

Reconocedor de Comandos Orales para Eliminar Barreras de Comunicación y Movilidad en Personas con Discapacidades Motrices y de Comunicación

Autor: Carlos Vaquero Avilés-Casco
Ingeniero de Telecomunicación

Director: Óscar Saz Torralba
Ingeniero de Telecomunicación

Ponente: Eduardo Lleida Solano
Dr. Ingeniero de Telecomunicación

*Departamento de Ingeniería Electrónica y Comunicaciones
Área de Teoría de la Señal y Comunicaciones
Centro Politécnico Superior
Universidad de Zaragoza*

Fecha de lectura: 11 de septiembre de 2006

Calificación: 9,3

RESUMEN DEL PROYECTO

Índice

Índice	2
1. Trabajo realizado.....	3
1.1. Origen	3
1.2. Objetivos	3
1.3. Desarrollo.....	4
1.3.1. Aplicación Vocaliza	4
1.3.2. Corpus Alborada	5
1.3.3. Adaptación al locutor MAP	6
1.3.4. Evaluación de pronunciación LR UV.....	6
1.4. Conclusiones	7
2. Originalidad	9
2.1. Tecnologías del habla en personas con problemas de dicción	9
2.2. Técnicas al servicio de la logopedia	10
3. Resultados.....	12
3.1. Resultados del análisis de la adaptación al locutor MAP.....	12
3.2. Resultados del análisis de evaluación de pronunciación mediante LR UV ..	13
4. Aplicabilidad	15
4.1. La aplicación Vocaliza	15
4.2. Corpus de habla patológica	15
4.3. La adaptación MAP en locutores con patologías en el habla	16
4.4. Verificación de pronunciación como método de evaluación	16

1. Trabajo realizado

1.1. Origen

El reconocimiento automático del habla (RAH), técnica que consiste en dotar a una máquina con la capacidad de detectar patrones fonéticos, cuando éstos son pronunciados por una persona, puede suponer una herramienta muy poderosa a la hora de salvar las barreras a las que las personas con discapacidades motrices se enfrentan en su día a día. Sin embargo, en muchos casos, dichas discapacidades motrices son consecuencia de alguna patología mayor, que puede conllevar igualmente patologías en el habla, impidiendo el adecuado funcionamiento de un sistema de RAH.

Existen dos vías de uso de un sistema de RAH como medio para ayudar a aquellas personas que sufren patologías que llevan asociadas discapacidades motrices y problemas de dicción:

- Sistemas de ayuda controlados por voz: consisten fundamentalmente en sistemas de RAH que reconocen comandos orales, que permiten a un sujeto con discapacidades motrices controlar dispositivos de su entorno únicamente con la voz.
- Asistencia a la logopedia: consiste en utilizar las tecnologías del habla como herramienta que ayude a superar los problemas de comunicación a los que se enfrenta un sujeto que padece problemas de dicción, colaborando en la mejora del habla del sujeto, consiguiendo que ésta sea más fácil de comprender.

Los sistemas de ayuda controlados por voz para personas con patologías en el habla requerirán, previsiblemente, un proceso de adaptación al locutor, esto es, la creación de un modelo acústico adaptado al sujeto, que permita al reconocedor discernir con mayor precisión las emisiones fonéticas del sujeto. Por tanto, será necesario estudiar el funcionamiento de las técnicas de adaptación habitualmente utilizadas con personas sin problemas de dicción, comprobando si son útiles para personas con patologías en el habla.

Los sistemas de asistencia a la logopedia deberían ofrecer métodos para que los sujetos con patologías en habla practicasen y mejoraran su forma de hablar, mostrando alguna medida de la evolución que está siguiendo la patología del usuario. En dichos métodos serán útiles las técnicas de RAH y de adaptación al locutor, pero también sería necesaria alguna técnica de evaluación de pronunciación, que permitiera evaluar la evolución del sujeto.

1.2. Objetivos

El presente proyecto se enmarca en la colaboración que el Grupo de Tecnologías de las Comunicaciones (GTC) y el Instituto de Investigación en Ingeniería de Aragón (I3A) mantienen con el Colegio Público de Educación Especial Alborada (CPEE Alborada), y tiene el objetivo de poner las tecnologías del habla al servicio de personas que sufren problemas de dicción, a través de sistemas de ayuda controlados por voz y de asistencia a la logopedia.

El objetivo principal de este proyecto es el desarrollo de una aplicación de software libre cuyo propósito fundamental es el de servir de apoyo a la logopedia, utilizando las tecnologías del habla para ayudar al usuario a mejorar su capacidad de comunicación, permitiendo de forma adicional la creación de modelos acústicos adaptados al locutor, para su integración en sistemas de ayuda controlados por voz que puedan facilitar la vida de aquellas personas que padezcan discapacidades motrices. Dicha aplicación se llamará "Vocaliza".

Para el desarrollo de la aplicación, es necesario analizar el funcionamiento de ciertas técnicas como son el RAH, la adaptación al locutor o la verificación de patrones pronunciados, cuando son utilizadas con habla patológica, para lo que será necesario, adicionalmente, la adquisición de un corpus o base de datos de habla patológica.

1.3. Desarrollo

En este apartado se describe el trabajo que se ha llevado a cabo en este proyecto.

1.3.1. Aplicación Vocaliza

En este proyecto se ha desarrollado Vocaliza 1.0, una aplicación informática diseñada para que aquellas personas que padecen alguna discapacidad en el habla puedan mejorar su capacidad de comunicación.

Para desarrollar dicha aplicación, se celebraron múltiples reuniones entre el GTC y el CPEE Alborada, con el objetivo de fijar los requisitos que debía cumplir Vocaliza 1.0. En ellas se decidió que la finalidad fundamental de la aplicación debía ser el apoyo a la logopedia, pero que la aplicación debía permitir extraer modelos acústicos del habla del sujeto, mediante técnicas de adaptación al locutor, para permitir su posterior integración en sistemas de ayuda controlados por voz, mejorando el funcionamiento de los mismos. Igualmente, la aplicación debía ser versátil y flexible, permitiendo tratar el mayor rango de patologías posible, pero a la vez debía ser atractiva visualmente y fácil de manejar, ya que muchos de sus usuarios iban a ser niños o personas sin conocimientos informáticos.

