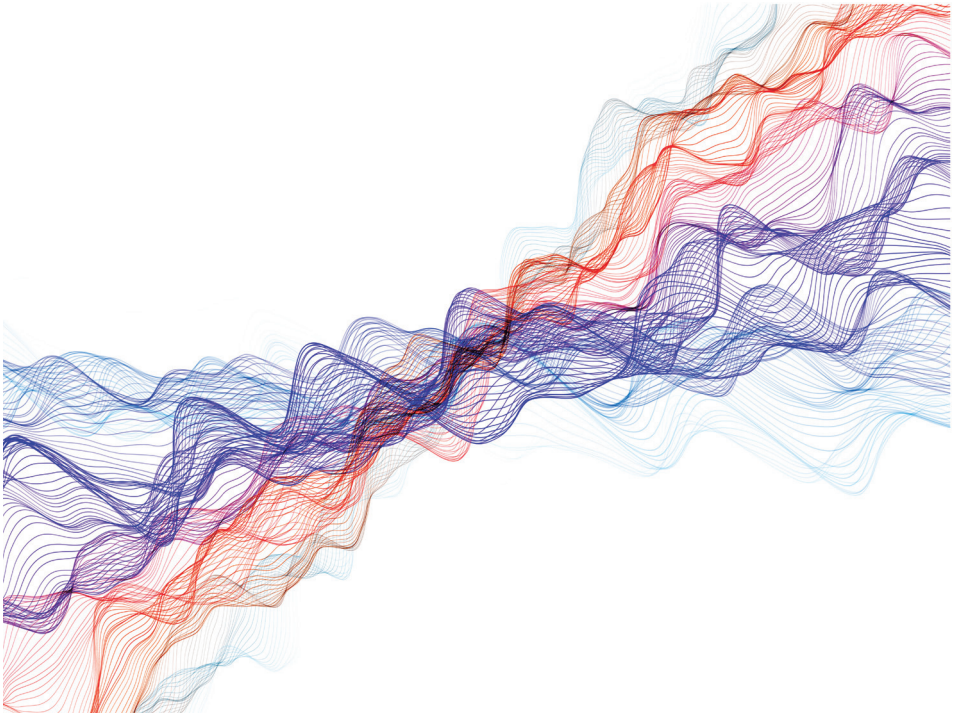
Sistemas de reconocimiento automático del habla:

Aplicación a la investigación lingüística



Ignacio Moreno-Torres

EDITORIAL COMARES

Interlingua

SISTEMAS DE RECONOCIMIENTO AUTOMÁTICO DEL HABLA

Ignacio Moreno-Torres

Sistemas de reconocimiento automático del habla:

Aplicación a la investigación lingüística

Granada, 2021

Colección indexada en la MLA International Bibliography desde 2005

EDITORIAL COMARES

INTERLINGUA

214

ENVÍO DE PROPUESTAS DE PUBLICACIÓN:

Las propuestas de publicación han de ser remitidas (en archivo adjunto, con formato PDF) a alguna de las siguientes direcciones electrónicas: anabelen.martinez@uco.es, psgines@ugr.es

Antes de aceptar una obra para su publicación en la colección INTERLINGUA, ésta habrá de ser sometida a una revisión anónima por pares. Para llevarla a cabo se contará, inicialmente, con los miembros del comité científico asesor. En casos justificados, se acudirá a otros especialistas de reconocido prestigio en la materia objeto de consideración.

Los autores conocerán el resultado de la evaluación previa en un plazo no superior a 60 días. Una vez aceptada la obra para su publicación en INTERLINGUA (o integradas las modificaciones que se hiciesen constar en el resultado de la evaluación), habrán de dirigirse a la Editorial Comares para iniciar el proceso de edición.

El presente trabajo se ha realizado gracias a la subvención recibida de la Junta de Andalucía-Universidad de Málaga (Proyecto UMA18-FEDERJA121) y del Ministerio de Ciencia, Innovación y Universidades (RTI2018- 094846-B-I00).

