conocimiento de los autores, se tiene en Gurlekian *et al.* (2004), realizado sobre español de la Argentina. En lo que respecta al euskara no existe ninguno.

Habiendo detectado la carencia de estudios de caracterización articulatoria a partir de las modernas técnicas experimentales que hay en el mundo científico acerca del euskara y español dos equipos de la Universidad de Deusto llevan trabajando desde 2005 en un proyecto de investigación conjunto, cada uno desde su especialidad. Uno de los equipos está compuesto por investigadores de fonética, que conocen este campo de estudio y son los que guían la investigación. El otro está compuesto por investigadores de ingeniería, especializados en ingeniería del software, matemáticas, inteligencia artificial y procesamiento digital de imagen. Gracias al trabajo de ambos grupos se han realizado varias investigaciones de descripción articulatoria sobre sonidos del euskara y español (Elejabeitia *et al*, 2009; Iribar *et al*, en prensa, 2013; Pagola *et al*, 2012; Iribar, 2012; Iribar, 2013) y se ha creado el marco de trabajo tecnológico que se presenta en este trabajo, sobre la base de la herramienta software integrada de producción propia *DicomPas*, de cara al estudio de la caracterización articulatoria de los sonidos en ambos idiomas, aunque evidentemente puede ser también utilizado para estudiar sonidos de otras lenguas.

En cuanto a las muestras adquiridas, sopesando todos los factores comentados anteriormente relativos a la adquisición de imágenes, tanto los estudios llevados a cabo hasta la fecha como el marco de trabajo tecnológico construido se han basado en secuencias de imágenes 2D tomadas en el plano medio sagital. No obstante, en el futuro se mejorará la descripción articulatoria escalando el marco de trabajo para posibilitar la realización de estudios sobre imágenes 3D, como se comentará más adelante.

Esta publicación ha sido redactada desde el punto de vista del marco de trabajo tecnológico. No es, por tanto, el objetivo aquí el profundizar en los temas fonéticos ni en los resultados específicos que sobre diferentes sonidos se han obtenido, sino la descripción del sistema para la obtención de los datos cuantitativos o, lo que es lo mismo, en cómo posibilita este marco de trabajo la realización de investigaciones de caracterización articulatoria sobre los sonidos.

Las aportaciones que se presentan en este trabajo son, por tanto, el propio marco de trabajo tecnológico, en el que se describe una metodología en la que se establecen formalmente los pasos para la realización de investigaciones de caracterización articulatoria y, sobre todo, la herramienta software innovadora *DicomPas*, que proporciona el soporte tecnológico para dicho marco, implementando buen número de funcionalidades y algoritmos de procesamiento digital de imagen y tratamiento de datos.

Se trata de un trabajo innovador, no existiendo en la literatura científica, al menos en el conocimiento de los autores, ninguna herramienta software similar a *DicomPas*, ni ningún marco de trabajo de propósito general de este tipo para la realización de investigaciones de caracterización articular. En todos los trabajos de descripción articular del estado del arte que se han estudiado, los diferentes autores realizan las tareas apoyándose en programas de tipo general de procesamiento digital de imagen y de tratamiento de datos y programando rutinas *ad hoc* en los casos necesarios. Si disponen de una infraestructura tecnológica de este tipo no la dan a conocer.

Este trabajo se ha dividido en los siguientes apartados: en el apartado 2 se explican las bases de datos de imágenes MRI que se manejan; en el apartado 3 se describe el diseño del sistema en su conjunto; en el apartado 4 se ilustra cómo se realizan las investigaciones de caracterización articular usando este marco de trabajo, en casos reales; en el apartado 5 se presentan las conclusiones; finalmente, en el apartado 6 se plantean las líneas futuras de trabajo.

2. BASE DE DATOS DE IMÁGENES MRI

Este apartado se ha dividido en dos subapartados; en primer lugar se explican las técnicas que se han empleado y en segundo lugar se describe la base de datos de imágenes que se tiene actualmente para la realización de los estudios, consistente en dos repositorios distintos de imágenes y un tercero en construcción, justificando su diseño a la luz de lo comentado en la introducción.

2.1. Técnicas empleadas

2.1.1. MRI

MRI (*Magnetic Resonance Imaging*) es una técnica de imagen frecuentemente usada en medicina, que utiliza el fenómeno de la resonancia magnética para obtener información sobre la estructura y composición de la parte del cuerpo a analizar, por medio de un escáner MRI, que a continuación es procesada por ordenadores y transformada en imágenes del interior de lo que se ha analizado y permite visualizar la información en 3D o desde diferentes perspectivas 2D (Hornak. 1996).

Se trata de un gran avance con respecto a las antiguas radiografías 2D de rayos X, habiendo cobrado en los últimos años gran importancia en la descripción articular.

latoria, ya que cumple simultáneamente las condiciones de ofrecer muy buenos resultados de resolución espacial y no presentar problemas de radiación, como es el caso de la CT (*X-ray Computed Tomography*) (Herman, 2009), una técnica de imagen de características similares, que lleva a ofrecer incluso mejores calidades de imagen, pero que es demasiado invasiva para el paciente, que es obligado a recibir grandes dosis de rayos X, y que por supuesto no se debe ni plantear su uso en los estudios de caracterización articularia.

Las imágenes generadas son almacenadas en formato DICOM.

2.1.2. DICOM

DICOM (*Digital Imaging and Communication in Medicine*) es el estándar a nivel mundial para el almacenamiento y transmisión de imágenes médicas. Los ficheros DICOM consisten en una cabecera y un cuerpo. En la cabecera se almacenan datos relativos a la adquisición de imágenes, como los datos del paciente, la fecha en la que se ha realizado el estudio, el aparato que se ha usado para la captura, la configuración que se ha utilizado, los datos de resolución espacial de las imágenes contenidas, etc. En el cuerpo se almacenan las propias imágenes tomadas, admitiéndose diferentes estándares de formatos de compresión (NEMA, 2013).

2.1.3. Opciones para el escáner MRI: generación 3D y diferentes perspectivas 2D

Para la generación de volúmenes 3D se capturan con el escáner múltiples imágenes 2D, a una distancia muy pequeña entre ellas, que actúan como múltiples rodajas o cortes 2D, que superpuestas permiten generar la información 3D del objeto.

No obstante, tal y como se ha comentado anteriormente, la generación 3D tarda mucho tiempo, por lo que frecuentemente se capturan imágenes 2D, pudiéndose configurar el escáner para obtener la perspectiva 2D deseada del objeto. Se pueden elegir perspectivas axiales (también llamadas horizontales o transversales), coronales (también llamadas frontales) o sagitales, o incluso oblicuas con respecto a dichos planos; no obstante, la perspectiva 2D más comúnmente utilizada en el caso del análisis del tracto vocal es el plano medio sagital. En la figura 1 se pueden ver las diferentes perspectivas y en la figura 2 el plano medio sagital.

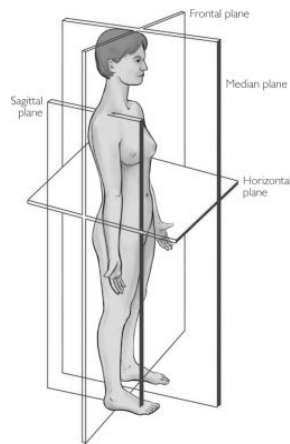


Figura 1. Planos del cuerpo humano. Fuente: *The Free Dictionary* (2013a)

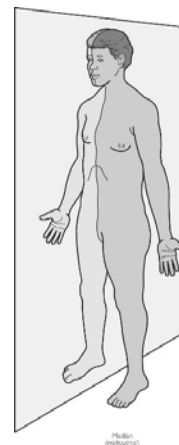


Figura 2. Plano medio sagital. Fuente: *The Free Dictionary* (2013b)

2.1.4. Cine-MRI

Como se ha comentado anteriormente, el cine-MRI consiste en la toma de una secuencia de imágenes MRI en un lapso de tiempo, lo cual permite analizar el comportamiento espacio-temporal del aparato articulario en la producción del sonido. En el presente trabajo se capturan secuencias de imágenes 2D del plano medio sagital. Un ejemplo de imagen se tiene en la figura 3:

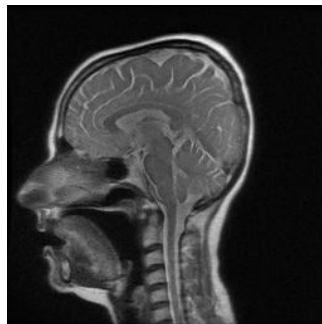


Figura 3. Imagen en el plano medio sagital. Informante pronunciando el sonido [l].

