

Tecnologías del habla y análisis de la voz. Aplicaciones en la enseñanza de la lengua*

LEONARDO CAMPILLOS LLANOS

Laboratorio de Lingüística Informática - Universidad Autónoma de Madrid
leonardo.campillos@uam.es

RESUMEN: El artículo presenta una revisión de los sistemas de visualización o análisis de la voz y las tecnologías del habla (reconocimiento de voz, síntesis y sistemas de diálogo) que se han empleado en la enseñanza de la lengua materna o extranjera. Se abordan tanto programas comerciales como prototipos de investigación, con especial atención a los recursos para el español. Asimismo, se consideran las recomendaciones y las evaluaciones de los programas expuestas por investigadores y expertos respecto a los procedimientos de corrección, los contenidos pedagógicos o el diseño de la interfaz. Por último, ofrecemos una referencia orientativa sobre cada tipo de aplicación más adecuada.

Palabras clave: tecnologías del habla, reconocimiento de voz, síntesis de voz, sistemas de diálogo, enseñanza de lengua, español como lengua extranjera.

ABSTRACT: The article presents a review of the systems for speech analysis and visualization and the speech technologies (voice recognition, text-to-speech synthesis and dialogue systems) which have been used in first or second language teaching. Both commercial programs and research prototypes are considered, especially those for the Spanish language. Besides, we have taken into consideration the recommendations and the evaluations of the systems made by researchers and experts regarding the correction methods, the pedagogic contents and the interface design. Finally, we include a brief guiding reference for every type of application.

Keywords: speech technologies, speech recognition, speech synthesis, dialogue systems, language teaching, Spanish as a foreign language.

0. INTRODUCCIÓN

La enseñanza/aprendizaje de lenguas es un área de desarrollo de las tecnologías del habla con cierto atractivo por sus productos de investigación. No cabe duda de que estos no se conciben como sustitutos de la instrucción presencial en el proceso de aprendizaje, sino como un complemento que se utiliza fuera del aula para el refuerzo de aspectos externos al currículo del curso o para tratar dificultades particulares de cada alumno.

La mayoría de los investigadores en la enseñanza de lenguas asistida por ordenador (en adelante, ELAO) constata el factor estimulante del uso de los programas informáticos (Ruipérez, 2004). Además, considerando la práctica de la destreza oral, un sistema automático de evaluación permite el aprendizaje en contextos de no inmersión lingüística y suple o facilita la tarea de corrección por parte del profesor. Con todo, su uso posee limitaciones no sólo de tipo técnico sino también pedagógico (*vid.* los aspectos positivos y negativos recogidos por Pennington, 1999: 430).

* Este trabajo ha sido financiado por la Consejería de Educación de la Comunidad de Madrid y el Fondo Social Europeo (FSE) a través de un contrato predoctoral. Quiero expresar mi agradecimiento al Dr. Rafael Martínez Olalla (Universidad Politécnica de Madrid), quien revisó con detalle la primera versión, y al Dr. Antonio Moreno Sandoval, por sus indicaciones para la mejora del artículo.

En este trabajo recogemos una panorámica de herramientas, programas y prototipos del ámbito investigador o comercial que aplicados a la enseñanza de la lengua materna o extranjera (designación que no distinguimos aquí de la de *segunda lengua*, en adelante L2). En algunos casos, dichos desarrollos también se han empleado para la terapia de la voz, como indicaremos cuando corresponda; no obstante, su tratamiento específico excede los límites de este trabajo. Centrándonos en los recursos para la práctica de la producción oral, tampoco abordaremos las aplicaciones del procesamiento del lenguaje natural para la enseñanza, ni los programas de ELAO que únicamente recogen bancos de datos y grabaciones de habla nativa para la enseñanza de la pronunciación, la fonética y la fonología, o la práctica de la comprensión auditiva.

En primer lugar, presentamos un resumen de los tipos de herramientas y tecnologías de habla, junto a una tabla que sintetiza sus características principales, para distinguir cada sistema y orientarse en este panorama de innovaciones (sección 1). Este apartado se estructura como sigue: en la sección 1.1 abordamos los sistemas de grabación y reproducción; en la sección 1.2, los programas de visualización y análisis acústico de la voz; en la sección 1.3, el reconocimiento del habla; en la sección 1.4, la conversión texto-habla; y en la sección 1.5, los sistemas de diálogo. En la sección 2 trataremos aspectos como la evaluación de los sistemas o recomendaciones para su diseño. Para concluir, en la sección 3 recogeremos una serie de recomendaciones para elegir la aplicación más adecuada, y en la sección 4 resumiremos los puntos más significativos de esta revisión bibliográfica. Los programas citados a lo largo del artículo se acompañan de un número de referencia que remite a las direcciones de Internet listadas al final del texto, después de bibliografía (sección 5). Apenas se han incluido imágenes de la interfaz de los programas informáticos que se abordan, ya que las capturas de pantalla se proporcionan en la documentación correspondiente indicada en dicho apartado.

La revisión bibliográfica de artículos, capítulos de libros y materiales en Internet (evaluaciones, reseñas, etc.) ha intentado ser lo más completa posible, aunque el objetivo principal es divisar las tendencias de investigación actuales más que detallar exhaustivamente todos los proyectos. Las fuentes de consulta han sido actas de congresos, informes de proyectos y libros y revistas especializadas en lingüística aplicada, enseñanza de lenguas asistida por ordenador o procesamiento de habla¹, desde mediados de la década de 1990 hasta el año 2010.

1. TIPOS DE HERRAMIENTAS Y TECNOLOGÍAS DE HABLA

Entre los sistemas destinados a la práctica de la destreza oral del alumno –ya sea corrigiendo sus producciones o incorporando un procesamiento más avanzado de la señal de voz– podemos diferenciar, por una parte, las *herramientas de visualización o análisis de la voz*, y por otra, las denominadas *tecnologías del habla* (el reconocimiento de voz, la síntesis o los sistemas de diálogo). Para el tratamiento pedagógico de la producción del habla, Llisterri (2006) distingue con más detalle entre²:

¹ Se han consultado especialmente revistas de enseñanza de lenguas asistida por ordenador (*Computer Assisted Language Learning*, *CALICO*, *Language Learning and Technology*, *ReCALL*, *Computer Speech and Language*, *Apprentissage des Langues et Systèmes d'Information et de Communication*), lingüística aplicada (*System*, *Language Learning*) o procesamiento del habla (*Speech Communication*); y actas de congresos especializados (*AESLA*, *ASELE*, *EUROCALL*, *Fonetik*, *InSTILL/ICALL*, *INTERSPEECH*). Cabe destacar los números monográficos de *Language Learning and Technology* sobre tecnología y aprendizaje de la pronunciación (octubre de 2009, vol. 13, 3) y de *Speech Communication* sobre tecnologías de habla para la educación (octubre de 2009, vol. 51, 10).

² Una explicación ilustrada con capturas de pantalla de diferentes programas se ofrece en la página web de este investigador: http://liceu.uab.es/~joaquim/applied_linguistics/L2_phonetics/EAO_Pron.html

- sistemas de grabación y reproducción: su aproximación es semejante a los ejercicios empleados en los laboratorios de idiomas;
- sistemas que usan información visual: pueden presentar dos tipos de visualización:
 - información acústica de la onda sonora, por ejemplo, espectrogramas, oscilogramas o una representación de la curva melódica; y,
 - información articulatoria de la posición de los órganos fonadores, que a veces se acompaña de una animación, un vídeo o ejemplos sonoros;
- sistemas de reconocimiento de habla: pueden proporcionar una puntuación al alumno y abordar dos aspectos de la producción oral:
 - los contenidos de aprendizaje del programa o curso en que se incluyen (del nivel gramatical, léxico o pragmático, especialmente la práctica de la conversación); y,
 - la práctica y la enseñanza de la pronunciación, que puede localizar errores y aportar consejos para mejorar la articulación de los sonidos.

Evidentemente, cada enfoque no es aislado, y existen programas con diferentes aproximaciones; además, hay que añadir los que incorporan síntesis de habla y sistemas de diálogo. Como resumen, ofrecemos una tabla con las características de las aplicaciones tratadas. Cuando no se ha localizado información sobre un rasgo, se ha dejado el campo en blanco. Las abreviaturas usadas para cada aspecto son las siguientes:

- Len. (lengua): A: alemán; Ar: árabe; C, chino; E, español; N: neerlandés; I, inglés; It: italiano; J: japonés; S: sueco; l. es contracción de 'lenguas'.
- Tecnología/enfoque: G/R: grabación/reproducción; Vis: visualización de espectrograma (esp.), espectro (sp.), oscilograma (osc.), entonación (ent.), órganos articulatorios (órg. art.); carta de formantes (cf); tv; triángulo vocálico y/o carta de formantes; RH: reconocimiento de habla; S: síntesis de voz; SD: sistema de diálogo; pt: pantalla táctil.
- Dest. ling. (destreza lingüística que se practica): A: comprensión auditiva; E: producción escrita; Fon: aprendizaje de la fonética y la fonología; L: comprensión lectora; O: destrezas orales; Prg: destrezas pragmáticas; Pron: pronunciación; G: gramática; V: vocabulario; C: aspectos culturales.
- Aplicac. (aplicación ya llevada a cabo o posible de realizar): EL: enseñanza de lenguas; T: terapia de personas con necesidades especiales; TE: test de evaluación.
- Int. (interfaz): L: diseño lúdico; C: diseño para el análisis científico; A: actividades de práctica y evaluación.
- Disp. (disponibilidad): Com: comercial; Inv: prototipo o proyecto de investigación; LD: libre distribución.
- Us. (usuario): A: adultos; N: niños.

SISTEMA	LEN.	TECNOLOGÍA/ ENFOQUE	DESTR. LING.	APLICAC.	INT.	Us.	DISP.
Accent coach	I	Vis. ent, órg. art., RH	Pron, Fon	EL	C	A	Com
Ancalvoz	-	Vis. osc., ent, cf	Pron, Fon	EL, T	C	A	Inv
Baldi	6 l.	SD	O, V, Pron	EL / T	L	A, N	Inv
BetterAccentTutor	I	Vis. esp, osc., ent.	Pron	EL, T	C	A	Com
Bortolini (2002)		S	A, L	EL / T		N	Inv
Brown (2004)	I	RH	O, G	EL	L	A, N	Inv
CallJ	J	RH	Pron, G, L	EL			Inv
CandleTalk	I	RH	O	EL	A	A	Inv
Cassell (2004)	I	RH	O, Prg			N	Inv
Cmp. Sp. Lb (CSL)	-	An. acústico de voz	Pron	EL, T	C, L	A, N	Com
Colorado Lit. Tutor	I	SD	O, L	EL / T	L	A, N	Inv
CSLU Toolkit	I, E	SD	O, Pron	EL / T	L, A	A, N	LD
DARWARS	Ar	SD	O, C, V	EL	L	A	Com
DEAL	I + 5 l	SD	V, G	EL	L	A	Inv
EduSpeak	9 l.	RH	Pron	EL		A, N	Com
English for kids	I	RH	Pron	EL	L, A	N	Com
Español interactivo/en marcha	E	G/R	A, L, E, O	EL	A	A	Com
EyeSpeak	I	Vis. osc., ent, RH	Pron, Fon	EL	L	A	Com
GETARUN/SLIM	I, It	S, RH	A, Pron	EL			Inv
Gómez <i>et al.</i> , 1997	I	Vis. osc., ent, cf	Pron, Fon	EL	L	A, N	Inv
HearSay	I	RH	Pron	EL	L	A, N	Inv
Hwe (1997)	E	G/R	Pron, Fon	EL	C	A	Inv
I SEE	I	pt, S	V	EL	A	N	Inv
IBM Speech Viewer	-	An. acústico de voz	Pron	EL, T	L		Com
ISLE	I	RH	Pron	EL	A	A	Inv
Learn to Speak Span.	E	RH	O, Prg, V, G	EL	L, A	A	Com
Let's go	I	SD	O	EL		A	Inv
LISTEN	I	RH	Pron, L, L	EL	A	N	Inv
Microworld	I	RH	O, Prg	EL	L	A	Inv
MyET / MyCT	I. C	V. osc., ent., órg art., RH	Pron,	EL	A	A	Com
NativeAccent	I	RH	Pron	EL	L, A	A	Com
Neri <i>et al.</i> , (2008)	N	RH	Pron, O	EL	L, A	A	Inv
PARLING	I	RH	Pron, V, L	EL	L	N	Inv
PLASER	I	RH, vis. órg. art.	Pron	EL	C	A	Inv
Praat	-	Vis esp, osc., ent., sp, S	Pron	EL, T	C	A	LD
Pron. y Fonét. 2.0.	E	G/R, vis. órg. art.	Pron, Fon	EL	C	A	Com
Pronto	E	RH	Pron	EL	L	A, N	Inv
ProNunciation	I	G/R, vis osc., órg. art.	Pron	EL	C	A	Com
RosettaStone	30 l.	RH	O	EL	L, A	A	Com
Saybot	I	RH	O	EL	A	A	Com
Seneff <i>et al.</i> (2004)	C	SD	O, V, G	EL	A	A	Inv
Seneff <i>et al.</i> (2007)	I	SD	O	EL	L	A	Inv
SPACE	N	S/RH	L	EL		N	Inv
Speech Analyzer	-	Vis. osc., ent, sp, S	Pron, Fon	EL, T	C	A	LD
Speech Filing Syst.	-	Vis. esp, osc., ent., S	Pron	EL, T	C	A	LD
STAR	I	RH	Pron, V, L	EL	A	N	Inv

STRAIGHT/SNACK	-	S	Pron, Fon	EL	C	A	Inv
Subarashii	J	RH/SD	Pron O V G	EL		A	Inv
TAIT	E A	RH	O, Pron	EL		A	Inv
TBALL	I	RH, pt	L	EL	A	N	Inv
Tell me more/TTM	8 l.	V. osc, ent, órg art, RH	Pron, O, A.	EL	L, A	A, N	Com
TraciTalk	I	RH	O, Prg	EL	L	A	Inv
Versant	I E Ar	RH	O, Pron	TE		A	Com
VICK	-	Vis. esp, osc, ent, sp.	Pron	EL, T	L	A	Inv
Ville	S	Vis. esp.	Pron, G/R	EL	A	A	Inv
VISHA	-	Vis. osc., ent, sp	Pron, Fon	EL, T	C, L	A, N	Inv
VisiPitch	-	Vis. esp, osc, ent, sp., tv	Pron	EL, T	C, L	A, N	Com
WASP	-	Vis. esp, osc., ent., sp.	Pron	EL, T	C	A	LD
Watch Me! Read	I	RH	O, L	EL	A	N	Inv
WaveSurfer	-	Vis. esp, osc., ent., sp	Pron	EL, T	C	A	LD
WinPitch	-	Vis. esp, osc., ent., tv, S	Pron	EL, T	C	A	Com
WinSnoori	-	Vis esp, osc., ent., sp, S	Pron	EL, T	C	A	LD
Word War	C	SD	O, V, Pron	EL	L	A	Inv
Zengo Sayu	J	RH	G, V	EL	L		Inv

Tabla 1 - Características de los sistemas y los programas analizados para la enseñanza de la lengua

1.1. Sistemas de grabación y reproducción. Algunos programas educativos para la práctica de la producción oral o la enseñanza de la fonética y la fonología permiten al usuario grabar sus propias producciones orales y compararlas con el modelo de lengua que incluyen (Llisterri, 2006). Su base pedagógica estriba en el sistema de audición y repetición propio de los enfoques de enseñanza estructuralistas y audio-linguales de los años 50.

No obstante, la efectividad de este método no resulta clara. Mientras que los experimentos de Akahane-Yamada, Tohkura, Bradlow y Pisoni (1996) muestran que únicamente el entrenamiento de la percepción puede ser efectivo, los resultados de investigación de Celce Murcia y Goodwin (1991; *apud* Eskenazi, 1999) indican que la repetición de sonidos no parece ser una forma eficaz de aprender a articularlos correctamente. Su debilidad principal puede estibar en que no se realiza una corrección apropiada de la pronunciación del usuario, siendo él quien debe evaluar la forma como pronuncia. Esto ocurre en la aplicación desarrollada por Hwu (1997) para la enseñanza y la práctica de la fonética española, o en los programas ProNunciation (*vid.* reseña de Brown, 2000), o Pronunciación y Fonética v. 2.0. (*vid.* reseña de Corsbie y Gore, 2002). En un sistema más reciente para el aprender el sueco, Ville (Wik y Hjalmarsson, 2009), se emplea este enfoque en la versión para el nivel inicial, pero en otros niveles ya se incorpora análisis acústico del habla y corrección de la producción fonética. De entre todos estos programas, destacamos Español interactivo (*vid.* reseña de Adams, 1998) y Español en marcha (Gimeno Sanz, 1998), desarrollados en la Universidad Politécnica de Valencia, por incluir grabaciones de nativos en actividades que simulan interacciones reales, integrando el aprendizaje de contenidos fonéticos, gramaticales y léxicos. Quizá, como proponía Jones (1997) para la enseñanza de la pronunciación, la tendencia es superar un enfoque de audición y repetición mecánica integrando la práctica en actividades más libres de tipo comunicativo, que simulen situaciones auténticas.

1.2. Visualización y análisis acústico de la señal sonora.

1.2.1. Oscilograma, curva melódica y espectrograma. Las primeras aplicaciones de visualización del habla, que datan de los años 60, surgieron para la rehabilitación de personas sordas, y posteriormente se aplicaron a la enseñanza de lenguas; ejemplos de ello son VisiPitch y Computerized Speech Lab (CSL), ambos de Kay Elemetrics (actualmente, Kay Pentax), IBM

Speech Viewer, o el más reciente VICK (VISual FeedbaCK) (Nouza, 1998). Estos sistemas permiten visualizar el resultado de un procesamiento avanzado de la voz del estudiante o el modelo que imitar, entre otros:

- el oscilograma: con la información de la intensidad de la señal sonora, importante para la detección del acento;
- la estimación de la curva melódica: para la visualización de la entonación; y,
- el espectrograma: para el análisis visual de los timbres vocálicos y de las características acústicas de las consonantes.

La herramienta WinPitch (Germain y Martin, 2000; Martin, 2005) presenta todos los tipos de información visual anteriores y permite incluso anotar la señal sonora del usuario, reproducirla a velocidad más lenta o resintetizarla con la prosodia correcta, aunque intentando simplificar la representación gráfica de la curva melódica. La manipulación y síntesis de la onda original también se puede realizar en programas de libre distribución como WaveSurfer, Praat (*vid.* figura 1 abajo), WinSnoori, o Speech Filing System (SFS) 4, que ofrecen el espectrograma de la voz, el oscilograma, la curva melódica, la detección de formantes o el espectro de una porción de la señal³. La conveniencia de un programa u otro dependerá de la cantidad de información que requiera el usuario sobre la onda sonora. Así, por ejemplo, en el ámbito investigador se emplea para el análisis acústico el lenguaje de programación MATLAB (figura 2), para el cual ya existe una herramienta llamada COLEA, que permite, entre otros, la grabación de una señal o la visualización de una gran riqueza de datos (espectro y espectrograma, análisis de tono y formantes, etc.).

