

LLISTERRI, J. (2007) "El papel de la fonética en las tecnologías del habla", in *Actas do 3º Congreso Internacional de Fonética Experimental*. Santiago de Compostela, 24-26 de outubro de 2005. Santiago de Compostela: Xunta de Galicia. pp. 23-37. ISBN: 978-84-453-4451-4

[http://liceu.uab.es/~joaquim/publicacions/Llisterri\\_05\\_Fonetica\\_Tecnologias\\_Habla.pdf](http://liceu.uab.es/~joaquim/publicacions/Llisterri_05_Fonetica_Tecnologias_Habla.pdf)

# EL PAPEL DE LA FONÉTICA EN LAS TECNOLOGÍAS DEL HABLA

JOAQUIM LLISTERRI

*Departamento de Filología Española, Universitat Autònoma de Barcelona*

"Can we leave it to the computer to learn about speech or shall we insist on developing our own insights in the many dimensions of the speech code?" (Fant 1983: 17)

"Computing power can not substitute crucial knowledge" (Fant 2004: 11)

## 1. LA RELACIÓN ENTRE LA FONÉTICA Y LAS TECNOLOGÍAS DEL HABLA

Cuando en 1983 Gunnar Fant se dirigía a los participantes en el décimo Congreso Internacional de Ciencias Fonéticas, planteaba una pregunta, seguramente retórica, a la que daba respuesta en 2004, afirmando que, frente al dilema entre las aproximaciones al estudio del habla basadas en el conocimiento y las que se fundamentan en el uso de técnicas informáticas, la potencia de cálculo no puede sustituir al conocimiento (Fant 1983, 2004). Ésta es, en esencia, una de las cuestiones centrales cuando se aborda la relación entre la fonética y las tecnologías del habla, al igual que sucede al considerar el papel de otras disciplinas lingüísticas en el procesamiento del lenguaje natural (Rodríguez 2004).

Si bien en sus inicios las tecnologías del habla pretendían incorporar varios niveles de conocimiento fonético, parece existir en la actualidad un cierto consenso para reconocer que los últimos años no han sido precisamente los más fértiles en lo que se refiere al uso de información fonética en la síntesis y en el reconocimiento; así lo señalan, entre otros, Strik (2005: 168) - "...in the last decades we have witnessed a decrease in the amount of phonetic knowledge used in ASR and TTS" - y Barry *et al* (2005: 1) - "...the linguistic approach soon lost terrain, in recognition applications at least, to (nonlinguistically oriented) engineers who were less concerned with formal linguistic insights, treating the signal as a pattern just like any other, and this with outstanding success" -, esgrimiendo las razones que discutiremos más adelante.

Otro de los hechos que más llaman la atención cuando se analiza la relación entre las dos disciplinas que nos ocupan es el contraste entre las constantes referencias a la necesidad de colaboración entre fonetistas y tecnólogos y la situación que se desprende de las publicaciones: si éstas reflejan acertadamente la realidad, no parece que, al menos en nuestro contexto más inmediato, sobren motivos para el optimismo en lo que a la interdisciplinariedad se refiere.

A modo de ejemplo, un estudio de las comunicaciones presentadas en los congresos anuales de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN) celebrados entre 2000 y 2005 revela que, de 71 contribuciones que podemos considerar relacionadas con las tecnologías del habla, únicamente tres (4%) aparecen firmadas conjuntamente por autores vinculados a departamentos tecnológicos y a departamentos de filología; si añadimos a ello que las tres proceden del mismo equipo (González y García 2000; González *et al.* 2001; Rodríguez *et al.* 2002), podemos darnos cuenta de que lo que en teoría se considera deseable es, en España, prácticamente inexistente<sup>1</sup>. Por ello no sorprende, tampoco, que los cuatro ponentes en el taller sobre “Tecnologías del habla: pasado, presente y futuro. Particularización sobre tecnología del español” (*sic*), que tuvo lugar en ocasión del congreso de la SEPLN el 2003, pertenecieran a Escuelas Técnicas Superiores de Ingenieros de Telecomunicación.

Es también sintomático que en los congresos de Fonética Experimental celebrados en 1999, 2001 y 2005, únicamente 9 (8,5%) del total de las 106 comunicaciones presentadas se centraran en las tecnologías del habla; si computamos solamente aquellas cuyos autores están vinculados a una universidad española, el porcentaje baja al 4,7% (5 comunicaciones). Otro dato que no deja de ser significativo es la presencia de 14 autores de comunicaciones que, entre 1999 y 2005, han participado tanto en los congresos de la SEPLN como en los de Fonética Experimental; esto representa aproximadamente un 10% de los investigadores que presentaron contribuciones a la SEPLN y un 13% de los que lo hicieron a los congresos de Fonética Experimental; sin embargo, entre ellos se cuentan diez investigadores asociados al mismo proyecto (Battaner *et al.* 2005a, b) y dos que firman conjuntamente sus trabajos (Escudero y Cardeñoso 2001, 2002) con lo que, en realidad, el número de equipos presentes en ambos congresos se reduce a tres<sup>2</sup>. La división entre la comunidad dedicada a la fonética y la centrada en las tecnologías del habla parece pues, en lo que se refiere a nuestro país, claramente marcada.

En las páginas que siguen, presentaremos, en primer lugar, algunos de los campos en los que las tecnologías del habla pueden beneficiarse del conocimiento fonético y exploraremos, a continuación, las razones que obstaculizan la integración entre ambas disciplinas. Esbozaremos, para concluir, algunas propuestas orientadas al futuro.