2.2. Base de datos

2.2.1. Descripción

En la construcción de la base de datos de imágenes se han tenido en cuenta las consideraciones realizadas en la introducción, y en base a esto se han generado hasta ahora dos repositorios diferentes de imágenes, a los que se añade un tercero en construcción.

Respecto a la elección entre realizar una adquisición de información 2D o 3D, tras un estudio detallado del estado de la cuestión, y tras sopesar los diferentes factores comentados en la introducción, se resolvió que, al ser las medidas tomadas en imágenes 2D del plano medio sagital buenas aproximaciones a lo que sucede en 3D, y reportando además la gran mayoría de los estudios de caracterización articulatoria valores cuantitativos relativos a imágenes de este tipo, lo más adecuado era trabajar con imágenes 2D del plano medio sagital, al menos como punto de partida en nuestras investigaciones.

Sobre el tipo de captura, se eligió una adquisición dinámica, de cine-MRI, intentando obtener la máxima calidad posible, con el menor intervalo de tiempo entre imagen e imagen, teniendo en cuenta que la resolución y la velocidad de captura son variables inversas entre sí. Los parámetros más adecuados se determinaron tras realizar un estudio previo de análisis de valores.

Es en el diseño de las pruebas donde existen diferencias entre los dos repositorios de imágenes de la base de datos con los que se está trabajando actualmente, tal y como se comentará a continuación. En las pruebas correspondientes al primer repositorio de imágenes se ha realizado la captura durante la producción del sonido en el interior de una palabra, de cara a analizar su pronunciación en la manera en que normalmente se produce. Es la manera más natural, y es obvio que en principio debería ser la adecuada para realizar este tipo de experimentos, no obstante presenta inconvenientes de índole técnica puesto que la adquisición de imágenes no puede hacerse de manera ultrarrápida, con múltiples capturas por segundo, y por tanto es difícil obtener la información relativa a la articulación del sonido en su momento central. Por esta razón el diseño de las pruebas del segundo repositorio de imágenes se ha hecho de tal manera que, en la medida de lo posible, se pudiese capturar el momento central del sonido correspondiente el máximo número de veces, de cara a obtener el máximo de muestras relativas a la articulación de dicho sonido. Para ello, se ha intentado que la pronunciación de cada uno de los sonidos se haga en posición fonética normal y de manera sostenida durante el mayor tiempo posible, requiriendo a los informantes que, mientras pronunciaban el sonido,

visualizasen éste mentalmente como pronunciado dentro de una palabra. El problema que se tiene es que, obviamente, este tipo de articulación es algo más artificial.

Por tanto, ambos conjuntos de imágenes tienen sus ventajas e inconvenientes, y se complementan entre sí, constituyendo una base de datos completa y única, de gran valor para el estudio del euskara y español, no existiendo bases de datos de características similares en ambos idiomas, al menos en el conocimiento de los autores. Para darse cuenta de la magnitud de lo que estamos hablando, señalaremos que la mayoría de los estudios de caracterización articulatoria sobre imágenes MRI tienen un número muy pequeño de informantes. A modo de ejemplo: uno (Story *et al*, 1996; Yang, 1999; Engwall y Badin, 1999; Takemoto *et al*, 2006; Martins *et al*, 2008; Gurlekian *et al*, 2004; Badin y Serrurier, 2006; Takemoto *et al*, 2004), dos (Tiede *et al*, 2000), tres (Romano y Badin, 2009) y cuatro (Narayanan y Alwan, 1995; Narayanan, Alwan y Haker, 1997; Alwan, Narayanan y Haker, 1997). Actualmente nuestra base de datos, tiene un total de 18 informantes diferentes, con gran número de imágenes por cada uno.

La figura 4 ilustra gráficamente el esquema de la base de datos de imágenes MRI.

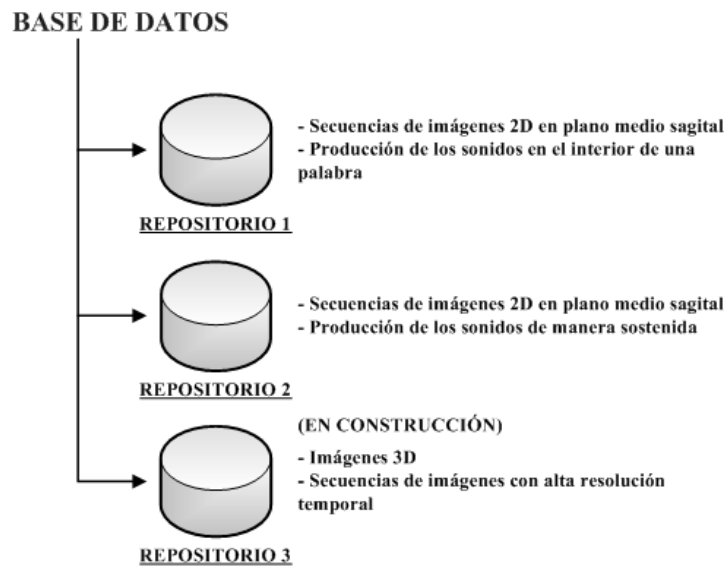


Figura 4. Repositorios de la base de datos.

Además, se está empezando a construir un tercer repositorio de imágenes, en el que se está realizando la adquisición de información 3D con el objetivo de extender la descripción articuladora a 3D y, por otra parte, se está realizando la adquisición de imágenes con alta resolución temporal, con el objetivo de poder mejorar la caracterización del comportamiento dinámico en la producción del sonido. En cualquier caso, se trata de un conjunto de imágenes en construcción, que aún no es usado en este marco de trabajo tecnológico, aunque sí lo será en futuras versiones a las que escale.

2.2.2. Repositorio de imágenes 1

Las grabaciones se realizaron a seis informantes (4 hombres y dos mujeres de distintos dialectos vascos), todos ellos adultos. En todos los casos se llevaron a cabo análisis acústicos previos que mostraron la idoneidad de la producción lingüística; todos los informantes fueron además adecuadamente preparados para la sesión de grabación. Se realizó una sesión previa de grabación con un informante para adecuar correctamente todos los parámetros (técnicos, lingüísticos y procedimentales). La encuesta consistió en la producción de logatomos del tipo [a]+[cons]+[a], con todas las consonantes de la variedad estándar del euskara, y acentuación oxítona.

La grabación se realizó en las instalaciones de Osatek, S.A, en el hospital público de Galdakao (Bizkaia). El equipo de resonancia magnética utilizado fue un 1,15 T *SIEMENS Magnetom Symphony Quantum Maestro Class* (2002, con actualizaciones posteriores). Las imágenes se adquirían a 6 segundos por fotograma de cine-MRI, en formato DICOM. El informante repetía el logatomo durante el tiempo de exposición a la resonancia. Se grabaron imágenes en el plano medio sagital, sin adquisición de audio.

2.2.3. Repositorio de imágenes 2

Se seleccionaron 6 hablantes nativos de español para las pruebas en español y 6 hablantes nativos de euskara para las pruebas en euskara, todos ellos individuos adultos, estudiantes y profesores de la Universidad de Deusto, con los cuales se realizaron sesiones previas de entrenamiento en el laboratorio de fonética. A cada uno de los informantes se le sometió a una batería de pruebas, cada una de ellas relativa a la articulación de un sonido diferente, siendo 34 el número total de sonidos.

En cada una de estas pruebas, tanto para las vocales como para las consonantes, el individuo repetía una palabra que contenía el sonido, articulando a continuación el sonido de manera sostenida durante un lapso de tiempo, y visualizando éste mentalmente como pronunciado dentro de dicha palabra. Se capturaban diferentes instantes de esta acción en una secuencia de 24 fotogramas de cine-MRI, en formato DICOM y sin adquisición de audio.

Las sesiones de grabación de las imágenes de MRI se realizaron en el Hospital Quirón de Bizkaia, mediante un equipo GE 1,5T. –HealthCare–, que obtiene secuencias dinámicas SSFSE (*Single Shot Fast Spin Eco*) con TR de 4,8 y TE de 1,3 con 24 fotogramas por minuto en el plano medio sagital.

2.2.4. En construcción: Repositorio de imágenes 3

Para el escalamiento de las investigaciones de caracterización articulatoria, y por supuesto del propio marco de trabajo tecnológico, se está construyendo un tercer repositorio de imágenes, con el objetivo de recopilar dos tipos de estudios.

En primer lugar, se quiere realizar la adquisición de información 3D de diferente tipo con el objetivo de realizar una caracterización en tres dimensiones de la producción del habla.

