Procesamiento del Lenguaje Natural, núm. 35 (2005), pp. 293-300 recibido 29-04-2005; aceptado 01-06-2005

Análisis y síntesis de expresión emocional en cuentos leídos en voz alta

Virginia Francisco
Universidad Complutense de Madrid
28015
vfrancisco@gmail.com

Pablo Gervás
Universidad Complutense de Madrid
28015
pgervas@sip.ucm.es

Raquel Hervás
Universidad Complutense de Madrid
28015
raquelhb@fdi.ucm.es

Resumen: Un reto importante para los conversores texto-voz es conseguir que la voz sintética suene lo mas parecido posible a la voz humana. La voz generada por estos sistemas actualmente suena artificial y esta es la principal causa de rechazo por parte del público general. Para conseguir que el sintetizador aparente "vida" interesa generar voz con distintos estados anímicos.

El desafío fundamental de la generación de voz emocional es tratar de generar una emoción suficientemente clara para que no haya confusión en el oyente. Existen muchas teorías para definir una escala emocional. La elección de una escala concreta determina las emociones que se pretenden distinguir. Otro desafío importante es analizar las características acústicas de los distintos estados emocionales para intentar posteriormente regenerar las mismas a través del sintetizador (Montero, 2003).

Este trabajo se plantea explorar la viabilidad de modelar las cadencias propias de la narración de historias a través de los parámetros de control de un conversor texto-voz. Para lograr estos parámetros se realizará un análisis de material de audio emocional y una vez modeladas cada una de las emociones se realizará una evaluación del material obtenido.
Palabras clave: Síntesis de voz, emoción, análisis de emociones, evaluación.

Abstract: An important challenge for text-to-speech is to get a synthesized voice that sounds as like as possible to the human voice. The voice synthesized by these systems sounds artificial and this is the most principal cause of rejection by the public at the moment. In order to obtain a lively synthesized voice it is necessary to generate a voice with emotions.

The main goal of the generation of emotional voice is try to generate an emotion so clear that there will be no confusion in the listener. There are a lot of theories in order to define an emotional scale. The choice of a specific scale determines the emotions that we try to distinguish. Another important challenge is analyse the acoustic characteristics at different emotional states in order to try to regenerate the same characteristics by the synthesizer (Montero, 2003).

This project raises to explorer the possibility of model the lack of the tales through control parameters in the synthesizer. In order to obtain these parameters we have to carry out an analysis of emotional audio and then, once we have obtained a model, we have carried out a test.
Keywords: Voice synthesis, emotion, analysis of emotions, test.

ISSN: 1135-5948

1 Expresión emocional y síntesis de voz

Los primeros estudios sobre generación de voz con emoción son realizados por Fairbanks y Pronovost (1939) a finales de los años 30. Aunque esta línea de trabajo ha dado lugar a muchas investigaciones y artículos publicados, aun quedan por cubrir una gran variedad de aspectos relevantes. La complejidad de este área de investigación comienza en el concepto mismo de emoción. Existen hoy en día múltiples teorías de las emociones, cada una con un interés distinto, difíciles de integrar y que en ocasiones se contradicen.

1.1 Teorías de las emociones

Las emociones se definen como un mecanismo flexible de adaptación a un ambiente cambiante (Scherer, 1979). Pueden distinguirse los siguientes tipos fundamentales de emoción (Cowie y Cornelius, 2003):

- *Emociones extremas:* Este término denota una emoción totalmente desarrollada, la cual típicamente es intensa e incorpora la mayoría de los aspectos que se consideran relevantes en el síndrome de la emoción.
- *Emociones subyacentes:* Denotan el tipo de colorante emocional que es parte de la mayoría o de todos los estados mentales.

De cara a seleccionar un subconjunto de emociones que permita realizar una tarea experimental, se han tenido en cuenta las siguientes clasificaciones teóricas:

1. *Categorías emocionales:* Emplea palabras a la hora de clasificar las emociones. Esta teoría engloba a otras tres:
   - *Emociones básicas:* El número de emociones básicas suele ser pequeño (en los primeros estudios menos de 10, en los más recientes entre 10 y 20)
   - *Emociones súper ordinarias:* Existen una serie de categorías que son más fundamentales que otras en el sentido de que incluyen en sí mismas a las otras (Ortony, Clore y Collins, 1988) (Scherer, 1984)
   - *Emociones esenciales del día a día:* La teoría se ejemplifica en el trabajo de Cowie y Cornelius (2003). Comenzando con una lista de términos emocionales de la literatura se insta a los sujetos a seleccionar un subconjunto que represente de manera apropiada las emociones relevantes de la vida diaria.
2. *Descripciones basadas en psicología:* Según esta teoría el aspecto esencial de una emoción es el estado del cuerpo que tiene asociado (Alter et al.,2000)(Smith, 1989)
3. *Descripciones basadas en la evaluación:* Estas teorías describen las emociones desde el punto de vista de las evaluaciones que implican (Scherer,1984)
4. *Dimensiones emocionales:* Las dimensiones emocionales son una representación simplificada de las propiedades esenciales de las emociones. Evaluación (positiva / negativa) y activación (activa / pasiva) son las dimensiones mas importantes, en algunas ocasiones se complementan con la dimensión poder (dominante / sumiso) (Wundt, 1989).

El trabajo de modelar la generación de voz con emoción empieza por seleccionar el subconjunto de emociones que se desea modelar. Este pequeño conjunto inicial de emociones puede ser posteriormente completado hasta obtener un conjunto de emociones más extenso. Esta tarea se simplifica si se elige empezar por clasificaciones basadas en emociones básicas, emociones súper ordinarias o las emociones del día a día. Una vez modeladas estas emociones pueden incorporarse emociones menos extremas, basadas en las restantes clasificaciones: descripciones basadas en psicología, basadas en evaluación o dimensiones emocionales. Otra forma de actuar sería comenzando desde un principio con cualquiera de las clasificaciones más extensas.

1.2 Estudios empíricos de la prosodia de la emoción

La identificación de la prosodia asociada a cada emoción debe realizares empíricamente. Distintas fuentes se han utilizado en el pasado para obtener una base de datos de voz con emoción:

- *Actores:* La técnica más antigua y más frecuentemente usada consiste en obtener los datos de grabaciones de actores.
- *Lectura expresiva de material emocional:* Se trata de una variante al uso de actores propuesta por Nick Campbell (2000) que propone tener lectores que lean textos con un contenido verbal apropiado con la emoción que se desea transmitir.
- *Ocurrencias naturales:* Un estudio que trabaja con la generación espontánea de estados de ánimo es el dirigido por Scherer,

Ladd y Silverman (1984). Grabaron entrevistas entre trabajadores sociales y actores que simulaban ser clientes. Tan solo se utilizaba el material grabado de los trabajadores sociales.

- *Producción de emociones:* Se trata de inducir a las personas por medio de distintos métodos: hipnosis, imaginación, películas, música… a expresar una emoción.

Los métodos mostrados varían en cuanto al control que se puede ejercer sobre la señal del habla, se podría decir que de más a menos control se ordenarían de la siguiente manera: actores, lectura de textos emocionales, producción de emociones, observación del habla natural.

Cada uno de los métodos anteriormente mencionados encaja mejor o peor dependiendo del dominio del estudio que se este realizando:

- *Investigación de emociones extremas:* El método mas adecuado será la utilización de actores.
- *Investigación de emociones subyacentes:* Lo mejor es observar el habla natural.
- *Estudios centrados en el hablante:* La mejor elección será la producción de emociones.

Existen numerosos estudios para obtener las reglas prosódicas que intervienen en la generación de voz con emoción. Estas reglas se han obtenido de diversas maneras:

- Extrayéndolas de la literatura.
- Haciendo un análisis de un corpus lingüístico en particular.
- Obteniendo los valores óptimos a través de la variación sistemática de parámetros en la síntesis.

En todos estos estudios los parámetros globales como son el nivel de la frecuencia fundamental, la escala de frecuencia fundamental y la velocidad del habla de la prosodia se tratan como universales, al menos siempre que el número de categorías emocionales sea pequeño.

Para la síntesis de voz las variables acústicas mas interesantes son aquellas que se pueden controlar en los sistemas de síntesis de voz.

A la hora de generar un sistema capaz de emitir voz con emoción será necesario contar con una correspondencia entre las distintas emociones que se desean modelar y las características de cada uno de los parámetros que intervienen en la prosodia.