- 
- 1 De este dato, sin embargo, no puede desprenderse que no existan otros equipos o proyectos en los que colaboran lingüistas y expertos en informática o en ingeniería de telecomunicación; más bien cabe pensar que, en algunos casos, la participación de los lingüistas no siempre se refleja en la autoría de las publicaciones (Hernando *et al.* 2002).
  - 2 Los dos anteriormente mencionados y Pérez *et al.* (2002)

## 2. LA INCORPORACIÓN DE CONOCIMIENTOS FONÉTICOS A LAS TECNOLOGÍAS DEL HABLA

Las ventajas y los problemas derivados de incorporar información fonética a los sistemas desarrollados en el marco de las tecnologías del habla y la relación entre ambas disciplinas son cuestiones que aparecen de modo recurrente en diversos trabajos, tanto en reflexiones generales (Fant 1983, 1989, 2004, 2005; Moore 1995; Rossi 1996; Pols 1999, 2001; Greenberg 2001a, b, 2005; Öhman 2001; Ainsworth 2005; Barry *et al.* 2005; Strik 2005), como en aportaciones centradas en la síntesis (Carlson y Granström 1991; Fant 1991; Huckvale 2002; van Santen 2005), el reconocimiento (Greenberg 1998; Dusan y Rabiner 2005) o los sistemas de diálogo (Shriberg 2005).

Como resultado de la colaboración en diversos proyectos en el campo de las tecnologías del habla, hemos intentado describir, en varias publicaciones de nuestro grupo, el papel del fonetista en este contexto (Aguilar *et al.* 1997; Garrido *et al.* 2000; Llisterri 1990, 2003a, b; Llisterri *et al.* 2003a, b, 2004). Esencialmente, y coincidiendo también con parte de las tareas que en Acero (1995) se describen como propias de un especialista en fonética o en lingüística –siempre, naturalmente, en estrecha colaboración con el resto de los profesionales que intervienen en el desarrollo de un sistema–, podríamos mencionar los ámbitos que se detallan a continuación.

### (1) Conversión de texto en habla

- Reglas para el procesamiento previo del texto que contemplen la expansión de los signos de puntuación sin valor lingüístico, las expresiones numéricas, las siglas y las abreviaturas.
- Reglas para el procesamiento morfológico y sintáctico en los casos en que el conversor contempla un análisis lingüístico automático del texto de entrada o supervisión manual del resultado de un proceso de etiquetado automático.
- Reglas de transcripción fonética automática, que establecen la correspondencia entre grafías y alófonos, la silabación y la acentuación, complementadas por diccionarios de pronunciación para el tratamiento de las excepciones.
- Modelos de duración segmental que consideren los diversos factores que influyen en la duración y basados en datos procedentes de corpus representativos.
- Modelos de intensidad segmental que, igualmente, consideren los factores que inciden en la intensidad y procedan de corpus representativos.
- Modelos de asignación de pausas que contemplen tanto las marcadas mediante signos de puntuación como las no marcadas y que establezcan, además, diferencias de duración entre los distintos tipos de pausas.
- Modelos de entonación que permitan generar una curva melódica natural, teniendo en cuenta factores fonéticos, sintácticos, semánticos y pragmáticos.

- Establecimiento del inventario de fonemas y alófonos de la lengua para la constitución del diccionario de unidades de síntesis.
  - Diseño del corpus de unidades de síntesis teniendo en cuenta las restricciones fonotácticas de la lengua y la frecuencia de aparición de las unidades.
  - Selección del locutor para la grabación del corpus de síntesis.
  - Supervisión de la grabación del corpus de síntesis para asegurar una realización adecuada de los elementos segmentales y suprasegmentales, tanto en los sistemas de síntesis por concatenación como en los basados en selección de unidades.
  - Segmentación (o supervisión de una segmentación semiautomática) del corpus de unidades de síntesis.
  - Evaluación objetiva de los distintos módulos del conversor, en un proceso iterativo que permita la corrección de errores.
  - Diseño de pruebas de evaluación subjetiva tanto de la inteligibilidad como de la naturalidad del resultado de la conversión de texto en habla.
- (2) Reconocimiento del habla
- Definición del inventario de fonemas y alófonos de la lengua para determinar las unidades del sistema de reconocimiento.
  - Diseño del corpus de entrenamiento teniendo en cuenta el inventario de unidades previamente definido y las restricciones sobre su aparición.
  - Selección de la muestra de población para la grabación del corpus de entrenamiento, considerando factores de variación individual, geográfica, social y de registro.
  - Segmentación (o supervisión de una segmentación automática) del corpus de entrenamiento del reconocedor.
  - Realización o validación de los diccionarios de pronunciación, que incorporen las formas canónicas y las variantes encontradas en el corpus.
  - Análisis fonético acústico de corpus significativamente amplios para obtener información sobre los factores que condicionan la variabilidad segmental y suprasegmental, contemplando tanto el nivel fonético como otros niveles más altos del análisis lingüístico.
- (3) Sistemas de diálogo
- Transcripción, anotación y estudio de corpus de interacciones naturales entre personas para definir el dominio del sistema, diseñar posibles estrategias de gestión del diálogo y establecer los escenarios que se emplearán en el corpus de entrenamiento.
  - Selección de la muestra de población para la obtención del corpus de entrenamiento, teniendo en cuenta factores de variación individual, geográfica, social y de registro.
  - Transcripción, anotación y análisis de corpus de interacciones simuladas (obtenidas mediante el protocolo del Mago de Oz) para el entrenamiento del sistema.

- Diseño de estrategias de acceso, de salida y de confirmación adecuadas desde el punto de vista pragmático.
- Estudio de los fenómenos propios del habla espontánea para modelarlos adecuadamente en el sistema de diálogo.
- Análisis de la relación entre el nivel fonético y el nivel pragmático, especialmente en lo que se refiere a las manifestaciones prosódicas de los actos que habla y a los correlatos acústicos de las emociones.
- Verificación del grado de corrección y adecuación lingüística del diálogo.