Una vez fijados los requisitos y durante el desarrollo de la aplicación, se siguieron celebrando reuniones semanalmente para permitir que los logopedas y pedagogos del CPEE Alborada tuvieran un seguimiento cercano de Vocaliza y pudieran aportar nuevas ideas.

De esta forma, se desarrolló una aplicación que permite practicar y mejorar la capacidad de comunicación del usuario mediante una serie de juegos, centrados en distintos niveles del lenguaje: nivel fonológico, semántico y sintáctico. Dichos juegos sólo exigen que el usuario interactúe utilizando su habla, de forma que el usuario únicamente necesita un micrófono para poder utilizar la aplicación, no siendo necesario que sepa utilizar un ratón o un teclado. Los mencionados juegos se basan en un sistema de RAH que verifica que el usuario ha pronunciado lo que la aplicación esperaba en cada momento de forma de correcta.

Todos los juegos son altamente configurables, de forma que un logopeda puede adecuarlos para atacar distintas patologías. Ello requiere un periodo previo de configuración de la aplicación, pero una vez realizado, el usuario con patologías en el habla puede practicar y mejorar su capacidad de comunicación con los distintos juegos sin supervisión, incluso aunque no tenga conocimientos de informática.

Como apoyo a aquellas personas que padecen discapacidades motrices, la aplicación permite la creación de modelos acústicos adaptados al habla del sujeto que posteriormente pueden ser integrados en sistemas de ayuda controlados por voz. La generación de modelos es posible gracias a la integración de un sistema de adaptación al locutor basado en la técnica del Máximo a Posteriori (MAP), que se detalla en el apartado 1.3.3.

Igualmente, la aplicación permite evaluar la progresión del habla del usuario, indicando la calidad relativa de su pronunciación respecto a pronunciaciones anteriores del mismo usuario, mediante la técnica de Verificación de Pronunciación mediante Ratio de Verosimilitud (*Likelihood Ratio Utterance Verification*, LR UV), técnica que se detalla en el apartado 1.3.4.

Para que el nivel de personalización exigido por las prestaciones de la aplicación sea posible, ésta permite la creación de perfiles de usuario que almacenan la información relativa a los modelos acústicos y la evolución del habla de cada usuario.

De esta forma, la aplicación Vocaliza constituye una potente herramienta que puede ser utilizada como apoyo para los logopedas y pedagogos que trabajen con personas con discapacidades motrices y de comunicación, así como para que aquellas personas con dificultades en el habla practiquen los juegos que atacan más directamente sus patologías sin supervisión y de forma entretenida, tras una configuración previa por parte de un logopeda.

Una vez desarrollada la aplicación Vocaliza 1.0, ésta fue distribuida entre los distintos logopedas y pedagogos del CPEE Alborada para su evaluación.

La evaluación consistió en un pequeño cuestionario dirigido tanto a logopedas y pedagogos como a alumnos del centro. Dicho cuestionario mostró resultados muy satisfactorios. Todos los logopedas y pedagogos que evaluaron la aplicación la encontraron muy útil y afirmaban que la utilizarían en sus labores cotidianas, mientras que todos los alumnos que la probaron la encontraron fácil de utilizar y muy entretenida.

1.3.2. Corpus Alborada

En paralelo al desarrollo de la aplicación se realizó la adquisición de un Corpus o base de datos de habla patológica. La motivación principal para la adquisición de dicho Corpus fue el hecho de que no existe ninguna base de datos de habla patológica en castellano, y es necesario disponer de una para poder analizar las prestaciones de las técnicas incluidas en Vocaliza 1.0 como son el RAH, la adaptación al locutor o la evaluación de pronunciación.

La adquisición de un Corpus de habla patológica presenta una dificultad añadida respecto a la adquisición de una base de datos de habla común, y es que las personas con problemas de dicción encuentran especialmente complicado y fatigoso el proceso de adquisición. Ello obliga a realizar sesiones de adquisición cortas, alargando notablemente el proceso.

Esta dificultad crea un compromiso a la hora de elegir el conjunto de datos que deben adquirirse. En general un Corpus ofrecerá más posibilidades de estudio cuanto mayor sea el conjunto de datos que lo forman, siempre que este conjunto sea fonéticamente balanceado, sin embargo, un conjunto de datos grande prolongaría el proceso de grabación y supondría un esfuerzo tremendo para las personas que deban grabar la base de datos. Por ello, para solventar este compromiso se tomaron como datos a adquirir las 57 palabras del “Registro Fonológico Inducido” (RFI), que es el registro que se utiliza en el campo de la logopedia para detectar problemas en la pronunciación de fonemas de forma rápida y precisa, ya que las 57 palabras constituyen un conjunto pequeño y fonéticamente balanceado, es decir, contienen la mayoría de fonemas y combinaciones de los mismos existentes en el castellano.

Como locutores para el proceso de adquisición fueron elegidos 6 alumnos del CPEE Alborada, que padecían patologías en el habla variadas y de diversa gravedad, pero que además tenían un nivel cognitivo suficientemente alto como para comprender y seguir el proceso de adquisición.

De cada locutor se grabaron 4 series de las 57 palabras del RFI, lo que supuso un total de 228 palabras por locutor y 1368 palabras en conjunto. Los registros fueron adquiridos con una frecuencia de muestreo de 16 KHz y cuantificados con 16 bits, lo que supone calidad suficiente para grabar la voz humana.

Una vez adquiridos, los registros fueron clasificados en función del locutor, serie de grabación y palabra grabada, para facilitar su uso en los estudios siguientes.

1.3.3. Adaptación al locutor MAP

La adaptación de un sistema de RAH al locutor consiste en la estimación de un modelo acústico adaptado a los parámetros del habla del locutor. Ello ofrece una considerable mejora en las prestaciones de los sistemas de RAH cuando son utilizados por ese locutor concreto.

No obstante cabe plantearse si la adaptación al locutor será útil en personas con patologías en el habla, ya que el principal problema de este habla es la variabilidad, lo que puede ser difícil de modelar con los métodos tradicionales.