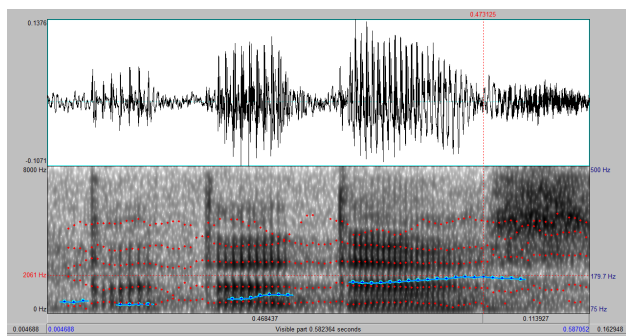


Figura 1 – Oscilograma (arriba) y espectrograma (abajo) de una señal de voz obtenidos con Praat. Entre otros, se muestran el contorno melódico de la entonación y la estimación de los formantes vocálicos.

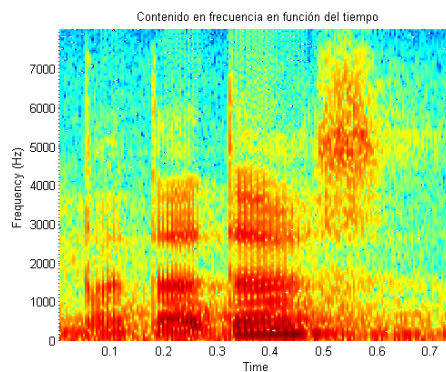


Figura 2 – Espectrograma obtenido con MATLAB

Las herramientas de visualización se han venido utilizando en el análisis contrastivo de las producciones nativas y no nativas, ya sea por parte del profesor o fonetista con fines puramente de investigación, ya sea por el propio alumno, para que aprenda los sonidos visualizándolos (*vid.* la sección siguiente, §1.2.1.1). No obstante, su uso directo en el aula de idiomas puede representar dificultades si el docente carece de rudimentos de fonética acústica, además de que podría *asustar* en cierto modo al aprendiz. Como señalan Gómez Vilda *et al.* (2008), estas herramientas precisan vencer el *salto semántico* mediante el diseño de interfaces apropiadas, actividades lúdicas o procedimientos de corrección sencillos.

Por ello, una aproximación más didáctica es la integración de este tipo de datos acústicos en un programa elaborado propiamente para la enseñanza de la lengua. Es el caso de ProNunciation

³ Una completa relación de programas y recursos disponibles puede consultarse en la página del Speech and Hearing Institute (www.speechandhearing.net). También se ofrecen capturas de pantallas y funcionalidades de muchos programas en la siguiente página web de J. Llisterri: http://liceu.uab.cat/~joaquim/phonetics/fon_anal_acus/herram_anal_acus.html

(*vid.* reseña de Brown, 2000), para la práctica de la pronunciación, que incorpora el oscilograma de la señal, o de Speaker (Cazade, 1998), que muestra la curva melódica de la entonación. Un sistema más reciente, BetterAccentTutor (Kommissachirk y Kommissachirk, 2000), incluso proporciona una corrección de la curva melódica, el acento o el ritmo del enunciado. Accent Coach (*vid.* reseña de Taylor, 1999), además de mostrar la curva melódica, incorpora reconocimiento de habla (para más detalles véase el artículo de Martin, 2005). Otros programas comerciales, como Tell me more o Talk to me, incluyen reconocimiento de voz y también ofrecen el oscilograma o la curva melódica para que el alumno visualice su producción sonora y la compare con el modelo nativo.

1.2.1.1. Visualización del oscilograma, la curva melódica y el espectrograma de la voz de producciones nativas y no nativas. Gracias a las herramientas de visualización, cada vez más potentes y accesibles, se han realizado estudios fonéticos tanto de muestras de nativos como de no nativos. Como ejemplo de esto último se puede citar la aportación de Molholt y Hwu (2008), quienes se valen del uso de espectrogramas de las consonantes aspiradas sordas y sonoras del hindi, pronunciadas por nativos o por estudiantes americanos, para comparar las diferencias en la pronunciación de cada uno.

Pero una aplicación más allá de los fines teóricos o descriptivos es el uso de este tipo de sistemas para el aprendizaje de la pronunciación por parte del alumno, quien visualiza y realiza el análisis de sus propias producciones. Por lo que respecta a la visualización de espectrogramas, ya en los años 90 Labrador Gutiérrez y Fernández Juncal (1994) propusieron el uso del sistema VISHA (Visualizador del Habla)⁴ para la enseñanza de la fonética y la pronunciación del español. Esta herramienta consta de varios módulos, entre los que se incluyen ISOTON, para la práctica de la entonación, la intensidad o rasgos de los sonidos como la sonoridad o la fricatividad, y Pc-VOX, que permite visualizar el espectrograma, la forma de onda o la intensidad de una frase, junto a otros parámetros más detallados. El método de trabajo del alumno pasa primero por visualizar con PcVOX los sonidos conflictivos y luego repetir y grabar las propias producciones; posteriormente, el alumno imita la producción del profesor con el programa ISOTON. Los autores citados lo aplicaron a alumnos anglosajones que practicaban la pronunciación de las vocales /e/ y /o/ (que tienden a diptongar en [ɛɪ] y [oʊ]), la consonante fricativa velar /x/ (que suele aspirarse y producir el sonido [h]), la /r/ y la /r/ (que generalmente se realizan como aproximante, [ɾ]) y las oclusivas /p t k/ (que se pronuncian aspiradas). Los alumnos experimentaron, aunque en diferentes grados, una mejora generalizada de la pronunciación.

La visualización de la forma de onda de los enunciados puede ser útil para la adquisición del valor fonológico de la duración de los sonidos. Por ejemplo, MotohashiSaigo y Hardison (2009) proponen su uso para el aprendizaje de la percepción y la distinción de las consonantes y vocales geminadas en japonés (p. ej., *sasu*, *sassu*, con geminación de /s/, y *saasu*, con alargamiento de /a/), y comentan un experimento con resultados significativos en la identificación de estos sonidos tras el entrenamiento.

Respecto a la visualización de la curva melódica, cabe citar las propuestas de Chun (1998: 81-87) para integrarla en la enseñanza de la entonación: aportar corrección visual, dotar a los estudiantes de habla auténtica y variada, grabar y analizar interacciones entre hablantes, y realizar un seguimiento del progreso del estudiante. Hardison (2004), por ejemplo, muestra un experimento de uso de Computerized Speech Lab (CSL) para el aprendizaje de la prosodia del francés, con resultados positivos en la generalización a nuevos enunciados; y otro test para esclarecer la relación

⁴ La herramienta VISHA fue desarrollada por la Escuela Técnica Superior de Ingenieros de Telecomunicación de la Universidad Politécnica de Madrid en colaboración con el departamento de Filología de la Universidad Nacional de Educación a Distancia.

de la prosodia y la adquisición del léxico en el aprendizaje a largo plazo. Por su parte, Molholt y Hwu (2008) abordan su uso en el aprendizaje de los tonos del chino.

Levis y Pickering (2004) también plantean los beneficios del apoyo visual para el aprendizaje de la entonación, pero ilustran mediante un experimento la necesidad de abordar la enseñanza y la práctica de la misma no solamente en la oración, sino también en el nivel discursivo. En efecto, de modo semejante a como las tabulaciones o convenciones tipográficas marcan los límites entre párrafos en los textos escritos, el ascenso del tono marca el inicio de cada grupo de enunciados que trata un nuevo tópico en el discurso (que Levis y Pickering denominan *paratono* o *párrafo entonativo*). Asimismo, la entonación dentro y entre dichos párrafos de habla puede mostrar una actitud de convergencia hacia el oyente o de distanciamiento. Así, los hablantes no nativos pueden necesitar comprender dichos patrones melódicos para mejorar su entonación, en la que suelen predominar patrones descendentes y suspendidos, resultando en un habla menos dinámica o con menor fluidez (*vid.* Hincks, 2005a, para un estudio de la fluidez y la variación de la entonación de las producciones orales de estudiantes de inglés y una propuesta de evaluación automática; o el estudio de Hincks y Edlund, 2009, acerca del efecto de la corrección visual sobre la variación en la entonación).

Por último, Toledo (2005) expone aplicaciones de la visualización de todos los aspectos acústicos anteriores mediante Speech Analyzer, que puede ser usado en combinación con otros programas como WinCecil (de tipo experimental) y Phonology Assistant (todos estos fueron desarrollados por el Summer Institute of Linguistics). Speech Analyzer ofrece el espectrograma de una señal sonora, la variación de la frecuencia fundamental, el análisis espectral o medidas de duración, y también permite la manipulación de la onda para realizar tests perceptivos. Con el fin de corregir la pronunciación de aprendices de español cuya lengua materna es el francés, Toledo muestra espectrogramas y oscilogramas de las vibrantes simple (/r/) y múltiple (/r/), así como de la fricativa velar sorda /x/ y de la oclusiva velar sonora /g/ frente a su alófono aproximante intervocálico [ɣ]. Además, ofrece curvas de intensidad para distinguir palabras que únicamente se diferencian en el acento (p. ej.: *límite*, *limite*, *limité*)⁵ y curvas de entonación que diferencian oraciones interrogativas y declarativas.

Las comparaciones entre la forma de onda nativa y no nativa pueden ser de gran utilidad para la enseñanza empírica de la pronunciación, a partir de los datos acústicos de las producciones orales. No obstante, parece imprescindible la supervisión del estudiante por parte de un profesor o experto en fonética, por lo que recomendamos descartar su uso en el autoaprendizaje.

1.2.2. Carta de formantes y triángulo vocálico. Otros sistemas incluyen la visualización de la información acústica de las vocales en un gráfico similar a una carta de formantes. Como explica Gil Fernández (2007: 436), dicha representación dispone los valores en frecuencia del primer formante vocálico (F1) en el eje de ordenadas y los correspondientes al segundo formante (F2) en el eje de abscisas, de modo que el punto coordinado de los dos ejes representa el timbre propio de la vocal y permite localizarla respecto al resto de vocales (*vid.* figura 2)⁶.

⁵ Los estudiantes francófonos de español experimentan dificultades en la adquisición del patrón acentual castellano, pues tienden a acentuar la última sílaba por interferencia de su lengua materna.

⁶ A veces se representa el tercer formante (F3) en lugar del primero o el segundo, o incluso se puede emplear un tercer eje vertical para expresar los valores de F3 en un espacio tridimensional, como en las representaciones que muestran en su libro Bernal Bermúdez, Bobadilla Sancho y Gómez-Vilda (2000).

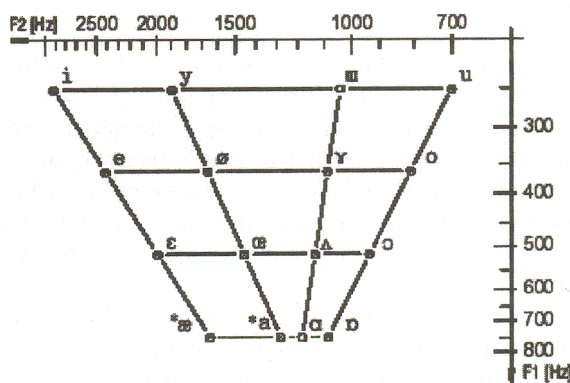


Figura 3 – Carta de formantes de las vocales cardinales primarias y secundarias (Delattre *et al.*, 1952; *apud* Gil Fernández, 2007: 437)

Este tipo de información visual (junto a la curva melódica, la forma de onda con la información acerca de la intensidad y el espectrograma) se presenta gráficamente en el sistema VisiPitch, el cual fue usado por Molholt y Hwu (2008) para exponer algunas cuestiones de corrección fonética de producciones no nativas. Dichos investigadores comparan el espectrograma o la posición de los sonidos en el triángulo vocálico de las vocales del inglés o del español producidas por el nativo y por el no nativo, de modo que se aprecien visualmente sus diferencias.

Además de la información anterior, otro tipo de interfaz también disponía de una ventana donde se presentaba la información de la evolución del movimiento de los formantes en el triángulo vocálico, o un esquema de rasgos consonánticos, lugar de articulación o grado de apertura (*vid.*, por ejemplo, el sistema desarrollado para el aprendizaje del inglés por Gómez Vilda *et al.*, 1997). La aplicación Ancalvoz, desarrollada por este equipo de investigadores y programada en el lenguaje MATLAB, presenta el oscilograma y el espectrograma junto a la información de la dinámica de los sonidos vocálicos en una tabla de formantes. Dicha forma de visualización es realmente valiosa para el fonetista, pero puede ser difícil de interpretar en el aprendizaje autónomo del alumno, que depende del apoyo explicativo del profesor (Hincks, 2003:5; Neri, Cucchiari, Strik, Boves, 2003:6). De esta forma, se ha propuesto el diseño de interfaces de usuario más lúdicas para la corrección del alumno: por ejemplo, para el aprendizaje de las vocales inglesas, Gómez Vilda y sus colaboradores crearon un entorno gráfico que simula una tirada de dardos, de manera que cuanto más se acerca la pronunciación del alumno a la nativa, más se aproxima el dardo al centro de la diana, indicándole además hacia dónde se ha desviado de la pronunciación modelo. Igualmente, estos investigadores desarrollaron un simulador de fórmula uno en el que el vehículo va conduciendo más centrado o se va saliendo más de la carretera conforme su pronunciación sea más semejante a la que ha de imitar (*vid.* más detalles en Gómez Vilda *et al.*, 2008).

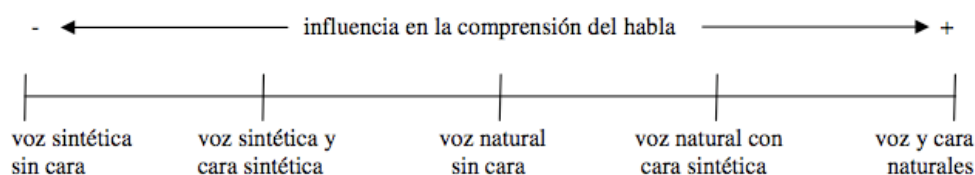
1.2.3. Visualización de órganos fonadores y movimientos articulatorios. Los aprendices de una L2 quizá dependan más de la información visual en la comprensión auditiva –al igual que los bebés en la adquisición de su lengua materna–, como indica Flege (1998: 372). Este hecho explicaría la gran dificultad que los estudiantes experimentan para comprender mensajes por teléfono. Lo que aún no queda del todo claro es el tipo de información visual más influyente en la percepción del habla, o más determinante para el aprendizaje de la pronunciación.

Por un lado, se han llevado a cabo estudios sobre el efecto de la retroalimentación visual en la adquisición de los rasgos articulatorios de una L2. Décadas atrás, el fonetista e investigador Flege (1988) empleó el glosómetro optoelectrónico, un dispositivo de visualización del movimiento de la lengua del propio locutor mientras habla. El glosómetro se introduce en la boca y consiste en una pieza de plástico que se sujeta al paladar y va provista de sensores, los cuales van registrando los

movimientos articulatorios. Flege usó este dispositivo para estudiar la neutralización articulatoria de las vocales /i/ y /a/ del inglés por parte de una hablante nativa de español, que tiende a realizarlas con los sonidos [i] y [a] respectivamente. A la participante en el experimento se le presentaban en una pantalla representaciones esquemáticas de las posiciones de la lengua que tenía que llegar a alcanzar a partir del punto de articulación de la vocal. Pese a que los resultados no permiten determinar la influencia determinante de la información visual, los problemas de pronunciación en una L2 parecieron eludirse en cierto grado con el apoyo de cierta forma de visualización.

Un curioso fenómeno que demuestra la influencia de la información visual en la percepción del habla es el llamado *efecto McGurk* (McGurk y MacDonald, 1976; MacDonald y McGurk, 1978). Se puede observar con un experimento sencillo que consiste en que a un individuo se le presenta un sonido /ba/ por un canal auditivo (por ejemplo, mediante unos auriculares), y simultáneamente, por un canal visual (por ejemplo, en un video), una persona articulando la sílaba /ga/. Debido al conflicto entre la información visual y auditiva, el sujeto tiende a percibir /da/, un sonido intermedio entre ambos. La importancia de la percepción visual en la comprensión oral resulta, pues, innegable, aunque algunos investigadores consideran que su influencia parece reducirse al área de la boca y quizá la parte interior de los labios (Flege, 1988: 370). Esto parece haberse confirmado en estudios más recientes de percepción del habla que han empleado modelos virtuales (obtenidos mediante articulografía electromagnética) para reproducir los movimientos articulatorios de los labios y la lengua (Badin *et al.*, 2010). Los resultados obtenidos por el citado grupo de investigadores constatan la predominancia de la lectura de los labios en la comprensión del habla, aunque podría complementar este proceso leer los movimientos de la lengua. Así, también se han hecho estudios sobre la articulación lingual; por ejemplo, comparando la visualización de movimientos reales y sintéticos (Engwall y Wik, 2009a y 2009b).

Precisamente, otros investigadores han correlacionado el grado de comprensión del habla y lo natural que resulta cualquier forma de visualización que acompaña a la producción oral. En efecto, la interacción de una persona con otra no produce el mismo efecto que si la comunicación se realiza con un personaje virtual, posiblemente por el grado de cercanía o naturalidad que experimentamos en la comunicación entre humanos. Especialmente interesante al respecto es el test que explican Beskow *et al.* (1997); sus resultados indican que la influencia de cada tipo de información se ordena de este modo:



En todo caso, sea cual sea el alcance o la importancia de la información visual en el habla, muchos sistemas para el aprendizaje de la lengua han ido incorporando algún tipo de apoyo visual. Muchas aplicaciones –especialmente para la práctica de la pronunciación– cuentan con una representación articulatoria del tracto vocal y los órganos fonadores que complementan a la información acústica. Es el caso de los programas Tell me more o Talk to me, o My English Tutor (MyET) y My Chinese Tutor (MyCT), respectivamente para el aprendizaje del inglés y el chino (ref. n.º 27). Otros sistemas ofrecen explicaciones más detalladas de la posición y el movimiento de los órganos de articulación (como *Pronunciación y Fonética v. 2.0.*, *vid.* reseña de Corsbie y Gore, 2002), o incluso animaciones que describen los movimientos articulatorios (véase la aplicación desarrollada por Hwu, 1997).

No obstante, como indica Llisterri (1997, 2001, 2006), la representación visual del tracto vocal o de los movimientos articulatorios, además de requerir un procesamiento de la señal sonora no exento de complejidad, puede resultar ineficaz. En efecto, debido al fenómeno de compensación articulatoria, el hablante puede llegar a producir un sonido con una configuración de los órganos fonadores distinta a la del modelo. Otros investigadores (Neri, Cucchiari, Strik y Boves, 2003: 2) han propuesto incluir en los sistemas la visualización del movimiento de los labios para mejorar la producción y la percepción del sonido, en consonancia con la reflexión citada de Flege (1988: 370). Probablemente sea positiva la inclusión de ambos tipos de representaciones.

- Por un lado, los órganos fonadores y articulatorios internos, que resultan más fáciles de exponer en una figura o animación artificial o incluso un avatar virtual. De hecho, Eriksson *et al.* (2005) recomiendan lo siguiente⁷: aportar la referencia visual del paladar y la mandíbula en la representación de los movimientos articulatorios (preferiblemente con imágenes tridimensionales), así como destacar (por ejemplo, con otro color) los rasgos o puntos importantes de la articulación.
- Por otro lado, la articulación externa de los sonidos también parece importante, y la mejor manera como se puede incorporar es mediante un video que muestre a auténticos nativos pronunciando una palabra o frase, centrándose en la zona de la boca y el movimiento de los labios.

Todo ello parece adecuado siempre que los modelos visuales de los órganos fonadores resulten sencillos y con bajo nivel de detalle para no abrumar al usuario (Eriksson *et al.*, 2005). Para mejorar las representaciones en estos sistemas, seguramente resultarán provechosos los conocimientos y progresos recientes en fonética articulatoria.