Cabe recordar, naturalmente, que parte del conocimiento que se adquiere mediante las tareas que acabamos de enumerar puede obtenerse a través de técnicas de aprendizaje automático aplicadas a grandes corpus de habla. Como bien apunta Greenberg (2005: 111) "Speech technology can proudly point to its apparent success with speech recognition and concatenative synthesis in defense of its machine-learning-centric approach"; sin embargo, precisa más adelante que "...imperfect science is capable of providing an effective foundation for technology - as long as the demands of the market are not exceedingly stringent or profound". Por tanto, existen no sólo argumentos teóricos para insistir en la necesidad de incorporar conocimiento, sino también consideraciones prácticas para intentar superar las limitaciones con las que actualmente nos encontramos.

### 3. LOS OBSTÁCULOS PARA LA INTEGRACIÓN DEL CONOCIMIENTO FONÉTICO EN LAS TECNOLOGÍAS DEL HABLA

Pese a que se trata, sin duda, de una cuestión compleja, podrían distinguirse dos clases de obstáculos que surgen en el momento de incorporar conocimiento fonético a las tecnologías del habla: por una parte, los derivados de la propia naturaleza de la información fonética disponible y, por otra, los que obedecen a las distintas tradiciones académicas o al contexto en el que se lleva a cabo la investigación.

Strik (2005: 177) formula el problema de un modo muy directo: "...phonetics does not provide ready-made quantitative models that can be plugged directly into a system", lo que nos lleva al primer tipo de dificultades al que hacíamos referencia. En efecto, en muchas ocasiones la información fonética sobre un determinado fenómeno no es suficientemente detallada, no está cuantificada o no se expresa con el formalismo adecuado para un entorno computacional. Si, por ejemplo, nos ceñimos a la descripción del español observamos que, pese a la proliferación de estudios acústicos, estamos lejos de disponer de los datos que se requieren en muchas de las aplicaciones de las tecnologías del habla (Gil y Llisterrri 2004). Siguiendo la distinción que establecen Barry *et al.* (2005: 10) entre hacerse una idea de un fenómeno, obtener datos cuantitativos sobre el mismo y presentarlos en un formato que permita incorporarlos a una aplicación, podríamos decir que nos encontramos, tal vez, todavía en el primer estadio.

Otra de las limitaciones de los estudios fonéticos que habitualmente se señalan es que los datos provienen, en su mayor parte, de trabajos de laboratorio, obtenidos me-

dian­te un diseño experimental en el que se establece un control de las variables y, a menudo, en condiciones muy distintas a las del uso real de las tecnologías del habla. Por tal motivo se ha propuesto que el fonetista podría emplear con provecho los grandes corpus recogidos para el entrenamiento de sistemas de reconocimiento u otras bases de datos que respondan a situaciones más realistas. Greenberg (2005), por citar un ejemplo, ofrece una muestra de este tipo de estudio, que responde a los planteamientos de Barry *et al.* (2005: 11): “The greater access phonetically trained researchers have to the databases and tools used in mainline technology applications, the more likely is that quantitative answers to phonetic questions can be presented in a way which can be useful for speech technology applications”. Sin embargo, es conocido el problema de la disponibilidad de los corpus, tanto por motivos económicos como por los derivados de políticas científicas que favorecen el desarrollo de recursos que no se reutilizan (Llisterri 2004b). Para el español existen las bases de datos distribuidas por ELDA (*Evaluation and Language resources Distribution Agency*) y por el LDC (*Linguistic Data Consortium*) además de las creadas en el marco de diversos proyectos que han gozado de financiación pública (Llisterri *et al.* 2005) pero aun así, no siempre es fácil localizar y conseguir acceder a estos recursos.

Un tercer aspecto inherente al propio carácter de la fonética como disciplina científica es la proliferación de modelos que, en el terreno que nos ocupa, se hace especialmente patente en el campo de la prosodia. Desde la perspectiva del desarrollo de las tecnologías del habla, “There is too much emphasis on theoretical concepts and on the discussion of which one is better suited for the description of a special language or of languages in general.” (Batliner y Möbius 2005: 25), lo que a menudo introduce, en opinión de los autores citados, niveles de abstracción que tal vez no son necesarios en el contexto de las aplicaciones, de modo que se pierde parcialmente la distinción entre el conocimiento básico y el que, de alguna manera, está mediatizado por un modelo<sup>3</sup>. Quizás por este motivo, en las comunicaciones presentadas en los congresos de la SEPLN a los que hacíamos referencia anteriormente encontramos trabajos sobre modelos prosódicos llevados a cabo por investigadores vinculados a departamentos de informática (Escudero y Cardeñoso 2001) o de ingeniería de telecomunicación (Fernández y Rodríguez 2000; Navas *et al.* 2002; Agüero y Bonafonte 2003) sin que en ellos intervengan especialistas en fonética y en los que, como es de esperar, se emplean técnicas de aprendizaje automático basadas en corpus prescindiendo de aportaciones procedentes del campo de la fonética.

La segunda categoría de obstáculos es de naturaleza académica —o, en ocasiones, el resultado de una determina política científica— y responde a la tradicional separación entre estudios humanísticos y tecnológicos. Una consecuencia importante es la falta de formación interdisciplinar que dificulta la formación de equipos mixtos y, a la vez, no facilita la incorporación de lingüistas o filólogos a un ámbito que podría ofrecer salidas profesionales. Compárese, por ejemplo, el perfil que define Acero (1995: 175) con los

---

3 A modo de ejemplo, Batliner y Möbius (2005: 26) señalan que “Phonological systems like the ToBI approach only introduce a quantisation error: the whole variety of F0 values available in acoustics is reduced to a mere binary opposition L vs. H, and to some few additional, diacritic distinctions”.

planes de estudio vigentes -y, probablemente, futuros- en las facultades de letras españolas: "... a successful phonetician working on a spoken language system will need some knowledge of computers, algorithms, statistics and signal processing [...] Also desired is proficiency with common computing environments such as Windows, UNIX and Macintosh, text editors, and speech analysis packages". Insiste el actual responsable de la investigación en el área de tecnologías del habla de Microsoft que el fonetista no debe ser un experto en todos estos campos, pero sí debe tener un conocimiento básico suficiente que le permita incorporarse a un equipo. Al no darse estas condiciones, "One of the reasons why there are not more linguists working in building Spoken Language Systems is that in many cases, lack of training in these other disciplines prevent them to be as effective in the team as an engineer or a programmer" (Acero 1995: 175).