Cuando se dispone de suficientes datos, el método ideal para la estimación de modelos acústicos adaptados a un locutor es el método de Máxima Verosimilitud (*Maximum Likelihood*, ML), sin embargo, dicho método requiere gran cantidad de grabaciones del habla del locutor para conseguir estimar un buen modelo adaptado, lo que hace inviable su utilización en el caso de locutor con habla patológica. Por ello se decidió utilizar, para su integración en la aplicación Vocaliza 1.0, el método de estimación del Máximo a Posteriori (MAP), que aprovechando un modelo acústico independiente del locutor como punto de partida, permite estimar un buen modelo adaptado al locutor a partir de serie reducida de muestras del habla del mismo.

El método MAP presenta la ventaja de necesitar pocas realizaciones del habla del locutor para ofrecer un buen modelo adaptado frente a ML, y de obtener como resultado un modelo adaptado que converge al que se obtendría mediante ML cuando el conjunto de datos es suficientemente grande.

Utilizando el Corpus Alborada adquirido en este mismo proyecto, se puede analizar la viabilidad de utilizar adaptación al locutor, y en concreto la estimación MAP, para mejorar las prestaciones de los sistemas de RAH en el caso de hablantes con problemas de dicción. Para ello se utilizan un conjunto de series de palabras adquiridas para cada hablante para estimar un modelo adaptado al mismo, y el resto de las series para evaluar el modelo estimado, utilizando todas las combinaciones posibles con las 4 series disponibles, lo que supone estimar un total de 14 modelos adaptados a cada locutor, con 1, 2 y 3 series de palabras.

Los resultados obtenidos de este análisis demuestran que la adaptación al locutor mediante estimación MAP puede utilizarse para mejorar las prestaciones de un sistema de RAH en el caso de personas con patologías en el habla, reduciendo la tasa de error de palabra (*Word Error Rate*, WER) promedio de todos los usuarios en más de un 50%, de forma que, para una tarea dada, el sistema de RAH combinado con la adaptación al locutor presentada consigue ofrecer unas prestaciones a una persona con patologías en el habla muy similares a las que ofrece el mismo sistema de RAH sin adaptación al locutor a una persona que no padece ninguna patología en el habla.

1.3.4. Evaluación de pronunciación LR UV

Una de las funcionalidades más interesantes de la aplicación Vocaliza 1.0 es la posibilidad de evaluar la calidad de la pronunciación del locutor. Dicha funcionalidad permite trabajar de forma precisa el nivel fonológico del lenguaje, así como realizar un seguimiento de la evolución de la patología del usuario.

Aunque no existe ningún método en el ámbito de las tecnologías del habla que permita evaluar directamente la calidad de la pronunciación del locutor, existen técnicas que ofrecen medidas de confianza para cada una de las palabras en las que el sistema de RAH segmenta una observación de parámetros del habla del locutor, en base a la proximidad a un modelo acústico dado. En concreto, la técnica de Verificación de Pronunciación mediante Ratio de Verosimilitud (*Likelihood Ratio Utterance Verification*, LR UV), utiliza dos modelos acústicos para obtener dicha medida de confianza, indicando si la palabra está más próxima a uno u otro.

De esta forma, si se supone que un modelo acústico independiente del locutor modela el habla que se considera correcta, ya que es válido para mayoría de hablantes que no presentan patologías en el habla, y se admite que un modelo acústico adaptado al usuario que padece problemas de dicción modela su habla patológica, se podría utilizar la técnica LR UV para decidir si cada palabra que pronuncia el usuario está más próxima al habla correcta que a su habla patológica, pudiendo evaluar así su pronunciación. Así, si el usuario evoluciona y mejora su pronunciación, el modelo adaptado a su patología quedará obsoleto y cabe esperar que las palabras que pronuncie estén más próximas al modelo de habla correcta.

El problema que plantea este método es la dependencia del usuario. Puesto que se utiliza un modelo adaptado al usuario, cabe esperar que las medidas de confianza resultantes para distintos usuarios no sean comparables, sino que dichos resultados sólo indicarán mejora o empeoramiento en la pronunciación del usuario realizando una comparación con otros resultados previos obtenidos por el mismo usuario.

Para estudiar el funcionamiento la técnica LR UV se utiliza el Corpus Alborada, y como modelo de habla patológica se utiliza, para cada usuario, los distintos modelos estimados en el estudio de la adaptación al locutor MAP, comentado en el apartado 1.3.3, mientras que como modelo de habla correcta se utiliza un modelo independiente del locutor. De esta forma se extraen medidas de confianza para todas las palabras de cada usuario que no hayan sido utilizadas en la estimación del modelo adaptado que se esté considerando, para mantener la independencia entre test y entrenamiento, obteniendo una gran cantidad de medidas.

El comportamiento de dichas medidas es bastante interesante. En primer lugar, se comprueba que para un usuario dado, las medidas de confianza obtenidas siguen una distribución gaussiana, lo que induce a pensar en la posibilidad de establecer un criterio de calificación a partir de las mismas. Además, se observa que las funciones de distribución obtenidas para los distintos usuarios y modelos adaptados presentan medias y varianzas muy similares. De hecho, se puede comprobar que la distancia de Kullback-Leibler (KL) entre las funciones de distribución es muy reducida. Esto permite aproximar las distintas funciones de distribución por una única campana de Gauss, que permitirá establecer un criterio de calificación único para todos los usuarios.

Así, se establece un criterio de evaluación de pronunciación que indicará al usuario la evolución en la calidad de su pronunciación cada vez que utilice la aplicación Vocaliza 1.0, mostrándole una calificación en función de la medida de confianza obtenida mediante LR UV, independientemente del usuario que utilice la aplicación.

1.4. Conclusiones

El trabajo realizado en este proyecto permite concluir que el uso de las tecnologías del habla para ayudar a personas con problemas de dicción es perfectamente viable, tanto en sistemas de ayuda controlados por voz como en sistemas de asistencia a la logopedia.

