

Diseño experimental de un sistema de conversión de texto a voz para lenguas menores*

Harold Somers

School of Informatics, University of Manchester
PO Box 88, Manchester M60 1QD
CHarold.Somers@manchester.ac.uk

Resumen: Los sistemas de síntesis de voz o sistemas *text-to-speech* (TTS) están disponibles para muchas lenguas mayores, pero no para las miles de otras lenguas menores existentes debido a que no hay tecnología para ellas. Mientras se espera el desarrollo de ésta, intentamos utilizar un sistema TTS ideado para una lengua mayor (lengua origen o LO) para crear un sistema de habla artificial que funcione con una lengua menor (lengua meta o LM). Este artículo describe primero el diseño de un experimento en el que se busca una LO adecuada para la LM del pitjantjatjara, una lengua aborígen australiana. Después, se evalúa la aplicación real del experimento al ofrecer la tecnología necesaria a hablantes de la LM que no conocen bien la LO. El ejemplo propuesto es una visita al médico.

Palabras clave: texto a voz, TTS, SLT, pitjantjatjara.

Abstract: Speech synthesis or text-to-speech (TTS) systems are currently available for a number of the world's major languages, but for thousands of the world's 'minor' languages no such technology is available. While awaiting the development of such technology, we would like to try the stop-gap solution of using an existing TTS system for a major language (the base language) to 'fake' TTS for a minor language (the target language). This paper describes the design for an experiment which involves finding a suitable base language for the Australian Aboriginal language Pitjantjatjara as a target language, and evaluating its usability in the real-life situation of providing language technology support for speakers of the target language whose understanding of the local majority language is limited, for example in the scenario of going to the doctor.

Keywords: text-to-speech, TTS, SLT, pitjantjatjara.

1. Introducción

Los sistemas de síntesis de voz, concretamente los sistemas de conversión de texto a voz o *text-to-speech* (TTS), pueden leer en voz alta un texto de ordenador. Además, están muy extendidos por todo el mundo y ofrecen una calidad aceptable, aunque esto sólo se aplica a las llamadas lenguas mayores. Por ejemplo, el programa Agent de Microsoft incluye el inglés (tanto americano como británico), español, francés, alemán, italiano, portugués, japonés, coreano, ruso y neerlandés. Además de los anteriores, RealSpeak de Scansoft ofrece también euskera, can-

tonés, mandarín, dos variedades del neerlandés, inglés australiano e indio, francés canadiense, noruego, polaco y dos variedades del portugués. La lista es larga, aunque aún faltan miles de lenguas.

Nuestro objetivo es ofrecer la tecnología necesaria para que los hablantes de lenguas menores¹ puedan desenvolverse en situaciones en las que el desconocimiento de otra lengua es una desventaja importante. Nos hemos centrado en el caso de los inmigrantes que necesitan atención médica, aunque las posibilidades son infinitas. Ésta es la situación que propone el marco de CANES² descrito en So-

* Agradecemos a Andrew Longmire del Departamento del Centro del Medio Ambiente y Patrimonio Cultural, Uluru-Kata Tjuta National Park, Yulara NT, y a Bill Edwards de la Unaipon School, University of South Australia, Adelaide, por el interés mostrado en el experimento y esperamos que finalmente también puedan ayudarnos en su dirección.

¹Cualquier lengua que no sea mayor es una lengua menor. No debe confundirse lengua menor con lengua minoritaria, ya que el número de lenguas menores incluye algunas de las lenguas más habladas en todo el mundo como el bengalí, el hindi, el urdu y el árabe.

²*Computer Assistance for Non-English Speakers* (Asistencia por ordenador para hablantes no ingleses).

mers y Lovel (2003) y Somers *et al.* (2004). Este marco propone el uso de software para obtener información general sobre problemas de salud, pedir una cita para el hospital, entender folletos informativos e instrucciones sobre tratamiento y toma de medicamentos y, por supuesto, conversaciones con médicos o enfermeros. Otros trabajos realizados en este campo se han centrado en la traducción oral (SLT en inglés) de conversaciones entre paciente y médico (Rayner *et al.*, 2003; Narayanan *et al.*, 2004). Reconocemos que, mientras que las conversaciones cara a cara juegan un papel importante en la atención médica, hay otras formas de comunicarse (algunas basadas en texto) que también cobran cierta importancia. En tales casos, observamos que los sistemas TTS son una tecnología esencial no sólo para las personas que apenas saben inglés, sino también para las que tienen dificultades para leer en su propio idioma debido a analfabetismo, dislexia o problemas de visión.