Cierto es que algunas universidades ofrecen cursos de postgrado en los que se pretende proporcionar una formación integral como la que acabamos de describir. La oferta actual en España no parece ser especialmente amplia; por ejemplo, el máster en Lingüística y Aplicaciones Tecnológicas que programa la Universitat Pompeu Fabra, ofrece únicamente 3 créditos de fonética y fonología y 3 créditos sobre tratamiento del habla, mientras que en el *European Masters in Language and Speech* que se imparte en la Universitat Politècnica de Catalunya es obligatoria una asignatura sobre procesamiento del habla (6 créditos) y pueden cursarse, como asignaturas de libre elección, una introducción a la fonética y a la fonología (6 créditos) y una materia sobre percepción del lenguaje (3 créditos), ambas ofrecidas por profesores de la Universitat de Barcelona. Por el momento, pues, la formación de postgrado en fonética y en tecnologías del habla no tiene un espacio propio en nuestro país, pese a la existencia de grupos consolidados y productivos en ambos campos (Llisterrí 2004<sup>a</sup>; Rubio y Hernáez 2005).

De un modo acorde a lo que sucede en la formación, la investigación interdisciplinar en España no parece estar especialmente favorecida, al menos en la práctica cotidiana. Al margen de los problemas burocráticos que suele plantear la colaboración entre departamentos, los propios mecanismos mediante los que se evalúa la investigación financiada con fondos públicos hacen posible que proyectos que en principio requerirían, por su temática, la colaboración de expertos procedentes de campos diversos, se lleven a cabo entre equipos de la misma especialidad (Llisterrí 2004a)

Todos los factores considerados llevan a la separación entre las dos "culturas", la lingüística y la tecnológica, que intentábamos mostrar con algunos datos en el apartado 1. El resultado lo describe claramente van Santen (2005: 149) "...the phonetics community has not focused on questions most relevant for speech technology while the speech technology community has not developed algorithms and data structures that are optimally receptive for the incorporation of phonetic knowledge".

#### 4. ALGUNAS PERSPECTIVAS DE FUTURO

Como mencionábamos anteriormente, algunas razones para abordar de nuevo la incorporación de conocimiento fonético son puramente prácticas, y responden a las limi-

taciones que se encuentran en los sistemas de reconocimiento de habla espontánea y en la falta de flexibilidad propia de la síntesis por selección de unidades.

En cuanto al reconocimiento, es preciso recordar, como hace Ainsworth (2005), que se obtienen buenos resultados en función de un corpus de entrenamiento de gran tamaño y de diccionarios de pronunciación que incorporan la variación documentada en el corpus, pero que la adaptación a nuevas situaciones no siempre se realiza con éxito. En la síntesis por selección de unidades, pese a su elevada naturalidad en dominios restringidos, la dependencia entre la fuente y el filtro reduce las posibilidades expresivas y prácticamente obliga a disponer de un nuevo corpus cada vez que se requiere una nueva voz o una nueva aplicación (Fant 2004).

Existen también áreas de investigación emergentes y cada vez más populares como la síntesis y el reconocimiento de las emociones, que probablemente no se están abordando de un modo completamente adecuado por falta de conocimiento lingüístico sobre la interacción comunicativa humana. Dada la dificultad de recopilar un corpus realista, se suele recurrir a actores y a un repertorio de emociones básicas que no parecen ser las que aparecen habitualmente en el habla, tal como recoge Campbell (2004: X): "... there was very little expression of the big-six emotions. Instead, there were a great variety of different speaking styles that changed as a consequence of listener and subject differences". Partiendo de estos datos relativamente artificiales, se aplican algoritmos de aprendizaje automático hasta encontrar el que es capaz de obtener mejores resultados en el reconocimiento. Es interesante destacar que mientras que las contribuciones sobre emociones en los últimos congresos de la Sociedad Española para el Procesamiento del Lenguaje Natural aparecen firmada por autores procedentes de departamentos de ingeniería de telecomunicación o de informática (Adell *et al.* 2005; Francisco *et al.* 2005; Luengo *et al.* 2005), en los tres congresos de Fonética Experimental no hemos podido encontrar ninguna comunicación que contenga la palabra *emoción* en el título.

Además de estos nuevos desarrollos, se plantean también en la actualidad una serie de tareas relacionadas con la mejora de las tecnologías existentes. Así, Batliner y Möbius (2005: 38-39) proponen una línea de investigación en prosodia que pueda resultar útil tanto para la síntesis como para el reconocimiento, basada en la realización de inventarios de funciones lingüísticas, paralingüísticas, léxicas y sintáctico/semánticas de la prosodia, en el diseño de un sistema de anotación motivado por consideraciones prácticas y orientado a la forma más que a la función, en el establecimiento de procedimientos para modelar rasgos prosódicos a partir de bases de datos que no representen necesariamente a un hablante específico y, finalmente, en el reconocimiento de que los parámetros prosódicos no pueden modelarse independientemente, ya que en el habla se producen de forma conjunta.