De hecho, en estos momentos se ha realizado la toma de imágenes 3D a una hablante de español, a la cual se le hicieron diferentes pruebas de captura de imágenes 3D de alta calidad, realizando diferentes configuraciones del escáner en la toma de imágenes, y comprobando la calidad de las mismas en cada estudio, a partir del análisis de su resolución espacial y el grado de contraste entre los tejidos que forman el aparato articulatorio. Realmente la prueba fue bastante incómoda para esta informante, que tuvo que mantener de manera sostenida la posición de articulación durante varios minutos en cada una de las diferentes pruebas realizadas. A partir de ahora, en la construcción de este repositorio se estudiará la configuración óptima para intentar capturar información 3D en un tiempo menor, evidentemente a costa de una peor calidad, intentando buscar el equilibrio adecuado entre calidad y tiempo. En cualquier caso, esta primera captura estática se considera de gran valor, ya que permitirá tener un modelo 3D de referencia para los posteriores análisis.

En segundo lugar, se quieren tomar imágenes 2D y 3D con la máxima resolución temporal posible, con el objetivo de modelar el comportamiento dinámico. Tal y

como se ha comentado con anterioridad, la calidad de imagen y la velocidad de captura son variables inversas entre sí, es decir, a mayor calidad menor velocidad y viceversa. Por ello, se trabajará en la conjugación de ambos factores, combinando imágenes lo más rápidas posibles con imágenes de mayor calidad, de cara a modelar lo más fielmente posible el comportamiento dinámico de la producción del sonido.

3. DISEÑO DEL SISTEMA

En este apartado se presenta el diseño del marco de trabajo tecnológico desarrollado por nuestro equipo de trabajo. Este diseño se presenta en cinco sub-apartados: en primer lugar se dará una vista de alto nivel del sistema en su conjunto, explicando su funcionamiento, en segundo lugar se describirá la herramienta *DicomPas*, detallando sus principales funcionalidades, en tercer lugar se explicará la toma de medidas usando dicha herramienta, en cuarto lugar se mostrará la funcionalidad de los algoritmos *ad hoc* contenidos en dicha herramienta y, finalmente, en quinto lugar, se tratará brevemente la extracción del conocimiento a partir de los datos.

3.1. Vista de alto nivel

Este marco de trabajo tecnológico está diseñado para posibilitar la realización de estudios de caracterización articular sobre imágenes MRI.

En la figura 5 se presenta una vista de alto nivel. Tal y como se puede ver, una vez obtenidas las imágenes, proceso descrito en el apartado correspondiente a la base de datos, la realización de un estudio consta de 3 etapas:

1. Toma de medidas: se toman medidas en las imágenes MRI.
2. Algoritmos *ad hoc* para el procesamiento de los datos: se ejecutan algoritmos diseñados para el procesamiento de los datos para ese problema en concreto. Esta etapa es opcional.
3. Extracción del conocimiento: a partir de los datos obtenidos se extrae el conocimiento, apoyándose en técnicas estadísticas o de inteligencia artificial.

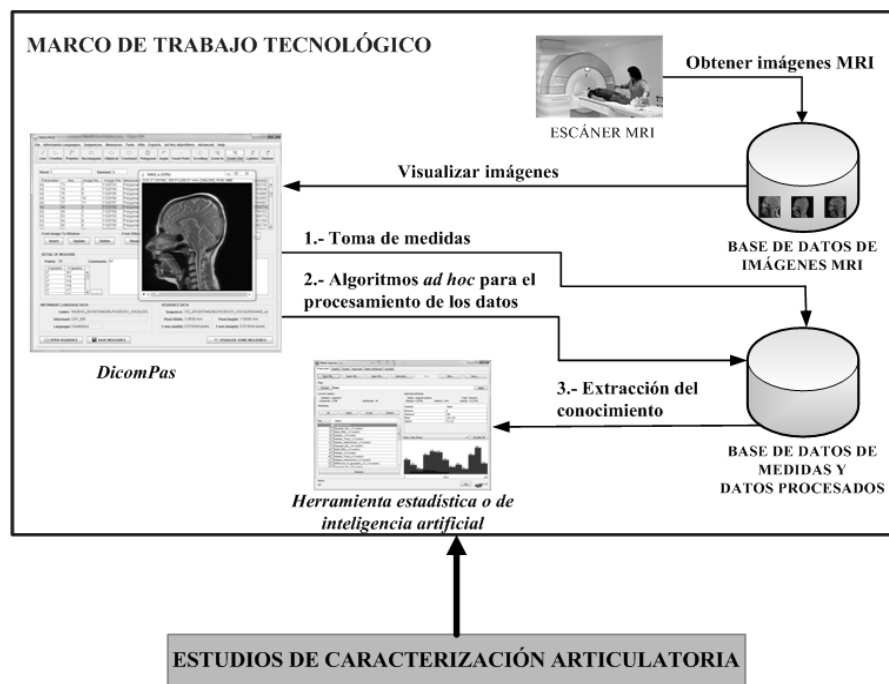


Figura 5. Vista de alto nivel del marco de trabajo tecnológico.

Existen dos bases de datos, la primera destinada al almacenamiento de imágenes MRI, ya descrita en el apartado anterior, y la segunda destinada al almacenamiento de las medidas tomadas y de los datos calculados por los algoritmos *ad hoc*, que será utilizada en la posterior extracción de conocimiento. Las tareas de la toma de medidas y la ejecución de los algoritmos *ad hoc* son realizadas por la herramienta software *DicomPas*, además de otras funcionalidades complementarias, mientras que la extracción del conocimiento es realizada utilizando aplicaciones software para estadística o inteligencia artificial, como el SPSS (IBM, 2013) o el WEKA (The University of Waikato, 2013).

3.2. Herramienta software *DicomPas*

La herramienta software *DicomPas* es una aplicación informática destinada a los fonetistas cuyas principales tareas son, como se ha comentado, la toma de medidas

y la ejecución de los algoritmos *ad hoc* para el procesamiento de datos. Además, provee de funcionalidades de gestión básicas y diferentes utilidades para facilitar el trabajo. Está escrita en lenguaje de programación Java (Oracle-Sun Microsystems, 2013), utilizando para el manejo de imágenes la plataforma ImageJ (U.S. National Institutes of Health, 2013).

Para que el lector se pueda hacer una idea de la apariencia de la herramienta software, en la figura 6 se muestra la pantalla principal de la aplicación. Las principales funcionalidades que proporciona la herramienta son:

1. *Gestión de los informantes y secuencias*: Se gestionan los informantes y las secuencias de imágenes. En el menú, las opciones: «Informant-Language» y «Sequences».
2. *Toma de medidas*: Se gestiona la toma de medidas sobre las imágenes. En el menú, las opciones: «Measures» y «Tools». La explicación detallada de la toma de medidas se abordará más adelante.
3. *Procesamiento digital de imagen*: Para facilitar la tarea de toma de medidas, la herramienta software provee de utilidades de procesamiento digital de imágenes, como son el zoom, el brillo, la detección de bordes, la inversión, y diferentes tipos de filtros. En el menú, la opción: «Utils».
4. *Export*: La herramienta software permite exportar los valores de las medidas a diferentes formatos, con diferentes criterios de selección y ordenación de los datos. En el menú, la opción: «Export». Se tiene:
 - a. *XML*: a XML (*World Wide Web Consortium W3C*, 2013b) para el intercambio de información con otras plataformas. Se sigue una sintaxis (XML-Schema, *World Wide Web Consortium W3C*, 2013c) definida en este proyecto.
 - b. *Excel*: a Excel (Microsoft, 2013a), que posibilita la realización de análisis estadísticos, tanto con Excel como con otras herramientas de análisis de datos como SPSS (IBM, 2013).
 - c. *JPG*: a formato JPG (*World Wide Web Consortium W3C*, 2013a), las medidas, bien solas, bien pintadas encima de las imágenes, una o varias, con el fin de poderse consultar gráficamente.

- d. *RTF*: a formato RTF (Microsoft, 2013b).
5. *Algoritmos ad hoc*: Algoritmos *ad hoc* sobre los datos de medidas en bruto, obteniéndose datos calculados. En el menú, la opción: «Ad Hoc Algorithms». La explicación detallada se abordará más adelante.
6. Otras opciones del menú son: «File» (operaciones de fichero básicas), «Advanced» (opciones avanzadas de configuración) y «Help» (ayuda).



Figura 6. Ventana principal de la aplicación DicomPas.

3.3. Toma de medidas

La toma de medidas sobre las imágenes de las secuencias MRI se realiza utilizando la herramienta software *DicomPas*.