A continuación se muestran las distintas opciones a la hora de seleccionar un paradigma de evaluación:

1. *Elección forzada:* Este tipo de evaluación ha sido empleado en gran cantidad de estudios de generación de voz con emoción (Cahn, 1990) (Montero, 2003) (Mozziconacci, 1998). Consiste en facilitar a los sujetos que van a realizar la evaluación un conjunto finito de posibles respuestas que engloban todas las emociones que han sido modeladas.
2. *Elección libre:* En este caso la respuesta no se restringe a un conjunto cerrado de emociones (Murray y Arnott, 1995) (Schröder, 1999). Esta especialmente indicado para encontrar fenómenos no esperados durante el experimento.
3. *Elección libre modificada:* Murray y Arnot (1995) y luego Stallo (2000) introdujeron una serie de modificaciones al sistema anterior: introdujeron categorías de distracción y la categoría “otros”.

1.3 Herramientas para el análisis y la síntesis de voz

Para este trabajo se ha seleccionado una herramienta para analizar secuencias de voz (Praat) y una herramienta para sintetizar voz (JavaSpeech API, FreeTTS).

El objetivo de *Praat*[1] se centra en el análisis y síntesis del habla, de modo que logremos manipular grabaciones de voz y crear gráficos.

Para nuestro trabajo, la mayor ventaja que nos ofrece Praat es la generación de gráficos de gran calidad en los que podemos visualizar la frecuencia fundamental, el espectrograma, la intensidad, los formantes y los pulsos de la grabación. A partir de estos gráficos en distintas grabaciones de cuentos realizadas por actores podremos establecer en que medida variar las características de la voz implicadas en las emociones.

Java Speech API define un estándar para el desarrollo de aplicaciones de voz. A través de él se soportan las dos tecnologías de voz principales: reconocimiento y síntesis de voz.

Se trata de una extensión de la plataforma Java. Las extensiones son paquetes de clases escritas en Java que los desarrolladores pueden emplear para extender la funcionalidad de Java.

[1] http://www.fon.hum.uva.nl/praat/

Existe un paquete denominado javax.speech que define una representación abstracta del "motor de voz". El sintetizador y el reconocedor de voz son ambos instancias del motor de voz.

El sintetizador de Java Speech puede variar los siguientes parámetros de la voz: volumen, velocidad, tono y rango de variación del tono, que son los parámetros de la prosodia que hemos seleccionado para modelar nuestras emociones.

JavaSpeech API define un estándar para realizar aplicaciones de voz, la implementación de sus métodos depende de los distintos motores de síntesis de voz. Entre todos los motores que implementan la interfaz JavaSpeech API para nuestro trabajo hemos empleado FreeTTS[2], se trata de un motor de síntesis escrito completamente en Java, basado en los sintetizadores Flite[3] y derivado de Festival[4] y FestVox[5], proporciona soporte para varias voces, emplea concatenación de dialofonos y permite su manejo a través del estándar JavaSpeech.

2 Expresión de emociones en la narración de cuentos

Se ha elegido la narración de cuentos infantiles como dominio de aplicación por considerar que presenta un marco donde las emociones claramente participan en el esfuerzo de comunicación. Estos cuentos tratan de resumir las emociones que experimentan la mayoría de los niños en su camino a la madurez: alegría, tristeza, enfado, miedo, celos…

Cuando se cuenta un cuento se tiende a exagerar. La voz de la persona o personas que cuentan el cuento será un instrumento tan importante como las palabras para que el niño pueda inferir las emociones de los personajes. Podemos afirmar, por tanto, que las emociones que se expresan en los cuentos son emociones extremas, no emociones subyacentes.

2.1 Marcado de emociones en cuentos

La clasificación de las emociones expresadas en el habla que más se adecua a las necesidades de un cuenta-cuentos es la clasificación de las emociones básicas, ya que son este tipo de emociones las que se intentan transmitir a la hora de contar un cuento. Cuando contamos un cuento tendemos a exagerar las emociones por lo que con un pequeño conjunto de emociones extremas tendremos suficiente. Para comenzar como emociones básicas hemos seleccionado las siguientes: enfado, alegría, tristeza, miedo y sorpresa.

De cara a simplificar la obtención de texto con emociones hemos adaptado un módulo ya existente de generación de texto en lenguaje natural en el campo de los cuentos fantásticos (Gervás et al., 2004). A partir de una entrada que representa las acciones que forman la trama del cuento y de información semántica sobre los personajes, lugares y objetos involucrados en las mismas, se genera el texto correspondiente en lenguaje natural.

En el proceso de generación de texto es necesario utilizar información semántica sobre los elementos involucrados en la historia. En nuestro generador tenemos esta información semántica en una base de conocimiento (Peinado, Gervás y Díaz-Agudo, 2004) donde se almacenan la lista de personajes, lugares y objetos que pueden aparecer en los cuentos, sus atributos y las relaciones que existen entre ellos. Esta base de conocimiento almacena también la información necesaria para marcar el texto con emociones, asociada en nuestro trabajo a dos tipos de elementos: los personajes y las acciones.