Por su parte, van Santen (2005: 162-163) establece un conjunto de tareas relevantes para la conversión de texto en habla: estudio de la percepción de las discontinuidades espectrales propias de la concatenación; percepción de las discontinuidades en los contornos melódicos; análisis de los aspectos subsegmentales en la organización temporal de la producción del habla; modelado de la reducción vocálica; estudio de la variación inter e intralocutor; determinación de los umbrales diferenciales en la percepción de

curvas melódicas; estudio de la percepción de las emociones generadas mediante síntesis; posibilidad de diseñar un modelo alternativo a ToBI para la descripción fonológica de la entonación; y, coincidiendo plenamente con Batliner y Möbius, análisis y modelado multidimensional de la interacción entre rasgos prosódicos.

Podríamos tal vez tener argumentos para pensar que nos encontramos en una etapa de transición en lo que se refiere a la relación entre la fonética y las tecnologías del habla. Por una parte, existe una gran cantidad de conocimiento fonético útil -paradójicamente "in part hidden in text-to-speech programs" (Fant 2002: 10)-, aunque no se encuentre en el formato adecuado; también siguen siendo válidas las aproximaciones basadas en reglas<sup>4</sup> y, si se dieran las condiciones favorables, el fonetista podría tener a su disposición corpus con los que realizar estudios específicamente adaptados a las necesidades de las tecnologías del habla, sin por ello renunciar a los modelos empleados en fonética; esta es, en parte, la perspectiva que sugiere Ainsworth (2005: 17): "the way to integrate phonetic knowledge into speech technology is not by deriving the detailed acoustic structure of phones from sets of phonetic rules, but by basing both speech recognition and speech synthesis on more realistic models of speech production. The details are probably best derived from speech databases as at present", coincidiendo con Barry *et al.* (2005: 11) en lo que se refiere al uso adecuado de los recursos lingüísticos.

Es factible, por tanto, realizar los necesarios avances en el estudio del habla como código siguiendo los objetivos tradicionales de la fonética, como propone Fant (2002: 10), y pensar con Greenberg (2005: 129) que "Over the coming decades this tension [entre la fonética y las tecnologías del habla] is likely to dissolve into a collaborative relationship melding linguistic knowledge with machine-learning and statistical methods as a means of developing mature science and technology".

Desde un punto de vista más práctico, investigadores como van Santen (2005: 164) proponen medidas concretas: incorporar materias relacionadas con la fonética y la lingüística en la formación de los tecnólogos y cursos de matemáticas, informática y procesamiento de señales en la de los fonetistas; organizar postgrados especializados en tecnologías del habla o licenciaturas conjuntas entre departamentos de lingüística, de ingeniería de telecomunicación y de informática para propiciar la formación de expertos en lo que se podría denominar "fonética computacional" (Moore 1995); y, por último, organizar congresos en los que, contrariamente a lo que ahora sucede, se programen sesiones en las que participen simultáneamente especialistas con formación humanística y tecnólogos.

No cabe duda de que la fonética puede y debe jugar un papel relevante en el desarrollo de las tecnologías del habla; si bien es cierto que existen obstáculos nada desdeñables, algunos de ellos pueden superarse apropiándose de nuevos enfoques y problemas, estableciendo mecanismos de interacción y coordinación que favorezcan la discusión en foros conjuntos y, especialmente, recordando a menudo las palabras de Fant con las que encabezábamos este trabajo.

---

4 Véase como muestra, la afirmación de Fant (2004: 9): "From detailed acoustic phonetic studies of text reading during the last 15 years, we have now been able to develop quite efficient prosodic rules for text-to-speech synthesis".

## BIBLIOGRAFÍA<sup>5</sup>

- ACERO, A. (1995) "The role of phoneticians in speech technology", in BLOOTH-OFT, G.- HAZAN, V.- HUBER, D.- LLISTERRI, J. (Eds.) *European Studies in Phonetics and Speech Communication*. Utrecht: OTS Publications. pp. 170-175.  
<http://research.microsoft.com/srg/papers/1995-alexac-esca.pdf>
- ADELL, J.- BONAFONTE, A.- ESCUDERO, D. (2005) "Analysis of prosodic features: towards modelling of emotional and pragmatic attributes of speech", *Procesamiento del Lenguaje Natural* 35: 277-283.  
[http://gps-tsc.upc.es/veu/research/pubs/download/Ade\\_Ana\\_05.pdf](http://gps-tsc.upc.es/veu/research/pubs/download/Ade_Ana_05.pdf)
- AGÜERO, P.D.- BONAFONTE, A. (2003) "Phrase break prediction: a comparative study", *Procesamiento del Lenguaje Natural* 31: 107-114.  
<http://www.sepln.org/revistaSEPLN/revista/31/31-Pag107.pdf>
- AGUILAR, L.- GARRIDO, J.M.- LLISTERRI, J. (1997) "Incorporación de conocimientos fonéticos a las tecnologías del habla", in SERRA, E.- GALLARDO, B.- VEYRAT, M.- JORQUES, D.- ALCINA, A. (Eds.) *Panorama de la investigación lingüística a l'Estat Espanyol. Actes del I Congrés de Lingüística General. Volum III. Comunicacions: Fonètica i Fonologia. Semàntica i Pragmàtica*. València: Universitat de València. pp. 5-13.  
[http://liceu.uab.es/~joaquim/publicacions/valencia\\_94.html](http://liceu.uab.es/~joaquim/publicacions/valencia_94.html)
- AINSWORTH, W.A. (2005) "Can phonetic knowledge be used to improve the performance of speech recognisers and synthesisers?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 13-20.
- BARRY, W.J.- van DOMMELEN, W.- KOREMAN, J. (2005) "Phonetic knowledge in speech technology - and phonetic knowledge from speech technology?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 1-12.  
<http://www.coli.uni-saarland.de/~koreman/Publications/2005/Eurospeech2001.pdf>
- BATLINER, A.- MÖBIUS, B. (2005) "Prosodic models, automatic speech understanding, and speech synthesis: Towards the common ground?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 21-44.
- BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.J.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2005a)

---

<sup>5</sup> Las direcciones de Internet citadas en la bibliografía se han consultado en octubre de 2005.