Es importante señalar que, previamente a la labor de toma de medidas, y de acuerdo a los objetivos de la investigación de caracterización articular que se esté desarrollando, se debe realizar una labor de reflexión sobre lo que se quiere medir (distancia entre labios, área de la faringe, etc.), lo que en el vocabulario del presente marco de trabajo tecnológico se denomina «parámetros articulatorios». Existe un repositorio de parámetros articulatorios ya definidos tras diferentes investigaciones.

Para la toma de una medida, una vez abierta la secuencia de imágenes, tal y como se ve en la figura 6, se elige el tipo de medida y se toma manualmente sobre la imagen seleccionada. Existen medidas de tipo *line*, *freeline*, *polyline*, *rectangular*, *elliptical*, *freehand*, *polygonal*, *angle* y *fixed point*, que se explicarán más adelante. Una vez tomada la medida, ésta se guarda, almacenándose en la base de datos de medidas. Las medidas de tipo *freeline* y *freehand* no se guardan como tal, sino que, como se explicará más adelante, se convierten antes de ser guardadas a medidas equivalentes, de tipo *polyline* y *polygonal* respectivamente. Tal y como se puede ver en la figura 7, por cada medida se guardan:

1. *Campos generales*. Estos campos son comunes a todo tipo de medidas.
2. Campos para el *almacenamiento del tipo de medida*. Estos campos son diferentes para cada tipo de medida.
3. Campos que contienen las *características de la medida*. Estos campos son diferentes para cada tipo de medida.

Los campos generales son aquellos comunes a todo tipo de medidas, relativos al informante, secuencia, imagen, tipo de medida y parámetro articulatorio que se quiere medir.

Los campos para el almacenamiento de la medida son aquellos especificados por los diseñadores de la herramienta para guardar cómo ha sido tomada la medida. En la tabla 1 se puede ver la composición de dichos campos para los diferentes tipos de medidas.

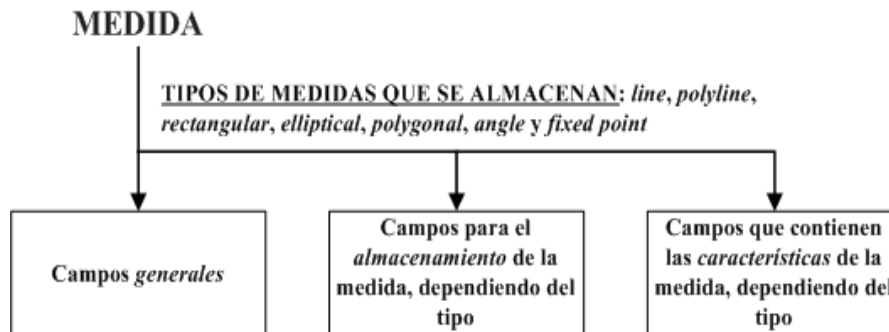


Figura 7. Tipos de campos de una medida.

Las características de la medida son propiedades cuantitativas de la medida, que permiten caracterizar matemáticamente en una imagen el parámetro articulatorio relativo a dicha medida. En la tabla 1 se pueden ver las características de los diferentes tipos de medidas. El manejo estadístico o por técnicas de inteligencia artificial de los valores de las características permite la extracción del conocimiento. No obstante, en muchas ocasiones, si se desea llegar a extraer conclusiones de cierta complejidad, será obligado realizar algún tipo de procesado *ad hoc*, como se verá más adelante.

A continuación, una vez explicados los conceptos básicos, se explicarán los diferentes tipos de medida que pueden ser tomados utilizando la herramienta *DicomPas*, detallando su almacenamiento y las características asociadas:

1. *Line*: líneas rectas tomadas entre dos puntos, que se usan para su almacenamiento. Su característica es la *longitud*.
2. *Freeline*: curvas tomadas a mano alzada. Para su almacenamiento se realiza un proceso de conversión de *freeline* a *polyline*, pudiéndose elegir el número de puntos que la describen. Por tanto, no existen medidas guardadas de tipo *freeline* como tal.
3. *Polyline*: curvas determinadas por N puntos, que se usan para su almacenamiento. Su característica es la *longitud*.

4. *Rectangular*: regiones con forma de rectángulos. Para su almacenamiento se usa un punto y la anchura y altura del rectángulo. Sus características son el *perímetro* y el *área*.
5. *Elliptical*: regiones con forma de elipses. Para su almacenamiento se usa un punto y la anchura y altura del rectángulo que circunscribe a la elipse. Sus características son el *perímetro* y el *área*.
6. *Freehand*: regiones tomadas a mano alzada. Para su almacenamiento se realiza un proceso de conversión de *freehand* a *polygonal*, pudiéndose elegir el número de puntos que la describen, de la misma forma que se hace en el caso de *freeline*. Por tanto, no existen medidas guardadas de tipo *freehand* como tal.
7. *Polygonal*: regiones que se corresponden con polígonos de N puntos, que se usan para su almacenamiento. Sus características son el *perímetro* y el *área*. Para poder realizar una descripción cuantitativa de otras propiedades de la medida se han añadido, como características avanzadas: el *diámetro de Feret*, la *relación de aspecto*, la *circularidad* y la *solidez*. Se trata de propiedades de tipo geométrico que calculan diferentes aspectos de la región poligonal. Se comentan brevemente más adelante.
8. *Angle*: ángulos tomados. Para su almacenamiento se usan los 3 puntos que se usan para definir el ángulo. Su característica es el *ángulo*.
9. *Fixed point*: se trata de un punto de interés, que se usa para su almacenamiento. No tiene características.

Respecto a las características avanzadas de las medidas de tipo *polygonal*, antes de comenzar su descripción se definirán tres conceptos:

1. *Región convexa*: es una región que cumple la condición de que todos sus puntos pueden ser unidos por una línea recta contenida en dicha región.
2. *Envolvente convexa de un polígono*: es el polígono convexo más pequeño que contiene dicho polígono.
3. *Elipse ajustada a un polígono*: es la mínima elipse que contiene la envolvente convexa del polígono.

A partir de estos conceptos, se describen las características avanzadas:

1. *Diámetro de Feret*: mide el diámetro máximo del polígono. Formalmente, se define como la máxima de las distancias de los segmentos que unen los puntos del polígono, tal y como se puede ver en la fórmula:

$$D_F = \max\{d(p_1, p_2) \mid p_1, p_2 \in P\}$$

donde la función d es la función de distancia entre puntos y P es el polígono.

2. *Relación de aspecto*: mide la relación entre la anchura y la altura del polígono. Formalmente, se define como la relación entre la anchura y la altura de la elipse ajustada al polígono, tal y como se puede ver en la fórmula:

$$R_A = \text{anchura}(\varepsilon(P)) / \text{altura}(\varepsilon(P))$$

donde ε es la elipse ajustada y P es el polígono.

3. *Circularidad*: es el grado de circularidad del polígono. Formalmente, se puede describir con la fórmula:

$$C = 4 * \pi * \text{area}(P) / \text{perimetro}(P) * \text{perimetro}(P)$$

donde P es el polígono.

4. *Solidez*: es el grado de convexidad que tiene el polígono. Formalmente, se puede describir con la fórmula:

$$C = \text{area}(P) / \text{area}(EC(P))$$

donde EC es la envolvente convexa y P es el polígono

Respecto a las unidades de medida en las que se computan las características, la *longitud*, el *perímetro* y el *diámetro de Feret* se miden en mm, el área en mm², el *ángulo* en grados y la *relación de aspecto*, la *circularidad* y la *solidez* son coeficientes calculados sin unidades de medida.

En la tabla 1 se pueden ver gráficamente los diferentes tipos de medidas que se almacenan, junto con los campos de almacenamiento y las características asociadas, ilustrando cada una de ellas con la vista de una figura asociada:








	Almacenamiento	Características	Vista
Line	2 puntos	<i>longitud</i>	
Polyline	N puntos	<i>longitud</i>	
Rectangular	1 punto, anchura y altura	<i>perímetro, área</i>	
Elliptical	1 punto, anchura y altura	<i>perímetro, área</i>	
Polygonal	N puntos	<i>perímetro, área</i> <i>AVANZADAS: diámetro de Feret, relación de aspecto, circularidad, solidez</i>	
Angle	3 puntos	<i>ángulo</i>	
Fixed point	1 punto		

Tabla 1. *Tipos de medidas que se almacenan.*

Finalmente, es importante señalar que, para agilizar la toma de medidas, la herramienta software provee al usuario de buen número de utilidades, como la posibilidad de copiar medidas o trozos de medidas, a partir de otras similares tomadas anteriormente, o incluso realizar transformaciones de medidas de un tipo en otro, reutilizando partes de éstas.