Colección fundada por: Emilio Ortega Arjonilla y Pedro San Ginés Aguilar

Editorial Comares, 2021
Polígono Juncaril • C/ Baza, parcela 208 • 18220 Albolote (Granada) • Tlf.: 958 465 382
http://www.comares.com • E-mail: libreriacomares@comares.com
https://www.facebook.com/Comares • htpps://twitter.com/comareseditor
https://www.instagram.com/editorialcomares

ISBN: 978-84-1369-200-5 • Depósito legal: Gr. 1415/2021

Impresión y encuadernación: comares

Sumario

Prólogo XI

1. Introducción a la representación y análisis de señales 1
   I. Introducción 1
   II. El dominio del tiempo 1
      1. Señales periódicas 3
      2. Discretización 4
      3. Envolvente temporal 7
   III. El dominio de la frecuencia 8
      1. Transformada de Fourier 10
      2. Enventanado 13
      3. Filtros en frecuencia 15
   IV. Vocoders 17

2. Producción y percepción del habla 19
   I. Introducción 19
   II. Producción de la señal lingüística 19
      1. Anatomía del sistema de producción del habla 19
      2. Modulación de la señal lingüística 21
         A. Respiración 22
         B. Fonación 23
         C. Articulación 25
            a. *Pistas espectrales* 26
            b. *Pistas temporales* 27
            c. *Pistas espectrales dinámicas* 28
   III. Percepción de la señal lingüística 31
      1. Anatomía del sistema auditivo periférico 31
      2. Análisis acústico 32
         A. Intensidad 32
         B. Análisis frecuencial 32
         C. Banco de filtros cocleares 34
         D. Saturación 34

3. Técnicas de reconocimiento automático del habla 37
I. Introducción 37
1. Tipos de sistemas de RAH 37
2. Estructura de un sistema de RAH 38
II. Sistemas de RAH con cadenas de Markov 40
1. Cadenas de Markov de primer orden 40
2. Cadenas de Markov Ocultas (CMO) (En inglés Hidden Markov Models o HMM) . . 42
3. Caso práctico 44
III. Parametrización de señales 46
1. MFCC 46
2. Otros parámetros 51
IV. Otros sistemas de RAH 52
1. Redes neuronales 52
A. Aproximación a las redes neuronales 52
B. Redes neuronales recurrentes (RNN) 54
2. Máquinas de vectores de soporte (Support Vector Machine, SVM) 55
A. Prosodia emocional 56
B. Nasalización 57
V. Entornos para la creación de sistemas de RAH 58
1. HTK 58
2. Kaldi 58

4. Aplicando la tecnología a la investigación lingüística 59
I. Introducción 59
1. Estudiando el proceso de producción 60
2. Estudiando la variación lingüística 62
II. ¿Cómo podemos saber si los resultados de una simulación son fiables? 64

5. Dificultades para el estudio del lenguaje en implantados cocleares 67
I. Introducción 67
II. El implante coclear 69
III. El desarrollo lingüístico del niño con implante coclear 71
1. Principales hitos en el desarrollo 71
2. Preguntas sin respuesta 73
IV. Percepción en ruido 74
1. Efectos del ruido 75
A. Enmascaramiento visual 75
B. Enmascaramiento acústico 76
C. Pistas acústicas no desenmascarables 78
2. Estudio empírico 78
A. Metodología 79
B. Resultados en sujetos normo-oyentes 80
C. Resultados en implantados 82
D. Preguntas sin respuesta 85

6. Percepción auditiva con Implantes Cocleares: explorando diferentes metodologías . . . 87
I. Introducción 87
II. Percepción de oclusivas sonoras en ruido en normo-oyentes expuestos a señales vocodizadas 88

1. Introducción 88
2. Método 88
A. Estrategia de codificación 88
B. Estímulos. 89
C. Procedimiento. 90
3. Resultados 90
4. Discusión 91
5. Preguntas no respondidas 92
III. Simulando la percepción en implantados con un sistema de RAH: exploración preliminar 92
1. Introducción. 92
2. Método. 93
A. Diseño del estudio 93
B. Experimentos 94
C. Estímulos. 95
D. Sistema de RAH 96
E. Datos de referencia de implantados. 96
3. Resultados 97
A. Experimento 1 97
B. Experimento 2 99
4. Discusión 100
5. Preguntas sin respuesta 102

7. Percepción auditiva con Implantes Cocleares: análisis de pistas acústicas. 103
I. Introducción 103
1. Sonoridad en las oclusivas en español. 105
2. La señal recibida por los implantados cocleares 106
3. Organización de este estudio 109
II. Materiales y método 111
1. Base de datos original. 111
2. Vocodizado 111
A. Manipulación de la estructura espectral. 112
B. Manipulación de la envolvente temporal 112
C. Arquitectura y configuración del sistema de RAH 113
D. Cálculo de la tendencia a la sonorización/ensordecimiento 114
III. Resultados. 115
1. ¿Aparece la tendencia a la sonorización si la señal se degrada?. 115
2. ¿Se puede revertir la tendencia a la sonorización?. 118
3. ¿Por qué se producen los cambios de tendencia en los errores de sonoridad? . . 120
IV. Discusión 122
1. Tendencias de error en el sistema de RAH 122
A. ¿Qué condiciones llevan a una tendencia a la sonorización? 122
B. ¿Es posible revertir el efecto de sonorización? 123
C. ¿A qué se debe que la combinación de enmascaramiento energético e informativo produzca la tendencia a la sonorización?. 124
2. Relación entre los resultados en las simulaciones y los datos de implantados cocleares 125
A. ¿ Se pueden explicar los resultados de los implantados cocleares mediante estas simulaciones? 125