1.3. Reconocimiento automático del habla. El uso del reconocimiento automático del habla en la enseñanza de lenguas es un área en la que han proliferado no pocas aplicaciones informáticas, investigaciones científicas e incluso tesis doctorales⁸. Aparte de la enseñanza y práctica de la pronunciación (aspecto que trataremos después en §1.3.1), uno de los enfoques de uso es la integración en las lecciones de un curso completo de aprendizaje de una lengua (junto al vocabulario o la gramática). Los sistemas suelen incorporar grabaciones de producciones nativas para la práctica de la comprensión oral, reconocimiento de voz para interactuar con el alumno, y corrección visual sobre la pronunciación. Precisamente este aspecto se corrige mediante indicaciones sobre el sonido erróneo o la posición de la vocal en una ilustración de los órganos articulatorios, donde se muestra su grado de altura o anterioridad/posterioridad. Igualmente, se suele ofrecer la forma de la onda producida por el hablante para indicar rasgos como la duración.

Así sucede en productos comerciales para la enseñanza del inglés, como Saybot (dirigido a estudiantes chinos; *vid.* referencia n.º 30), RossettaStone (disponible para más de treinta lenguas, entre ellas el español peninsular e hispanoamericano; *vid.* referencia n.º 29), EyeSpeak (Ferguson, 2005), o Tell me more y Talk to me (también disponible para el español y otras seis lenguas). Estos dos últimos programas permiten que el estudiante mantenga diálogos interactivos con el programa en situaciones comunicativas que se aproximan al uso real del lenguaje. También simulan intercambios comunicativos auténticos programas como Microworld, entorno del sistema MILT

⁷ Aunque dichas recomendaciones se recogieron para el diseño de un sistema destinado a personas con déficits auditivos, nos han parecido también adecuadas para una aplicación de enseñanza de lenguas.

⁸ Por ejemplo, se han presentado las siguientes tesis doctorales sobre el tema: *Use of Speech Recognition in Computer-Assisted Language Learning*, de S. M. Witt (Cambridge University, Reino Unido, 1999); *Computer Support for Learners of English*, de R. Hincks (KTH School of Computer Science and Communication, Estocolmo, 2005); y *The pedagogical effectiveness of ASR-based computer assisted pronunciation training*, de A. Neri (University Nijmegen, 2007).

(Military Language Trainer), o el entorno didáctico TraciTalk (*vid.* referencias de ambos en el trabajo de Gamper y Knapp, 2002). Asimismo, el curso para el aprender español Learn to Speak Spanish incorpora situaciones de diálogo bastante realistas (*vid.* reseña de Gill, 1999). Algunos sistemas, además de usar reconocimiento de habla para la práctica de destrezas orales y la evaluación de la pronunciación, incorporan la comprensión auditiva y lectora; por ejemplo, VILTS (Voice Interactive Training System), desarrollado por la empresa Nuance para el aprendizaje del inglés en varios niveles (desde inicial a avanzado), y la versión francesa ECHOS (Rypa y Price, 1999). Otros dignos de mención son NativeAccent, comercializado por Carnegie Speech (ref. n.º 3) a partir del proyecto de investigación FLUENCY en la universidad Carnegie Mellon (que desarrolló el reconocedor CMU SPHINX); o CandleTalk, prototipo del ámbito investigador desarrollado en Taiwan para la práctica de diálogos (Liou, Chiu y Yeh, 2006). Sin duda, como indican Neri, Cucchiari, Strik y Boves (2003), estas aplicaciones no sólo incluyen las ventajas de los juegos para el aprendizaje, sino que también permiten la adquisición de la lengua mediante el aprendizaje por tareas.

Una aproximación diferente, como plantea Brown (2004), es el uso de tutores inteligentes guiados mediante la voz. Por ejemplo, para el aprendizaje de las preposiciones del inglés, el estudiante consulta un mapa y debe ir guiando a un personaje mediante las expresiones de espacio adecuadas. Otro tutor inteligente que incorpora el reconocimiento y la grabación de voz es Intelligent Tutor, de la empresa DinEd (antes Dynamic English), también para el aprendizaje del inglés (ref. n.º 25).

El reconocimiento del habla se está experimentando para automatizar la evaluación de la producción oral en tests de nivelación (*vid.* Wet *et al.*, 2009, para la lengua inglesa). En algún caso, se ha llegado a incluso implementar en exámenes automáticos a hablantes extranjeros; por ejemplo, para el inglés existe el sistema SpeechRaterSM (Zechner, Higgins, Xia y Williamson, 2009) o el test por teléfono Versant —antes llamado PhonePass o SET10—, que sólo necesita 10-12 minutos (*vid.* más detalles en Bernstein y Chen, 2008). Esta prueba es gestionada por la empresa Ordinate (*vid.* ref. n.º 16). El sistema integra un reconocedor desarrollado por el equipo de Bernstein, que procedía inicialmente del campo de investigación de las aplicaciones para las patologías del habla. Los resultados de evaluación de Versant parecen tener una correlación muy cercana a la corrección realizada por evaluadores profesionales (*vid.* Bernstein y Chen, 2008). Este equipo de investigadores también implementó un sistema semejante para la evaluación del árabe y del español (*vid.* Bernstein *et al.*, 2004).

1.3.1. Reconocimiento automático del habla y enseñanza de pronunciación. La utilidad de los programas para la enseñanza de la pronunciación asistida por ordenador (*Computer Assisted Pronunciation Teaching*, o *CAPT*) se fundamenta en la hipótesis de que la simple exposición a la lengua extranjera no asegura el desarrollo de la pronunciación ni la producción oral correcta, como indican varios investigadores o desarrolladores de aplicaciones (*vid.* Hwu, 1997; Neri, Cucchiari, Strik, 2002). Además, también se justifica por la necesidad de completar el aprendizaje de una lengua con el dominio de la pronunciación, pero fuera de la instrucción presencial, donde no suele haber tiempo material para practicarla (Strik, Neri, Cucchiari, 2008).

Llisterri (2001, 2007) ofrece una exhaustiva bibliografía reciente sobre el uso del reconocimiento automático del habla en este ámbito⁹. Respecto a la tecnología integrada, como proponen Gamper y Knapp (2002), se puede distinguir entre los *sistemas de reconocimiento de habla discreta* (que analizan patrones simples y se emplean generalmente para la enseñanza de la

⁹ Una bibliografía actualizada se puede consultar en la página personal de Joaquim Llisterri: http://liceu.uab.es/~joaquim/applied_linguistics/L2_phonetics/CALL_Pron_Bib.html

pronunciación o la mejora de la fluidez) y los *sistemas de reconocimiento de habla continua* (para el habla más espontánea). Estos últimos aún presentan deficiencias importantes. Los resultados más rápidos y fiables parecen obtenerse con frases prefabricadas, que dejan poca libertad creativa al usuario, en dominios controlados y con vocabulario reducido.

Cuando se consideran los errores de pronunciación, se diferencia entre los que afectan a los *segmentos* (fonemas mal pronunciados) y a los *suprasegmentos* (entonación, ritmo, acento de intensidad o fluidez de habla). Ambos niveles (el segmental y el suprasegmental) han sido considerados para su corrección automática (por ejemplo, una propuesta de corrección automática de la entonación se presenta en Arias *et al.*, 2010). Asimismo, los dos parecen tener igual rango de importancia en la comprensión del habla (Neri, Cucchiari y Strik, 2002). Sin embargo, con la tecnología actual parece existir la necesidad de abordar desde perspectivas diferentes los errores fonético-fonológicos. Así, el reconocimiento automático del habla puede corregir la pronunciación en el nivel segmental, pero puede plantear un reto de procesamiento analizar a la vez la variación en el tono y la entonación, el ritmo y la duración, aunque llegue a medir con éxito la velocidad de habla (Hincks, 2003). Precisamente este último factor ha sido relacionado con el grado de competencia en el habla extranjera por varios estudios (Cucchiari, Strik y Bobes, 2000); esto es, parece que se tiende a percibir que un no nativo domina un idioma con fluidez si habla bastante rápido, aunque su gramática esté plagada de errores.

Por todo ello, respecto al grado de corrección de la pronunciación por parte del sistema, se pueden emplear dos enfoques, como explican Strik, Neri y Cucchiari (2008) y Eskenazi (2009):

- detección de un error individual (de la pronunciación de un único fonema); o,
- evaluación de la pronunciación (la impresión global de la fluidez del habla).

Para la detección del error individual se han empleado técnicas de enfoque en errores más frecuentes, métodos probabilísticos propios del reconocimiento del habla, o clasificadores fonético-acústicos (*vid.* una comparación de clasificadores aplicados para el neerlandés en Strik *et al.*, 2009). Por ejemplo, uno de los más frecuentes es el algoritmo *goodness of pronunciation (GOP)* propuesto por Witt (1999). Los aspectos técnicos del ajuste automático de niveles corrección de la pronunciación se pueden consultar en Neumeyer *et al.* (1999), Witt y Young (2000) o en Li *et al.* (2006).

Básicamente, un modelo de sistema para la corrección de la pronunciación asistida por ordenador tendría las siguientes fases, como explican Neri, Cucchiari y Strik (2003):

- reconocimiento de habla: la fase más importante, de la que dependen las siguientes;
- puntuación: evalúa la corrección de la pronunciación del hablante a partir de sus propiedades acústicas o temporales, y tomando como modelo un enunciado nativo;
- detección del error: es importante que el sistema no localice falsos positivos, esto es, errores que realmente no son tales; y que no deje de corregir errores auténticos;
- diagnóstico del error: el sistema identifica el tipo de error y sugiere cómo mejorar la pronunciación usando modelos de errores típicos almacenados previamente; y,
- presentación de la corrección: puede realizarse mediante una escala numérica, mediante una barra con diferentes grados, o indicando el sonido o la sílaba donde aparece la incorrección mediante un color diferente (como hace Tell to me). Gamper y Knapp (2002) añaden otras formas de corrección: expresar comentarios orales, rechazar respuestas no entendidas por el sistema o mostrar una indicación visual.

Este último es el diseño que incorpora el sistema para el aprendizaje del neerlandés desarrollado por el equipo de Neri, Cucchiari y Strik (2008; Cucchiari *et al.*, 2009). Otra herramienta que sigue un esquema semejante es PLASER (Mak *et al.*, 2003), para la práctica de la pronunciación del inglés por hablantes de chino cantonés originarios de Hong Kong. El sistema

ofrece ilustraciones de los órganos articulatorios y vídeos que muestran los movimientos de la cara, y corrige a nivel de fonema (según un código de tres colores) y de palabra (una puntuación global). Otros programas incorporan, junto a la corrección de la pronunciación, ejercicios para aprendizaje de léxico o gramática; es el caso del sistema CallJ para iniciarse en el japonés (Waple *et al.*, 2007; Wang *et al.*, 2009), cuya particularidad es la generación dinámica de preguntas para variar cada lección.

1.3.1.1. Bases de datos de hablantes no nativos. Con el objeto de mejorar la corrección de los errores de pronunciación, se ha abordado el estudio del habla de no nativos mediante grabaciones de producciones controladas, poco espontáneas, obtenidas mediante pruebas de repetición de palabras o frases, corrección de estructuras gramaticales, lectura de textos, etc. Tal es el procedimiento que siguieron Raux y Kawahara (2002) con hablantes japoneses de inglés.

Otros investigadores han recogido bancos de datos más extensos o con un diseño más elaborado. El proyecto ISLE (International Spoken Learner English)¹⁰ desarrolló un corpus de grabaciones de estudiantes de inglés que tuvieran como lengua materna el italiano y el alemán. Se llevó a cabo la descripción de la interlengua¹¹ fonética de estos hablantes, y a partir del análisis de errores se pudo establecer un conjunto de reglas fonéticas que modelaban la producción de los no nativos, teniendo en cuenta los procesos de interferencia de su lengua materna (L1 en adelante): la adición de vocal [ə] tras sílaba final por parte de los italianos, el ensordecimiento de las oclusivas sonoras finales por los alemanes, etc. (Bonaventura, Herron y Menzel, 2000). No obstante, parece que el uso de dichas reglas es limitado y no puede modelar completamente la variabilidad propia de cada hablante extranjero, lo que explica que los resultados de reconocimiento hayan sido pobres (Menzel, Herron, Bonaventura y Morton, 2000)¹². También para el inglés, el equipo de Dalby y Kewley-Port (2008) realizó un análisis de errores de las grabaciones de no nativos en el desarrollo de los sistemas HearSay y Pronto para la práctica y la percepción de la pronunciación (Dalby y Kewley-Port, 1999). Igualmente, para la mejora del modelo acústico de un reconocedor destinado a exámenes orales de inglés, el equipo de Zechner, Higgins, Lawless *et al.* (2009) recogió grabaciones de frases leídas por no nativos con distinta L1 (japonés, chino, coreano y español), atendiendo a las dificultades fonéticas según su idioma materno.

Por lo que respecta a otras lenguas, el grupo de investigación de Neri, Cucchiarini y Strik (2006; Doremalen *et al.*, 2009) también ha empleado bancos de datos de no nativos para el diseño o entrenamiento de un sistema de reconocimiento de habla dirigido al aprendizaje de neerlandés en un nivel inicial. Destaca, por lo práctico que resulta, el método de obtención de grabaciones que proponen Wik y Hjalmarsson (2009) para reunir producciones de no nativos en una L2. Mientras los estudiantes utilizan el sistema Ville para el aprendizaje del sueco, su pronunciación es grabada e incorporada al banco de datos oral. Por lo que respecta al español, se ha recogido una base de datos de producciones leídas por norteamericanos y hablantes de inglés de origen latinoamericano (Bratt *et al.*, 1998; Precoda y Bratt, 2008). Este equipo de investigación terminó constituyendo una *spin-off*, EduSpeak (*vid.* ref. n.º 8), que actualmente comercializa motores de reconocimiento de habla para nueve lenguas (entre ellas, el inglés o el español latinoamericano).

¹⁰ Más información en la página: <http://nats-www.informatik.uni-hamburg.de/~isle/>

¹¹ La interlengua, según la definición que estableció Selinker del concepto (1972), es el sistema lingüístico variable que construye la persona que aprende una lengua durante estadio intermedio del proceso de adquisición.

¹² En la aplicación diseñada en este proyecto se corrige la pronunciación (sobre el fonema o el acento de la palabra mal pronunciado) mediante representaciones visuales (*vid.* Menzel *et al.*, 2001).

Más información acerca de otras bases de datos no nativas se presenta en el artículo de Zinovjeva (2005). El principal obstáculo para la explotación de estos recursos es el acceso a los mismos, por sus dificultades de elaboración, como indica Blake (2008: 64).

1.3.2. Reconocimiento de habla para niños y el aprendizaje de la lectura. Existen menos aplicaciones destinadas a niños, debido a que es necesario contar con bancos de datos específicos de la edad (cuya disponibilidad es más limitada que los corpus de voz adulta) para entrenar acústicamente los sistemas. Dicha necesidad parte del hecho de que la voz infantil posee unas propiedades acústicas particulares: en rasgos generales, unos valores típicos más altos de la frecuencia fundamental y de los formantes, una mayor variabilidad espectral (dadas las diferencias anatómicas y morfológicas del tracto vocal) y una velocidad de habla más lenta (*vid.* Price *et al.*, 2009, donde se indican más investigaciones al respecto). Además, como apunta Eskenazi (2009), los reconocedores deben afrontar la variabilidad propia del habla infantil, llena de reformulaciones, dudas, incorrecciones fonéticas, etc. Entre los estudios que han evaluado la precisión de los algoritmos de reconocimiento automático de habla infantil, se pueden consultar los que citan Neri, Mich, Gerosa y Giuliani (2008: 395).

Uno de los campos más activos es el de los llamados tutores de lectura (*reading tutors*). Russell *et al.* (1996) desarrollaron el sistema STAR (Speech Training Aid Research) para la práctica interactiva de la pronunciación del inglés por niños de entre 5 y 7 años. Utiliza un reconocedor basado en modelos ocultos de Markov (*Hidden Markov Models* o HMM), para cuyo entrenamiento fue necesario recopilar un corpus de voz infantil a partir de la lectura de un vocabulario seleccionado. Al usuario se le presentaba la imagen de una palabra que tenía que producir y su pronunciación era evaluada automáticamente. Los resultados del experimento de uso del sistema en la clase fueron positivos, y los niños encontraron la aplicación estimulante (*vid.* Russell *et al.*, 2000). También fue probado en el aula el prototipo desarrollado por el equipo de Mostow (Mostow *et al.*, 1994; Mostow, 2004; Aist y Mostow, 2009) en el marco del proyecto LISTEN de la universidad Carnegie Mellon. El tutor de lectura desarrollado emplea un reconocedor de voz de habla continua para evaluar la lectura en voz alta de una historia leída por un niño. Por su parte, Zechner, Sabatini y Chen (2009) aplicaron un sistema de corrección automática de la pronunciación en la lectura de pasajes de textos y palabras aisladas. Con respecto a la lengua neerlandesa, se ha diseñado SPACE, que incorpora reconocimiento y síntesis de habla para el seguimiento del proceso de lectura en el niño y evaluación del nivel lector (Duchateau *et al.*, 2009). Dicho sistema también puede ser empleado por niños con retraso lector.

Para el aprendizaje del inglés por parte de niños italianos, el equipo de Neri, Mich, Gerosa y Giuliani (2008) desarrolló PARLING (PARla INglese), dirigido a la práctica de la pronunciación a nivel de palabra. Se trata de un sistema modular en que cada componente contiene:

- una historia (p. ej., Hansel y Gretel);
- un juego de palabras, que se va adaptando al usuario según su progreso o el itinerario de tareas que realiza;
- un conjunto de palabras activas;
- un diccionario visual (con algunas palabras con hipervínculos a la imagen y su pronunciación);
- una herramienta para que el usuario cree su propio diccionario (con grabaciones suyas e ilustraciones propias); y,
- un menú de ayuda.

PARLING incorpora un reconocedor de voz que fue entrenado con un banco de datos de habla infantil del rango de edad de los usuarios a los que se dirige (10-11 años), tanto nativos como no nativos. El prototipo fue probado con un grupo de alumnos italianos: unos recibieron instrucción

tradicional, y otros mediante PARLING. Los resultados apuntan que, tanto con un sistema como con otro, los niños mejoraron la pronunciación de palabras difíciles o desconocidas, pero aquellos que practicaron con PARLING necesitaron menos tiempo de práctica. A pesar de dichos resultados, los investigadores no dejan de señalar las limitaciones de su trabajo, que necesitaría una población experimental más amplia y se enfoca únicamente a la práctica de palabras aisladas (lo cual resulta poco realista para aprender a hablar espontáneamente en una L2, dominando con naturalidad los efectos de coarticulación y la fonotáctica de la lengua meta).

El proyecto TBALL (*vid. Alwan et al., 2007*), desarrollado en la universidad de Southern California y en la universidad de Los Ángeles, tiene como objetivo la práctica y evaluación de las destrezas lectoras de niños norteamericanos nativos y no nativos procedentes de México. Se compone de los siguientes módulos:

- interfaz multimedia, con un diseño multimodal (presenta imágenes, texto y sonido, integrando también una pantalla táctil) y recoge datos de interacción con el usuario;
- módulo de corrección y evaluación de la pronunciación, que incorpora reconocimiento de voz; y,
- interfaz del profesor, que realiza un seguimiento del estudiante y permite consultar sus datos personales o académicos (nivel del idioma, edad, procedencia, etc.).