- “VILE: Estudio acústico de la variación inter e intralocutor en español”, *Procesamiento del Lenguaje Natural* 35: 435-436.  
[http://liceu.uab.es/~joaquim/phonetics/VILE/VILE\\_SEPLN05.pdf](http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_SEPLN05.pdf)
- BATTANER, E.- CARBÓ, C.- GIL, J.- LLISTERRI, J.- MACHUCA, M.J.- MADRIGAL, N.- MARRERO, V.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2005) “VILE: Estudio acústico de la variación inter e intralocutor en español”, in *Actas del III Congreso de Fonética Experimental*. Universidade de Santiago de Compostela (en prensa).  
[http://liceu.uab.es/~joaquim/phonetics/VILE/VILE\\_IICFE05.pdf](http://liceu.uab.es/~joaquim/phonetics/VILE/VILE_IICFE05.pdf)
- CAMPBELL, N. (2004) “Getting to the Hearth of the Matter: Speech is more than just the Expression of Text or Language”, in *LREC 2004. Proceedings of the 4th International Conference on Language Resources and Evaluation*. Paris: ELRA, European Language Resources Association. Vol. 5, pp. VII-X.  
<http://feast.atr.jp/nick/pubs/lrec-keynote.pdf>
- CARLSON, R.- GRANSTRÖM, B. (1991) “Speech synthesis development and phonetic research - a personal introduction”, *Journal of Phonetics* 19, 1: 3-8.  
<http://www.speech.kth.se/~rolf/papers/wwjphonint.pdf>
- DUSAN, S.- RABINER, L.R. (2005) “On integrating insights from human speech perception into automatic speech recognition”, in *EUROSPEECH 2005 - INTERSPEECH 2005. Proceedings of the 9th European Conference on Speech Communication and Technology*. pp. 1233-1236.  
<http://www.caip.rutgers.edu/~sdusan/2177anav.pdf>
- ESCUADERO, D.- CARDEÑOSO, V. (2001) “Modelo cuantitativo de entonación del español”, *Procesamiento del Lenguaje Natural* 27: 233-240.  
<http://www.sepln.org/revistaSEPLN/revista/27/27-articulo27.pdf>
- ESCUADERO, D.- CARDEÑOSO, V. (2002) “Una experiencia en reconocimiento automático de tipos de unidades melódicas a partir de su perfil de entonación”, in *Actas del II Congreso de Fonética Experimental*. Sevilla: Laboratorio de Fonética, Facultad de Filología, Universidad de Sevilla. pp. 161-166.  
<http://WWW.infor.uva.es/~descuder/investig/pdfs/cfeII.pdf>
- FANT, G. (1983) “Phonetics and Speech Technology”, *Speech Transmission Laboratory - Quarterly Progress and Status Report* 2-3: 20-35; in *Proceedings of the 10th International Congress of Phonetic Sciences*. Dordrecht: Foris, 1984. pp. 13-24.  
[http://www.speech.kth.se/qpsr/pdf/1983/1983\\_24\\_2-3\\_020-035.pdf](http://www.speech.kth.se/qpsr/pdf/1983/1983_24_2-3_020-035.pdf)
- FANT, G. (1989) “Speech Research in Perspective”, *Speech, Music and Hearing - Quarterly Progress and Status Report* 30, 4:1-7; in *EUROSPEECH 1989. European Conference on Speech Communication and Technology*. Edinburgh: CEP Consultants Ltd. Vol 1, pp. 3-4.  
[http://www.speech.kth.se/qpsr/pdf/1989/1989\\_30\\_4\\_001-007.pdf](http://www.speech.kth.se/qpsr/pdf/1989/1989_30_4_001-007.pdf)

- FANT, G. (1991) "What can basic research contribute to speech synthesis?", *Journal of Phonetics* 19, 1: 75-90.
- FANT, G. (2004) "More than half a century in phonetics and speech research", in FANT, G. *Speech Acoustics and Phonetics*. Dordrecht: Kluwer. pp. 1-14.  
<http://www.speech.kth.se/~gunnar/halfcentury.pdf>
- FANT, G. (2005) "Historical notes", *Speech, Music and Hearing - Quarterly Progress and Status Report* 47: 9-19.  
<http://www.speech.kth.se/qpst/tmh/2005/05-47-009-019.pdf>
- FERNÁNDEZ SALGADO, X.- RODRÍGUEZ BANGA, E. (2000) "Proposición de un marco adecuado para el estudio de contornos de F0 para síntesis de voz", *Procesamiento del Lenguaje Natural* 24: 175-182.  
<http://www.sepln.org/revistaSEPLN/revista/26/fernandez-salgado.pdf>
- FRANCISCO, V.- GERVÁS, P.- HERVÁS, R. (2005) "Análisis y síntesis de la expresión emocional en cuentos leídos en voz alta", *Procesamiento del Lenguaje Natural* 35: 293-300.
- GARRIDO, J.M.- ORTÍN, I.- QUAZZA, S.- SALZA, P.L.- MANCINI, F. (2000) "Desarrollo de un módulo de asignación de parámetros prosódicos para la versión en español del sistema de conversión texto-habla ACTOR<sup>®</sup>", *Procesamiento del Lenguaje Natural* 24: 183-190.  
<http://www.sepln.org/revistaSEPLN/revista/26/garrido-alminana.pdf>
- GIL, J.- LLISTERRI, J. (2004) "Fonética y fonología del español en España (1978-2003)", *Lingüística Española Actual* 26, 2: 4-44.  
[http://liceu.uab.es/~joaquim/publicacions/Gil\\_Llisterrí\\_LEA\\_2003.pdf](http://liceu.uab.es/~joaquim/publicacions/Gil_Llisterrí_LEA_2003.pdf)
- GONZÁLEZ REI, B.- CARDENAL LÓPEZ, A.- DOCÍO FERNÁNDEZ, L.- GARCÍA MATEO, C. (2001) "Problemática de la recogida y anotación de una base de datos oral para el gallego", *Procesamiento del Lenguaje Natural* 27: 37-44.  
<http://www.sepln.org/revistaSEPLN/revista/27/27-articulo4.pdf>
- GONZÁLEZ REI, B.- GARCÍA MATEO, C. (2000) "Diseño de una base de datos tipo SpeechDat para el idioma gallego", *Procesamiento del Lenguaje Natural* 26: 197-204.  
<http://www.sepln.org/revistaSEPLN/revista/26/gonzalez-rei.pdf>
- GREENBERG, S. (1998) "Recognition in a new key - Towards a science of spoken language", in *ICASSP 1998. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 1401-1405.  
[http://www.icsi.berkeley.edu/~steveng/PDF/Recognition\\_in\\_a\\_New\\_Key.pdf](http://www.icsi.berkeley.edu/~steveng/PDF/Recognition_in_a_New_Key.pdf)
- GREENBERG, S. (2001a) "From here tu utility - Melding phonetic insight with speech technology", in *Eurospeech 2001. Proceedings of the 7th European Conference on Speech Communication and Technology*. Vol 4. pp. 2485-2488.  
<http://www.icsi.berkeley.edu/ftp/global/pub/speech/papers/euro01-utility.pdf>