B. ¿Por qué se produce la tendencia a la sonorización? 126
C. ¿Tendencia fonológica o fonética? 127
3. Consideraciones finales 128

8. Variación individual y análisis de datos clínicos 131
I. Introducción 131
II. Evaluación del habla en pacientes afásicos 132
1. Medidas finas/groseras 136
2. Medidas de grupo/individuales 137
III. Evaluación de habla nasalizada 137
1. La nasalización 137
2. Midiendo la nasalización 138

9. Precisión articulatoria en la producción de consonantes y vocales en adultos jóvenes y de edad intermedia 141
I. Introducción 141
II. Método 145
1. Experimento de RAH: diseño del sistema 145
2. Experimento de RAH: corpus de entrenamiento y evaluación 145
3. Experimento de percepción en ruido 147
4. Análisis de datos 147
III. Resultados 148
1. Análisis manuales del corpus de palabras y pseudopalabras 148
2. Reconocimiento automático de palabras y pseudopalabras: resultados preliminares 150
3. Experimento de RAH: vocales, consonantes y rasgos fonológicos de las consonantes 151
4. Experimento de percepción en ruido 154
IV. Discusión 156
1. RAH como modelo del reconocimiento de voz humana 156
2. Precisión articulatoria y envejecimiento 158
3. Reflexiones finales 161

Referencias bibliográficas 163
Anexo 1 169
Anexo 2 175
Anexo 3 185

Prólogo

Los avances tecnológicos están en el origen de cientos de cambios en las facetas más diversas de nuestro día a día. Desde nuestra forma de cocinar hasta la forma de comunicarnos o de viajar se han visto afectadas en los últimos decenios por la tecnología. Lo mismo se podría decir de la investigación. Con respecto a esta última el efecto es doble: no solo cambian la forma en que investigamos, además cambia el foco de la investigación. Así ha ocurrido claramente en ámbitos como la biología, donde conforme la técnica ha avanzado, ha ido creciendo el conjunto de fenómenos que podían ser observados. Con respecto a la lingüística, los avances en ámbitos como el procesamiento de señales e inteligencia artificial, los cuales son la base de los sistemas de Reconocimiento Automático del Habla (RAH), abren todo un abanico de posibilidades aún no del todo conocidas.

Como expondremos con más detalle en el capítulo 4, este conjunto de técnicas abre dos puertas a la investigación lingüística. Por un lado, hace posible replicar en condiciones de laboratorio el proceso por el que un hablante aprende una lengua; entre otras cuestiones hoy podemos comprobar qué ocurre si modificamos la información acústica disponible y/o el inventario de fonemas, de unidades léxicas o de estructuras gramaticales durante el proceso de aprendizaje. Por otro lado, ese mismo conjunto de técnicas puede emplearse para estudiar las diferentes formas de variación lingüística (sociolingüística, dialectal, etc.) Por ejemplo, hoy es factible emplear un sistema de RAH para valorar hasta qué punto hay diferencias entre dos sistemas fonológicos o para analizar si las diferencias fonéticas entre dos variantes podrían tener un impacto fonológico. Tal vez lo más interesante de todo esto es que se trata de preguntas a las que en muchos casos sería muy difícil responder si usáramos una metodología más tradicional.

Ahora bien, conocer la tecnología requiere un esfuerzo y, lamentablemente, hay una larga tradición de ignorancia mutua entre tecnólogos y humanistas. Es una situación que no es nueva, la describía ya D. Hirst (2006) en su reseña del texto *Introducing Speech and Language Processing* (Coleman, 2005):

> «It is unfortunate that there is still today an enormous gap between the community of linguists and phoneticians on the one hand and that of engineers and computer scientists on the other. Each community needs the other and, in an ideal world, linguists would provide theoretical frameworks and data which are useful to engineers, while engineers would provide tools which are useful to linguists. The exchange between the two communities, however, is in practice very slow. It often takes decades for ideas which are current in one community to be adopted by the other» (D. Hirst, 2006).