3.4. Algoritmos *ad hoc* para el procesamiento de los datos

En el contexto de este marco de trabajo tecnológico, las medidas se consideran los datos en bruto. Tal y como se ha comentado anteriormente, a partir de las caracte-

rísticas correspondientes a esos datos en bruto se puede extraer conocimiento utilizando técnicas estadísticas o de inteligencia artificial. De hecho, hay estudios de caracterización articular en los cuales no se han desarrollado algoritmos *ad hoc*.

Sin embargo, en muchas ocasiones, hay aspectos que no pueden ser estudiados sin un tratamiento *ex profeso* para la investigación que se está manejando. Por ello, si se desea realizar algún tipo de procesado o si se desea calcular datos de cierta complejidad interesantes para una investigación de caracterización articular a partir de los datos de las medidas tomadas, será obligada la incorporación de algún algoritmo *ad hoc*, es decir, diseñado para ese problema en concreto, cuya ejecución se realizará utilizando la herramienta software *DicomPas*.

Estos algoritmos *ad hoc* se definen como un conjunto de una o varias operaciones *ad hoc*, tal y como se puede ver gráficamente en la figura 8:

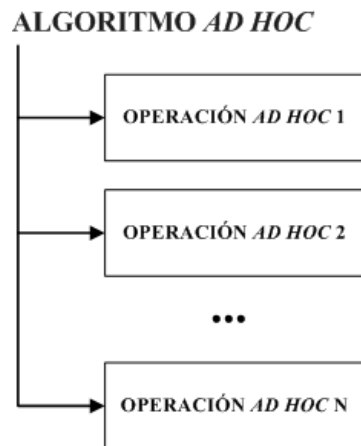


Figura 8. Descripción de un algoritmo *ad hoc* como conjunto de una o varias operaciones *ad hoc*.

Estas operaciones se ilustrarán a continuación a partir del análisis de la diferente casuística que se puede encontrar en la realización de las investigaciones. Tal y como se puede ver gráficamente en la figura 9, se han identificado cinco tipos de operaciones:

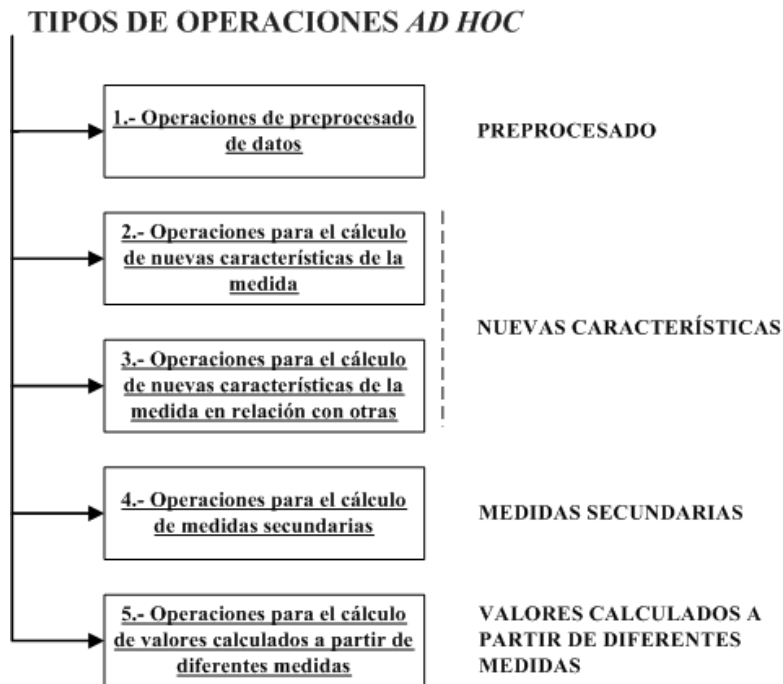


Figura 9. Tipos de operaciones ad hoc.

En primer lugar, en muchas ocasiones, cuando se quiere comparar datos de diferentes fuentes o informantes, existen diferentes particularidades en las diferentes imágenes, causadas por la configuración o las circunstancias de la toma de imágenes, por ejemplo cuando existen traslaciones o rotaciones. Ello exige de algún tipo de preprocesado de los datos de las medidas tomadas o de los valores calculados a partir de éstas. A este tipo de operaciones se las denomina como *1.- Operaciones de preprocesado de datos*.

En segundo lugar, tal y como se ha comentado anteriormente, usando las características de las medidas, tales como área, perímetro, etc., e incluso con las características avanzadas en el caso de las poligonales, quizás no sea suficiente para describir cierta información de una medida. Se diferenciarán aquí dos tipos de operaciones. Por un lado, se tienen las *2.- Operaciones para el cálculo de nuevas características de la propia medida*, que son operaciones para el cálculo de carac-

terísticas de la medida que se vean visualmente y que sean significativas, pero que no puedan ser descritas usando las características predefinidas. Por ejemplo, en el caso de las medidas poligonales, el grado de curvatura en la parte superior o el grado de estrechamiento en la parte derecha. Por otro lado, se tienen las 3.- *Operaciones para el cálculo de nuevas características de la medida en relación con otras*, que son operaciones para el cálculo de características de la medida que estén en relación con otras y sean significativas. Por ejemplo, en el caso de las poligonales, el desplazamiento del polígono medido con respecto al punto medio de la línea que une los labios, o el área del polígono medido con respecto al área del polígono correspondiente a la cavidad faríngea.

En tercer lugar, frecuentemente, a partir de medidas tomadas, que podríamos llamar primarias, se considera adecuado obtener medidas secundarias, llamadas así porque no son tomadas directamente sino calculadas por la herramienta software. A este tipo de operaciones se las denomina como 4.- *Operaciones para el cálculo de medidas secundarias*, y las podemos describir formalmente como operaciones para el cálculo de medidas secundarias que sean de interés en el contexto del problema planteado y cuya información sea relevante de cara a los cálculos posteriores. Es una operación muy común, por ejemplo, a partir de diferentes puntos tomados generar una medida secundaria que consista en un punto resultado de calcular el centro de gravedad de todos ellos.

Finalmente, en cuarto lugar, muchas veces se necesita calcular información correspondiente a un conjunto de medidas y valores. Se trata de 5.- *Operaciones para el cálculo de valores a partir de diferentes medidas*, que son operaciones de cálculo de valores interesantes para el estudio, extraídos a partir de un conjunto de medidas, tanto primarias como secundarias, o incluso a partir de resultados calculados previamente.

En resumen, los algoritmos *ad hoc* son algoritmos de alto nivel que realizan una o varias operaciones *ad hoc*, para realizar el procesamiento de los datos correspondientes a las medidas tomadas.

Hay que destacar además la íntima relación existente entre los algoritmos *ad hoc* y la posterior extracción de conocimiento, de hecho frecuentemente se desarrollan los algoritmos *ad hoc* cuando se está extrayendo el conocimiento a partir de las medidas, en los casos en que para el análisis cuantitativo de un aspecto concreto no basta con el análisis estadístico o de inteligencia artificial realizado a partir de éstas. Se puede ver gráficamente esto en la figura 10, en que se ilustra el ciclo de vida del sistema.

Otro aspecto que hay que tener en cuenta en la descripción de este marco de trabajo tecnológico es que, mientras que la toma de medidas y la posterior extracción de conocimiento pueden ser realizadas directamente por el investigador fonetista, a partir de la herramienta *DicomPas* y una herramienta estadística o de inteligencia artificial, el diseño y desarrollo de algoritmos *ad hoc* exige una labor de programación específica para el problema en concreto, lo cual en determinados momentos puede restar rapidez a la investigación, al implicar un solapamiento de tareas. No obstante, en el futuro se prevé agilizar esta cuestión ya que, tras la realización de buen número de investigaciones con este marco de trabajo, se habrá definido un conjunto completo de algoritmos *ad hoc* de propósito general, para la realización de cálculos y procesados que se hayan necesitado en diferentes problemas concretos, con lo cual éstos podrán ser ejecutados sin que tenga que realizarse previamente una labor de programación específica por parte del equipo de ingenieros. Esto permitirá dotar de mayor flexibilidad al fonetista en la realización de investigaciones con este marco de trabajo.

3.5. Extracción del conocimiento

Esta es la última etapa que se tiene en un estudio de caracterización articulatoria. Es el momento en que el investigador fonetista debe extraer el conocimiento a partir de los datos numéricos extraídos, en forma de generación de inferencias estadísticas o de inteligencia artificial. Evidentemente, esta tarea no es un conjunto de cálculos efectuados de manera independiente a las etapas anteriores. Ya anteriormente a la toma de medidas, en el momento de la definición de los parámetros articulatorios, se tiene presente la extracción del conocimiento que se hará finalmente, enfocando su diseño de cara a ello. Y tanto la toma de medidas como por supuesto el desarrollo de los algoritmos *ad hoc*, tal y como se ha comentado anteriormente, son efectuados con miras a las conclusiones cuantitativas y cualitativas a las que se desea llegar.