Las actividades para la práctica de la lectura son de distinto tipo: juntar y leer dos sílabas para formar una palabra, responder preguntas de tipo *sí/no* a partir de un texto leído en voz alta, identificar el sonido representado por un carácter alfabético, o determinar el nombre correcto de una letra. Es de señalar que cada tarea de evaluación incorpora diferentes versiones según el nivel de dificultad, permitiendo al usuario controlar en cada una el progreso de su aprendizaje (aunque con cierto control automático de la temporización o tiempo para realizar la actividad). Igual que para otros sistemas antes mencionados, fue necesario recoger un banco de datos de habla infantil nativa y no nativa para el entrenamiento acústico del reconocedor; para dicha tarea se utilizó una interfaz de tipo *magó de Oz*¹³ y se procedió después al análisis de errores de las producciones obtenidas, lo cual permitió añadir reglas de errores para extender el lexicón. El sistema fue llevado al aula durante el curso académico 2007-2008 (*vid. detalles en Price et al., 2009*), y dicho estudio de caso ha planteado cuestiones como la falta de acuerdo de los profesores en lo que considerar como correcto, la falta de objetividad en la evaluación del acento extranjero, o la dificultad de diagnosticar cada error, en el que a menudo confluyen factores de distintos niveles (ortografía, fonética, etc.).

Entre los programas comerciales para niños, se pueden señalar Talk to me, Tell me more Kids (Auralog), English for kids (*vid. reseña de Krajka, 2001*), los productos comercializados por la empresa Soliloquy Learning (*vid. referencia n.º 14*), o los reconocedores de EduSpeak (*vid. ref. n.º 8*), que también incluyen modelos acústicos de voz infantil para el inglés americano. Incluso IBM desarrolló el sistema Watch Me! Read para la mejora de las destrezas de lectura y presentación oral (*vid. ref. n.º 39*).

Por último, existen prototipos multimodales que combinan el procesamiento de voz con otros sistemas de entrada como texto, tacto o reconocimiento de escritura en pantalla. Un ejemplo de ello es el proyecto I SEE (*Xiao et al., 2002; Oviatt et al., 2004*), en el marco del cual se ha estudiado la interacción entre niños de 7 a 10 años con un programa que les realizaba preguntas sobre animales marinos mediante síntesis de voz. También se están investigando sistemas de audición de historias (*Story Listening Systems, SLS*) dirigidos a niños que cuentan cuentos. Por ejemplo, Sam the

¹³ En la prueba *magó de Oz* el usuario interactúa con una herramienta informática (por ejemplo, un sistema de diálogo o un programa de corrección de la pronunciación) que simula ser automático pero en realidad es controlado por un investigador. Su objetivo es probar el sistema para detectar errores de diseño antes de implementar la versión definitiva.

CastleMate, desarrollado por el equipo de Cassell (2004), que emplea una interfaz conversacional con reconocimiento de voz para dialogar con el usuario, y un módulo de visión informática que realiza un seguimiento de sus gestos o su postura, de modo que pueda interactuar con él convenientemente.

1.3.3. Sistemas de reconocimiento de habla para el aprendizaje del español. En primer lugar, abordamos los sistemas surgidos en el ámbito de la investigación académica. Uno de los primeros fue The Audio Interactive Tutor (TAIT), de los laboratorios de investigación Mitsubishi (Waters, 1995), como recogen Gamper y Knapp (2002) en su recopilación de sistemas inteligentes empleados en la enseñanza de lenguas asistida por ordenador (*Intelligent CALL* o *ICALL*). Para la enseñanza de la pronunciación del español a norteamericanos se desarrolló el programa Pronto, surgido del trabajo de investigación para la mejora de la pronunciación de niños con problemas de audición o articulación mediante el sistema ISTRa de reconocimiento de voz (Dalby y Kewley-Port, 1999). Dispone de una interfaz lúdica que simula juegos para corregir la pronunciación (por ejemplo, los bolos: cuanto más se aproxime el enunciado pronunciado al modelo nativo, más bolos se derriban). Fue desarrollado con el enfoque de reforzar especialmente la percepción fonético-fonológica. Las dificultades de pronunciación fueron determinadas a partir de estudios empíricos de corpus no nativos y análisis de errores entre pares de lenguas. Ignoramos los resultados de Pronto sobre el aprendizaje del español, aunque cabe señalar que un programa similar para el inglés (HearSay) resultó positivo para aprender a pronunciar las consonantes y vocales tanto de las palabras practicadas con el programa como de las desconocidas que incluían los contrastes problemáticos (*vid.* Dalby y Kewley-Port, 2008).

Para el español de México, se ha empleado el módulo de reconocimiento del CSLU Toolkit, desarrollado por el Center for Spoken Language Understanding, Oregon Graduate Institute (Kirschning, Aguas y Ahuactzin, 2000). El sistema evalúa la pronunciación del hablante comparando la señal que graba con el modelo nativo.

En el ámbito comercial existen los programas ya mencionados Talk to me y Tell me more (ambos de Auralog; *vid.* ref. n.º 36), RosettaStone (ref. n.º 29), el test oral por teléfono Versant de la empresa Ordinate (Bernstein y Chen, 2008), así como los reconocedores de habla comercializados por empresas como EduSpeak (ref. n.º 8). A estos podemos añadir Learn to Speak Spanish (*vid.* reseña de Gill, 1999), programa para el aprendizaje del español con reconocimiento de voz integrado en la práctica de ejercicios de vocabulario o diálogos para interactuar con un personaje virtual, de gran utilidad para la práctica de la conversación, aunque únicamente para el habla mexicana.

1.3.4. Evaluación de la efectividad de los programas que incorporan sistemas de reconocimiento de habla para el aprendizaje de lenguas. A menudo los sistemas de ELAO solamente son evaluados por los propios desarrolladores, lo cual proyecta una imagen incompleta de su eficacia. Además, la metodología empleada a veces no proporciona resultados que arrojen luz acerca del impacto que tienen en el proceso de aprendizaje en contextos reales (Felix, 2005), un aspecto sobre el cual tampoco abundan las evaluaciones (*vid.* al respecto la revisión de sistemas realizada por Stockwell, 2007).

Eskenazi y Brown (2006) exponen algunos de los aspectos que tener en cuenta en la evaluación de un programa de este tipo: lo estimulante y la facilidad de uso, los objetivos pedagógicos, las tecnologías empleadas, la forma de realizar la corrección, la manera como los estudiantes son evaluados o guiados, etc. También merece la pena considerar los criterios que sugiere Chapelle (2001) para la evaluación de un programa de ELAO: el potencial para el aprendizaje de la lengua, cómo se ajusta al perfil del estudiante, el foco de significado (esto es, el conocimiento lingüístico o el proceso en el que el aprendiz fija su atención mientras realiza la

tarea), el impacto en el proceso de aprendizaje, la autenticidad de la actividad, y la funcionalidad de la herramienta (o sea, lo práctico que resulta su uso)¹⁴.

La evaluación de la aplicación de sistemas de dictado automático (por ejemplo, Dragon Naturally Speaking, de Nuance; *vid.* referencia de Internet n.º 22) en la corrección de la pronunciación ha tenido resultados negativos, debido a que son sistemas que no se diseñaron para el reconocimiento del habla no nativa ni para la corrección de errores (*vid.* Strik, Neri y Cuchiarini, 2008). En cuanto a programas diseñados específicamente para la enseñanza de lenguas, se han realizado evaluaciones de programas como Talk to me (Auralog) para el aprendizaje del español (Lafford, 2004) y del inglés (Hincks: 2003, 2005b). En este último caso, se llevó a cabo un experimento controlado en que se comparó un grupo que fue instruido con el programa y otro que no. Tanto Lafford como Hincks apuntan que la herramienta se aleja un tanto de los principios de la enseñanza comunicativa y se centra en ejercicios de repetición propios de enfoques audio-linguales. También se ha evaluado con un experimento el sistema de corrección automática de la pronunciación para el aprendizaje del neerlandés desarrollado por el equipo de Neri, Cucchiarini y Strik (2008). Los resultados obtenidos indican la mejora significativa de la pronunciación, a pesar de la distancia tipológica entre la lengua materna de los estudiantes y la lengua meta. Las conclusiones de este equipo coinciden con los de Hincks (2003) en el hecho de que los hablantes con un acento más marcado obtuvieron una mejoría mayor, lo cual parece recomendar el uso de programas de ayuda para la pronunciación en los niveles más bajos. Aliaga-García (2007) también realizó un experimento con EyeSpeak para la enseñanza de la pronunciación de las oclusivas sordas del inglés a españoles y catalanes. Pese a los resultados positivos obtenidos, expresa su cautela sobre la influencia de estas aplicaciones, por lo que parecen necesarias más investigaciones al respecto.

Resultados no tan positivos fueron obtenidos por el equipo de Mayfield Tomokiyo (2000), que evaluó la eficacia del sistema FLUENCY (desarrollado en la Universidad Carnegie Mellon para el aprendizaje del inglés). El sistema detecta errores de pronunciación, señala al usuario dónde han ocurrido y le explica cómo pronunciar correctamente los sonidos. El procesamiento toma, entre otros, información sobre la duración vocálica (Eskenazi, 1999), aspecto importante en la fonética del inglés. Los investigadores llevaron a cabo un experimento con un grupo de hablantes de diferente L1 que recibió instrucción para la mejora de la pronunciación con esta herramienta, frente a otro grupo que recibió la misma enseñanza en clase. La prueba no mostró una mejora destacable por parte de los estudiantes de niveles iniciales, ni tampoco una mejora comparativamente significativa entre el grupo que recibió corrección por el sistema de reconocimiento y el grupo que fue corregido en clase. Además, en el grupo experimental pareció existir una gran variación en el nivel de mejoría de la pronunciación, a diferencia de quienes recibieron instrucción presencial, que perfeccionaron sus producciones de manera más homogénea. Barr *et al.* (2005) también aportan resultados negativos en la mejora de las destrezas orales en el aprendizaje del francés (nivel inicial) haciendo uso de un programa con reconocimiento de voz (Tell me more), aunque en su estudio aquel se integraba en un entorno multimedia más amplio. Dichos investigadores apuntan la necesidad de actividades que consigan una comunicación orientada al mensaje, propia de la comunicación cara a cara.

Respecto a la importancia del uso de información visual para el aprendizaje de la pronunciación, sería positivo plantear una evaluación como la que propuso Flege (1988: 402), esto es, comparar los resultados en la mejora de la pronunciación entre cuatro grupos de sujetos: uno que

¹⁴ Existen multitud de reseñas de programas en la revista *CALICO* (www.calico.org) o en asociaciones como *Information and Communications Technology (ICT) for Language Teachers* (www.ict4lt.org), en cuya página se presenta un modelo de parrilla de evaluación de software educativo para la enseñanza de lenguas (<http://www.ict4lt.org/en/evalform.doc>).

recibiera únicamente corrección visual, otro que recibiera corrección visual y auditiva, otro grupo que tuviera que imitar las producciones nativas, y otro que, además de tener que intentar aproximarse a ese modelo, fuera evaluado por un profesional.

1.3.5. Problemas del reconocimiento de habla para la enseñanza de la lengua. Para lograr un reconocimiento de calidad y una buena integración en un sistema completo de ELAO, aún hay que resolver deficiencias tecnológicas. Hincks (2003) y Eskenazi y Brown (2006) han indicado, entre otras, las siguientes.

- Parecen aún necesarios avances en el área del procesamiento del lenguaje natural y el reconocimiento del habla para emplearlas más allá de dominios restringidos. La aplicación del reconocimiento de habla libre, poco controlada, genera una gran tasa de error, por lo que se hace imprescindible predecir las producciones del usuario. Así, el estudiante puede llegar a asumir un papel pasivo, pues sólo repite un conjunto cerrado de frases o un vocabulario reducido. Para evitarlo, pueden adoptarse técnicas de obtención de enunciados que anticipan los que producirá el hablante, pero concediéndole cierto margen de libertad. Esta estrategia se realiza en el sistema FLUENCY (Eskenazi, 1999) o en el proyecto LISTEN (Aist y Mostow, 2009), ambos de la universidad Carnegie Mellon. Con todo, los fallos de reconocimiento de habla espontánea impiden implementar aún actividades muy atractivas (por ejemplo, una discusión abierta), o incluso dificultan el tratamiento de niveles lingüísticos como la sintaxis o la morfología (Cucchiari *et al.*, 2009).
- Probablemente se obtengan mejores resultados cuando se perfeccionen los sistemas de reconocimiento independientes de locutor. Actualmente, el reconocimiento funciona mejor cuanto más se haya entrenado el sistema con un mismo hablante, así que es difícil obtener buenos resultados en las primeras locuciones de un nuevo usuario. Por ejemplo, los sonidos correspondientes a fonemas ligeramente distintos (como /æ/ y /ε/ en inglés) se pueden solapar en el espacio acústico si los modelos de los sonidos se entrenan independientemente de un hablante específico (algo necesario para que cualquier hablante use el sistema), como explican Neri, Cucchiari, Strik y Boves (2003). Para evitarlo, se pueden incluir palabras que son fonéticamente muy diferentes, o información sobre la duración.
- Los futuros sistemas han de tener en cuenta de mejor manera el sexo del locutor o su edad, debido a la diferencia de timbre entre hombres adultos y mujeres o niños.
- Aún se ha de mejorar el reconocimiento de sonidos en determinados contextos; por ejemplo, en el nivel fonético a veces existen problemas de reconocimiento con los primeros sonidos de un enunciado tras un silencio, y con los sonidos de sílabas átonas (parece ser que las sílabas tónicas se reconocen más fácilmente).

Otros investigadores añaden que el reconocimiento de habla no nativa es menos precisa que respecto al habla nativa, debido en parte al alto número de disfluencias (titubeos, pausas, reparaciones, reinicios, etc.) en las producciones orales realizadas por extranjeros (van Doremalen *et al.*, 2009). Dichos fenómenos de la oralidad precisan ser incluidos en el modelo de lenguaje. Asimismo, las variantes de pronunciación debidas a la interferencia de la L1 necesitan una atención especial. Así, es necesario adaptar el modelo acústico del reconocedor (Mayfield Tomokiyo y Waibel, 2001), por ejemplo incorporando sonidos mal pronunciados, aunque se trata de un proceso aún no resuelto definitivamente (van Doremalen *et al.*, 2009, Zechner, Higgins, Lawless *et al.*, 2009) y en el que se están experimentando nuevos métodos (*vid.* por ejemplo Ohkawa *et al.*, 2009). Por otra parte, como indica Blake (2008: 64), la incorporación del reconocimiento automático de

habla a la enseñanza de lenguas tendrá un impacto significativo cuando se disponga de una herramienta de autor que haga más simple su uso por los no ingenieros.

1.4. Conversión texto-habla. La síntesis de habla es una tecnología que no ha sido muy considerada para las aplicaciones de enseñanza de lenguas, en cierto modo porque aún no presenta un grado suficiente de madurez tecnológica (cuando se compara, por ejemplo, con el reconocimiento de habla), de manera que la voz sintetizada resulta poco natural (Delmonte, 2008). Una de las críticas principales de su uso es precisamente la falta de autenticidad del habla generada. Como contrapartida, Delmonte (2008) expone la ventaja que supone poder sintetizar cualquier texto inmediatamente generado u obtenido para el estudiante, sin necesidad de que un hablante o un locutor profesional grabe la lectura del mismo, o sin que un tutor humano esté presente mientras realiza una actividad de aprendizaje. Con todo, podemos diferenciar tres estrategias de empleo de la síntesis de voz para la enseñanza de lenguas.

- El uso de la síntesis para percibir los resultados de la manipulación de la voz (duración, timbre, melodía prosódica...). Dicho enfoque es el más sencillo, pues puede prescindir de modelos acústicos para la síntesis, manipulando simplemente la grabación de un hablante y resintetizándola. Algunos programas de visualización del habla incluyen también módulos para transformar la señal y sintetizar el resultado: por ejemplo, Praat, WinSnoori, Speech Filing System (SFS) o WinPitch, como ya se ha explicado en el apartado 1.2.1. En esta línea de trabajo, se ofrecen ejemplos de aplicaciones didácticas para la enseñanza del acento de intensidad en Hincks (2002); de la prosodia, en Lahoz Bengoechea (2008); y también de la entonación y el ritmo, en Sundstrom (1998) y Martin (2005). Probst *et al.* (2002) han planteado que la resíntesis de la propia voz del hablante proporciona mejores resultados de aprendizaje, y este procedimiento es el que aplican Felps *et al.* (2009, manipulando rasgos prosódicos y segmentales) o Bissiria y Pfitzinger (2009, para la adquisición del acento léxico del alemán por parte de italianos). Análogamente, el sistema de síntesis STRAIGHT (Kawahara y Akahane-Yamada, 2006) y otras herramientas como SNACK Sound Toolkit se han empleado en experimentos de fonética perceptiva, con ejercicios de identificación de fonemas cuya duración, intensidad o timbre se ha manipulado, o experimentos de discriminación de patrones melódicos modificados (Sjoelander *et al.*, 1999). Otras pruebas de discriminación perceptiva que hacen uso de la síntesis de habla se explican en Berkovitz (1999) o en Vogel *et al.* (2009).
- El empleo de síntesis de voz en los sistemas de diálogo que, junto al módulo de reconocimiento y de gestión de la conversación, permiten la interacción comunicativa con el hablante para practicar la lengua. Por ejemplo, las herramientas del Center for Spoken Language Understanding (CSLU) Toolkit, desarrolladas en la Universidad de Oregón, integran el conversor texto-habla Festival (creado en la Universidad de Edimburgo), que puede sintetizar habla en inglés, pero también en galés y en español de variedad mexicana (*vid.* apartado 1.5.3).
- Herramientas que integran la conversión texto-habla en un sistema multimedia para el aprendizaje de una lengua, ya sea de los aspectos de la pronunciación o la ortografía (para dictados), la práctica de la comprensión lectora o la lectura en voz alta de cada entrada de un diccionario bilingüe (*vid.* al respecto Lyras *et al.*, 2009). A continuación trataremos este tipo de aplicaciones, que de manera preponderante se han desarrollado para el inglés, aunque ha sido creado algún prototipo para una lengua minoritaria como el bretón (Mercier, Guyomard y Siroux, 1999).

Delmonte (2002, 2008 y 2009) explica el uso de un conversor texto-habla en el sistema GETARUN (ahora integrado en el sistema SLIM, System for Interactive Language Learning) para el aprendizaje del inglés. Se trata de una aplicación diseñada para actividades de comprensión auditiva con tecnología de procesamiento del lenguaje natural (por ejemplo, en el análisis de las respuestas proporcionadas por el usuario). El módulo de síntesis de habla realiza la lectura en voz alta de un texto, sobre el cual el usuario responde a unas preguntas relacionadas con el contenido. Además, la corrección de las preguntas o las instrucciones también se pueden oír con voz sintetizada. La última versión (Delmonte, 2009) incorpora reconocimiento de habla y también permite la práctica de la pronunciación (en el nivel de la sílaba o la palabra) y de aspectos prosódicos (el ritmo, la entonación o el acento). SLIM es flexible respecto a la forma de trabajo, ya que permite el trabajo preestructurado y programado para alcanzar objetivos de aprendizaje concretos, o el autodirigido por el alumno. Delmonte (2008, 2009) también ha aplicado la síntesis del habla para hacer dictados, ejercicios de rellenar huecos o verdadero/falso, así como para la simulación de la interlengua del estudiante mediante reglas fonéticas y prosódicas (de estudiantes angloamericanos de italiano).