En el ámbito de la logopedia, los análisis de las tecnologías del habla ofrecen resultados muy interesantes, que permiten concluir que es posible tratar una amplia variedad de patologías, trabajando en tres niveles del lenguaje (fonológico, semántico y sintáctico) simplemente mediante la técnica de RAH, y que es posible, del mismo modo, medir el grado de evolución de una persona con patologías en el habla, mediante técnicas de adaptación al locutor (MAP) y verificación de pronunciación (LR UV).

Por otro lado, se ha demostrado que la técnica de adaptación al locutor MAP permite estimar modelos acústicos adaptados al habla de cualquier persona con problemas de dicción que, a su vez, permiten que dicha persona pueda aprovechar la funcionalidad que ofrece el RAH de forma similar a como lo haría una persona sin patologías, siempre que se dispongan de suficientes muestras del habla del usuario.

Por tanto, es posible utilizar reconocedores de comandos orales que faciliten la vida a personas que sufran alguna discapacidad motriz, además de patologías en el habla.

Otro de los logros de este proyecto ha sido el desarrollo de Vocaliza 1.0, una aplicación de software libre y en castellano de apoyo a la logopedia, que aprovecha las diversas tecnologías del habla estudiadas. Dicha aplicación está implantada y está siendo utilizada actualmente en el CPEE Alborada, y se está implantando en otros centros de educación especial de Zaragoza, para que los alumnos con patologías del lenguaje practiquen y mejoren su capacidad de comunicación. Igualmente se pretende colgar en Internet para que cualquier persona o institución pueda disponer de ella.

La gran aceptación demostrada, tanto por los logopedas y educadores del CPEE Alborada, como por los alumnos del colegio ante la aplicación implementada en este proyecto, demuestra que las tecnologías del habla no sólo son útiles para apoyar el tratamiento de personas con patologías en el habla, sino que además, los métodos que las utilizan resultan mucho más atractivos a dichas personas que los métodos convencionales, y por tanto serán muy utilizados.

Igualmente, se han conseguido importantes avances en la investigación en el marco de tecnologías del habla para personas con problemas de comunicación que el GTC realiza, no sólo por los análisis y estudios realizados en sí, sino también por la adquisición de un Corpus con muestras del habla de personas que padecen problemas de dicción, que servirá para realizar muchos otros estudios futuros sobre el habla de estas personas. La adquisición de dicho Corpus representa un importante progreso en la investigación en esta línea, ya que existen muy pocas bases de datos con registros de habla patológica en castellano. De hecho la compañía Estadounidense Nuance Communications, Inc. Líder en tecnologías del habla y creadora del sistema de RAH *Dragon NaturallySpeaking*, se ha puesto en contacto con el GTC mostrando un gran interés por el Corpus adquirido y previsiblemente planteará un proyecto de desarrollo interno en el que el GTC participará proporcionando el Corpus Alborada adquirido.

Además, la realización del proyecto ha estrechado la relación de colaboración entre el CPEE Alborada y el GTC, facilitando el desarrollo de futuros proyectos en este ámbito.

Cabe añadir que este proyecto ha contribuido con dos artículos a las IV Jornadas en Tecnología del Habla, congreso que se celebra a nivel nacional cada dos años. Dichos artículos se referencian a continuación y se adjuntan en formato electrónico en el mismo soporte que éste resumen, como anexos:

- Carlos Vaquero, Óscar Saz, Eduardo Lleida, José Manuel Marcos y César Canalís, "Vocaliza, an Application for Computer-Aided Speech Therapy in Spanish Language", IV jornadas en tecnologías del habla, Zaragoza, España, pp. 321-326, 2006.
- Óscar Saz, Carlos Vaquero, Eduardo Lleida, José Manuel Marcos y César Canalís, "Study of Maximum a Posteriori Speaker Adaptation for Automatic Speech Recognition of Pathological Speech", IV jornadas en tecnologías del habla, Zaragoza, España, pp. 395-398, 2006.

Igualmente, merece la pena destacar que este proyecto ha sido condecorado con el primer premio al Proyecto Fin de Carrera en la VII Edición de la Noche de las Telecomunicaciones de Aragón.

Como conclusión final, se puede afirmar que se han alcanzado todos los objetivos que se perseguían con la realización del presente proyecto, logrando numerosos progresos adicionales y abriendo la puerta a nuevos avances en esta línea de investigación.

2. Originalidad

La originalidad de este proyecto radica tanto en el uso innovador de las tecnologías del habla al servicio de personas que padecen patologías en el habla, como en el desarrollo de una aplicación fundamentalmente logopédica, libre y en castellano, que aunque no es única en su género, su planteamiento es muy diferente al que presentan las aplicaciones logopédicas tradicionales.

Los siguientes apartados introducen brevemente el estado del arte en el área de las tecnologías del habla utilizadas por personas con problemas de dicción y en el desarrollo de aplicaciones logopédicas, justificando en cada caso la originalidad de este proyecto.

2.1. Tecnologías del habla en personas con problemas de dicción

Los sistemas de RAH han alcanzado buenos resultados en situaciones controladas (ambiente sin ruido, pequeño vocabulario, usuario cooperativo...), pero las prestaciones caen rápidamente cuando se sale de esta situación. Ese es el caso cuando se trabaja con personas que presentan patologías en el habla.

El principal problema que se plantea a la hora de utilizar las tecnologías del habla en personas que padecen algún problema de dicción no es únicamente que el habla de estas personas no siga los patrones esperados, sino que además, la forma de hablar de estas personas presenta excesiva variabilidad, por lo que no siempre es posible que una persona con patologías en el habla pronuncie el mismo patrón fonético de la misma forma.

Otro gran problema es la inexistencia de datos suficientes para el adecuado entrenamiento de sistemas de RAH. Normalmente los sistemas de RAH que van a ser utilizados por personas con discapacidad de comunicación carecen de muestras suficientes del habla de la persona como para adecuarse a ella y minimizar el número de errores cometidos.

Estas dificultades han impedido obtener grandes avances en esta línea de trabajo, de forma que no existen sistemas actuales de RAH para personas con discapacidades en el habla, si bien existen proyectos de investigación (STARDUST) que pretenden desarrollarlos, pero sólo consiguen mejoras notables respecto a los sistemas tradicionales utilizando sistemas de palabras aisladas y vocabulario reducido, lo que limita considerablemente las prestaciones.