Una solución a largo plazo sería desarrollar sistemas TTS para más lenguas, algo que no parece viable. Actualmente, el desarrollo de un sistema TTS depende, en primer lugar, de un análisis fonológico extensivo de la lengua para identificar los sonidos de la voz de una persona (fonemas) y sus variantes (alófonos); en segundo lugar, del desarrollo de reglas de conversión de texto a fonemas para identificar cómo se corresponde la ortografía de una lengua con la fonología y del desarrollo de reglas que determinan el tono y la duración de las palabras (prosodia); y por último, y dependiendo del enfoque elegido, también depende de la grabación de un discurso real para extraer los miles de elementos característicos de una persona en concreto (difonos, trifonos o semisílabas) o modelar un número similar de elementos usando un sintetizador por formantes.

Mientras tanto, pretendemos utilizar un sistema TTS existente para una lengua mayor para crear un sistema de habla artificial funcione con una lengua menor, concretamente con el piptjantjatjara³, una lengua aborigen australiana. Somers *et al.* (2006) describe un experimento similar con el somalí.

Esta idea ya fue explorada brevemente por

³Este trabajo se inició cuando el autor se encontraba en el Centro para las Tecnologías del Habla de la Universidad de Macquarie en Sidney.

Evans *et al.* (2002), quienes denominan tal proceso como «farfullar»⁴. Al parecer, los sintetizadores de voz adecuados para su uso con lectores de pantalla se desarrollan en un tiempo mínimo y pueden ofrecerse al usuario sin coste alguno. Según Evans *et al.* (2002: 576), «el único requisito es que el habla sea consistente y comprensible pese a que no suene natural o sea lingüísticamente incorrecta».⁵

2. Síntesis de TTS

La mayoría de los sistemas TTS se dividen en dos etapas: en la primera se realiza la conversión de texto a fonemas y se determina la pronunciación del texto y en la segunda se generan los sonidos del habla. Aunque nuestro objetivo no es detallar el funcionamiento de los sistemas TTS, creemos necesario hacer una pequeña introducción para explicar su diseño básico con el fin de hacerlos ficticios.

2.1. Conversión de texto a fonemas

La primera etapa identifica principalmente los fonemas que se utilizarán, además del tono y la duración, para producir una prosodia adecuada (entonación y acentuación). Para ello, normalmente se siguen reglas para hacer un modelo de conversión de letras a sonidos junto con el uso de un diccionario para casos irregulares. Aparte de palabras que no siguen las reglas habituales de pronunciación, el diccionario debe ser capaz de pronunciar abreviaturas, números, símbolos⁶, etc. El módulo de conversión de texto a fonemas también debe tener reglas que indiquen lecturas poco frecuentes de secuencias de símbolos. Por ejemplo, se dice <cinco dólares> y no <dólar cinco> para \$5. Además, la mayoría de lenguas tiene homógrafos cuya pronunciación correcta puede necesitar un análisis com-

⁴«Farfullar» es hablar deprisa y de forma incongruente. En nuestro caso, lo que se introduce en el sistema TTS es incongruente desde el punto de vista del sistema, pero esperamos que tenga sentido para el oyente.

⁵“the minimum requirement is that the speech has to be consistent and understandable, but does not necessarily have to be the especially natural sounding or indeed linguistically accurate”.

⁶En este artículo se utilizan las siguientes convenciones: el texto y los grafemas se escriben en *cursiva*; los textos que lee el sistema TTS están entre <paréntesis angulares>; por último, los símbolos fonéticos del Alfabeto Fonético Internacional se escriben entre [corchetes] y los fonemas entre /barras/.

plejo que además puede obtener información de la prosodia.