Además del aprendizaje de lenguas, también aborda la terapia del habla el Sistema Interattivo Multimediale per l'Alfabetizzazione, desarrollado por el equipo de Umberta Bortolini de la Universidad de Padua (2002). La herramienta integra síntesis de habla y se dirige al desarrollo de las destrezas prelectoras de niños normales y discapacitados mediante entrenamiento auditivo y visual. Permite la práctica del análisis segmental de las palabras y el aprendizaje de la correspondencia entre grafemas y fonemas. El niño puede combinar libremente cualquier grafía y escuchar los resultados mediante síntesis de habla. También puede practicar la comprensión auditiva de una sílaba sintetizada que tiene que identificar entre varias opciones presentadas.

Más información sobre los avances más recientes en aplicaciones de síntesis de habla para la enseñanza de lenguas se puede encontrar en Black (2007), y algunas métricas de evaluación se presentan en Handley y Hamel (2005). Asimismo, Handley (2009) ofrece una evaluación de cuatro sistemas de síntesis para la enseñanza del francés y discute otras vertientes educativas de dicha tecnología. Por lo que se refiere a la valoración de los propios estudiantes, se puede consultar Kang *et al.* (2009).

1.5. Sistemas de diálogo. Tampoco existen muchas aplicaciones que optan por un sistema de diálogo para el aprendizaje de lenguas, a pesar de su potencial para simular intercambios comunicativos reales (por ejemplo, una llamada para realizar la reserva de un vuelo). Uno de los principales obstáculos es la complejidad de su tecnología: no solamente precisa un reconocedor (cuya calidad de reconocimiento aún no es perfecta), sino que además requiere más componentes y módulos específicos para cada estado del proceso. A grandes rasgos, estas diferentes fases son las siguientes:

- comprensión del lenguaje natural;
- gestión del diálogo; y,
- generación de lenguaje.

En todo caso, el objetivo de innovación que los sistemas de diálogo pretenden en la enseñanza es la superación de ejercicios tradicionales de repetición de sonidos, lectura o rellenar huecos mediante la simulación de situaciones comunicativas lo más auténticas posible.

Podemos diferenciar tres aplicaciones didácticas de los sistemas de diálogo: los que permiten la práctica de situaciones comunicativas concretas (explicados en §1.5.1), los integrados en el entorno de un videojuego (§1.5.2) y los agentes animados (§1.5.3).

1.5.1. Sistemas de diálogo para la práctica de situaciones comunicativas. Hemos reunido en este apartado sistemas de diálogo que en mayor o menor medida emulan la interacción en situaciones de comunicación determinadas. Por ejemplo, Raux y Eskenazi (2004) explican el sistema Let's go, que simula un diálogo en que el usuario solicita información sobre los horarios del autobús. Como estos investigadores comentan, un sistema de diálogo aplicado al aprendizaje de lenguas tiene dos objetivos: que el usuario complete la tarea requerida y que mejore su destreza lingüística y conversacional. Así, el diseño de las estrategias de corrección ha de cuidarse para que, cuando le pregunte el sistema, el usuario no realice una simple repetición o confirmación de información, sino que practique frases completas con unas dificultades determinadas. La corrección al usuario, además, ha de ser concisa y comprensible.

A medio camino entre una aplicación con reconocedor de voz y gestor de la conversación, y un sistema de diálogo, la herramienta Subarashii (Ehsani *et al.*, 2000) permite el aprendizaje del japonés en niveles iniciales, mediante la interacción con el programa en encuentros comunicativos como una presentación o concertar un plan de ocio. No solamente se corrige el nivel de la pronunciación, sino también errores gramaticales o de léxico.

El grupo de sistemas de diálogo del Massachusetts Institute of Technology ha desarrollado una herramienta para el aprendizaje del chino mandarín por hablantes de inglés (explicado en detalle en Seneff, Wang, Peabody y Zue, 2004). El diseño pedagógico de los contenidos parte de un dominio temático (los viajes, el tiempo o la información sobre vuelos) que se expande en escenarios y situaciones concretas (el hotel, el aeropuerto, etc.). Previamente a la práctica oral, se plantean ejercicios o juegos vía web que emplean traducción automática. El objetivo de estas prácticas es doble:

- preparar el vocabulario y la gramática de la lección; y,
- recoger producciones escritas (respuestas a preguntas de los contenidos) para el módulo de comprensión de lenguaje natural (*Natural Language Understanding* o *NUL*) y obtener grabaciones de habla para entrenar el reconocedor.

El sistema resulta muy flexible, pues el hablante puede interactuar mediante la voz o por escrito, y además permite el uso de la lengua nativa del usuario por medio del módulo de traducción automática (se traducen al chino las preguntas en inglés cuando el usuario no encuentra la forma de preguntar en la lengua meta). Este grupo de investigadores ha seguido la misma metodología para desarrollar otras aplicaciones que incorporan síntesis de habla (para oír la traducción que el usuario debe repetir), o que se han implementado para su uso por teléfono.

1.5.2. Sistemas de diálogo integrados en el entorno de un videojuego. Uno de los enfoques de investigación y desarrollo en el área de sistemas de diálogo educativos será el diseño de aplicaciones e interfaces cercanas a las de un videojuego, lo que añade un componente de competitividad y amplía la participación a más usuarios, como apunta Seneff (2007). En el artículo citado se presentan ejemplos de videojuegos que incorporan sistemas de diálogo: en uno de ellos el hablante tiene que colocar en el lugar donde se le indica formas geométricas (cuadrados, triángulos) con determinados colores, seleccionándolos mediante la voz; en otro, en el dominio temático del tiempo libre, el usuario tiene que conversar con el sistema sobre sus aficiones para cuadrar su horario y conseguir quedar con el personaje (*vid.* Seneff, Wang y Chao, 2007). El usuario es asistido en la actividad del diálogo con un tutor robótico que le ayuda a planificar la interacción. Otro juego más reciente, Word War (McGraw y Seneff, 2008; McGraw, Yoshimoto y Seneff, 2009), incorpora la interacción mediante la voz para el aprendizaje del vocabulario del chino mandarín. El usuario utiliza comandos de voz en esta lengua para mover al hueco indicado por el sistema imágenes de animales, plantas o comida (aunque puede crear otras nuevas según sus necesidades de

aprendizaje). Un rasgo importante de Word War es que se realiza un seguimiento continuo del usuario y sus pronunciaciones son grabadas para su posterior análisis.

Para la enseñanza del vocabulario y otros niveles lingüísticos junto a los aspectos culturales, se ha integrado un sistema de diálogo para interactuar con personajes virtuales en un videojuego: el DARWARS Tactical Language Training System (TLTS) desarrollado por el equipo de Johnson (2004). Dirigido al aprendizaje de árabe por soldados americanos destinados a Irak, recientemente ha sido comercializado por la empresa Alelo (Johnson y Valente, 2008). En la década anterior ya existía un sistema semejante en objetivos –también estaba destinado a militares– y diseño –era multimedia, aunque no en un entorno de videojuego– (Harless, Zier y Duncan, 1999). Este programa, llamado Virtual Conversations, también empleaba reconocimiento de voz para practicar el árabe en simulaciones de situaciones reales. Otro juego en desarrollo, DEAL, emplea un avatar con el que se interactúa mediante un sistema de diálogo, para el aprendizaje de la gramática o el vocabulario en el dominio del comercio (*vid.* Wik *et al.*, 2007; Wik y Hjalmarsson, 2009).

Por último, tampoco habrá que dejar de lado las posibilidades de integración de los sistemas de diálogo con la realidad virtual. La combinación de estos avances ya se ha llevado a cabo en el programa experimental Zengo Sayu para el aprendizaje del japonés (Rose y Billingham, 1995), o en el prototipo para la práctica de la comprensión auditiva en el aprendizaje del inglés que explican García-Ruiz *et al.* (2008). La adquisición de una lengua mediante la interacción y la actividad física del hablante –aunque limitada a un entorno virtual– encaja en la línea de enfoques de enseñanza como la respuesta física total (*Total Physical Response*) de J. Asher o el enfoque natural (*Natural Approach*) de T. Terrell y S. Krashen (para más detalles sobre ambos, *vid.* Richards y Rodgers, 2003).

1.5.3. Sistemas de diálogo y agentes animados. Otros desarrollos integran el uso de caras parlantes (*talking faces*) en los que se visualizan los movimientos articulatorios del habla. Por ejemplo, las herramientas integradas en el Center for Spoken Language Understanding (CSLU) Toolkit permiten la creación de agentes animados que interactúan con el usuario mediante reconocimiento y síntesis de voz. Una de las principales ventajas es el uso de información multimodal: auditiva y visual, sirviendo esta última de mayor apoyo a los hablantes de segundas lenguas o con déficits auditivos (Granström, 2004). La generación de voz está sincronizada con las imágenes tridimensionales que simulan los movimientos articulatorios de la boca o los órganos fonadores (vistos desde interior, desde una vista de perfil medio sagital, o incluso desde atrás), así como otros gestos de la cara. Estas imágenes se sintetizan, por ejemplo, utilizando datos de los órganos articulatorios obtenidos mediante electropalatografía o ultrasonidos en el proceso de fonación (Massaro, 2006).

Dichas tecnologías se han usado –al parecer, con buenos resultados– para que los estudiantes de segundas lenguas aprendan el vocabulario o la pronunciación. Un ejemplo de ello para el inglés es el agente animado Baldi (Massaro, 2006), que también se ha adaptado para el español, el francés, el italiano, el árabe o el mandarín. Cole *et al.* (1999) explican las ventajas del uso de un agente parlante: aportar una dimensión más humana a la interacción hombre-máquina, poder transmitir contenido emocional, o reunir mayor capacidad de transmitir información. Baldi también se ha destinado a usuarios con problemas de sordera, con trastornos del espectro autista o incluso dislexia. Sin embargo, algunos investigadores han destacado ciertas limitaciones: no proporciona instrucciones específicas sobre la propia pronunciación del hablante ni le corrige sus errores (Engwall *et al.*, 2006); tampoco se ha evaluado la usabilidad del sistema (*vid.* Eriksson *et al.*, 2004), y el uso de la síntesis de voz puede resultar artificial (Cole *et al.*, 1999) –aunque el profesor puede grabar cualquier enunciado y emplearlo en lugar del habla sintética–. Estos sistemas de agentes animados también se han destinado a la adquisición de las destrezas de lectura por parte de niños; es

el caso del Colorado Literacy Tutor (Hagen *et al.*, 2003), que incorpora un módulo de evaluación de resúmenes.

2. RECOMENDACIONES DE DISEÑO DE APLICACIONES DE VISUALIZACIÓN DE VOZ Y DE TECNOLOGÍAS DEL HABLA PARA LA ENSEÑANZA DE LENGUAS

Tanto en el ámbito comercial como investigador, sería beneficiosa la participación del usuario final (docente o aprendiz) en la fase de diseño de una aplicación didáctica, como expone el modelo de diseño de programas de ELAO sugerido por Colpaert (2004; *apud* Ward, 2006: 134). Junto a ello, consideramos positivas las recomendaciones recogidas en la bibliografía consultada, que exponemos a continuación, acerca de los procedimientos de corrección (§2.1), los contenidos pedagógicos (§2.2) o el diseño de la interfaz (§2.3).

2.1. Los procedimientos de corrección. Varios investigadores ya han indicado que el factor más importante de cualquier tecnología de habla o de visualización de la voz para la enseñanza de una lengua es que el alumno reciba una evaluación acerca de su enunciado o de cómo pronuncia, esto es, una puntuación dependiendo de si se aleja más o menos del modelo nativo (Hincks, 2003: 5; Neri, Cucchiarini, Strik, Boves, 2003: 6; Martin, 2005; Gómez Vilda *et al.*, 2008). No solamente es necesario marcar los aspectos negativos de su producción oral, sino también sus aciertos. Además de ello, lo óptimo sería recibir evaluación acerca del lugar donde se ha cometido el error de pronunciación (Hincks, 2003), para evitar la persistencia del mismo en el habla (o *fosilización*, en términos de Selinker, 1972). Por ejemplo, Tell me more y Talk to me muestran en otro color las palabras en que se han cometido errores, mediante un sistema de reconocimiento de habla que procesa la señal producida por el no nativo y la compara con el modelo de pronunciación nativa (Lafford: 2004). En cuanto a los aspectos prosódicos, el sistema BetterAccentTutor (Kommissachirk y Kommissachirk, 2000) presenta una corrección visual de la entonación, el acento y el ritmo.

Respecto a *qué* corregir, se pueden distinguir dos niveles en la pronunciación extranjera: el acento extranjero y la inteligibilidad (*vid.* Neri, Cucchiarini y Strik, 2002). Parece más razonable intentar alcanzar una pronunciación correcta (en cuanto a su grado de comprensión) que una completamente libre de acento extranjero, por lo que resulta más sensato concentrarse en la práctica de la pronunciación de los sonidos que más dificultan o impiden la comunicación. De esta forma, se torna imprescindible establecer jerarquías de errores para el aprendizaje de cualquier par de lenguas. Baremos que consideren tanto el nivel fonético-fonológico como el suprasegmental, ya que, por ejemplo, las variaciones de intensidad o de duración pueden ser más importantes y distintivas en una lengua que en otra (Eskenazi, 1999). En el sistema desarrollado por el equipo de Neri, Cucchiarini y Strik (2002, 2008; Cucchiarini *et al.*, 2009), los errores que corregir debían ser perceptivamente significativos, frecuentes, comunes entre hablantes de lenguas maternas diferentes, persistentes en el tiempo, que podrían dificultar la comunicación, y apropiados para ser detectados automáticamente. Respecto a la corrección automática, es importante que sea lo más parecida a la evaluación humana de la pronunciación. Para ello, en el desarrollo de un sistema, parece aconsejable realizar estudios en los que se comparan las valoraciones o juicios de inteligibilidad de varios evaluadores humanos acerca de los enunciados producidos por no nativos (por ejemplo, Warren *et al.*, 2009), a fin de establecer un parámetro de referencia para la corrección.

Todo ello plantea la necesidad de que la corrección sea, en la medida de lo posible, específica para la lengua materna del hablante extranjero. Para lograrlo, serán provechosos los estudios de base empírica (fundamentadas en corpus de producciones no nativas) sobre la interferencia fonética y fonológica en el proceso de aprendizaje, especialmente utilizando herramientas de visualización y análisis acústico de la voz.

En cuanto a *cómo* corregir, se ha sugerido que el método de corrección debe ser comprensible a primera vista y fácil de interpretar (Neri, Cucchiarini y Strik, 2002), sin ser demasiado repetitivo, insistente o con frases largas (Eskenazi: 1999; *apud* Wik y Hjalmarsson, 2009). Un método que parece eficaz y adecuado es el uso de sistemas de colores o iconos sencillos (como las luces de tráfico empleadas en Ville; vid. Wik y Hjalmarsson, 2009). También es importante no corregir excesivos errores para no desanimar al alumno; por ejemplo, en el sistema que exponen Neri, Cucchiarini y Strik (2008), como máximo se señalan tres errores en un mismo enunciado. Desde luego, como proponen Neri, Mich, Gerosa y Giuliani (2008), se hace necesario un estudio experimental entre varios grupos que reciban diferentes formas de corrección, para estudiar la influencia de cada uno.

2.2. El diseño de los contenidos pedagógicos. Los aspectos pedagógicos de los sistemas que incorporan tecnologías de habla suelen carecer de las pautas propias procedentes de la investigación en la adquisición de segundas lenguas (Neri, Cucchiarini, Strik y Bobes, 2002). Igualmente, pueden adolecer de un currículum limitado, o no parten de un marco teórico claro ni siguen un modelo determinado de pronunciación (Pennington, 1999: 432-433). De hecho, Eskenazi y Brown (2006) consideran importante para la formación de un especialista en tecnologías de habla educativas abordar una introducción a la teoría del aprendizaje cognitivo y los principios del diseño del software que se derivan de ella.

Siguiendo las recomendaciones de Neri, Cucchiarini y Strik (2003, 2008) es importante diseñar adecuadamente las actividades de aprendizaje para que las tareas de reconocimiento sean lo más simples posible y así se pueda aprovechar el estado actual de la tecnología, que aún no funciona correctamente del todo. Además, interesa la calidad de los contenidos incluidos, sobre todo en cuanto a la autenticidad de las actividades en relación con las situaciones de comunicación a las que se enfrentará el estudiante. En el uso del reconocimiento automático del habla aplicado al aprendizaje de una lengua extranjera se puede correr el riesgo de enseñar una lengua artificial, desligada de su uso real o centrada en la práctica formal, separada de sus aspectos comunicativos o pragmático-funcionales, como Lafford (2004) indica al respecto de Tell me more. No hay duda de que la presentación de ejercicios de pares mínimos en estos sistemas de enseñanza resulta efectiva, pero seguramente mejorarían más si se ofrecieran en enunciados más amplios o en situaciones naturales de comunicación, vinculando la pronunciación con otros niveles lingüísticos del habla (Pennington, 1999).

Por otro lado, la mayoría de los investigadores coinciden en que el usuario ha de ser expuesto a gran cantidad de habla nativa de diferentes variedades geográficas, sociales y de registro. No obstante, como Alwan *et al.* (2007) apuntan acerca de los sistemas para niños, el tipo de estímulo que debe emplearse depende tanto del tiempo de presentación del habla en la tarea o el curso, como el tipo de voz o estilo característico del texto (su variabilidad, dificultad, etc.).

Otras recomendaciones se pueden leer en el artículo de Pennington (1999: 433-ss.); entre otras, facilitar la toma de conciencia de los contrastes entre la lengua meta y la L1 del usuario, o trabajar de forma graduada los contenidos lingüísticos y mediante objetivos concretos de progreso, preferiblemente en un currículum de aprendizaje elaborado y en un enfoque comunicativo o basado en tareas.

2.3. El diseño de la interfaz y de la herramienta. En el diseño del aspecto, sería conveniente seguir las recomendaciones de *usabilidad* que Jakob Nielsen ha propuesto para el diseño web: principalmente, economizar la información presentada en pantalla, llamando la atención de los contenidos clave¹⁵. La interfaz ha de ser estimulante para el usuario (Neri, Cucchiarini y Strik,

¹⁵ Más información se ofrece en la página de Nielsen: <http://www.useit.com/alertbox/>.

2002), especialmente cuando se trata de una aplicación para niños, que ha de conseguir mantener su atención e impedir que se distraigan, motivar en su aprendizaje e incorporar instrucciones sencillas de funcionamiento (Alwan *et al.*, 2007). Además, Eskenazi y Brown (2006) exponen las ventajas del aprendizaje multimodal, en que la interfaz incorpore sonido y video. Dado que cada hablante posee estrategias de aprendizaje diferentes, Eskenazi (1999) recomienda que la interfaz ofrezca la posibilidad de que el usuario elija el tipo de información visual que prefiere (acústica, articularia, o simplemente la corrección de su enunciado). Para ayudarle en la elección de la misma, se puede emplear algún tipo de test con apariencia de juego.

Los mejores sistemas parecen ser los que realizan un seguimiento del usuario y le van presentando contenidos de manera gradual según sus propias dificultades (Eskenazi, 1999). Por ejemplo, la aplicación Pronto o el sistema TAIT se van adaptando a cada estudiante teniendo en cuenta sus propias dificultades de pronunciación y percepción (Dalby y Kewley-Port, 1999). Igualmente, el progreso en el aprendizaje es un aspecto fundamental que habría que cuidar para evitar que los alumnos más capaces se aburran con el sistema, o a la inversa, que se queden rezagados los que necesitan más tiempo de práctica (Alwan *et al.*, 2007). Por ello, parece adecuado un diseño que permita al usuario controlar su propio ritmo de aprendizaje, no sin dejar de controlarlo en cierto grado (por ejemplo, con la temporización automática de las actividades).