- GREENBERG, S. (2001b) "Whither Speech Technology? - A Twenty-First Century Perspective", in *Eurospeech 2001. Proceedings of the 7th European Conference on Speech Communication and Technology*. Vol 1, pp. 3-6.  
<http://www.icsi.berkeley.edu/ftp/global/pub/speech/papers/euro01-whither.pdf>
- GREENBERG, S. (2005) "From here to utility", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer pp. 107-132.  
[http://www.icsi.berkeley.edu/%7Esteveng/PDF/Phonetic\\_Insight.pdf](http://www.icsi.berkeley.edu/%7Esteveng/PDF/Phonetic_Insight.pdf)
- HERNANDO, J.- PADRELL, J.- RODRÍGUEZ, H. (2002) "Sistema de información meteorológica automática por teléfono ATTEMPS", *Procesamiento del Lenguaje Natural* 29: 311-312.  
<http://www.sepln.org/revistaSEPLN/revista/29/29-Pag311.pdf>
- HUCKVALE, M. (2002) "Speech Synthesis, Speech Simulation and Speech Science", in *ICSLP 2002 - INTERSPEECH 2002. Proceedings of the 7th International Conference on Spoken Language Processing*. pp. 1261-1264.  
<http://www.phon.ucl.ac.uk/home/mark/papers/icslp02synth.pdf>
- LLISTERRI, J. (1990) "Algunes reflexions sobre el paper de la lingüística en la tecnologia de la veu", *Límits. Revista d'Assaig i d'Informació sobre les Ciències del Llenguatge* 8: 19-32.  
[http://liceu.uab.es/~joaquim/publicacions/llisterri\\_88.html](http://liceu.uab.es/~joaquim/publicacions/llisterri_88.html)
- LLISTERRI, J. (2003a) "Las tecnologías del habla: Entre la ingeniería y la lingüística", in *Actas del Congreso Internacional "La Ciencia ante el Público. Cultura humanística y desarrollo científico y tecnológico"*. Salamanca: Instituto Universitario de Estudios de la Ciencia y la Tecnología. Edición en CD-ROM. pp. 44-67.  
[http://liceu.uab.es/~joaquim/publicacions/TecnolHab\\_Salamanca\\_02.pdf](http://liceu.uab.es/~joaquim/publicacions/TecnolHab_Salamanca_02.pdf)
- LLISTERRI, J. (2003b) "Lingüística y tecnologías del lenguaje", *Lynx. Panorámica de Estudios Lingüísticos* 2: 9-71.  
[http://liceu.uab.es/~joaquim/publicacions/TecnoLing\\_Lynx02.pdf](http://liceu.uab.es/~joaquim/publicacions/TecnoLing_Lynx02.pdf)
- LLISTERRI, J. (2004a) "Las tecnologías del habla para el español", in SEQUERA, R. (Ed.) *Ciencia, tecnología y lengua española: la terminología científica en español*. Madrid: Fundación Española para la Ciencia y la Tecnología. pp. 123-141.  
[http://liceu.uab.es/~joaquim/publicacions/TecnolHablaEsp\\_FECyT03.pdf](http://liceu.uab.es/~joaquim/publicacions/TecnolHablaEsp_FECyT03.pdf)
- LLISTERRI, J. (2004b) "Las tecnologías lingüísticas en España", in *El español en el mundo. Anuario del Instituto Cervantes 2004*. Madrid: Instituto Cervantes – Círculo de Lectores – Plaza & Janés. pp. 229-251.  
[http://cvc.cervantes.es/obref/anuario/anuario\\_04/llisterri/default.htm](http://cvc.cervantes.es/obref/anuario/anuario_04/llisterri/default.htm)
- LLISTERRI, J.- CARBÓ, C.- MACHUCA, M. J.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2003a) "El papel de la lingüística en el desarrollo de las tecnologías del habla",