Los problemas para crear este módulo artificial surgen, en primer lugar, de las diferencias existentes en las reglas para hacer un modelo de conversión de letras a sonidos entre la lengua para la que se diseña un sistema TTS (lengua original o LO) y la lengua que se intenta producir (lengua meta o LM). Por ejemplo, mientras que la *j* se pronuncia como [dʒ] en inglés, en español se pronuncia como [x]. Además, es posible que algunas palabras de la LM se escriban igual en la LO, pero que se pronuncien de forma diferente, como *train* en inglés [trem] y [trɛ̃] en francés. Normalmente, las reglas de lectura de símbolos son específicas de la lengua con la que se trabaja, al igual que la prosodia. Por otro lado, puede ocurrir que la LM no utilice el mismo sistema de escritura que la LO o que ni siquiera tenga uno.

En todos estos casos se puede intentar reescribir la palabra de la LM para que siga las reglas de la LO. Así, podemos escribir la palabra francesa *train* como <tran>.

Algunos sintetizadores de TTS aceptan fonemas como datos de entrada en vez de texto. Según el programa, puede usarse un tipo de transcripción fonética (<dh-ah-s> para *thus*) o símbolos del Alfabeto Fonético Internacional (AFI). En el planteamiento de Evans *et al.* (2002: 578), «[l]as reglas [para la conversión de texto a fonemas] se almacenan en una tabla para que un programa adecuado pueda aplicarlas».⁷

De hecho, tal planteamiento es una alternativa notable al nuestro con sus ventajas y desventajas. Mientras que ellos han de especificar el modelo de todos los fonemas, nosotros sólo especificamos los fonemas que son diferentes, aunque también tienen más control sobre la entonación y puede utilizar un diccionario con pronunciaciones atípicas como números y símbolos.

2.2. Conversión de fonemas al habla

En la segunda etapa se generan los sonidos del habla mediante la concatenación de sonidos humanos ya grabados o con un sintetizador por formantes. El principal problema

⁷The rules [for text-to-phoneme translation] are contained in a text-based table and applied by a generic piece of software that is capable of applying any appropriately specified rule set

que supone crear un sistema de habla artificial se debe a que es muy difícil que el conjunto de sonidos de la LM coincida con el de la LO. Además, de ser así, las reglas de variación alofónica serán probablemente diferentes. De este modo, faltarán algunos fonemas de la LM, mientras que otros serán similares a los de la LO, pero con diferencias en la producción de alófonos diferentes.

La clave está en elegir una LO en la que apenas exista este problema, si bien es posible que haya que invertir un tiempo considerable para ello. Uno de los objetivos de la presente investigación es evaluar qué LO de entre varias es la más apropiada para una LM determinada e identificar qué factores hacen más fácil identificar la mejor LO.

3. Crear un sistema TTS artificial para el pitjantjatjara

El pitjantjatjara es la lengua aborigen australiana que hablan los anangu, conocidos también como los dueños de la zona del Uluru del Territorio del Norte, antes llamada Ayers Rock. El pitjantjatjara lo hablan unas 1.300 personas, aunque es uno de los dialectos ininteligibles que forman el grupo de lenguas del Desierto Oeste de Australia, que tiene 4.000 hablantes. El pitjantjatjara es una de las lenguas aborígenes más importantes. Según Eckert y Hudson (1991), el pitjantjatjara se ha escrito en alfabeto romano desde los años cuarenta. Su ortografía está más o menos normalizada desde la reunión que hubo en 1979 entre sus literatos y con la publicación en 1987 de un diccionario pitjantjatjara-inglés por el Instituto para el Desarrollo Aborigen en Alice Springs.

3.1. Fonología del pitjantjatjara

El pitjantjatjara, al igual que muchas lenguas aborígenes, tiene un conjunto muy simple de fonemas consonantes (Cuadro 1). Se utilizan cinco posiciones para su articulación: bilabial, alveolar, retroflexa, palatal y velar, con un único fonema oclusivo y nasal en cada posición. Hay tres fonemas laterales y tres aproximantes, además de un sonido vibrante de /r/. No hay fricativas y, como en otras lenguas aborígenes, no se distingue entre consonantes sonoras y sordas.