Para hacer frente a estos problemas, en este proyecto se ha adquirido un corpus de habla patológica en castellano suficientemente grande como para analizar el funcionamiento de un sistema de RAH, así como para estudiar la viabilidad de aplicación de otras técnicas como son la adaptación al locutor y la verificación de pronunciación. El estudio de dichas técnicas es muy innovador, ya que no existe ningún estudio similar para habla patológica en castellano y existe muy poca investigación a nivel internacional en tecnologías del habla para habla patológica.

Concretamente, el uso de una técnica de verificación de pronunciación como método de evaluación del habla del usuario es completamente innovador ya que hasta la publicación de este proyecto, no existía ninguna línea de investigación que se hubiera planteado la posibilidad de utilizar las tecnologías del habla para evaluar la evolución del habla del usuario.

Igualmente, la adquisición de la base de datos de habla patológica en sí es algo original en el marco de un proyecto fin de carrera, pero además es algo muy innovador en el campo de investigación a nivel nacional de la aplicación de las tecnologías del

habla para la ayuda a personas con problemas de dicción, ya que no existe ningún corpus de habla patológica en castellano del tamaño del corpus que se ha adquirido.

2.2. Técnicas al servicio de la logopedia

El desarrollo de sistemas de logopedia asistidos por ordenador ha sido siempre demandado por los profesionales y educadores del ámbito de la Educación Especial. Sin embargo, el desarrollo de estos sistemas desde un punto de vista comercial ha sido muy limitado; sin duda, el sistema más popular es el Speech Viewer, desarrollado por IBM, pero su última versión, Speech Viewer III, data del año 2000, y el soporte dado por este software a la logopedia en lengua española es mínimo, no existiendo ninguna versión del mismo para nuestro idioma. Por otra parte, la rigidez del sistema y la incapacidad de adaptarse a las necesidades concretas de cada profesional de la logopedia en su trabajo diario impide su uso y aprovechamiento de forma intensiva en centros educativos.

La investigación científica en el área de las ayudas logopédicas en el marco de los proyectos de investigación en Europa ha sido muy extensa. Sólo desde mediados de los 90, se han llevado a cabo los siguientes proyectos con financiación europea: proyecto HARP, consorcio ISAEUS, proyecto SPECO y proyecto Orto Logo-Paedia. El gran inconveniente de los proyectos de investigación científica pura es la poca transmisión del conocimiento a sistemas plenamente funcionales y disponibles para los educadores especializados en logopedia. Si bien algunos de estos proyectos sí que han acabado siendo programas de distribución final, como es el caso de SPECO o del sistema Orto Logo-Therapy (germen del proyecto Orto Logo-Paedia), su distribución ha sido mínima y siempre está presente el gran problema de la no existencia de sistemas para la lengua española.

Por ello, la originalidad de la aplicación Vocaliza 1.0 radica en que cubre las necesidades que demandan los logopedas españoles, ya que ha sido desarrollado con la supervisión de expertos logopedas y pedagogos que trabajan en un colegio de educación especial, que son quienes van a utilizar finalmente la aplicación. Ello ha permitido romper la tradicional barrera que existe entre la investigación pura y el desarrollo de sistemas útiles, que es el principal problema que presentan los proyectos europeos mencionados.

Frente a las aplicaciones comerciales disponibles cabe destacar el hecho de que Vocaliza está totalmente en castellano, y que es mucho más flexible y configurable que la mayoría de las aplicaciones existentes, por dos motivos fundamentales.

- En primer lugar, la aplicación Vocaliza 1.0 trabaja niveles altos del lenguaje, como son el nivel fonológico, semántico y sintáctico, mientras que las aplicaciones tradicionales suelen trabajar únicamente el prelenguaje, centrándose en aspectos básicos como la respiración, el tono o el volumen del habla. Este avance de trabajar niveles altos del lenguaje y no quedarse únicamente en el prelenguaje se consigue gracias al uso de técnicas avanzadas como son el RAH, la adaptación al locutor y la evaluación de pronunciación, técnicas que las aplicaciones comerciales disponibles no utilizan.
- En segundo lugar, la aplicación Vocaliza fue diseñada para tratar un amplio rango de patologías, y en base a esta idea ofrece opciones como la creación de distintos perfiles de usuario para los distintos usuarios que utilicen la aplicación, o la configuración del comportamiento de los distintos juegos para cada usuario, de forma que se pueda atacar con mayor precisión cada patología concreta.

Igualmente, cabe destacar la funcionalidad de la adaptación al locutor, introducida pensando en el uso del modelo adaptado a cada usuario dentro de la aplicación, pero también en sistemas de ayuda controlados por voz, para aquellas personas que padecen discapacidades motrices, lo que supone un planteamiento innovador que ninguna aplicación de carácter logopédico había tenido en cuenta.

Finalmente, merece la pena reiterar el hecho de que ninguna aplicación comercial dispone de un método de evaluación de pronunciación, ya que es algo innovador en la aplicación Vocaliza 1.0.

3. Resultados

Como resultados tangibles de este proyecto cabe destacar la aplicación Vocaliza 1.0 y el Corpus Alborada, por la utilidad futura que ambos pueden tener, que se comenta con más detalle en la sección 4. Sin embargo, este apartado se centra en los resultados numéricos obtenidos en los análisis de la adaptación al locutor mediante el método de estimación MAP y de la Verificación de Pronunciación mediante Ratio de Verosimilitud (LR UV), técnicas que han sido estudiadas en este proyecto para comprobar su adecuado funcionamiento con habla patológica y decidir si finalmente se incluían en la aplicación.