El marcado del texto en nuestro generador se realiza en la fase de lexicalización, donde para cada elemento del texto se escoge la etiqueta léxica correspondiente del vocabulario disponible. En los primeros experimentos de marcado de cuentos para síntesis de voz hemos decidido marcar los textos a nivel de oraciones. Para decidir la emoción con la que marcar cada frase se tienen en cuenta tanto los personajes involucrados en la misma, como la acción en sí.

Los cuentos generados por el cuenta-cuentos son cuentos de unas 40 frases, cuya duración aproximada es de 1 o 2 minutos. Las frases son cortas y en ellas el único personaje que habla con su propia voz es el narrador. El narrador intentará imprimir la emoción con la que haya sido marcada la frase, de modo que cuando este contando un pasaje triste del cuento su voz transmitirá tristeza y alegría cuando se trate de un pasaje alegre.

[2] http://freetts.sourceforge.net
[3] http://www.speech.cs.cmu.edu/flite
[4] http://www.cstr.ed.ac.uk/projects/festival
[5] http://www.festvox.org

2.2 Fijación de los parámetros prosódicos asociados a las emociones.

Para obtener los parámetros de la prosodia hemos analizado grabaciones de cuentos leídos por actores para identificar la relación entre los parámetros de la voz y las distintas emociones. Se ha optado por la utilización de actores por ser la mejor elección cuando lo que se busca es la investigación de emociones extremas.

Los aspectos de la voz que actúan como marcadores de personalidad son: frecuencia fundamental, volumen, cualidad de la voz y fluidez. Para obtener el valor de estos parámetros en cada una de las emociones nos hemos basado fundamentalmente en: los estudios de Scherer (1979), el análisis de material emocional, y la obtención de los valores óptimos a través de la variación sistemática de parámetros en la síntesis.

Scherer en sus trabajos realiza una descripción acerca de la relación entre los parámetros de la voz y las emociones:

- El enfado se caracteriza por un tono medio alto y una velocidad de locución rápida.
- La alegría se manifiesta con un incremento del tono medio y en su rango, así como un incremento en la velocidad de locución y la intensidad.
- El tono triste exhibe un tono medio mas bajo de lo normal, un estrecho rango de variación y una velocidad de locución lenta.
- El miedo presenta un tono medio más elevado, un rango mayor y una velocidad de locución rápida.
- La sorpresa presenta un tono de voz mayor que la voz neutra, velocidad igual a la voz neutra y un rango amplio de variación.

Los cuentos contados por actores se han analizado utilizando la herramienta Praat. Se han seleccionado tres cuentos narrados por actores, dos de ellos leídos en español y otro en inglés. La duración de estos tres cuentos tomados como base para nuestro análisis es de unos 15 minutos. Se han escogido cuentos largos para facilitar la aparición de todas las emociones que se deseaban modelar. En primer lugar se ha realizado una división por actores, para obtener emociones con las mismas características en la voz base. A partir de estas frases de cada actor se ha realizado una segunda división por emociones. Estas divisiones proporcionan dos o tres frases por actor y por emoción, haciendo posible un análisis de las características concretas de cada emoción en los distintos actores. Este análisis proporciona para cada emoción y cada actor valores para las siguientes características de la voz: frecuencia fundamental media, rango de variación y velocidad. Esto constituye una primera aproximación a las características de la voz en cada emoción.

Para obtener resultados más genéricos se calcula la variación de los valores de los parámetros en cada una de las emociones con respecto a la voz base. Por último, se adaptaron estos parámetros a las características de la voz que tomamos como base, la voz de nuestro sintetizador, y se realizo un ajuste a través de la variación sistemática de los parámetros hasta obtener unos valores óptimos.

En la Tabla 1 se muestran los valores de los parámetros de la prosodia que hemos variado para obtener las emociones.

	Volumen	Velocidad (Pal/min.)	Pitch	Rango Pitch
Enfado	1.0	145	100	30
Sorpresa	1.0	120	125	20
Alegría	1.0	155	135	14
Tristeza	0.8	110	90	7
Miedo	1.0	135	175	24

Tabla 1: Parámetros de la prosodia en las emociones

2.3 Síntesis expresiva de cuentos

El sintetizador utilizado para nuestro cuenta-cuentos - el paquete javax.speech.synthesis, del motor FreeTTS - recibe como entrada uno de los cuentos generados, con las emociones de cada frase ya marcadas, y dependiendo de cada emoción fijara los parámetros de la voz para, a continuación, sintetizar la frase.