El pitjantjatjara tiene seis fonemas vocálicos: [i], [u] y [a], todos largos y cortos. En la ortografía, las vocales largas se duplican (ii, uu, aa). Las sílabas están muy limitadas,

	bilabial	alveolar	retroflexa	palatal	velar
oclusiva	[p] <i>p</i>	[t] <i>t</i>	[ɽ] <i>ɽ</i>	[c] <i>tj</i>	[k] <i>k</i>
nasal	[m] <i>m</i>	[n] <i>m</i>	[ɳ] <i>ɳ</i>	[ɲ] <i>ny</i>	[ŋ] <i>nj</i>
vibrante		[r] <i>r</i>			
lateral		[l] <i>l</i>	[ɭ] <i>ɭ</i>	[ʎ] <i>ly</i>	
aprox.	[w] <i>w</i>		[ɻ] <i>ɻ</i>	[j] <i>y</i>	

Cuadro 1: Consonantes del pitjantjatjara. Símbolos fonéticos entre [paréntesis] y ortografía estándar en *cursiva*.

pues todas son (C)V(C). No hay grupos consonánticos salvo cuando hay límites de sílabas. La mayoría de monosílabos son CV. El acento de las palabras suele recaer en la primera sílaba y las siguientes tienen la misma prominencia.

3.2. Elegir una LO para crear un pitjantjatjara artificial

La parte más difícil del experimento es quizás escoger la mejor LO para nuestro sistema TTS artificial. Hoy día los desarrolladores de sistemas TTS comercializan cada vez más productos de una calidad sorprendentemente buena. Hay que decir que no podemos evaluarlos con nativos de todas las lenguas, por lo que necesitamos disponer de criterios que nos permitan limitar el campo. Tres factores posiblemente conflictivos guían nuestra decisión: conjuntos de fonemas, prosodia comparativa y modelo ortográfico.

3.2.1. Conjuntos de fonemas

El requisito más obvio a tener en cuenta es encontrar una LO que contenga un conjunto de fonemas idéntico o similar al de la LM. Este problema tiene dos caras: por un lado, el fonema de la LM puede no tener equivalente en la LO; por otro, el fonema supuesto equivalente puede ser en realidad muy diferente.

El primer caso no plantea tantos problemas como puede parecer, ya que, por ejemplo, el pitjantjatjara tiene consonantes retroflexas y palatales que no se encuentran en ninguna de las LO candidatas. No obstante, la solución a esto puede ser el uso de secuencias de fonemas de la LO que suenen lo suficientemente parecido a los de la LM. Otra alternativa es que si se fuerza a la LO a combinar una distinción fonética, el resultado puede ser simplemente un sonido artificial, como si tuviera un acento marcado en vez de generar algo incomprensible.

El segundo caso es más problemático. Si pensamos de nuevo en sonidos similares a la *r* en pitjantjatjara y en sus sonidos retroflexos, no parece que utilizar el francés o el portugués como LO sea apropiado debido a que el fonema /r/ es una fricativa uvular. Los fonemas de las vocales podrían plantear problemas en otros usos, aunque por suerte el sistema vocálico del pitjantjatjara es uno de los más simples del mundo.

3.2.2. Prosodia

La prosodia es un buen indicio para ver lo realista que es un sistema TTS, por lo que sus desarrolladores ponen especial interés en identificarla adecuadamente. Los mejores sistemas TTS son capaces de identificar la entonación y acento correctos, incluso si hay cierto «farfuleo». La prosodia del pitjantjatjara es bastante uniforme, por lo que cualquier LO que tenga una prosodia variable, como el sueco o el danés, podría funcionar. Las reglas del pitjantjatjara que indican dónde recae el acento en una palabra son también muy simples, ya que normalmente siempre es en la primera sílaba. Por ello, toda lengua que tenga pocas palabras cuya sílaba tónica es la primera queda descartada, como es el caso del inglés. Así pues, una LO como el euskera puede ser una apuesta segura debido a su entonación uniforme y ritmo de sílabas a pesar de que en euskera el acento suele recaer en la última sílaba.

3.2.3. Modelo de conversión de texto a fonemas

Frente a la equivalencia de fonemas y prosodia se encuentran las reglas del modelo de conversión de texto a fonemas. Son dos las preguntas que se deben plantear: ¿cómo se pronuncian las letras en la LO y cuál es la probabilidad de que las palabras de la LM sean iguales en la LO pero con una pronunciación diferente?