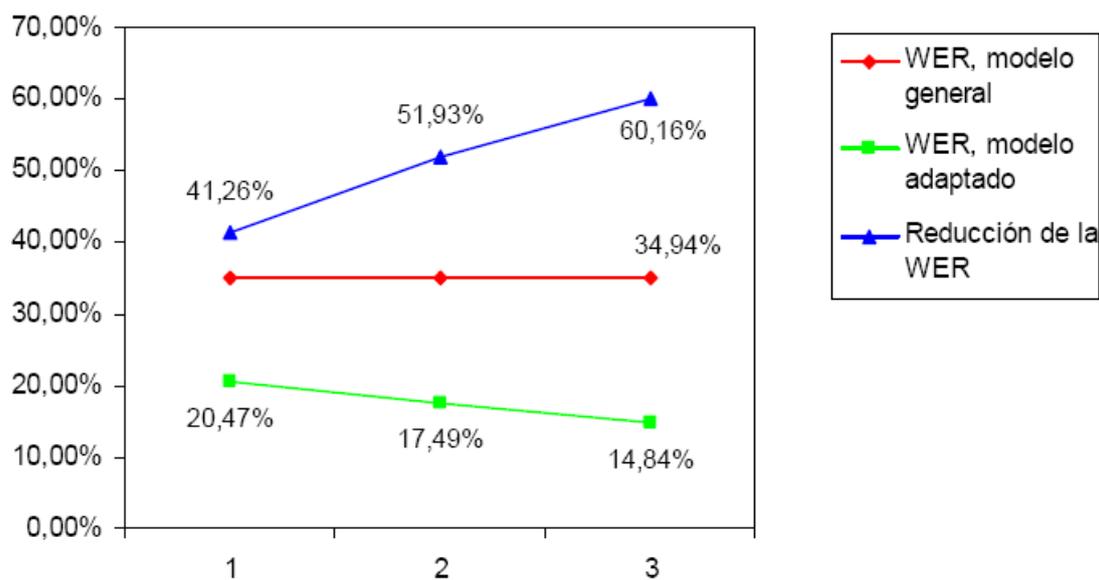
3.1. Resultados del análisis de la adaptación al locutor MAP

La adaptación al locutor mediante el método de estimación MAP permite que la aplicación Vocaliza 1.0 pueda generar modelos acústicos adaptados al habla del usuario, de forma que estos puedan ser utilizados posteriormente en sistemas de ayuda controlados por voz que faciliten el día a día de un usuario que padezca discapacidades motrices, ayudándole a superar las barreras de movilidad a las que se enfrenta. Sin embargo, todo ello está sujeto al correcto funcionamiento de la estimación MAP cuando el locutor padece patologías en el habla.

Para analizar la viabilidad de utilizar el método de estimación MAP como técnica de adaptación al locutor cuando el habla de éste es patológica, se utiliza el Corpus Alborada adquirido en este mismo proyecto. De dicho corpus se toma un conjunto de series de palabras adquiridas de cada hablante para estimar un modelo adaptado al mismo, y el resto de las series de cada hablante para evaluar el modelo estimado, utilizando todas las combinaciones posibles con las 4 series disponibles por hablante, lo que supone estimar un total de 14 modelos adaptados a cada locutor, con 1, 2 y 3 series de palabras, que serán evaluados con 3, 2 y 1 series de palabras, respectivamente.

Para evaluar los modelos adaptados se calcula tasa de error de palabra (*Word Error Rate*, WER) que obtiene un sistema de RAH con esos modelos a la hora de reconocer las palabras pertenecientes a las series no utilizadas en la estimación de cada uno de ellos, independizando de esta forma las secuencias de entrenamiento y de test del modelo. La WER obtenida con cada modelo se compara con la WER que se obtendría con las mismas secuencias de test considerando un modelo independiente del locutor, y también se compara con la WER que obtendría un hablante sin problemas de dicción utilizando el mismo sistema de RAH con un modelo independiente del locutor.

Los resultados obtenidos son muy satisfactorios: en general, la estimación MAP consigue mejorar las prestaciones del sistema de RAH, reduciendo la WER de un 34,94% a un 17,59% en promedio para todos los usuarios, lo que supone una reducción de un 51,23%. Igualmente, se observa que al aumentar la cantidad de datos de entrenamiento, la WER promedio de los usuarios disminuye, llegando a situarse en un 14,84%, como se muestra en la Figura 1.



Número de series de 57 palabras utilizadas en la estimación MAP

Figura 1: Evolución de la WER y su reducción en función del número de series utilizadas en la estimación MAP

Aunque 14,84% es todavía una WER muy alta en comparación con la obtiene una persona sin problemas de dicción en el mismo sistema, que es de 6,67%, la adaptación al locutor ha conseguido una mejora notable en las prestaciones del RAH, lo que demuestra que es viable utilizar este tipo de sistemas para ayudar a personas con patologías en el habla, tanto en sistemas de apoyo a la logopedia como en sistemas de reconocimiento de comandos orales o de ayuda controlados por voz.

3.2. Resultados del análisis de evaluación de pronunciación mediante LR UV

La evaluación de pronunciación mediante LR UV permite que Vocaliza 1.0 pueda indicar al usuario como progresa su habla conforme éste utiliza la aplicación, ofreciendo una medida de confianza que muestra lo próxima que está el habla del usuario al habla que se considera correcta respecto a lo próxima que está el habla del mismo al habla característica de su patología. Sin embargo, antes de integrar un sistema de evaluación de pronunciación en la aplicación conviene analizar su funcionamiento, con el objetivo de establecer un criterio de evaluación que permita ofrecer al usuario una calificación coherente con la calidad de la pronunciación del mismo, en función de la patología y la mejora que presenta.

Para estudiar el funcionamiento la técnica LR UV se necesita, además de un conjunto de observaciones del habla de cada usuario, al menos un modelo del habla patológica de cada usuario y un modelo común que represente el habla correcta. De esta forma, para realizar el análisis de la técnica LR UV se utiliza el Corpus Alborada, y como modelo de habla patológica se utiliza, para cada usuario, los distintos modelos estimados en el estudio de la adaptación al locutor MAP, mientras que como modelo de habla correcta se utiliza un modelo independiente del locutor, que funciona adecuadamente con la mayoría de los usuarios que no presentan habla patológica.