De hecho, en las investigaciones frecuentemente se trabaja usando un modelo iterativo, en el que a partir de la toma de medidas, y si procede la ejecución de algoritmos *ad hoc*, se realiza la extracción del conocimiento, y tras la obtención de resultados se vuelven a tomar medidas y se realizan nuevos algoritmos *ad hoc*, para posteriormente volverse a extraer el conocimiento, y así sucesivamente durante las iteraciones necesarias hasta conseguir los resultados deseados.

En la figura 10 se muestra gráficamente este modelo iterativo de ciclo de vida en el que se puede ver que, en un estudio de caracterización articulatoria, frecuentemente se vuelve sobre los pasos anteriores de cara a realizar nuevos avances.

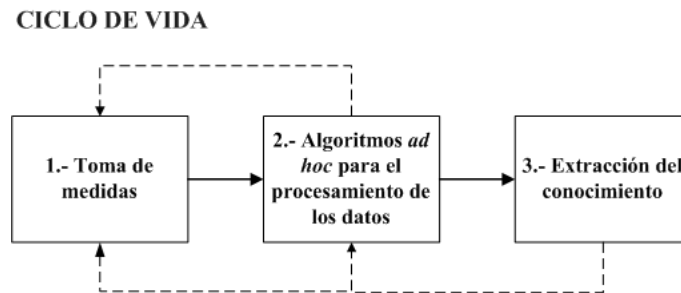


Figura 10. *Ciclo de vida iterativo de un estudio de caracterización articulatoria usando el marco de trabajo tecnológico.*

Finalmente, no se entrará aquí a tratar el tema de las diferentes técnicas estadísticas o de inteligencia artificial que se pueden utilizar.

4. REALIZACIÓN DE INVESTIGACIONES USANDO EL MARCO DE TRABAJO TECNOLÓGICO. ILUSTRACIÓN CON DOS CASOS REALES

En este apartado se va a explicar cómo se realizan las investigaciones de caracterización articulatoria usando el marco de trabajo tecnológico, ilustrándolo con dos casos de investigaciones efectuadas. No se explicarán aquí estos estudios de manera pormenorizada, sino que el objetivo es ver cómo en su realización se utiliza el marco de trabajo tecnológico.

4.1. Ejemplo 1

El primer ejemplo (Iribar, Pagola y Túrrez, 2013) amplía el trabajo realizado en Iribar, Pagola y Túrrez (en prensa). En este trabajo se realiza la descripción articulatoria de la pronunciación de [l] en euskara y español, y se analizan sus diferencias. Además, en el caso del euskara, se pretende dar solución a un problema abierto relativo a la tendencia a la retroflexión en la [l] del euskara (Txillardegí, 1980; Pagola, 1992).

Se han tomado datos de 4 informantes de euskara y 4 de castellano pronunciando el sonido [l] de manera sostenida en un lapso de tiempo y capturándose de cada uno de ellos una secuencia de imágenes MRI. Estas capturas son parte del repositorio 2 de la base de datos de imágenes MRI.

Tras realizarse un análisis visual de las imágenes obtenidas se han descubierto un buen número de características articulatorias que varían, en distinto grado y consideración, según los informantes e idiomas. Con el objetivo de realizar la descripción articulatoria se han definido diferentes parámetros articulatorios, que se han dividido en las tres categorías siguientes: parámetros de área de lengua, parámetros de área de cavidad y parámetros de distancia y longitud.

A partir de esta especificación se realiza la toma de medidas, siendo las relativas a los parámetros de área de lengua medidas de tipo *polygonal* tomadas sobre la lengua, las relativas a los parámetros de área de cavidad medidas de tipo *polygonal* tomadas sobre diferentes cavidades y las relativas a los parámetros de distancia y longitud medidas de tipo *line* y *polyline* que se han considerado significativas. La toma de medidas se realiza sobre las imágenes en las que se ha capturado bien la articulación del sonido.

En este estudio se ha estimado que no es necesario calcular datos suplementarios, ni obtener medidas secundarias. Por ello, en este estudio no se realizan algoritmos *ad hoc* y se trabaja directamente sobre los datos de las medidas obtenidas.

A partir de dichas medidas, y en base a los valores de las características obtenidos a partir de dichas medidas, se ha realizado la extracción de información, utilizando técnicas estadísticas, y llegándose a una serie de conclusiones de tipo cuantitativo y cualitativo. Además, y como añadido a lo anterior, se ha realizado una comparación entre los datos obtenidos de las imágenes MRI con datos fonéticos y de EPG.

4.2. Ejemplo 2

En el segundo ejemplo (Iribar, 2013) el autor presenta una parte de los resultados obtenidos en su tesis doctoral (Iribar, 2012). En este trabajo se realiza la caracterización articulatoria de los sonidos vocálicos en español.

Se han tomado datos de 4 informantes de español, quienes pronunciaron los 5 sonidos vocálicos de manera sostenida en un lapso de tiempo, capturándose para cada informante y vocal una secuencia de imágenes MRI. Estas capturas son parte del repositorio 2 de la base de datos de imágenes MRI.

Con el objetivo de realizar la descripción articulatoria se define como parámetro articulatorio para cada vocal un punto característico: el punto más alto del dorso lingual. A partir de esta especificación se realiza la toma de medidas de tipo *fixed*

point correspondientes, sobre las imágenes en las que se ha capturado bien la articulación del sonido.

Estas son las medidas obtenidas, a partir de las cuales se ejecuta un algoritmo *ad hoc*, con el objeto de obtener medidas secundarias y datos calculados que se consideran de interés. Este algoritmo *ad hoc* es ejecutado para cada uno de los informantes, y consta de los siguientes 4 pasos:

1. Para cada una de las vocales se recogen todos los puntos tomados y se calcula el punto medio («P_a» para [a], «P_e» para [e], «P_i» para [i], «P_o» para [o] y «P_u» para [u]), generándose la medida secundaria de tipo fixed point correspondiente para cada una de las vocales.
2. Se genera una medida secundaria de tipo polygonal, llamada «Tr_iaou» que representa el triángulo vocálico generado uniendo los puntos calculados en el paso 1 (realmente no es un triángulo, es un pentágono, pero seguimos la nomenclatura clásica). La construcción del polígono se realiza en el orden i-e-a-o-u. Se calcula el valor del área.
3. Se generan medidas secundarias de tipo line, correspondientes a las uniones de los puntos calculados en el paso 1 («L_ae», «L_ai», «L_ao», «L_au», «L_ei», «L_eo», «L_eu», «L_io», «L_iu» y «L_ou», para las parejas de vocales correspondientes). Se calculan los valores de las longitudes.
4. Se genera una medida secundaria de tipo polygonal, llamada «Tr_aiu», que representa el triángulo vocálico generado por la [a] y las dos vocales extremas [i] y [u]. La construcción del polígono se realiza en el orden a-i-u. Se calcula el valor del área.

A partir de los datos obtenidos de la toma de medidas y la ejecución del algoritmo *ad hoc* se realiza la extracción de información, utilizando técnicas estadísticas, y llegándose a una serie de conclusiones de tipo cuantitativo y cualitativo.

5. CONCLUSIONES

Se ha presentado un marco de trabajo tecnológico innovador para la realización de investigaciones fonéticas relativas a la caracterización articulatoria de los sonidos

de un idioma. Hasta ahora se ha aplicado sobre una base de datos de secuencias de imágenes 2D de cine-MRI correspondientes a la articulación de los sonidos del euskara y español de Euskadi, tomadas en el plano medio sagital, que consta de dos repositorios de imágenes, correspondientes a sendos tipos de adquisiciones de imágenes. En el primer repositorio se ha capturado la producción de cada sonido en el interior de una palabra, mientras que en el segundo repositorio se ha capturado la producción de cada sonido realizada de manera sostenida.

Como elemento fundamental del marco de trabajo tecnológico se tiene la herramienta software *DicomPas*, destinada a los usuarios fonetistas, que permite la realización de diferentes tareas, entre las cuales las más importantes son la toma de medidas de los parámetros articulatorios sobre las imágenes y la ejecución de algoritmos *ad hoc* sobre dichas medidas, de cara al procesamiento de los datos, con vistas a la posterior extracción del conocimiento, en forma de generación de inferencias estadísticas o de inteligencia artificial, usando herramientas estadísticas o de inteligencia artificial. Los usuarios han encontrado la herramienta *DicomPas* muy valiosa y fácilmente usable, y los algoritmos y funcionalidades desarrolladas han sido probados de manera formal, lo que asegura la fiabilidad del sistema en su conjunto.