Dado que la ortografía pitjantjatjara se basa en la inglesa, existe un buen modelo de conversión de texto a fonemas entre el inglés y el pitjantjatjara. El sistema de escritura también es bastante «fonético», es decir, que hay un modelo sencillo entre letras y fonemas. En otras combinaciones lingüísticas, las letras más problemáticas son la *c*, *g*, *j*, *v*, *w*, *x*, *y*, *z* y todas las vocales y dígrafos. El pitjantjatjara no plantea problemas al respecto, pues sólo usa doce consonantes y tres vocales. El sistema de escritura tiene cuatro dígrafos

<i>p</i> <p>	<i>t</i> <t>	<i>ṭ</i> <rt>	<i>tj</i> <tt>	<i>k</i> <k>
<i>m</i> <m>	<i>n</i> <n>	<i>ṇ</i> <rn>	<i>ny</i> <n̄>	<i>ng</i> <ng>
	<i>r</i> <rr>			
	<i>l</i>	<i>ḷ</i> <rl>	<i>ly</i> <ll>	
<i>w</i> <w>		<i>ṛ</i> <rr>	<i>y</i> <i>	

Cuadro 2: Modelo de la ortografía pitjantjatjara en *cursiva* sobre las letras de euskera en <paréntesis angulares>.

(*tj*, *ny*, *ly*, *ng*), cuatro letras que representan sonidos retroflexos (*ṭ*, *ṇ*, *ḷ*, *ṛ*) y tres dígrafos de vocales largas. Ningún dígrafo del inglés como *ch*, *sh*, o *th* aparece en palabras (nativas) del pitjantjatjara.

Teniendo en cuenta lo dicho sobre la prosodia, si elegimos el euskera como LO tenemos que considerar su ortografía. Afortunadamente, ésta se normalizó en 1964, por lo que su modelo de conversión de letras a fonemas es bastante simple. El Cuadro 2 muestra las diferencias del modelo entre los fonemas del pitjantjatjara y la escritura del euskera. Hay que tener en cuenta que la letra *y* debe evitarse en caso de que la palabra sea un préstamo de otra lengua, ya que entonces se pronunciaría como en español: [i]. Los sonidos palatales resultan óptimos, pues <tt> se utiliza para los diminutivos y la <n̄> y <ll> para préstamos del español, lo que nos da los sonidos palatales *ny* y *ly*.

3.3. Implementación

Todas las transliteraciones del Cuadro 2 se pueden aplicar al texto en pitjantjatjara con un simple programa de reemplazo de caracteres. Los ejemplos 1 y 2 son textos en pitjantjatjara con su ortografía normalizada (a) y cómo se introducen en el sistema TTS de euskera (b).

1. Lleva este niño al hospital.
 - a) *Tjitji nyanga kati atjupitlakutu.*
 - b) *Ttitti ñanga kati attupitlakutu.*
2. ¿Cuántas veces debo tomarlo?
 - a) *Yaaltjitu arana tjikinma?*
 - b) *Iaalttitu ararna ttikinma?*

4. Evaluación

Evans *et al.* (2002) describen una modesta evaluación de un prototipo de sintetizador griego mediante el uso del inglés y del español

como LO con tres griegos nativos que hablan inglés fluidamente. Una primera evaluación utilizaba una variante del Test de Rimas Modificado o MRT por sus siglas en inglés (House *et al.*, 1965; Miner y Danhauer, 1976; Logan *et al.*, 1989; Goldstein 1995) en la que los sujetos tenían que señalar la palabra que creían haber oído de una lista de 5 palabras. Los sujetos participaron en 15 pruebas de 5 palabras cada una. Evans *et al.* (*ibíd.*) comentan que hubo un total de 7 errores en 45 pruebas realizadas con un sintetizador de español. Con el sintetizador de inglés hubo 10 errores (96 %). Una segunda evaluación se basó en el uso de palabras absurdas, cuyo porcentaje de reconocimiento fue mucho más bajo (50 % en los dos sistemas). En una tercera evaluación, los sujetos se sometieron a una prueba de reconocimiento de frases completas, incluidas frases simples y comunes o trabalenguas. El porcentaje de reconocimiento fue del 100 % en este caso, aunque no está claro si los sujetos tuvieron que identificar lo que escucharon o simplemente elegir de entre un número de alternativas.