De esta forma, se extraen medidas de confianza para todas las palabras de cada usuario que no hayan sido utilizadas en la estimación del modelo adaptado que se esté considerando, para mantener la independencia entre test y entrenamiento. Se consideran únicamente aquellas medidas de confianza asociadas a las palabras que el

sistema de RAH haya dado por válidas, considerando el modelo adaptado al usuario, ya que la aplicación sólo evalúa la pronunciación del usuario si considera que el usuario ha pronunciado aquello que se le exigía (hecho que ocurre la mayoría de las veces, si se utiliza un modelo adaptado al usuario y el usuario trata de pronunciar aquello que cada juego de la aplicación exige). Así se obtienen un total de 7815 medidas de confianza para el total de los usuarios, lo que supone un promedio de 1302,5 medidas por usuario, que constituyen un conjunto de medidas suficiente para realizar un análisis estadístico de las mismas.

El comportamiento de dichas medidas es bastante interesante. En primer lugar, se comprueba que para un usuario dado, las medidas de confianza obtenidas siguen una distribución gaussiana, lo que induce a pensar en la posibilidad de establecer un criterio de calificación a partir de las mismas. Además, se observa que las funciones de distribución (fdp) obtenidas para los distintos usuarios y modelos adaptados son muy similares y presentan medias y varianzas muy parecidas. De hecho, se puede comprobar que la distancia de Kullback-Leibler (KL) entre las distintas fdp es muy reducida, lo que permite aproximar todas ellas por una única campana de Gauss, que permitirá establecer un criterio de calificación único para todos los usuarios.

Así, se puede establecer un criterio de evaluación de pronunciación que indicará al usuario la calidad de su pronunciación cada vez que utilice la aplicación Vocaliza 1.0, mostrándole una calificación en función de la medida de confianza obtenida mediante LR UV, independientemente del usuario que utilice la aplicación.

La Figura 2 muestra la campana de Gauss que mejor aproxima las fdp obtenidas en los distintos casos de estudio, y sobre ella, el criterio de calificación y las posibles calificaciones que el usuario puede obtener en función de la medida de confianza obtenida. Los nombres de las calificaciones se eligen para no desanimar al usuario y para ser orientativas, incluso en el caso de usuarios muy jóvenes.

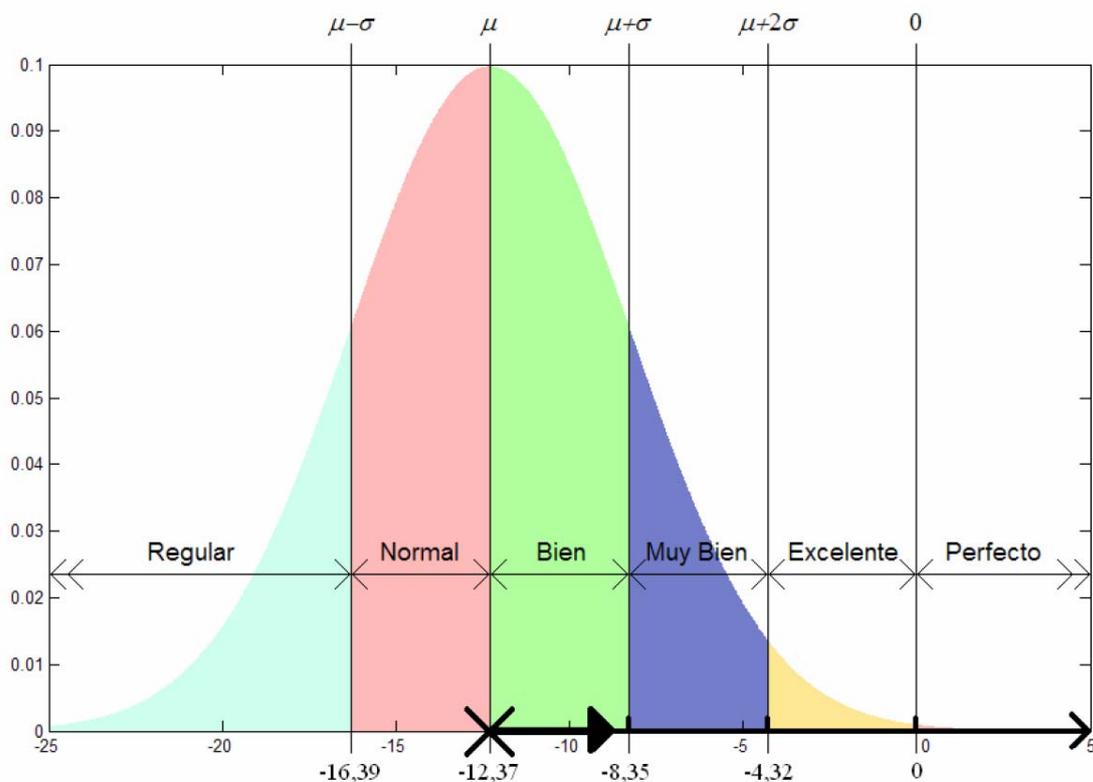


Figura 2: Distribución de las medidas de confianza y criterio de calificación único elegido

4. Aplicabilidad

La amplitud del proyecto realizado, en el sentido de que ha abarcado desde el estudio y análisis de diversas técnicas dentro de las tecnologías del habla, hasta la implementación de una aplicación de asistencia a la logopedia que las utiliza, ha propiciado que la aplicabilidad del mismo se dé en diversas áreas y de distintas formas. Por ello, en los siguientes apartados se comenta la aplicabilidad de cada una de las tareas desarrolladas en este proyecto por separado.

4.1. La aplicación Vocaliza

Sin duda, el resultado de este proyecto que mayor aplicabilidad tiene es la aplicación Vocaliza en sí. Dicha aplicación está implantada y está siendo utilizada actualmente en el CPEE Alborada, y se está implantando en otros centros de educación especial de Zaragoza, para que los alumnos con patologías del en el habla practiquen y mejoren su capacidad de comunicación. Igualmente se pretende colgar en Internet para que cualquier persona o institución pueda disponer de ella.