No habría que desdeñar tampoco cuestiones de *accesibilidad* de las programas, esto es, el grado de facilidades que presenta una herramienta para ser utilizada por personas con limitaciones visuales, auditivas, etc. Hasta donde hemos profundizado, de todos los sistemas expuestos parece que los únicos que las han tenido en cuenta son los que también han sido empleados con personas con déficits auditivos, articularios o cognitivos; por ejemplo, Pronto (Dalby y KewleyPort, 1999), el Sistema Interattivo Multimediale per l'Alfabetizzazione (Bortolini, 2002) o el agente animado Baldi (Massaro, 2006).

3. ¿QUÉ HERRAMIENTA ELEGIR?

Muchas de las aplicaciones anteriormente expuestas permanecen en el ámbito investigador o en estado de prototipo (a menudo por la falta de financiación), o tienen un enfoque altamente tecnológico, abordando aspectos puntuales de la enseñanza (Gamper y Knapp, 2002: 10). Seguramente, la eficiencia de muchos recursos podría mejorar si se compartiesen aplicaciones y contenidos ya desarrollados en diferentes grupos de investigación, como sugiere Stockwell (2007: 117). En cualquier caso, la variabilidad del usuario final es tan amplia que los desarrollos del mundo investigador no tienen por qué competir con los productos comerciales; quizá, simplemente, el canal de acceso a ellos es más limitado.

Si bien la elección de cada herramienta dependerá de los criterios personales del docente o el aprendiz, hemos recogido una serie de recomendaciones generales que nos parecen adecuadas para el uso de estos sistemas. A continuación se exponen agrupadas para cada tipo de aplicación, aunque sin abordar las que emplean el reconocimiento de habla infantil o los tutores de lectura, por desconocerlos de forma directa.

1. Sistemas de grabación y reproducción:

- Ya que en este tipo de enfoque no suele corregir la pronunciación, desaconsejamos su uso en el aprendizaje autónomo, siendo más adecuados en contextos de aprendizaje guiado por un docente que explica o corrige los fallos.
- Parecen más adecuados para la práctica de la pronunciación (de palabra o fonema) en niveles iniciales.

- De los programas expuestos, destacamos Español en marcha (para niveles intermedios; Gimeno Sanz, 1998) y Español interactivo (para niveles iniciales; *vid.* reseña de Adams, 1998), por incluir grabaciones auténticas de nativos y actividades que promueven la interacción en situaciones cercanas a la realidad.
2. Herramientas de visualización del habla:
- Para los usuarios con conocimientos de fonética acústica, o los estudiantes de Lingüística (tanto de nivel de grado como de postgrado), ofrecen una gran riqueza de información herramientas como Computerized Speech Lab, MATLAB, Praat, VisiPitch o Win Pitch.
 - Para quienes se inician en el análisis acústico y fonético, o para aquellos profesores de idiomas que empleen estos sistemas con sus alumnos, resultan recomendables por su sencillez programas de libre distribución como WASP (Waveforms Annotations Spectrograms & Pitch) y WaveSurfer, o cualquier otro con una interfaz que emplee una representación simplificada de la señal.
 - Con respecto a la visualización de las propias producciones para la práctica de la pronunciación, parece más apropiado el oscilograma (en la adquisición del acento de intensidad) y la curva melódica (para la entonación, especialmente por aprendices con dificultades con este rasgo, como chinos y japoneses). La información que proporcionan las cartas de formantes no nos parece adecuada (requiere conocimientos avanzados de fonética), a menos que se integre en algún tipo de videojuego educativo.
3. Visualización de los órganos fonadores y articuladores:
- En la comprensión del habla parece tener igual importancia la visualización de los órganos fonadores y articuladores tanto internos como externos; por ello, resulta positivo incorporar ambos tipos de información. Se puede tomar como modelo el recurso Fonética: Los sonidos del español (Dispensa *et al.*, 2001; ref. n.º 7), en el que se ofrece una vista medio sagital con las posiciones de la lengua y paladar, y un vídeo con la articulación de la boca y los labios.
 - El programa más adecuado debería ofrecer este tipo de información lo más simplificada posible para no desbordar al usuario.
 - Nos parece tan importante ofrecer la información articulatoria como la corrección de la pronunciación, en la línea de programas como PLASER (Mak *et al.*, 2003), Tell me more o Talk to me (ref. n.º 36).
4. Reconocimiento de habla:
- Varios investigadores se mantienen cautelosos sobre el grado de efectividad de estos programas. Dado el estado actual de esta tecnología, parecen recomendables únicamente como complemento a la instrucción presencial –pero nunca un sustituto–, sobre todo, en contextos de enseñanza donde el aprendiz no está suficientemente expuesto al *input* nativo (por ejemplo, en el aprendizaje de una lengua fuera del país donde esta se habla).
 - Teniendo cuenta que los mejores resultados de reconocimiento se obtienen en sistemas con un vocabulario o un conjunto de frases reducido, el reconocimiento de voz resultaría especialmente aconsejable para niveles iniciales (sobre todo, para practicar la pronunciación), en los que el alumno no posee aún los recursos para construir un discurso más elaborado o espontáneo.
 - Los programas idóneos son los que simulan situaciones comunicativas reales, integrando las destrezas pragmáticas junto a contenidos gramaticales o léxicos; por ejemplo, Learn to speak Spanish. Dicha autenticidad apenas se consigue, según Lafford (2004), en programas como Talk to me o Tell me more (ref. 36).

- Aunque ignoramos la efectividad de los tutores inteligentes guiados mediante la voz (ref. n.º 25), el tipo de interactividad que ofrecen promete ser positiva, más aún si se integran en interfaces de tipo videojuego.
 - Acerca de la evaluación por medio de test con reconocimiento de voz, desconocemos los resultados de primera mano, pero sospechamos que podrían generar aún rechazo en el usuario, especialmente si se trata de un examen que puede tener gran repercusión académica.
5. Conversión texto-habla y sistemas de diálogo:
- Pese a lo atractivo del uso de personajes virtuales o avatares (especialmente si pueden ser personalizados o controlados mediante la voz), el principal escollo de estas aplicaciones es la falta de naturalidad del habla.
 - Quizá las tecnologías más maduras, que realizan una síntesis más natural en el nivel de la palabra, son ya adecuadas para la integración en herramientas como diccionarios electrónicos, o cualquier otra que emplee vocabularios reducidos.
 - Resulta muy interesante el campo de aplicación en entornos de videojuego o que contengan un componente de competitividad entre usuarios (como el prototipo Word War; McGraw *et al.*, 2009), siempre y cuando se mejore la robustez y la flexibilidad de estos sistemas para que se adapten rápidamente a voces distintas.
 - Las aplicaciones más útiles nos parecen las que simulan escenarios reales, como Let's go (Raux y Eskenazi, 2004) o Subarashii (Ehsani *et al.*, 2000).

En la tabla inferior se sintetizan las ideas clave explicadas para cada tipo.

TIPO DE SISTEMA	RECOMENDACIONES DIDÁCTICAS	SISTEMAS ACONSEJADOS
Sistemas de grabación y reproducción	Niveles iniciales. Necesaria la corrección del docente.	Los que simulen situaciones comunicativas reales (p.ej. Español interactivo, Español en marcha, etc.).
Herramientas de visualización del habla	Para la práctica de la pronunciación, aconsejable la supervisión por el profesor.	Para iniciarse: con información simplificada (WASP, Wavesurfer...). Para usuarios avanzados: Computerized Speech Lab, MATLAB, Praat, etc.
Visualización de órganos fonadores/articulatorios	Corrección al mismo tiempo que visualización de la articulación.	Tell me more, PLASER, etc.
Reconocimiento de habla	Niveles iniciales. Complemento de la instrucción formal (sobre todo, fuera del país donde se habla la lengua meta).	Los que simulan interacciones reales (p. ej., Learn to speak Spanish) o tienen carácter lúdico.
Conversión texto-habla y sistemas de diálogo	Aún precisa mejorar la naturalidad del habla sintetizada o la robustez del reconocimiento de usuarios distintos.	Los que simulen situaciones comunicativas reales o tengan un componente lúdico.

Tabla 2 – Recomendaciones para la elección de cada tipo de sistema

4. CONCLUSIONES

Se han revisado distintas tecnologías de visualización, análisis acústico y procesamiento del habla para el aprendizaje de la lengua materna o extranjera. Los sistemas de visualización de la voz resultan una poderosa ayuda para el análisis acústico de las producciones no nativas, facilitando la investigación de aspectos fonéticos que permitan crear modelos acústicos para los sistemas de procesamiento del habla, ayudar a la enseñanza de la fonética y la fonología, etc. (dichas herramientas, además, han facilitado el diagnóstico o la terapia de patologías del habla). Sin

embargo, requieren conocimientos de fonética acústica o el apoyo de un profesor para ser utilizadas directamente o de forma autónoma por el estudiante.

El formato más atractivo de una tecnología educativa (sobre todo si se dirige a jóvenes o a niños) puede ser una interfaz de tipo videojuego. Sin olvidar que en el diseño del programa parece positivo incluir también un módulo de seguimiento del usuario, o la posibilidad de que el aprendiz pueda establecer su propio ritmo de aprendizaje.

Además, en la enseñanza de la pronunciación parece necesario abordar varios niveles de la oralidad: el nivel segmental (por ejemplo, mediante ejercicios de pares mínimos, de corrección de la articulación, etc.) y el nivel suprasegmental (identificación y producción correcta de la curva melódica). Junto a ello, resulta imprescindible la práctica de las destrezas del nivel pragmático de la comunicación (por ejemplo, ejercicios interactivos con personajes virtuales en simulaciones de situaciones comunicativas reales).

En el desarrollo de estos programas, puede ser eficaz adoptar una perspectiva multidisciplinar formada por ingenieros del habla, lingüistas, logopedas, pedagogos o profesores de lenguas. Sin olvidar, por otra parte, las recomendaciones propias de los usuarios y las necesidades específicas de los estudiantes a los que van destinados (según su edad, si aprenden la lengua materna o un segundo idioma, los errores específicos por interferencia de su lengua materna, etc.).

BIBLIOGRAFÍA

[Los documentos de Internet estaban accesibles el 27 de diciembre de 2010]

ARTÍCULOS Y LIBROS

- ADAMS, C.R. (1998), "CALICO Software Review: *Español interactivo 1.01*", *Calico Software Reviews*, 11/98. <https://calico.org/p-164-%20Español%20Interactivo.html>
- AIST, G. y MOSTOW, J. (2009), "Designing Spoken Tutorial Dialogue with Children to Elicit Predictable but Educationally Valuable Responses", *Proceedings of INTERSPEECH 2009*, 588-591. Documento disponible en: http://www.cs.cmu.edu/~listen/pdfs/2009InterspeechAistMostow_final.pdf
- AKAHANE-YAMADA, R., TOHKURA, Y., BRADLOW, A. R. y PISONI, D. B. (1996), "Does training in speech perception modify speech production?", *Proceedings of the Fourth International Conference on Spoken Language Processing, ICSLP 96*, 606609. Disponible en: www.asel.udel.edu/icslp/cdrom/vol2/277/a277.pdf
- ALIAGA-GARCÍA, C. (2007), "The role of phonetic training in L2 speech learning", *Proceedings of the Phonetics Teaching & Learning Conference, UCL, August 24-26, 2007*. Disponible en: www.phon.ucl.ac.uk/ptlc/proceedings/ptlcpaper_32e.pdf
- ALWAN, A., BAI, Y., BLACK, M., CASEY, L., GEROSA, M., HERITAGE, M., ISELI, M., JONES, B., KAZEMZADEH, A., LEE, S., NARAYANAN, S., PRICE, P., TEPPERMAN, J. y WANG, S. (2007), "A System for Technology Based Assessment of Language and Literacy in Young Children: the Role of Multiple Information Sources" *IEEE 9th Workshop on Multimedia Signal Processing (MMSP 2007)*, 2630. Disponible en: http://diana.icsl.ucla.edu/Tball/publications/tball_mmmsp07.pdf
- ARIAS, J. P., BECERRA YOMA, N. y VIVANCO, H. (2010), "Automatic intonation assessment for computer aided language learning", *Speech Communication*, 52(3), 254-267.
- BADIN, P., TARABALKA, Y., ELISEI, F. y BAILLY, G. (2010) "Can you 'read' tongue movements? Evaluation of the contribution of tongue display to speech understanding", *Speech Communication*, 52(6), 493-503.

- BARR, D., LEAKEY, J. y RANCHOUX, A. (2005), "Told like it is! An evaluation of an integrated oral development pilot project", *Language Learning and Technology*, 9 (3), 55-78. Disponible en: <http://llt.msu.edu/vol9num3/barr/>
- BERKOVITZ, R. (1999), "Design, development and evaluation of computer-assisted learning for Speech Science education", *Proceedings of MATISSE*. 9-16, Londres.
- BERNAL BERMÚDEZ, J., BOBADILLA SANCHO, J. y GÓMEZ VILDA, P. (2000), *Reconocimiento de voz y fonética acústica*, Madrid, Editorial Ra-Ma.
- BERNSTEIN, J., BARBIER, I., ROSENFELD, E. y JONG, J. D. (2004), "Development and validation of an automatic spoken Spanish test", *Proceedings of InSTIL/ICALL Symposium 2004*. ISCA
- BERNSTEIN, J. y CHEN, J. (2008), "Logic and Validation of a Fully Automatic Spoken English Test", en M. Holland y F. Pete Fisher (eds.) *The Path of Speech Technologies in Computer Assisted Language Learning. From Research Toward Practice*, London, Routledge.
- BISSIRIA, M^a. P. y PFITZINGER, H. R. (2009), "Italian speakers learn lexical stress of German morphologically complex words", *Speech Communication*, 51(2), 933-947.
- BLACK, A. W. (2007), "Speech Synthesis for Educational Technology", *Proceedings of Speech and Language Technology in Education (SLaTE2007)*, 104-107.
- BLAKE, R. J. (2008), *Brave New Digital Classroom*, Washington, Georgetown Un. Press
- BONAVENTURA, P., HERRON, D. y MENZEL, W. (2000), "Phonetic Rules for Diagnosis of Pronunciation Errors", *KONVENS 2000*, 225-230. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.42.9705&rep=rep1&type=pdf>
- BONAVENTURA, P., HOWARTH, P. y MENZEL, W. (2000), "Phonetic annotation of a non-native speech corpus", *Proceedings International Workshop on Integrating Speech Technology in the (Language) Learning and Assistive Interfac, InStil 2000, Dundee*, 10-17. Documento disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.1229&rep=rep1&type=pdf>
- BORTOLINI, U. (2002), *Sistema Interattivo Multimediale per l'Alfabetizzazione (SIMpA)*, Padua, CNR-Servizio Pubblicazioni.
- BRADIN, C. (1999), "Review of *Oral Language Archive*", *Language Learning & Technology*, vol. 2, n.º 2, 16-22. <http://llt.msu.edu/vol2num2/pdf/review1.pdf>
- BRATT, H., NEUMEYER, L., SHRIBERG, E. y FRANCO, H. (1998), "Collection and Detailed Transcription of a Speech Database for Development of Language Learning Technologies", *Proc. of ICSLP 98*. <ftp://ftp.speech.sri.com/pub/papers/icslp98-ratings.ps>
- BROWN, I. (2000), "CALICO Software Review: Pro-nunciation. The English Communication Toolkit", *Calico Software Reviews*, 10/00. <https://calico.org/p-39-Pro-nunciation.html>
- BROWN, J. (2004), "Integrating Tools for the Creation of Speech-Enabled Tutors", CMU LTI Technical Report CMU-LTI-04-186. www.cs.cmu.edu/afs/cs/Web/People/jonbrown/publications/Brown_TechReport.pdf
- CASSELL, J. (2004), "Towards a model of technology and literacy development: Story listening systems", *Applied Developmental Psychology*, 25, 75-105. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.75.2092&rep=rep1&type=pdf>
- CAZADE, A. (1999), "De l'usage des courbes sonores et autres supports graphiques pour aider l'apprenant en langues", *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, vol. 2, n.º 2. Documento disponible en: http://alsic.u-strasbg.fr/Num4/cazade/alsic_n04-rec1.htm
- CELCE MURCIA, M. y GOODWIN, J. M. (1991), "Teaching pronunciation". En M. Celce Murcia (Ed.) *Teaching English as a Second Language*, New York, Heinle and Heinle.

- CHUN, D. M. (1998), "Signal Analysis Software For Teaching Discourse Intonation", *Language Learning & Technology*, vol. 2, n.º 1, 74-93. Disponible en: <http://llt.msu.edu/vol2num1/article4/>
- COLE, R., MASSARO, D. W., VILLIERS, J., RUNDLE, B., SHOBAKI, K., WOUTERS, J., COHEN, M., BESKOW, J., STONE, P., CONNORS, P. y SOLCHER, D. (1999), "New tools for interactive speech and language training: Using animated conversational agents in the classrooms of profoundly deaf children", *Proc. ESCA/SOCRATES Workshop on Method and Tool Innovations for Speech Science Education, London*, 45-52. Disponible en: http://cslu.cse.ogi.edu/toolkit/pubs/ps/cole_MATISSE_99.ps
- COLPAERT, J. (2004), *Design of online interactive language courseware: conceptualization, specification and prototyping: research into the impact of linguistic-didactic functionality on software architecture*, doctoral dissertation, Antwerpen, Universiteit Antwerpen, Faculteit Letteren en Wijsbegeerte, Departement Taalkunde.
- CORSBIE, C. y GORE, J. (2002), "Review of Pronunciación y Fonética Ver. 2.0", *CALICO Software Review. CALICO Journal*, Vol. 20 (3), 621-631. Disponible en: [https://www.calico.org/p-148-Pronunciación%20y%20Fonética%20\(42002\).html](https://www.calico.org/p-148-Pronunciación%20y%20Fonética%20(42002).html)
- CUCHIARINI, C., STRIK, H. y BOBES, L. (2000), "Different aspects of expert pronunciation quality ratings and their relation to scores produced by speech recognition algorithms", *Speech Communication*, 30, pp.109–119.
- CUCCHIARINI, C., NERI, A. y STRIK, H. (2009), "Oral Proficiency Training in Dutch L2: the Contribution of ASR-based Corrective Feedback", *Speech Communication*, 51(10), 853-863. Disponible en: <http://lands.let.ru.nl/~strik/2-div/KUL-LEA-0809/Cucchiarini-EtAl-DutchCAPT.pdf>
- DALBY, J. y KEWLEY-PORT, D. (1999), "Explicit pronunciation training using Automatic Speech Recognition Technology", *CALICO Journal*, vol. 16 (3), 425-445. Disponible en: https://www.calico.org/html/article_622.pdf
- DALBY J. y KEWLEY-PORT, D. (2008), "Design Features of Three Computer-Based Speech Training Systems", en M. Holland y F. Pete Fisher (eds.) *The Path of Speech Technologies in Computer Assisted Language Learning. From Research Toward Practice*. London: Routledge.
- DELMONTE, R. (2002), "Feedback generation and linguistic knowledge in 'SLIM' automatic tutor", *ReCALL*, 14 (2) 209-234.
- (2008), "Speech Synthesis for Language Tutoring Systems", en M. Holland y F. Pete Fisher (eds.), *The Path of Speech Technologies in Computer Assisted Language Learning. From Research Toward Practice*, London, Routledge.
- (2009), "Prosodic tools for language learning", *International Journal of Speech Technology*, 12 (4), 161-184. Documento disponible en: <http://lear.unive.it/bitstream/10278/1460/1/ProsodicTools3.pdf>
- DOREMALEN, J., STRIK, H. y CUCCHIARINI, C. (2009), "Optimizing non-native speech recognition for CALL applications", *Proceedings of INTERSPEECH-2009*, 592-595. Documento disponible en: http://taaluniecentrum-nvt.taalunie.info/taal/technologie/stevin/documenten/disco_interspeech2009.pdf
- DUCHATEAU, J., KONG, Y. O., CLEUREN, L., LATA CZ, L., ROELENS, J., SAMIR, A., DEMUYNCK, K., GHESQUIÈRE, P., VERHELST, W. y VAN HAMME, H. (2009), "Developing a reading tutor: Design and evaluation of dedicated speech recognition and synthesis modules", *Speech Communication*, 51(10), 985-994. Documento disponible en: http://www.esat.kuleuven.be/psi/spraak/cgi-bin/get_file.cgi?/duchato/specom09/corrected.pdf&pdf