Evans *et al.* (*ibíd.*) admiten las deficiencias de esta evaluación, como el reducido número de personas participantes y el hecho de que todos los sujetos hablaran inglés fluidamente y que llevaran tiempo residiendo en Reino Unido, lo que los hacía familiares con los fonemas ingleses y su prosodia. Además, subrayan que las pruebas realizadas no reflejaban la verdadera aplicación del software utilizado, que en este caso era un lector de pantalla.

Nuestra evaluación intenta superar estas deficiencias: a) con un grupo de suficientes sujetos para que los resultados sean significativos estadísticamente; b) prescindiendo del inglés y del euskera; c) sabiendo que el contacto con el inglés es, más o menos, constante para todos los hablantes del pitjantjatjara; y d) intentando simular mejor la situación en la que podría usarse el software.

4.1. Una situación realista

Nuestra metodología se basa en la utilizada por Somers y Sugita (2003) en su evaluación de software SLT. La investigación en SLT se centra casi sin excepción en un tipo particular de aplicación, concretamente en diálogos orientados a tareas, como por ejemplo pedir cita o preparar un viaje. En consecuencia, Somers y Sugita decidieron evaluar el software SLT mediante la traducción de

frases de una cinta de audio para turistas. Su principal interés estaba en la habilidad mostrada por los sujetos para identificar correctamente la intención del mensaje más que el estilo o sintaxis de la traducción. Del mismo modo, nuestro interés se centra sobre todo en saber si la producción oral es inteligible y no en la naturalidad y precisión fonética a menos que afecten a la comprensión.

En nuestro experimento, un investigador cuya lengua nativa es la misma que la de los sujetos participantes les pedirá a éstos que se imaginen que han ido a la consulta de un médico por sufrir alguna molestia, como por ejemplo problemas respiratorios, y que todo lo que les diga el médico será traducido y «hablado» por el ordenador. Una vez hecho esto, se les pedirá que escuchen el habla sintetizada y que le digan al investigador qué han entendido. Dado que la sintaxis o calidad de la traducción no nos interesan, se considerará aceptable que repitan literalmente lo que han escuchado. El investigador decidirá sobre lo que han entendido y hará preguntas aclaratorias si es necesario. Todo el experimento se grabará como prueba.

Al igual que en el experimento de Somers y Sugita (2003), habrá 5 situaciones diferentes con 5 preguntas en cada una, lo que da un total de 25 frases (ver Anexos 2 y 3). Cada situación durará unos 15 minutos. Las frases se han adaptado para que ofrezcan la información esencial necesaria para que el investigador sepa cuando el sujeto repita la frase si la ha entendido. Al final de cada sesión, el investigador hará un informe de las reacciones y opiniones de los sujetos en una entrevista privada.

En caso de haber suficientes sujetos, se utilizarán grabaciones realizadas por un humano en vez de una grabación sintetizada para una selección de frases de forma que tengamos más control. Por otro lado, sería interesante contrastar el uso del euskera como LO, habiendo adaptado antes su ortografía, en vez de utilizar la implementación barata pero de baja calidad de programas TTS gratuitos.

En nuestros experimentos con el somalí (Somers *et al.*, 2006) probamos varias implementaciones diferentes combinando la propuesta del presente artículo con la del «farfalleo» de Evans *et al.* (2002). Lo que hicimos fue decirle a los sujetos que el ordenador tendría varias voces diferentes según la situación y personal relacionado con ella

(un recepcionista, un médico, una enfermera, etc.). Así pues, la organización experimental es de un número cuadrado de voces latinas \times número de respuestas, lo que determina el número mínimo de sujetos necesarios. Cuando aparece una voz diferente, se introduce y reproduce un ejemplo.

5. Estudios piloto

En estos momentos estamos intentando organizar un experimento con nativos de pitjantjatjara por medio de varias agencias. Gracias a las observaciones realizadas en experimentos con nativos de somalí (Somers *et al.*, 2006), descubrimos ciertos factores interesantes que nos han permitido reajustar el diseño del experimento.