No entraremos en la situación en el campo de la ingeniería. Pero con respecto a nuestro campo, la lingüística, sí creemos que sigue habiendo un cierto escepticismo hacia todo lo que suene a tecnológico. El caso es que el rechazo podría estar en parte justificado: al fin y al cabo, y a diferencia de áreas como la biología, buena parte de los datos con los que trabaja el humanista no son cuantificables, y los ordenadores se llevan mal con todo lo que no sea cuantificable. Aún así, creemos que hay dos motivos por los que los lingüistas deberíamos valorar una buena formación técnica.

En primer lugar, como ya hemos apuntado antes, los avances tecnológicos en procesamiento de señales y, especialmente, en inteligencia artificial, abren la posibilidad de responder preguntas que hasta ahora quedaban sin respuesta. Si queremos responderlas, no nos queda otra que conocer, al menos de forma rudimentaria, esta nueva tecnología. Eso sí, no se trata de que nuestros estudiantes se vuelvan ingenieros que dominen un lenguaje completamente nuevo, basta con que adquieran un conocimiento pasivo que les permita sacar partido de las herramientas desarrolladas por los ingenieros. En segundo lugar, la tecnología que estudiamos aquí es la que sirve de base para crear una nueva comunidad de hablantes virtuales que en número tal vez supere ya al de seres humanos. Nos referimos a los sistemas de interacción humano-máquina que emplean, entre otros, nuestros dispositivos móviles. No debemos olvidar que la creación de estos hablantes digitales implica una serie de decisiones que podríamos considerar parte de la política lingüística: estos dispositivos no hablan / reconocen «una lengua»; más bien debíamos decir que hablan / reconocen «unas determinadas variedades lingüísticas de un selecto conjunto de lenguas». Esto quiere decir que, posiblemente guiados por criterios empresariales, las empresas que crean estos dispositivos están de facto fijando las pautas de un proceso de normalización lingüística; si los lingüistas queremos tomar parte de este proceso, sería conveniente que antes conociéramos la tecnología que hay detrás del mismo.

Este libro resume algunas de las técnicas empleadas en el campo del procesamiento de señales y RAH, y muestra algunos ejemplos de cómo se pueden aplicar estas técnicas en investigación lingüística. No es un libro técnicamente muy detallado; de hecho, es más bien divulgativo en algunos apartados; pero esperamos que contenga los suficientes detalles como para apreciar algunas de las posibilidades que ofrece la tecnología, y animar al lector a aplicar estas nuevas técnicas a la investigación.

El libro tiene tres limitaciones claras, al menos para alguien que pretenda aprender lo suficiente para hacer investigación lingüística aprovechando la tecnología. La primera es que apenas explica las bases matemáticas de las diferentes técnicas que presenta. En lugar de ello hacemos una presentación intuitiva y destacamos, cuando podemos, las aplicaciones y limitaciones. Así lo hacemos por ejemplo al describir los coeficientes MFCC, tan usados en los sistemas de RAH, o la noción de filtro, muy útil para explicar el funcionamiento de nuestro sistema auditivo. Por ello, quien desee profundizar sobre estas y otras cuestiones sería deseable que haga una aproximación más formal que le permita comprender y aprovechar de forma más completa los avances tecnológicos.

La segunda limitación se refiere a la ausencia de referencias a ningún lenguaje de programación. Se trata de una cuestión importante. La tecnología no es más que un instrumento en manos del investigador, pero el aprovechamiento de ese instrumento requiere de un lenguaje de programación que le permita adaptarla a sus necesidades. Por ello, es necesario alcanzar un cierto dominio de lenguajes de programación como Python o R, o incluso el lenguaje incorporado dentro del programa Praat.

La tercera limitación se refiere al hecho de que este libro no trata de toda la tecnología que podría emplear el lingüista, ni tampoco de todas las posibles aplicaciones. En cuanto a la tecnología nos centramos en dos aspectos, y los vemos de forma muy somera. Estos dos aspectos son las técnicas de procesamiento de señales y los sistemas de Reconocimiento Automático del Habla (RAH). La presentación será por tanto de utilidad, o eso esperamos, para quienes se centren en el ámbito de la fonética y la fonología, pero podría ser insuficiente para alguien interesado en el léxico, la gramática, la semántica o la pragmática. En cuanto a las aplicaciones veremos solo un caso particular relacionado con aprendizaje de lenguas y otro relativo a la variación lingüística. Esperamos que, a pesar de que estamos ofreciendo al lector una pequeña muestra de las posibilidades de la tecnología, sea suficiente para despertar su curiosidad.