- EHSANI, F., BERNSTEIN, J. y NAJMI, A. (2000), "An interactive dialog system for learning Japanese", *Speech Communication*, 30, 167-177.
- EHSANI, F. y KNOTT, E. (1998), "Speech Technology in Computer-Aided Language Learning: Strengths and Limitations of a new CALL Paradigm", *Language Learning & Technology*, vol. 2, n.º 1, julio 1998, 45-60. <http://lt.msu.edu/vol2num1/pdf/article3.pdf>
- ENGWALL, O., BÄLTER, O., ÖSTER, A. M. y KJELLSTRÖM, H. (2006), "Designing the user interface of the computer-based speech training system ARTUR based on early user tests", *Journal of Behaviour and Information Technology*, vol. 25, n.º 4, pp.353-365.
- ENGWALL, O. y WIK, P. (2009a), "Real vs. rule-generated tongue movements as an audiovisual speech perception support", *Proceedings of FONETIK 2009*, 30-35. http://www2.ling.su.se/fon/fonetik_2009/030%20engwall_wik_fonetik2009.pdf
- ENGWALL, O. y WIK, P. (2009b), "Are real tongue movements easier to speech read than synthesized?", *Proceedings of INTERSPEECH 2009*, 824-827. Documento disponible en: www.speech.kth.se/prod/publications/files/3348.pdf
- ERIKSSON, E., BÄLTER, O., ENGWALL, O., ÖSTER, A. M. y KJELLSTRÖM, H. (2005), "Design Recommendations for a Computer-Based Speech Training System Based on End-User Interviews", *Proceedings of the Tenth International Conference on Speech and Computers, Patras, Greece*. 483-486. Disponible en: www.speech.kth.se/prod/publications/files/1270.pdf
- ESKENAZI, M. (1999), "Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype", *Language Learning and Technology*, 2 (2), 62-76. Disponible en: <http://lt.msu.edu/vol2num2/pdf/article3.pdf>
- (2009), "An overview of spoken language technology for Education", *Speech Communication*, 51, 832-844. Documento disponible en: <http://ml.hss.cmu.edu/courses/jones/82-888/Eskenazi-SpeechInEducation.pdf>
- ESKENAZI, M. y BROWN, J. (2006), "Teaching the creation of software that uses speech recognition", en P. Hubbard y M. Levy (eds.) *Teacher education in CALL*. John Benjamins.
- FELIX, U. (2005), "Analyzing recent CALL effectiveness research: Towards a common agenda", *Computer Assisted Language Learning*, 18 (1&2), 1-33. Disponible en: <http://arts.monash.edu.au/lcl/newmedia-in-langlearn/phd-calljournal05-final.pdf>
- FELPS, D., BORTFELD, H. y GUTIÉRREZ-OSUNA, R. (2009), "Foreign accent conversion in computer assisted pronunciation training", *Speech Communication*, 51(10), 920-932. Documento disponible en: http://research.cs.tamu.edu/prism/publications/sc09_felps.pdf
- FLEGE, J. E. (1988), "Using visual information to train foreign language vowel production", *Language Learning* 38, 3, 365-407.
- GAMPER, J. y KNAPP, J. (2002), "A review of Intelligent CALL Systems", *Computer assisted language learning*, vol. 15, n.º 4, 329-342. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.3960&rep=rep1&type=pdf>
- GARCÍA-RUIZ, M. A., EDWARDS, A., EL-SEOUD, S. A. y AQUINO-SANTOS, R. (2008), "Collaborating and learning a second language in a Wireless Virtual Reality Environment" *International Journal of Mobile Learning and Organisation archive*, Vol. 2 (4), 369-377. Disponible en: www.hrl.uoit.ca/~miguelga/Collaborating_and_learning_a_second_language_in_a_Wireless_Virtual_Reality_Environment.pdf
- GERMAIN, A. y P. MARTIN (2000), "Présentation d'un logiciel de visualisation pour l'apprentissage de l'oral en langue seconde", *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, vol. 3, n.º 1, 61-76. http://alsic.u-strasbg.fr/Num5/germain/alsic_n05-rec7.htm

- GIL FERNÁNDEZ, J. (2007), *Fonética para profesores de español*, Madrid, Arco/Libros S.L.
- GILL, B. (1999), "Review of *Learn to Speak Spanish*", *CALICO Journal*, vol. 17 (2), 320-333. <https://calico.org/p-161-Speak%20Spanish%20%2851999%29.html>
- GIMENO SANZ, A. (1998), "El aprendizaje de español/LE asistido por ordenador. *CAMILLE: Español en marcha*", en Jiménez Juliá, T., Losada Aldrey, M.C., Márquez, J.F., y Sotelo, S. (eds.), (1999) *Actas IX congreso ASELE*. Santiago de Compostela.
- GODWIN-JONES, R. (2009), "Speech Tools and Technologies", *Language Learning and Technology*, 13(3), 4-11. Disponible en: <http://llt.msu.edu/vol13num3/emerging.pdf>
- GÓMEZ VILDA, P., ÁLVAREZ, A., NIETO, V., RODELLAR, V., AGUILERA, S., LESTANI, J., BOBADILLA, J., BERNAL, J. y PÉREZ, M. (1995), "A User Interface to integrate Speech Audio-Feedback in CALL Systems", *Proceedings of the EUROCALL'95, Valencia, 7-9 September, 1995*, 47-48.
- GÓMEZ VILDA, P., ÁLVAREZ, A., MARTÍNEZ, R., BOBADILLA, J. BERNAL, J., V. RODELLAR, V. y NIETO, V. (2008), "Applications of Formant Detection in Language Learning", en M. Holland y F. Pete Fisher (eds.), *The Path of Speech Technologies in Computer Assisted Language Learning. From Research Toward Practice*. London: Routledge.
- GRANSTRÖM, B. (2004), "Towards a virtual language tutor", *Proc InSTIL/ICALL2004 – NLP and Speech Technologies in Advanced Language Learning Systems*, 1-8. Disponible en: www.speech.kth.se/ville/publications/instill04_bg.pdf
- HAGEN, A., PELLOM, B. y COLE, R. (2003), "Children's Speech Recognition With Application To Interactive Books And Tutors", 2003 IEEE Workshop on Automatic Speech Recognition and Understanding, 2003. ASRU '03. En: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.4106&rep=rep1&type=pdf>
- HANDLEY, Z. y HAMEL, M. (2005), "Towards establishing a methodology for benchmarking speech synthesis for Computer Assisted Language Learning", *Language and Technology Journal*, vol. 9, n.º 3. Documento disponible en: www.cs.bham.ac.uk/~mg1/cluk/papers/handley.pdf
- HANDLEY, Z. (2009), "Is text-to-speech synthesis ready for use in computer-assisted language learning?", *Speech Communication*, 51 (10), 906-919.
- HARDISON, D. M. (2004), "Generalization Of Computer-Assisted Prosody Training: Quantitative And Qualitative Findings", *Language Learning and Technology*, vol. 8 (1). Documento disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.131.6103&rep=rep1&type=pdf>
- HARLESS, W. G., ZIER, M. A. y DUNCAN, R. C. (1999), "Virtual Dialogues with Native Speakers: The Evaluation of an Interactive Multimedia Method", *CALICO*, vol. 16 (3). Disponible en: https://www.calico.org/html/article_617.pdf
- HINCKS, R. (2002), "Speech synthesis for teaching lexical stress", *Proceedings of Fonetik 2002, TMHQPSR*, 44, 153-156. Documento disponible en: www.speech.kth.se/~hincks/papers/fon02bh.pdf
- (2003), "Speech Technologies for pronunciation feedback and evaluation", *ReCALL*. 15 (1), 3-20. Disponible en: www.speech.kth.se/~hincks/papers/REcall.pdf
- (2005a), "Measures and perceptions of liveliness in student oral presentation speech: A proposal for an automatic feedback mechanism", *System*, vol. 33 (4), 575-591. Disponible en: www.speech.kth.se/~hincks/papers/System.pdf
- (2005b), *Computer Support for Learners of English*. Doctoral Thesis. Stockholm. KTH School of Computer Science and Communication, Department of Speech, Music and Hearing. Disponible en: www.speech.kth.se/~hincks/papers/hincksthesis.pdf

- HINCKS, R. y EDLUND, J. (2009), "Transient visual feedback on pitch variation for Chinese speakers of English", *Proceedings of FONETIK 2009*, 102-108. Publicado en *Language Learning & Technology*, 13 (3), 32-50. <http://lt.msu.edu/vol13num3/hincksedlund.pdf>
- HWU, F. (1997), "Providing an Effective and Affective Learning Environment for Spanish Phonetics with a Hypermedia Application", *CALICO Journal*, vol. 14 (2-4). Documento disponible en: https://www.calico.org/html/article_358.pdf
- ISLE D3.3 (1999), "Recognition of Learner Speech", *ISLE Deliverable D3.3*. Disponible en: <http://nats-www.informatik.uni-hamburg.de/~isle/public/D33/D33.pdf>
- JONES, R. H. (1997), "Beyond "Listen And Repeat": Pronunciation Teaching Materials And Theories Of Second Language Acquisition", *System*, vol. 25, n.º 1, 103-112. Accesible en: http://humanities.ucalgary.ca/lrc/sites/ucalgary.ca.lrc/files/beyond_listen_and_repeat.pdf
- JOHNSON, W. L., MARSELLA, S. y VILHJÁLMSSON, H. (2004), "DARWARS Tactical Language Training System (TLTS)", *Proc. of Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2004*. Disponible en: www.ru.is/faculty/hannes/publications/IITSEC2004.pdf
- KANG, M., KASHIWAGI, H., TREVIRANUS, J. y KABURAGI, M. (2009), "Synthetic speech in foreign language learning: An evaluation by learners", *International Journal of Speech Technology*, 11(2), 97-106.
- KAWAHARA, H. y AKAHANE-YAMADA, R. (2006), "STRAIGHT as a research tool for L2 study: How to manipulate segmental and suprasegmental features", *Journal of the Acoustical Society of America*, 120 (5), 3137-3137. www.wakayama-u.ac.jp/~kawahara/Resources/L2toolSTRAIGHT.pdf
- KIRSCHNING, I., AGUAS, N. y AHUACTZIN, A. (2000), "Aplicación de tecnología de voz en español en la enseñanza del español", *HAVOL 2000, 1er Taller Internacional de Tratamiento del Habla, Procesamiento de Voz y el Lengua*. México DF, Agosto de 2000. Disponible en: <http://ict.udlap.mx/people/ingrid/ingrid/HAVOL2000a.pdf>
- KOMMISACHIRK, J. y KOMMISACHIRK, E. (2000), "BetterAccent Tutor – Analysis and Visualization of Speech Prosody", *Proc. of Speech Technology in Language Learning*. August 2000, Dundee, Scotland, 86-89. Disponible en: <http://www.betteraccent.com/papers/BetterAccent%20STILL%20Paper.doc>
- KRAJKA, J. (2001), "ENGLISH+KIDS", *CALICO Software Review, CALICO Journal*, vol. 20, n.º 2, 393-404. [www.calico.org/p-33-English%2BKIDS%20\(102001\).html](http://www.calico.org/p-33-English%2BKIDS%20(102001).html)
- LABRADOR GUTIÉRREZ, T. y C. FERNÁNDEZ JUNCAL (1994), "Aplicaciones del visualizador de habla en la enseñanza del español L/E", *Actas del IV Congreso de la Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE)*. http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/04/04_0267.pdf
- LAFFORD, B. (2004), "Review of *Tell me more Spanish*", *Language Learning and Technology*. September 2004, vol. 8 (3), 21-34. Disponible en: <http://lt.msu.edu/vol8num3/pdf/review1.pdf>
- LAHOZ BENGOCHEA, J. M. (2008), "Audio en Campus Virtual: la enseñanza de la fonética y la comprensión auditiva", *IV Jornada Campus Virtual UCM: Experiencias en el Campus Virtual (Resultados)*. Editorial Complutense, Madrid, 63-69. Documento disponible en: <http://eprints.ucm.es/7789/1/campusvirtua72-781.pdf>
- LI, C. L., LIU, J. y XIA, S. H. (2006), "Perceptual Evaluation of Pronunciation Quality for Computer Assisted Language Learning", in *Technologies for E-Learning and Digital Entertainment First International Conference, Edutainment 2006, Hangzhou, China, April*

- 16-19, 2006. *Proceedings. Lecture Notes In Computer Science Series*. Springer Berlin/Heidelberg
- LIU, H. C., CHIU, T. L. y YEH, Y. L. (2006), “Effects of web-based oral activities enhanced by automatic speech recognition”, comunicación presentada en el congreso *CALICO 2006 annual symposium: online learning, come ride the wave*, University of Hawaii at Manoa, May. 16-20. Disponible en: http://candle.fl.nthu.edu.tw/newcandle/chi/publi/CALICO_Stan.ppt
- LYRAS, D. P., KOKKINAKIS, G., LAZARIDIS, A., SGARBAS, K. y FAKOTAKIS, N. (2009), “A Large Greek-English Dictionary with Incorporated Speech and Language Processing Tools”, En *Proceedings of INTERSPEECH-2009*, 891-1894. Disponible en: http://www.wcl.ece.upatras.gr/alaza/KORAIS_Lazaridis.pdf
- LLISTERRI, J. (1997), “Nuevas tecnologías y enseñanza del español como lengua extranjera”, *Actas del VIII Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera. La Enseñanza del Español como Lengua Extranjera: del Pasado al Futuro. Alcalá de Henares, 1997*. Disponible en: http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/08/08_0043.pdf
- (2001), “Enseñanza de la pronunciación, corrección fonética y nuevas tecnologías”, *Es Espasa, Revista de Profesores*, 28 de noviembre de 2001. Disponible en: http://liceu.uab.cat/~joaquin/publicacions/CorrFon_NT_2001.pdf
- (2006), “La enseñanza de la pronunciación asistida por ordenador”, en *Actas del XXIV Congreso Internacional de AESLA. Aprendizaje de lenguas, uso del lenguaje y modelación cognitiva: perspectivas aplicadas entre disciplinas*, Madrid, Universidad Nacional de Educación a Distancia - AESLA, Asociación Española de Lingüística Aplicada. 91-120. Documento disponible en: http://liceu.uab.cat/~joaquin/publicacions/Llisterri_06_Pronunciacion_Tecnologias.pdf
- MACDONALD, J. y MCGURK, H. (1978), “Visual influences on speech perception processes”, *Perception & Psychophysics*, 24, 253-257.
- MAK, B., SIU, M., NG, M., TAM, Y. C., CHANY, Y. C., CHAN, K. W., LEUNG, K. Y., HO, S., CHONG, F. H., WONG, J. y LO, J. (2003), “PLASER: Pronunciation Learning via Automatic Speech Recognition”, *Proc. HLT-NAACL*, 23–29. Disponible en: www ldc.upenn.edu/acl/W/W03/W03-0204.pdf
- MARTIN, P. (2005), “WinPitch LTL, un logiciel multimédia d’enseignement de la prosodie”, *Apprentissage des Langues et Systèmes d’Information et de Communication*. Vol. 8, 95-108. Documento disponible en: http://hal.archives-ouvertes.fr/docs/00/10/96/13/PDF/alsic_v08_13-rec7.pdf
- MASSARO, D. W. (2006), “The psychology and technology of talking heads: Applications in Language Learning”, en O. Bernsen, L. Dybkjaer, y J. van Kuppevelt (Eds.), *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems*, 183-214. Dordrecht, The Netherlands: Kluwer Academic Publishers. Disponible en: <http://mambo.ucsc.edu/pdf/massaro4.pdf>
- MAYFIELD TOMOKIYO, L., WANG, L. y ESKÉNAZI, M. (2000), “An Empirical Study of the Effectiveness of Speech-Recognition-Based Pronunciation Tutoring”, en *Interspeech 2000, Proceedings of the 6th International Conference on Speech and Language Processing. October 2000, Beijing, China*. Disponible en: <http://www.cs.cmu.edu/~laura/Papers-PS/icslp-fluency.ps>
- MAYFIELD TOMOKIYO, L. y WAIBEL, A. (2001), “Adaptation Methods For Non-Native Speech”, *Proceedings of Multilinguality in Spoken Language Processing. Aalborg, September*,

2001. Documento accesible en: http://reference.kfupm.edu.sa/content/p/r/processing_86286.pdf#page=39
- McGRAW, I. y SENEFF, S. (2008), "Speech-enabled Card Games for Language Learners", Association for the Advancement of Artificial Intelligence Disponible en: http://wami.csail.mit.edu/papers/wordwar_mcgraw_seneff_aaai08.pdf
- McGRAW, I., YOSHIMOTO, B. y SENEFF, S. (2009), "Speech-enabled card games for incidental vocabulary acquisition in a foreign language", *Speech Communication*, 51 (10), 1006-1023. Documento accesible en: http://wami.csail.mit.edu/papers/wordwar_mcgraw_yoshimoto_seneff_speechcomm08.pdf
- McGURK, H. y J. MACDONALD (1976), "Hearing lips and seeing voices", *Nature*, 264, 746-748.
- MENZEL, W., HERRON, D., BONAVENTURA, P. y MORTON, R. (2000), "Automatic detection and correction of non-native English pronunciations", *Proceedings of InSTILL 2000. Dundee, Scotland*, 49-56. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.38.5834&rep=rep1&type=pdf>.
- MENZEL, W., HERRON, D., MORTON, R., PEZZOTTA, D., BONAVENTURA, P. y HOWARTH, P. (2001), "Interactive pronunciation training", *ReCALL*, 13 (1), 67-78.
- MERCIER, G., GUYOMARD, M. y SIROUX, J. (1999), "Synthèse de la parole en breton: Didacticiels pour une langue minoritaire", *Proceedings of InSTIL: Speech Technologies Applications in CALL*, 57-61.
- MOLHOLT, G. y HWU, F. (2008), "Visualization of Speech Patterns for Language Learning", en V. M. Holland y F. P. Fisher (2008), *The Path of Speech Technologies in Computer Assisted Language Learning. From Research Toward Practice*, Londo, Routledge.
- MOSTOW, J., ROTH, S., HAUPTMANN, A. G. y KANE, M. (1994), "A prototype reading coach that listens", *Proc. 12th Natl. Conf. on Artificial Intelligence (AAAI-94)*, Seattle, WA, 785-792. Documento disponible en: www.cs.cmu.edu/~listen/pdfs/aaai94_online.pdf
- MOSTOW, J. (2004), "Some Useful Design Tactics for Mining ITS Data", *Proceedings of the ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes, August, 2004*, 20-28. Disponible en: http://www.ri.cmu.edu/pub_files/pub4/mostow_jack_2004_2/mostow_jack_2004_2.pdf
- MOTOHASHI-SAIGO, M. y D. M. HARDISON (2009), "Acquisition Of L2 Japanese Gemimates: Training With Waveform Displays", *Language Learning and Technology*, vol. 13, n.º 2. Disponible en: <http://lt.msu.edu/vol13num2/motohashisaigohardison.pdf>
- NERI, A., CUCCHIARINI, C. y STRIK, H. (2002), "Feedback in Computer Assisted Pronunciation Training: when technology meets pedagogy", *Proc. of the 10th Int. CALL Conference on CALL professionals and the future of CALL research*, University of Antwerp, 179-188. <http://lands.let.ru.nl/%7Eestrik/publications/a95.pdf>
- NERI, A., CUCCHIARINI, C., STRIK, H. y BOVES, L. (2002), "The pedagogy-technology interface in Computer-Assisted Pronunciation Training", *Computer-Assisted Language Learning* 15 (5), 441-467. <http://lands.let.ru.nl/%7Eestrik/publications/a99.pdf>
- NERI, A., CUCCHIARINI, C. y STRIK, H. (2003), "Automatic Speech Recognition for second language learning: How and why it actually works", *Proceedings of 15th International Congress of Phonetic Sciences. Barcelona, Spain*. 1157-1160. Disponible en: <http://lands.let.kun.nl/literature/neri.2003.1.pdf>
- NERI, A., CUCCHIARINI, C. y STRIK, H. (2006), "Selecting segmental errors in L2 Dutch for optimal pronunciation training", *International Review of Applied Linguistics*. 44, 357-404. Disponible en: <http://lands.let.kun.nl/literature/neri.2006.3.pdf>