Esta aplicación puede cambiar el modo de trabajo de los logopedas en estos centros, permitiendo que un único logopeda que disponga de varios ordenadores pueda trabajar con varios alumnos a la vez, simplemente dedicando un mínimo tiempo a configurar la aplicación y a crear los perfiles de usuario que mejor se adapten a la patología particular de cada usuario, y después manteniendo una supervisión mínima de seguimiento para cada usuario, lo que conseguirá mayor eficiencia en el trabajo del logopeda.

Igualmente, el hecho de que los métodos utilizados por la aplicación sean sencillos, atractivos y entretenidos, y hayan tenido tan buena aceptación entre las personas jóvenes que padecen patologías en el habla, que en definitiva son los usuarios finales de la aplicación, permite que los alumnos de estos centros puedan practicar su habla, una vez creado su perfil, sin la supervisión de un logopeda y, puesto que el software es de libre distribución, podrán practicar en casa solos y sin supervisión, cosa que es habitualmente complicada ya que los métodos tradicionales resultan aburridos para las personas jóvenes, y los adultos que les supervisan (padres y tutores) no siempre tienen tiempo suficiente para obligarles a realizarlos.

Puesto que la aplicación de las tecnologías del habla en el campo de la logopedia es un área de investigación en la que el GTC de la Universidad de Zaragoza continúa trabajando, se está ofreciendo cierto soporte a los centros en los que se integra la aplicación, soporte que resulta muy interesante ya que, a la vez, se recibe una realimentación de los logopedas que trabajan en dichos centros, que incluye sugerencias y mejoras en la aplicación, tanto a nivel de facilidad de uso, como a nivel de herramienta logopédica.

Actualmente el GTC sigue trabajando en algunas de esas mejoras, que se centran fundamentalmente en aumentar la capacidad logopédica del programa, incluyendo juegos que practiquen el prelenguaje de forma similar a como lo hacían las aplicaciones comerciales tradicionales, así como juegos que practiquen el nivel más alto del lenguaje: el nivel pragmático.

4.2. Corpus de habla patológica

El corpus adquirido en el CPEE Alborada representa una herramienta muy útil a la hora de investigar en la línea de poner las tecnologías del habla al servicio de personas con discapacidad de comunicación. Dicho corpus permite hacer gran cantidad de estudios y análisis sobre la aplicación de las tecnologías del habla en personas con problemas de dicción, ayudando a conocer un poco más los patrones

que caracterizan el habla asociada a las distintas patologías y así poder tratarlas más eficientemente.

De hecho, la compañía Estadounidense Nuance Communications, Inc. Líder internacional en tecnologías del habla y creadora del sistema *Dragon NaturallySpeaking*, se ha puesto en contacto con el GTC mostrando un gran interés por el corpus adquirido y previsiblemente planteará un proyecto de desarrollo interno relacionado con el uso de las tecnologías del habla para ayudar a personas con discapacidad de comunicación, en el que el GTC participará proporcionando el Corpus Alborada.

Sin embargo, existen técnicas relacionadas con el RAH cuyo análisis requiere mayor cantidad de datos. Por ello se pretende ampliar el corpus, realizando nuevas sesiones de adquisición, tanto en el CPEE Alborada como en otros centros de educación especial.

4.3. La adaptación MAP en locutores con patologías en el habla

Como se ha concluido, la adaptación MAP permite estimar modelos acústicos adaptados a un locutor con patologías en el habla para que el reconocimiento automático del habla funcione de forma excelente con dicho locutor. Estos modelos adaptados son de gran utilidad en el desarrollo de proyectos de implementación de sistemas de ayuda controlados por voz que faciliten el día a día de personas que sufren discapacidades motrices y de comunicación, tales como reconocedores de comandos orales para determinados entornos, o tableros de comunicación controlados por voz, ya que permiten incrementar notablemente las prestaciones del sistema de reconocimiento del habla.

La colaboración con el CPEE Alborada permitiría probar los mencionados sistemas y que los usuarios a los que van destinados pudieran demandar lo que necesiten y evaluar su funcionalidad.

4.4. Verificación de pronunciación como método de evaluación

La existencia de un método de evaluación basado en tecnologías del habla capaz de calificar la pronunciación del usuario y evaluar su evolución permite que personas ajenas a la logopedia puedan hacer un seguimiento de forma sencilla de cómo evoluciona un usuario de la aplicación.

De esta forma, cada usuario de la aplicación puede trabajar solo, o con la supervisión de una persona ajena a la logopedia, y tanto en un caso como en otro saber si su pronunciación va evolucionando. Esto supone un gran avance de forma que el periodo de trabajo en el habla de estas personas no se limita únicamente a los momentos en los que están con el logopeda, permitiéndoles trabajar mucho más el habla, lo que reportará una mejora sustancial en la evolución de estas personas.

Obviamente, la experiencia y los conocimientos de un logopeda siempre son necesarios, y no pueden ser sustituidos por este método de evaluación. El método de evaluación propuesto sólo pretende ser orientativo para el usuario, de forma que éste pueda saber si realmente está progresando o no, y notificar su evolución al logopeda. Al final, es el logopeda el que debe orientar a la persona con patologías en el habla sobre qué debe practicar para mejorar su capacidad de comunicación y cómo debe hacerlo, utilizando los medios que proporciona la herramienta Vocaliza 1.0.

La verificación de pronunciación como método de evaluación planteada en este proyecto es algo innovador y sin antecedente alguno, y aunque se han logrado resultados muy interesantes, todavía queda mucho que avanzar en este campo.

Un estudio muy interesante que se podría plantear sería el desarrollo de una técnica de verificación de pronunciación que ofreciera una medida de calidad absoluta

de la pronunciación del paciente, lo que permitiría diagnosticar la gravedad de cada patología en una escala única (la técnica de verificación de pronunciación planteada en este proyecto ofrece una medida de calidad relativa condicionada a la patología y la progresión del usuario). Una posible forma de realizarlo sería utilizar como medida de confianza únicamente la verosimilitud dado el modelo de habla “correcta”, sin realizar ningún ratio. Sin embargo, esto requiere un análisis profundo, pues en este mismo proyecto se descartó dicha posibilidad por no ofrecer resultados que directamente se adecuaran a la gravedad de la patología de cada usuario.