Así pues, el destinatario ideal de este libro sería un investigador con formación lingüística y curiosidad científica, que desee empezar a conocer nuevas tecnologías para el estudio de la fonética o la fonología. Eso sí, no se asume un fin concreto en la investigación: puede servir al interesado en el aprendizaje de segundas lenguas, en la rehabilitación del habla, en la sociofonética…

El libro se divide en dos partes. En la primera hacemos un recorrido por los dos conjuntos de técnicas antes indicados: las de procesamiento de señales y las relacionadas con el reconocimiento de patrones o, de forma más específica, los sistemas de RAH. Como ya hemos indicado, se obvia el aparato matemático que da soporte a dicha tecnología, aunque sí se discuten en la medida de lo posible algunas limitaciones. En la segunda parte examinamos cómo aplicar esta tecnología. No ocultamos que presentaremos exclusivamente dos ejemplos sobre los cuales tenemos cierta

experiencia acumulada. Además, dedicaremos un capítulo completo a contextualizar cada objeto de estudio. Ello no es una decisión arbitraria, sino que responde nuestra filosofía de fondo con respecto a la relación entre la tecnología y el conocimiento científico. La tecnología es un instrumento que ofrece miles de opciones, y es fácil dejarse seducir por el canto de sirenas de la simulación por ordenador: podríamos pasar meses o años haciendo un estudio que sea un alarde de tecnología pero que no logre aportar nada desde un punto de vista científico o social. Por ello, antes de aplicar la tecnología debemos comprender el problema que abordamos, y debemos valorar qué metodología es la más apropiada en cada caso.

Por último, debemos decir que este trabajo no habría sido posible sin la inestimable ayuda de diversas personas. En primer lugar, la de nuestros compañeros y amigos, y además expertos en la tecnología que intentamos usar, Enrique Nava y Pablo Otero, cuya ayuda y amistad ha sido un inestimable apoyo y sin la cual no habría podido llegar aquí. También debo agradecer la ayuda de los técnicos e investigadores que en diferentes momentos han estado trabajando dentro del laboratorio Calíope, Salvador Florido en un primer momento, y posteriormente María Cristina Armero y Andrés Lozano. En un plano diferente, la presencia de mi familia, y especialmente de la recién llegada Sol Moreno-Torres Lee, aún en la distancia marcada por la pandemia, siempre ha sido el mejor estímulo. Naturalmente, los errores son de mi propia cosecha.

colección:

INTERLINGUA

214

Dirigida por:

Ana Belén Martínez López y Pedro San Ginés Aguilar

Los avances tecnológicos tienen un gran impacto sobre la comunicación lingüística. Por ejemplo, permiten crear nuevos hablantes *virtuales* y nuevas situaciones comunicativas. Desde otra perspectiva, esa misma tecnología se ha mostrado sumamente útil para la investigación lingüística. Así, hoy es factible simular en un ordenador el proceso de adquisición de una lengua, lo que permite hacer una descripción mucho más precisa de la que obtendríamos empleando una metodología clásica.

Ahora bien, los planes de estudio de las facultades de filología apenas incluyen asignaturas de corte técnico. Por ello, no es de extrañar que como colectivo los graduados de estas facultades queden en segundo plano en un ámbito en el que deberían tener un papel destacado. Para cambiar esta situación debemos intentar que, aunque sea de forma pasiva, nuestros estudiantes conozcan con cierto detalle las nuevas tecnologías lingüísticas. Este texto se centra en dos de estas tecnologías: el procesamiento de señales y el reconocimiento automático del habla.

El principal objetivo que perseguimos es hacer una presentación relativamente sencilla de los principales conceptos técnicos. A ello dedicaremos dos capítulos de este libro. Uno de ellos se centra en el análisis de señales, que es la tecnología que se encarga de convertir las señales sonoras en objetos (vectores) susceptibles de una manipulación posterior. En otro capítulo se explica el proceso por el cual, partiendo de esos vectores, podemos construir un sistema de reconocimiento del habla. Un segundo objetivo de este libro es el de mostrar cómo dicha tecnología se puede aplicar en la investigación lingüística. Para ello planteamos dos casos prácticos: análisis del impacto de las pérdidas auditivas sobre la adquisición de la lengua, y cuantificación de las diferencias fonológicas entre dos variedades lingüísticas. Esperamos que, aunque solo describimos una parte pequeña de la nuevas tecnologías lingüísticas y sus posibles aplicaciones, este texto sirva para crear la curiosidad del lector y ayude a reducir la brecha entre la tecnología, por un lado, y la filología y la lingüística, por otro.

COMARES
editorial

ISBN 978-84-1369-200-5
9 788413 692005