- NERI, A., CUCCHIARINI, C. y STRIK, H. (2008), "The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch", *ReCALL*, 20 (2), 225-243. Disponible en: <http://lands.let.ru.nl/%7Estrik/publications/a139-CAPT-ReCALL.pdf>
- NERI, A., MICH, O., GEROSA, M. y GIULIANI, D. (2008), "The effectiveness of computer assisted pronunciation training for foreign language learning by children", *Computer Assisted Language Learning*, 21 (5), 393-408. Disponible en: <http://dx.doi.org/10.1080/09588220802447651>
- NEUMEYER, L., FRANCO, H., DIGALAKIS, V. y WEINTRAUB, M. (1999), "Automatic Scoring of Pronunciation Quality", *Speech communication*, vol. 30 (2-3), 83-93. Documento disponible en: www.speech.sri.com/people/hef/papers/SpeechCommPronuncScoring.ps
- NOUZA, J. (1998), "Training Speech Through Visual Feedback Patterns", *Proceedings of ICSLP'98, Sydney, December 1998*. Documento disponible en: <http://www.shlrc.mq.edu.au/proceedings/icslp98/PDF/SCAN/SL981139.PDF>
- OHKAWA, Y., SUZUKI, M., OGASAWARA, H., ITO, A. y MAKINO, S. (2009), "A speaker adaptation method for non-native speech using learners' native utterances for computer-assisted language learning systems", *Speech Communication*, 51(10), 875-882.
- OVIATT, S., DARVES, C. y COULSTON, R. (2004), "Toward Adaptive Conversational Interfaces: Modeling Speech Convergence with Animated Personas", *ACM Tcs on Computer-Human Interaction (TOCHI)*. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.78.550&rep=rep1&type=pdf>
- PENNINGTON, M. C. (1999), "Computer-Aided Pronunciation Pedagogy: Promise, Limitations, Directions", *Computer Assisted Language Learning*, 12 (5), 427-440. http://wiki.umd.edu/teamill/images/7/70/Computer-Aided_Pronunciation_Pedagogy.pdf
- PRECODA, K. y BRATT, H. (2008), "Perceptual underpinnings of Automatic Pronunciation", en M. Holland y F. Pete Fisher (eds.), *The Path of Speech Technologies in Computer Assisted Language Learning. From Research Toward Practice*. London: Routledge.
- PRICE, P., TEPPERMAN, J., ISELI, M., DUONG, T., BLACK, M., WANG, S., BOSCARDIN, C. K., HERITAGE, M., PEARSON, P. D., NARAYANAN, S. y ALWAN, A. (2009), "Assessment of emerging reading skills in young native speakers and language learners", *Speech Communication*, 51 (10), 968-984. Disponible en: www.ee.ucla.edu/~spapl/paper/alwan_speechcom_09.pdf
- PROBST, K., KE, Y. y ESKENAZI, M. (2002), "Enhancing foreign language tutors – In search of the golden speaker", *Speech Communication*, vol. 37 (3-4), 161-173.
- RAUX, A. y T. KAWAHARA (2002), "Automatic Intelligibility Assessment And Diagnosis Of Critical Pronunciation Errors For Computer-Assisted Pronunciation Learning", *ICSLP 2002. Proceedings of the 7th International Conferences on Spoken Language Processing. Denver, Colorado, September 16-20, 2002*. 737-740. www.cs.cmu.edu/~antoine/papers/icslp2002a.pdf
- RAUX, A. y ESKENAZI, M. (2004), "Using Task-Oriented Spoken Dialogue Systems for Language Learning: Potential, Practical Applications and Challenges", *Proceedings of InSTIL/ICALL 2004 Symposium on Computer Assisted Learning*. <http://www.cs.cmu.edu/~antoine/papers/rauxeskenazi-instil-04.pdf>
- RICHARDS, J. C. y RODGERS, T. (2003), *Enfoques y métodos en la enseñanza de idiomas*. Traducción de José M. Castrillo y edición española de Álvaro Gracia y Josep M. Mas. Cambridge: Cambridge University Press
- ROSE, H., y BILLINGHURST, M. (1995), *Zengo Sayu: An Immersive Educational Environment for Learning Japanese* (informe técnico), Seattle, WA: Human Interface Technology

- Laboratory, University of Washington. Disponible en: www.hitl.washington.edu/publications/r-95-4/
- RUIPÉREZ GARCÍA, G. (2004), “La enseñanza de lenguas asistida por ordenador (ELAO)”, en J. Sánchez Lobato e I. Santos Gargallo, (dirs.), *Vademécum para la formación de profesores. Enseñar español como segunda lengua (L2) / lengua extranjera (LE)*, Madrid, SGEL.
- RUSSELL, M., BROWN, C., SKILLING, A., SERIES, R., WALLACE, J., BONHAM, B. y BARKER, P. (1996), “Applications of automatic speech recognition to speech and language development in young children”, *Proc. Internat. Conf. on Spoken Language Processing, ICSLP'96*, Philadelphia, PA. Disponible en: <http://www.asel.udel.edu/icslp/cdrom/vol1/580/a580.pdf>
- RUSSELL, M., SERIES, R. W., WALLACE, J. L., BROWN, C. y SKILLING, A. (2000), “The STAR system: an interactive pronunciation tutor for young children”, *Computer Speech and Language*, 14, 161–175.
- RYPAN, M. E. y PRICE, P. (1999), “VILTS: A Tale of Two Technologies”, *CALICO*, n.º 16 (3). Disponible en: https://www.calico.org/html/article_620.pdf
- SELINKER, L. (1972), “Interlanguage”, *International Review of Applied Linguistics in Language Teaching*, 10 (3), 209-231. Existe traducción española en J. M. Liceras (ed.) (1991) *La adquisición de lenguas extranjeras*, 79-97, Madrid, Visor.
- SENEFF, S., WANG, C., PEABODY, M. y ZUE, V. (2004), “Second Language Acquisition through Human Computer Dialogue”, en *Proc. of the 4th International Symposium on Chinese Spoken Language Processing, 2004, Hong Kong, China*. Disponible en: <http://people.csail.mit.edu/wangc/papers/icslp04-muxing.pdf>
- SENEFF, S. (2007), “Web-based dialogue and translation games for spoken language learning2”, en *Proceedings of the SLATE Workshop on Speech and Language Technology in Education (SLATE-2007)*, 9-16. <http://groups.csail.mit.edu/sls/publications/2007/keynote.pdf>
- SENEFF, S., WANG, C. y CHAO, C. (2007), “Spoken Dialogue Systems for Language Learning”, *HLT-NAACL (Demonstrations)*, 13-14. Disponible en: <http://people.csail.mit.edu/wangc/papers/hlt2007-demo.pdf>
- SJOELANDER, K., BESKOW, J., GUSTAFSON, J., LEWIN, E., CARLSON, R. y GRANSTRÖM, B. (1999), “Web-Based Educational Tools For Speech Technology”, *Proceedings of the MATISSE Workshop*, 141-144. University College, London. Disponible en: http://www.speech.kth.se/~jocke/publications/icslp98_web.html
- STOCKWELL, G. (2007), “A review of technology choice for teaching language skills and areas in the CALL literature”, *ReCALL*, 19(2), 105-120. Disponible en: http://www.f.waseda.jp/gstock/Stockwell_ReCALL_2007.pdf
- STRIK, H., A. NERI, y C. CUCCHIARINI (2008), “Speech Technology for Language Tutoring”, *Proceedings of the Language and Speech Technology Conference LangTech 2008, Rome, Italy, February 28-29*. 73-76. Disponible en: <http://lands.let.ru.nl/~strik/publications/a142-ASR-CALL-Langtech08.pdf>
- STRIK, H., TRUONG, K., de WET, F. y CUCCHIARINI, C. (2009), “Comparing different approaches for automatic pronunciation error detection”. *Speech Communication*, 51(10), 845-852. Disponible en: <http://lands.let.ru.nl/~strik/2-div/KUL-LEA-0809/PED-Strik-EtAl-v02.pdf>
- SUNDSTROM, A. (1998), “Automatic Prosody Modification as a Means for Foreign Language Pronunciation Training”, *Proc. of STiLL - Speech Technology in Language Learning. May 25-27*. 49-52. Documento disponible en: http://www.speech.kth.se/ctt/publications/papers/STiLL98_49.pdf

- TAYLOR, R. P. (1999), "CALICO Software Review: *Accent Coach. English Pronunciation Trainer*", *Calico Software Reviews*, 7/99. Disponible en: <https://calico.org/p-48-Accent%20Coach.html>
- TOLEDO, G. (2005), "Uso del *Speech Analyzer* para la enseñanza de la ortofonía, la fonética y la fonología españolas", *Revista de Filología de la Universidad de La Laguna*, 23, 293-304.
- VOGEL, I., HESTVIK, A., BUNNELL, H. T. y SPINU, L. (2009), "Perception of English Compound vs. Phrasal Stress: Natural vs. Synthetic Speech", *Proceedings of INTERSPEECH 2009*, 1699-1702. Documento disponible en: http://hestvik-lab.cogsci.udel.edu/w/images/a/a8/Vogel_Hestvik_Bunnell_Spinu_%282009%29.pdf
- WANG, H., WAPLE, C. J. y KAWAHARA, T. (2009), "Computer Assisted Language Learning system based on dynamic question generation and error prediction for automatic speech recognition", *Speech Communication*, 51 (10), 995-1005.
- WAPLE, C. J., WANG, H., KAWAHARA, T., TSUBOTA, Y. y DANTSUJI, M. (2007), "Evaluating and Optimizing Japanese Tutor System Featuring Dynamic Question Generation and Interactive Guidance", *Proceedings of INTERSPEECH 2007*, 2177-2180. <http://www.ar.media.kyoto-u.ac.jp/lab/bib/intl/WAP-EUROSP07.pdf>
- WARD, M. (2006), "Using Software Design Methods in CALL", *Computer Assisted Language Learning*, vol. 19(2&3), 129-147.
- WARREN, P., ELGORT, I. y CRABBE, D. (2009), "Comprehensibility and prosody ratings for pronunciation software development", *Language Learning & Technology*, 13 (3), 87-102. Accesible en: <http://llt.msu.edu/vol13num3/warrenelgortcrabbe.pdf>
- WATERS, R. C. (1995), "The Audio Interactive Tutor", *Computer Assisted Language Learning*, 8(4), 325-354.
- WET, F., VAN DER WALT, C. y NIESLER, T. R. (2009), "Automatic assessment of oral language proficiency and listening comprehension", *Speech Communication*, 51 (10), 864-874.
- WIK, P., HJALMARSSON, A. y BRUSK, J. (2007), "DEAL: A Serious Game for CALL Practicing Conversational Skills in the Trade Domain", *Proceedings of SLaTE-Workshop on Speech and Language Technology in Education (SLaTE2007)*, 88-91. Documento disponible en: www.speech.kth.se/prod/publications/files/3111.pdf
- WIK, P., y HJALMARSSON, A. (2009), "Embodied conversational agents in computer assisted language learning", *Speech Communication*, 51 (10), 1024-1037. Documento disponible en: <http://www.speech.kth.se/prod/publications/files/3350.pdf>
- WITT, S. (1999), *Use of Speech Recognition in Computer Assisted Language Learning*. Tesis doctoral. Universidad de Cambridge.
- WITT, S.M., y YOUNG, S. J. (2000), "Phone-level pronunciation scoring and assessment for interactive language learning", *Speech Communication*, 30, 95-108.
- XIAO, B., GIRAND, C., y OVIATT, S. (2002), "Multimodal Integration Patterns in Children", *Proceedings of the 7th International Conference on Spoken Language Processing September 16-20, 2002. Denver, Colorado, USA*, 629-632. Disponible en: http://reference.kfupm.edu.sa/content/m/u/multimodal_integration_patterns_in_child_1828409.pdf
- ZECHNER, K., HIGGINS, D., XIA, X. y WILLIAMSON, D. M. (2009), "Automatic scoring of non-native spontaneous speech in tests of spoken English", *Speech Communication*, 51 (10), 883-895. http://www.mkzechner.net/SR_SpeComm09.pdf
- ZECHNER, K., HIGGINS, D., LAWLESS, R., FUTAGI, Y., OHLS, S. e IVANOV, G. (2009), "Adapting the Acoustic Model of a Speech Recognizer for Varied Proficiency Non-Native Spontaneous Speech Using Read Speech with Language-Specific Pronunciation

Difficulty”, *Proceedings of INTERSPEECH 2009*, 604-607. Disponible en: http://mkzechner.net/rdspAMadapt_interspeech09.pdf

ZECHNER, K., SABATINI, J. y CHEN, L. (2009), “Automatic Scoring of Children's Read-Aloud Text Passages and Word Lists”, *Proc. of the 4th Workshop on Innovative Use of NLP for Building Educational Applications, NAACL-HLT 2009 Workshops*, 1018. Disponible en: www.cs.rochester.edu/~tetreaul/bea4/Zechner-BEA4.pdf

ZINOVJEVA, N. (2005), “Use of Speech Technology in Learning to Speak a Foreign Language”, Term paper, Speech Technology, Autumn 2005. Documento disponible en: www.speech.kth.se/~rolf/NGSLT/gslt_papers_2005/Natalia2005.pdf

PÁGINAS DE INTERNET

- 1) Alelo. www.alelo.com
- 2) CandleTalk. http://candle.fl.nthu.edu.tw/newcandle/Home_E.asp
- 3) Carnegie Speech (<http://carnegiespeech.com/>) y NativeAccent (<http://www.carnegiespeech.com/products/nativeaccent.php>)
- 4) Chengo Chinese (2004) www.elanguage.cn
- 5) The Colorado Literacy Tutor. www.colit.org
- 6) DEAL - Role-playing and Dialogue System for Second Language Learners: <http://www.speech.kth.se/deal/>
- 7) Dispensa, M., L. Waite, D. Siebert, L. Wickelhaus, C. E. Piñeros, y C. Moon (2001) *Fonética: Los sonidos del español*, University of Iowa. <http://www.uiowa.edu/~acadtech/phonetics/spanish/frameset.html>
- 8) EduSpeak. www.eduspeak.com
- 9) Proyecto Fluency (Carnegie Mellon University) www.lti.cs.cmu.edu/Research/Fluency/
- 10) Jones, Chris. *The Spanish Oral Language Archive* <http://ml.hss.cmu.edu/mlrc/ola/spanish/index.html>
- 11) Llisterri, J. “Computer-Assisted Pronunciation Teaching Bibliography” http://liceu.uab.es/~joaquim/applied_linguistics/L2_phonetics/CALL_Pron_Bib.html
- 12) Llisterri, J. “La enseñanza de la pronunciación y la corrección fonética asistidas por ordenador”. http://liceu.uab.es/~joaquim/applied_linguistics/L2_phonetics/EAO_Pron.html
- 13) Llisterri, J. “Speech analysis and transcription software” http://liceu.uab.cat/~joaquim/phonetics/fon_anal_acus/herram_anal_acus.html
- 14) Soliloquy Learning. www.scilearn.com
- 15) Speech and Hearing Institute. www.speechandhearing.net
- 16) Versant (test de la empresa Ordinate). www.ordinate.com/versant/versant.jsp
- 17) West Point Heroico Spanish Speech (Linguistic Data Consortium, Philadelphia). <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2006S37>

PROGRAMAS

- 18) BetterAccentTutor. www.betteraccent.com
- 19) COLEA: A Matlab Software Tool for Speech Analysis <http://www.utd.edu/~loizou/speech/colea.htm>
- 20) Computerized Speech Lab (CSL), Multi-Dimensional Voice Program (MDVP) y VisiPitch. Kay Pentax (Kay Elemetrics). www.kayelemetrics.com
- 21) CSLU Toolkit for Spoken Dialogue Systems. <http://cslu.cse.ogi.edu/toolkit/index.html>
- 22) Dragon Naturally Speaking (Nuance). www.nuance.com/naturallyspeaking/
- 23) EyeSpeak. www.eyespeakenglish.com

- 24) The Festival Speech Synthesis System. www.cstr.ed.ac.uk/projects/festival/
- 25) Intelligent Tutor (DynEd). www.dyned.com/products/inteltutor.shtml
- 26) Project LISTEN (Literacy Innovation that Speech Technology ENables). A Reading Tutor that Listens. Universidad Carnegie Mellon. www.cs.cmu.edu/~listen/
- 27) My English Tutor (MyET). www.myet.com
- 28) Praat. Paul Boersma y David Weenink, Universidad de Amsterdam. www.fon.hum.uva.nl/praat/
- 29) Rossetta Stone. www.rosettastone.com
- 30) Saybot. www.saybot.com
- 31) SNACK Sound Toolkit. KTH (Royal Institute of Technology). www.speech.kth.se/snack/
- 32) SPACE (SPeech Algorithms for Clinical and Educational applications) <http://www.esat.kuleuven.be/psi/spraak/projects/index.php?proj=SPACE>
- 33) Speech Analyzer. Summer Institute of Linguistics. www.sil.org/computing/sa/index.htm
- 34) Speech Filing System (SFS) 4. www.phon.ucl.ac.uk/resource/sfs/
- 35) Speech Technology Applications Toolkit - STAPTK. Universidad de Sheffield <http://www.dcs.shef.ac.uk/spandh/projects/staptk/>
- 36) Tell me more y Talk to me. Auralog. <http://es.tellmemore.com/>
- 37) VILLE - The Virtual Language Tutor: <http://www.speech.kth.se/ville/>
- 38) WASP: Waveforms Annotations Spectrograms & Pitch. www.phon.ucl.ac.uk/resource/sfs/wasp.htm
- 39) Watch-me!-Read (IBM). www.ibm.com/ibm/ibmgives/grant/education/programs/reinventing/watch.shtml
- 40) WaveSurfer. www.speech.kth.se/wavesurfer/
- 41) WinPitch. www.winpitch.com
- 42) WinSnoori. www.loria.fr/~laprie/WinSnoori/