Ello está posibilitando a nuestro equipo de trabajo efectuar importantes avances en el estudio de la caracterización articulatoria, tanto en euskara como en español. De hecho, ya se han publicado varias investigaciones de descripción articulatoria en ambas lenguas utilizando este marco de trabajo tecnológico (Iribar *et al*, en prensa, 2013; Pagola, Iribar y Túrrez, 2012; Iribar, 2012; Iribar, 2013). En cualquier caso, evidentemente, éste puede ser también utilizado para estudiar sonidos de otros idiomas.

6. LÍNEAS FUTURAS DE TRABAJO

Actualmente, el trabajo se encuentra en un estado robusto y estable. No obstante, se continúa en el desarrollo del sistema; se seguirán añadiendo informantes e imágenes a la base de datos, se acometerán nuevas investigaciones y el marco de trabajo tecnológico escalará en diferentes frentes, que se comentarán a continuación.

En primer lugar, probablemente, la crítica más fuerte que se puede hacer a este marco de trabajo tecnológico es el hecho de que la toma de medidas ha de realizarse a mano, en un proceso lento y tedioso, que bien podría ser automatizado

o semi-automatizado añadiendo inteligencia al software desarrollado. Existe buen número de utilidades en la herramienta para agilizar la toma de medidas, entre las que se encuentra la posibilidad de copiar medidas o trozos de medidas, a partir de otras similares tomadas anteriormente, o incluso realizar transformaciones de medidas de un tipo en otro, reutilizando partes de éstas. No obstante, un paso importante en la mejora de este marco de trabajo tecnológico será la introducción de algoritmos inteligentes para la toma de medidas de parámetros articulatorios, usando técnicas avanzadas de procesamiento digital de imagen, a partir de información previamente introducida, relativa a la especificación de los parámetros articulatorios. Por ejemplo, el sistema debería ser capaz de tomar una medida de tipo poligonal de toda la cavidad bucal, o ser capaz de tomar una medida de tipo línea correspondiente a una línea recta entre los labios, todo ello declarativamente, de manera automática, es decir, sin necesidad de que el investigador tenga que tomar a mano la medida, o en otros casos semiautomática, es decir, a partir de unos pocos puntos que sean tomados de guía. Esto liberará a los investigadores de parte del tedioso trabajo de toma de medidas, y les permitirá centrarse en tareas de más alto nivel, lo cual permitirá avanzar las investigaciones en términos exponenciales.

En segundo lugar, la realización de nuevas investigaciones determinará la creación de nuevos algoritmos *ad hoc* para el procesamiento de los datos en bruto de las medidas, que se añadirán al marco de trabajo tecnológico, y cuando se observe que determinadas funcionalidades puedan ser aplicadas a diferentes problemas, se desarrollarán algoritmos *ad hoc* de propósito general, que sean aplicables a diferentes casos.

En tercer lugar, otro tema en el que se quiere avanzar es en el estudio de ejemplos en 3D de cara a analizar el comportamiento articulatorio en el espacio 3D. Esto supondrá que al marco de trabajo tecnológico se le incorporará la funcionalidad necesaria para el manejo 3D, para la segmentación de este tipo de imágenes, la toma de medidas y la ejecución de algoritmos *ad hoc*, de cara a la posterior extracción de conocimiento. De hecho, tal y como se comentó en el apartado correspondiente a la base de datos, ya se ha comenzado con la adquisición estática de información 3D, en un tercer repositorio de imágenes que se ha creado.

Finalmente, en cuarto lugar, se quiere además realizar capturas 2D y 3D lo más rápidas posibles, que se añadirán también al tercer repositorio, con el fin de poder modelar lo más fielmente posible el comportamiento dinámico de la producción del sonido. Para ello, el marco de trabajo tecnológico escalará, dando soporte a las nuevas funcionalidades que se presenten para dar solución a este tipo de problemática. Una consecuencia interesante de este estudio dinámico será la posibilidad de

obtener un modelo que permita obtener, utilizando cálculos de interpolación, la articulación en movimiento de la producción del sonido a partir de fotogramas tomados en instantes concretos de tiempo.

Existen otras posibles líneas futuras para el escalamiento de este marco de trabajo tecnológico, relativas a la integración en él de la relación de los resultados de caracterización articuladora obtenidos a partir de las imágenes MRI con los resultados obtenidos usando otras técnicas complementarias, como la descripción fonética de los sonidos, las señales EMG (*Electromyography*, que consiste en la evaluación y registro de la actividad eléctrica producida por los músculos utilizados en la producción de sonido), EPG o EMMA.

AGRADECIMIENTOS: Este proyecto se ha desarrollado durante el periodo 2010-2012 en la Universidad de Deusto gracias a la subvención del MICINN (referencia FFI2009-10706). Alain Sánchez González, del equipo de investigación de ingeniería, de la Universidad de Deusto, ha colaborado en la programación de la herramienta software DicomPas. En la construcción de este marco de trabajo tecnológico ha colaborado el equipo de investigación de fonética, de la Universidad de Deusto, compuesto por Rosa Miren Pagola Petrarena, Alexander Iribar Ibabe e Itziar Túrrez Aguirrezabal, más las colaboradoras Maialen Garbizu Piñeiro, Ainara Iturbe Gojenola y Raquel Gómez Dávila.

7. REFERENCIAS BIBLIOGRÁFICAS

- ALWAN, A.; S. NARAYANAN y K. HAKER (1997): «Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part II. The rhotics», *The Journal of the Acoustical Society of America*, 101, 2, pp.1078-1089.
- BADIN, P; G. Bailly; M. Raybaudi y C. Segebarth (1998): «A three-dimensional linear articulatory model based on MRI data», *Proceedings of the Third ESCA/COCOSDA International Workshop on Speech Synthesis*, Jenolan Caves House, Blue Mountains, NSW, Australia, pp. 249–254.
- BADIN, P.; G. Bailly; L. Reveret; M. Baciú; C. Segebarth y C. Savariaux (2002): «Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images», *Journal of Phonetics*, 30, 3, pp.533–553.
- BADIN, P. y A. SERRURIER (2006): «Three-dimensional modeling of speech organs: Articulatory data and models», *Transactions on Technical Committee*

of *Psychological and Physiological Acoustics*, *The Acoustical Society of Japan*, 36, 5, pp.421–426.

BAER, T. (1991): «Analysis of vocal tract shape and dimensions using magnetic resonance imaging: Vowels», *The Journal of the Acoustical Society of America*, 90, 2, pp.799–828.

BEAUTEMPS, D.; P. BADIN y G. BAILLY (1996): «Evaluation of an articulatory-acoustic model based on a reference subject», *Proceedings of the First ESCA Tutorial and Research Workshop on Speech Production Modeling - Fourth Speech Production Seminar*, Autrans, Francia, pp.45-48

DONOHO, D. L. (2006): «Compressed sensing», *IEEE Transactions on Information Theory*, 52, 4, pp.1289–1306.

ELEJABEITIA, A.; A. IRIBAR y R. M. PAGOLA (2009): «El cine-MRI aplicado a la descripción de las sibilantes vascas», *Estudios de fonética experimental*, XVIII, pp.145–160.

ENGWALL, O. y P. BADIN (1999): «Quarterly Progress and Status Report Collecting and analysing two- and three-dimensional MRI data for Swedish», *Dept. for Speech, Music and Hearing. Quarterly Progress and Status Report (TMH-QPSR)*, 40, 3-4, pp. 11–38.

ENGWALL, O. (2000): «Are static MRI measurements representative of dynamic speech? Results from a comparative study using MRI, EPG and EMA», en B. Yuan; T. Huang y X. Tang (eds.): *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, Pekín, China, pp. 17-20.

ENGWALL, O. (2003a): «A revisit to the Application of MRI to the Analysis of Speech Production-Testing our assumptions», en S. Palethorpe y M. Tabain (eds): *Proceedings of 6th International Seminar on Speech Production*, Sydney, Australia, pp.43-48.

ENGWALL, O. (2003b): «Combining MRI, EMA and EPG measurements in a three-dimensional tongue model», *Speech Communication*, 41, 2-3, pp.303-329.

FERNÁNDEZ PLANAS, A. M. (2008): «La electropalatografía (EPG) en el estudio articulatorio del habla. El WinEPG de Articulate Instruments Ltd», *Estudios de fonética experimental*, XVII, pp.285–299.