El primero de ellos tiene que ver con las traducciones. No es nada nuevo descubrir que el pitjantjatjara no tiene términos específicos para expresar conceptos como *contaminación* o *historial médico*. Incluso parece ser que el asma no tiene equivalentes. También hay algunas construcciones que son difíciles de expresar en pitjantjatjara, como por ejemplo los comparativos (*¿Es peor durante la noche?*). La conclusión a la que llegamos es que es difícil de comprender préstamos de otras lenguas.

Una segunda idea que surgió de los estudios piloto fue reajustar nuestro sistema de puntuación. Observamos que cada sujeto quería escuchar la grabación dos o tres veces y que la comprendía mejor al escucharla más veces. Por ello, estamos diseñando un sistema de puntuación que tenga en cuenta el número de veces que necesita cada sujeto para comprender el mensaje, ya que en una situación real puede darse el caso de tener que repetir algo por no haberlo comprendido totalmente.

Por último, nos dimos cuenta de que los sujetos se acostumbraban rápidamente a la voz del ordenador y que comprendían mejor hacia el final de la prueba. Debemos considerar la importancia de este descubrimiento, si bien es cierto que el diseño de un experimento con varias voces puede reducir considerablemente tal efecto.

Esperamos poder ofrecer los resultados de nuestro experimento en su debido momento.

Bibliografía

Eckert, P. y Hudson, J. 1991. *Wangka Wiru: A Handbook for the Pitjantjatjara Language*

- ge Learner*. Underdale, SA: University of South Australia.
- Evans, G., Polyzoaki, K. y Blenkhorn, P. 2002. An approach to producing new languages for talking applications for use by blind people. En Miesenberger, K. Klaus, J. y Zagler, W. (eds). *8th ICCHP, Computers Helping People with Special Needs*. (LNCS 2398), Berlin: Springer Verlag, pp. 575-582.
- Goldstein, M. 1995. Classification of methods used for assessment of text-to-speech systems according to the demands placed on the listener *Speech Communication*, 16:225-244.
- House, A.S., Williams, C.E., Hecker, M.H.L. y Kryter, K.D. 1965. Articulation-testing methods: Consonant differentiation with a closed-response set. *Journal of the Acoustic Society of America*, 37:158-166.
- Logan, J., Greene, B. y Pisoni, D. 1989. Segmental intelligibility of synthetic speech produced by rule. *Journal of the Acoustical Society of America*, 86:566-581.
- Miner, R. y Danhauer, J.L. 1976. Modified rhyme test and synthetic sentence identification test scores of normal and hearing-impaired subjects listening in multitalker noise. *Journal of the American Audiological Society*, 2:61-67.
- Narayanan, S., Ananthakrishnan, S., Belvin, R., Ettelaie, E., Gandhe, S., Ganjavi, S., Georgiou, P.G., Hein, C.M., Kadambe, S., Knight, K., Marcu, D., Neely, H.E., Srinivasamurthy, N., Traum, D. y Wang, D. 2004. The Transonics spoken dialogue translator: An aid for English-Persian doctor-patient interviews. In: T. Bickmore (ed.). *Dialogue Systems for Health Communication: Papers from the AAAI Fall Symposium*. Menlo Park, CA: The AAAI Press, pp. 97-103.
- Rayner, M., Bouillon, P., Van Dalsem, V., Isahara, H., Kanzaki, K. y Hockey, B.A. 2003. A limited-domain English to Japanese medical speech translator built using REGULUS 2. *Companion Volume, 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 137-140.
- Somers, H., Evans, G., y Mohamed, Z. 2003. Developing speech synthesis for under-resourced languages by 'faking it': An experiment with Somali. Artículo enviado al *LREC 2006* (Genoa).
- Somers, H. y Lovel, H. 2003. Computer-based support for patients with limited English. *Association for Computational Linguistics EACL 2003, 10th Conference of The European Chapter, Proceedings of the 7th International EAMT Workshop on MT and other language technology tools*, Budapest, pp. 41-49.
- Somers, H., Lovel, H., Johnson, M. y Mohamed, Z. 2004. Language technology for patients with limited English. *EACH International Conference on Communication in Healthcare*, Bruges, P04.09.
- Somers, H. y Sugita, Y. 2003. *MT Summit IX: Proceedings of the Ninth Machine Translation Summit*. New Orleans, pp. 370-377.