- in CASAS, M. (Dir.) - VARO, C. (Ed.) *VII Jornadas de Lingüística*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz. pp. 137-191.  
[http://liceu.uab.es/publicacions/Linguistica\\_TH\\_Cadiz02.pdf](http://liceu.uab.es/publicacions/Linguistica_TH_Cadiz02.pdf)
- LLISTERRI, J.- CARBÓ, C.- MACHUCA, M. J.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2004) "La conversión de texto en habla: aspectos lingüísticos", in MARTÍ, M. A. - LLISTERRI, J. (Eds.) *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona – Fundació Duques de Soria. pp. 145-186.  
[http://liceu.uab.es/publicacions/Linguistica\\_CTH\\_FDS02.pdf](http://liceu.uab.es/publicacions/Linguistica_CTH_FDS02.pdf)
- LLISTERRI, J.- MACHUCA, M.J.- de la MOTA, C.- RIERA, M.- RÍOS, A. (2005) "Corpus orales para el desarrollo de las tecnologías del habla en español", *Oralia. Análisis del discurso oral* 8 (en prensa).  
[http://liceu.uab.es/~joaquim/publicacions/Oralia\\_04.pdf](http://liceu.uab.es/~joaquim/publicacions/Oralia_04.pdf)
- LLISTERRI, J.- MACHUCA, M.J.- de la MOTA, C.- RIERA, M.- RÍOS, M. (2003b) "Entonación y tecnologías del habla", in PRIETO, P. (Ed.) *Teorías de la entonación*. Barcelona: Ariel. pp. 209-243.  
[http://liceu.uab.es/~joaquim/publicacions/Ariel\\_Aplicaciones.pdf](http://liceu.uab.es/~joaquim/publicacions/Ariel_Aplicaciones.pdf)
- LUENGO, I.- NAVAS, E.- HERNÁEZ, I.- SÁNCHEZ, J. (2005) "Reconocimiento automático de emociones utilizando parámetros prosódicos", *Procesamiento del Lenguaje Natural* 35: 13-20.
- MOORE, R. (1995) "Computational Phonetics", in *ICPhS 95. Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol 4, pp. 68-71.
- NAVAS, E.- HERNÁEZ, I.- SÁNCHEZ, J.M. (2002) "Modelo de duración para la conversión de texto a voz en euskera", *Procesamiento del Lenguaje Natural* 29: 147-152.  
<http://www.sepln.org/revistaSEPLN/revista/29/29-Pag147.pdf>
- ÖHMAN, S. (2001) "Why current speech technology is false phonetics", *Working Papers (Lund University, Department of Linguistics)* 49: 180-183.  
<http://www.ling.lu.se/disseminations/pdf/49/bidrag46.pdf>
- PÉREZ, G.- LÓPEZ, T.- QUESADA, J.F. (2002) "Modelado de los candidatos seleccionados por un reconocedor de voz mediante técnicas de análisis gramatical", in *Actas del II Congreso de Fonética Experimental*. Sevilla: Laboratorio de Fonética, Facultad de Filología, Universidad de Sevilla. pp. 279-285.
- POLS, L C.W. (1999) "Flexible, robust, and efficient human speech processing versus present-day speech technology", in *ICPhS 99. Proceedings of the 14th International Congress of Phonetic Sciences*. Vol 1, pp. 9-16.  
<http://www.fon.hum.uva.nl/pols/ICPHS'99.VS2.doc>
- POLS, L. (2001) "Acquiring and implementing phonetic knowledge", in *Eurospeech 2001. Proceedings of the 7th European Conference on Speech Communication and Techno-*

- logy. Vol 1. pp. K3-K6; in *IFA Proceedings (Institute of Phonetic Sciences, University of Amsterdam)* 24: 39-46.  
[http://www.fon.hum.uva.nl/Proceedings/Proceedings24/Proc24Pols\\_corrton.html](http://www.fon.hum.uva.nl/Proceedings/Proceedings24/Proc24Pols_corrton.html)
- RODRÍGUEZ BANGA, E.- CAMPILLO DÍAZ, F.- FERNÁNDEZ REI, E.- MÉNDEZ PAZÓ, F. (2002) "Sistema de conversión texto-voz en lengua gallega basado en la selección combinada de unidades acústicas y prosódicas", *Procesamiento del Lenguaje Natural* 29: 153-158.  
<http://www.sepln.org/revistaSEPLN/revista/29/29-Pag153.pdf>
- RODRÍGUEZ, H. (2004) "Lingüística y estadística, ¿incompatibles?", in MARTÍ, M. A. – LLISTERRI, J. (Eds.) *Tecnologías del texto y del habla*. Barcelona. Edicions de la Universitat de Barcelona – Fundació Duques de Soria. pp. 89-117.
- ROSSI, M. (1996) "Connaissances et traitement automatique de la parole", in MÉLONI, H. (Coord.) *Fondements et Perspectives en Traitement Automatique de la Parole*. Paris: Éditions AUPELF-UREF. pp. 19-30.  
<http://www.bibliotheque.refer.org/html/parole/rossi/rossi.htm>
- RUBIO, A.- HERNÁEZ, I. (2005) *Libro blanco de Tecnologías del Habla*. Granada: Universidad de Granada - Red Temática en Tecnologías del Habla.  
<http://www.rthabla.org/LibroBlancoTecnologiasDelHabla.pdf>
- SANTEN, J.P.H. van (2005) "Phonetic knowledge in text-to-speech synthesis", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 149-166.
- SHRIBERG, E. (2005) "Spontaneous Speech: How People Really Talk and Why Engineers Should Care", in *EUROSPEECH 2005 - INTERSPEECH 2005. Proceedings of the 9th European Conference on Speech Communication and Technology*. pp. 1781-1784.  
<http://www.speech.sri.com/papers/eurospeech2005-shriberg-keynote2005-10-22>
- STRIK, H. (2005) "Is phonetic knowledge of any use for speech technology?", in BARRY, W.J.- van DOMMELEN, W.A. (Eds.) *The Integration of Phonetic Knowledge in Speech Technology*. Dordrecht: Springer. pp. 167-180.  
<http://lands.let.kun.nl/literature/strik.2005.1.pdf>