-
- FITCH, W. T. y J. GIEDD (1999): «Morphology and development of the human vocal tract: A study using magnetic resonance imaging», *The Journal of the Acoustical Society of America*, 106, 3, pp.1511–1522.
- GURLEKIAN, J. A; N. ELISEI y M. ELETA (2004): «Caracterización articulatoria de los sonidos vocálicos del español de Buenos Aires mediante técnicas de resonancia magnética», *Revista Fonoaudiológica*, 50, 2, pp.7–14.
- HERMAN, G. T. (1980): *Fundamentals of Computerized Tomography*, Londres, Springer-Verlag, 2009².
- HOOLE, P. y C. MOOSHAMMER (2002): «Articulatory analysis of the German vowel system», en P. Auer; P. Gilles y H. Spiekerman (ed): *Silbenschnitt und Tonakzente*, Tübingen, pp. 129–152.
- HORNAK, J. (1996): *The Basics of MRI*, ScientificCommons.
<http://www.cis.rit.edu/htbooks/mri/> [5/2/2013]
- IBM (2013): *SPSS*
<http://www-01.ibm.com/software/analytics/spss/>. [5/2/2013]
- IRIBAR, A. (2013): «Apuntes para la caracterización articulatoria experimental del vocalismo del español», *Estudios de fonética experimental*. XXII, pp. 37-80.
- IRIBAR, A. (2012): *Caracterización fonética experimental del vocalismo vasco-románico*, tesis doctoral. Universidad de Deusto.
- IRIBAR, A.; R. M. PAGOLA e I. TÚRREZ (en prensa): «Observaciones sobre la articulación de la lateral alveolar en euskara y castellano», en *Actas del V Congreso de Fonética Experimental*, Cáceres 2011.
- IRIBAR, A.; R. M. PAGOLA e I. TÚRREZ (2013): «Caracterización articulatoria de ele en español y euskara», *Estudios de fonética experimental*, XXII, pp.129-171.
- KARTHIKESWARAN, D. y S. DINAKAR (2011): «Developing a scientific visualization tool for Inner articulators», en *Proceedings of the 2011 International Conference on Emerging Trends in Electrical and Computer Technology*, IEEE, Chunkankadai, Nargelcoil, India, pp. 480–488.
- KIM, Y. C.; S. S. NARAYANAN y K. S. NAYAK (2009a): «Accelerated 3D MRI of vocal tract shaping using compressed sensing and parallel imaging»,

Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Taipei, Taiwan, pp.389–392.

KIM, Y. C.; S. S. NARAYANAN y K. S. NAYAK (2009b): «Accelerated three-dimensional upper airway MRI using compressed sensing», *Magnetic resonance in medicine: official journal of the Society of Magnetic Resonance in Medicine*, 61, 6, pp.1434–40.

LUSTIG, M., D. DONOHO y J. M. PAULY (2007): «Sparse MRI: The application of compressed sensing for rapid MR imaging», *Magnetic Resonance in Medicine: Official Journal of the Society of Magnetic Resonance in Medicine*, 58, 6, pp.1182–95.

MARTINS, P.; I. Carbone; A. Pinto; A. Silva y A. Teixeira (2008): «European Portuguese MRI based speech production studies», *Speech Communication*, 50, 11-12, pp.925–952.

MICROSOFT (2013a): *Excel*, versión 2010
<http://office.microsoft.com/en-us/excel/>. [5/2/2013]

MICROSOFT (2013b): *RTF*, versión 1.9.1.
<http://www.microsoft.com/en-us/download/details.aspx?id=10725>.
[5/2/2013]

NARAYANAN, S.; K. Navak; S. Lee y D. Byrd (2004): «An approach to real-time magnetic resonance imaging for speech production», *The Journal of the Acoustical Society of America*, 115, 4, pp.1771–1776.

NARAYANAN, S.; A. ALWAN y K. HAKER (1997): «Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I. The laterals», *The Journal of the Acoustical Society of America*, 101, 2, pp.1064–1077.

NARAYANAN, S.; D. BYRD y A. KAUN (1999): «Geometry, kinematics, and acoustics of Tamil liquid consonants», *The Journal of the Acoustical Society of America*, 106, 4, pp.1993–2007.

NARAYANAN, S. S. y A. ALWAN (1995): «An articulatory study of fricative consonants using magnetic resonance imaging», *The Journal of the Acoustical Society of America*, 98, 3, pp.1325–1347.

-
- NEMA (2013): *DICOM specification*.
<http://medical.nema.org/dicom/>. [5/2/2013]
- ORACLE-SUN MICROSYSTEMS (2013): *Java*.
<http://java.sun.com>. [5/2/2013]
- PAGOLA, R. M. (1992): *Euskal fonetika Nafarroan*. Iruñea, Nafarroako Gobernua.
- PAGOLA, R. M.; A. IRIBAR e I. TÚRREZ (2012): «La descripción articulatoria de los sonidos en euskara y castellano: el proyecto DAELPACE», en M. Acillona (ed.): *Marcos interpretativos de la realidad social contemporánea*, Bilbao, Universidad de Deusto, pp.107–118.
- ROMANO, A. y P. BADIN (2009): «An MRI study of the articulatory properties of italian consonants», *Estudios de fonética experimental*, XVIII, pp.327–344.
- ROMERO, J. (2008): «La electromagnetometría en el estudio de la producción del habla», *Estudios de fonética experimental*, XVII, pp.359–374.
- SERRURIER, A. y P. BADIN (2005): «Towards a 3D articulatory model of velum based on MRI and CT images», *ZAS Papers in Linguistics (Speech production and perception: Experimental analyses and models)*, 40, 1, pp.195–211.
- SERRURIER, A. y P. BADIN (2008): «A three-dimensional articulatory model of the velum and nasopharyngeal wall based on MRI and CT data», *Journal of the Acoustical Society of America*, 123, 4, pp. 2335–2355.
- STORY, B. H.; I. R. TITZE y E. A. HOFFMAN (1996): «Vocal tract area functions from magnetic resonance imaging», *The Journal of the Acoustical Society of America*, 100, 1, pp.537–54.
- TAKEMOTO, H.; T. Kitamura; H. Nishimoto y K. Honda (2004): «A method of tooth superimposition on MRI data for accurate measurement of vocal tract shape and dimensions», *Acoustical Science and Technology*, 25, 6, pp.468–474.
- TAKEMOTO, H.; K. Honda; S. Masaki; Y. Shimada y I. Fujimoto (2006): «Measurement of temporal changes in vocal tract area function from 3D cine-MRI data», *The Journal of the Acoustical Society of America*, 119, 2, pp.1037–1049.
-

- THE FREE DICTIONARY (2013a): *Body planes*.
<http://medical-dictionary.thefreedictionary.com/coronal+planes>. [5/2/2013]
- THE FREE DICTIONARY (2013b): *Midsagittal plane*.
<http://medical-dictionary.thefreedictionary.com/midsagittal+plane>. [5/2/2013]
- THE UNIVERSITY OF WAIKATO (2013): *WEKA*.
<http://www.cs.waikato.ac.nz/ml/weka/>. [5/2/2013]
- TIEDE, M. K. S. MASAKI y E. VATIKIOTIS-BATESON (2000): «Contrasts in speech articulation observed in sitting and supine conditions», *Proceedings of the Fifth Seminar on Speech Production: Models and Data*, Kloster Seeon, Bavaria, Alemania, pp. 25–28.
- TXILLARDEGI (1980): *Euskal fonologia*. Donostia, Ediciones Vascas.
- U.S. NATIONAL INSTITUTES OF HEALTH (2013): *ImageJ*, versión 1.46.
<http://rsbweb.nih.gov/ij/>. [5/2/2013]
- WORLD WIDE WEB CONSORTIUM (W3C) (2013a): *JPEG*.
<http://www.w3.org/Graphics/JPEG/>. [5/2/2013]
- WORLD WIDE WEB CONSORTIUM (W3C) (2013b): *XML*.
<http://www.w3.org/XML/>. [5/2/2013]
- WORLD WIDE WEB CONSORTIUM (W3C) (2013c): *XML Schema*.
<http://www.w3.org/XML/Schema>. [5/2/2013]
- YANG, B. (1999): «Measurement and synthesis of the vocal tract of Korean monophthongs by MRI», en J. Ohala; Y. Hasegawa; M. Ohala; D. Granville y A. C. Bailey (eds.): *Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS) 1999*, San Francisco, E.E.U.U, pp. 2005–2008.
- ZHOU, X. (2009): *An MRI-based articulatory and acoustic study of American English liquid sounds/r/and/l/*, tesis doctoral. Universidad de Maryland, College Park.
- ZHOU, X. et al. (2010): «An MRI-based articulatory and acoustic study of lateral sound in American English», *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Dallas, E.E.U.U, pp. 4182–4185.