A. Anexo 1: Instrucciones del experimento

[El texto en cursiva lo dice el investigador en la lengua nativa del sujeto] *Va a escuchar algunas frases del ordenador. Me gustaría que me dijera con sus propias palabras qué es lo que dice el ordenador. No tiene que responder preguntas. Puedo repetir las frases hasta tres veces. No se preocupe si no entiende algo, usted no es el objeto de esta prueba. El ordenador tiene voces diferentes. Cada vez que cambie le daré un ejemplo y le diré lo que dice. Imagine que tiene que ver al médico porque tiene dificultades para respirar.*

B. Anexo 2: Preguntas y respuestas que se traducen al pitjantjatjara (inglés original)

[El texto en negrita representa la grabación de cada situación]

1. *The receptionist says: **First we have to make an appointment.***
 - a) What is your name?
 - b) Have you been to this clinic before?
 - c) Can you come next Thursday?
 - d) How about ten o'clock?
 - e) Do you want an appointment card?
2. *The doctor's assistant says: **I am going to ask you some questions about your circumstances.***

- a) Do you smoke, or have you ever smoked?
- b) Do you have any pets?
- c) Do you take regular exercise?
- d) Does anyone in your family have asthma?
- e) Is your home dry?

3. *The doctor says: I want to get some background details.*

- a) When did you first have difficulty with breathing?
- b) Is your breathing worse at any particular time of day?
- c) Does it prevent you from sleeping?
- d) Is it worse after strenuous activity?
- e) Does it depend on the weather?

4. *The nurse says: I will make some suggestions.*

- a) Your asthma may be triggered by the weather.
- b) Check your bathroom for mold.
- c) Use the brown inhaler if you have a mild attack.
- d) Use it once a day for a week after an attack.
- e) Take two puffs every four hours.

5. *The pharmacist says: Your prescription is ready.*

- a) Are you taking any other medication?
- b) What is your name and address?
- c) Take this three times a day
- d) Are you entitled to a free prescription?
- e) That will be six dollars fifty please.

C. Anexo 3: Preguntas y respuestas que se traducen al pitjantjatjara (traducción al español)

[El texto en negrita representa la grabación de cada situación]

1. *El recepcionista dice: Primero tenemos que darle cita.*

- a) ¿Cómo se llama?

- b) ¿Ha estado antes en esta clínica?
- c) ¿Puede venir el próximo jueves?
- d) ¿Qué tal a las diez?
- e) ¿Quiere una tarjeta con los detalles de la cita?

2. *El ayudante del médico dice: Le voy a hacer unas preguntas sobre su situación.*

- a) ¿Fuma o ha fumado alguna vez?
- b) ¿Tiene mascotas?
- c) ¿Hace ejercicio regularmente?
- d) ¿Tiene alguien de su familia asma?
- e) ¿Su casa es húmeda?

3. *El médico dice: Quiero algunos detalles sobre sus antecedentes.*

- a) ¿Cuál fue la primera vez que tuvo problemas respiratorios?
- b) ¿Hay algún momento del día en que se sienta peor?
- c) ¿La respiración le impide dormir?
- d) ¿Se siente peor tras hacer esfuerzos?
- e) ¿Influye el tiempo en su respiración?

4. *La enfermera dice: Le haré unas recomendaciones.*

- a) El clima influye en el asma.
- b) Compruebe que no haya moho en su cuarto de baño.
- c) Use el inhalador marrón en caso de un ataque leve.
- d) Úselo una vez al día durante una semana después de un ataque.
- e) Inhale dos veces cada dos horas.

5. *El farmacéutico dice: Su receta está lista.*

- a) ¿Toma algún otro medicamento?
- b) ¿Cuál es su nombre y dirección?
- c) Tómese esto tres veces al día.
- d) ¿Tiene algún tipo de receta gratuita?
- e) Son seis dólares con cincuenta, por favor.