Para cada una de las emociones el sintetizador del cuenta-cuentos varía los parámetros de la prosodia descritos en el apartado anterior: frecuencia fundamental, rango de la frecuencia, velocidad y volumen.

El sintetizador narra los cuentos en inglés, y las características de la voz son las de un varón adulto, se trata de una voz seria, con un tono de voz bajo y un rango de variación pequeño. A continuación en la Tabla 2 se muestran las características de la voz neutral de nuestro sintetizador:

	Volumen	Velocidad (Pal/min.)	Pitch	Rango Pitch
Neutral	0.9	120	100	11

Tabla 2: Parámetros de la prosodia en la emoción neutral

2.4 Evaluación de la expresividad conseguida

Se han tratado de medir dos aspectos fundamentalmente:

- Grado de reconocimiento de cada una de las emociones.
- Influencia del significado en el reconocimiento de emociones.

Para conseguir estas medidas se han hecho una serie de distinciones que atienden a los textos que intervienen y el tipo de pruebas realizadas.

Los textos que han formado parte de las pruebas son de dos tipos:

- Uno de los cuentos generados por el cuenta-cuentos, "Hansel y Gretel". Se le ha realizado una pequeña modificación que consiste en eliminar algunas de las frases neutras para obtener una menor duración. Se ha seleccionado este cuento y no otro por tratarse de un cuento que reunía todas las emociones modeladas y por ser el cuento que más frases emocionantes contenía de entre todos los generados.
- La misma frase leída con cada una de las 5 emociones modeladas. En este caso se ha seleccionado una frase al azar, sin ningún significado emocional: "*My name is Virginia and I need your help in order to try this product*".

Mediante esta división se pretende medir el impacto del significado en la adjudicación de emociones, así como el impacto que tienen la consecución de emociones intercaladas con texto neutro.

Se han realizado dos tipos de pruebas:

- Pruebas de elección forzada: Se da a elegir entre las 5 emociones modeladas y los usuarios deben seleccionar una de estas 5 emociones.
- Pruebas de elección libre: A la lista anterior se han añadido 3 emociones mas: culpa, vergüenza y envidia así como la categoría "Otras".

De este modo se pretende medir por un lado la capacidad de reconocimiento de las emociones modeladas así como la aparición de confusiones con emociones no modeladas. Tanto en las pruebas de elección forzada como las de elección libren existen frases sin ningún contenido emocional para lograr una medida del impacto de la prosodia en la percepción.

En las pruebas han intervenido un total de 10 personas, tanto hombres como mujeres. Se trataba de personas de entre 15 y 50 años, entre ellos había ingenieros informáticos, ingenieros en sonido e imagen, enfermeras, amas de casa, y estudiantes. En primer lugar se realizaron las pruebas de elección libre sobre un cuento y sobre la frase repetida 5 veces cada una con una de las cinco emociones modeladas. Mientras escuchaban el cuento iban marcando todas aquellas frases que consideraban "emocionantes" asignándoles la emoción que se correspondía con lo que estaban percibiendo en esos momentos. En el caso de las frases, se solicito a los evaluadores que fijasen para cada una de las cinco frases la emoción que considerasen oportuna. Una vez realizadas las pruebas de elección libre se repitió el cuento y las frases, pero esta vez con un conjunto reducido de posibles emociones a seleccionar, tan solo las 5 que se habían modelado.

Los resultados obtenidos se muestran en las Tablas 3, 4, 5 y 6. Las filas representan las emociones[6] que se pretendían modelar, y las columnas representan las emociones percibidas por los evaluadores.

	E	S	A	T	M	V	C	EN
E	90%	0%	0%	0%	0%	0%	10%	0%
S	23%	53%	5%	0%	12%	0%	2%	5%
A	25%	10%	55%	0%	0%	0%	0%	10%
T	0%	0%	0%	88%	0%	10%	2%	0%
M	35%	15%	0%	0%	40%	0%	0%	10%

Tabla 3: Cuento con elección libre

	E	S	A	T	M	V	C	EN
E	60%	0%	30%	0%	0%	0%	0%	10%
S	0%	10%	10%	20%	10%	30%	20%	0%
A	20%	20%	10%	0%	20%	10%	10%	10%
T	0%	0%	0%	70%	0%	20%	10%	0%
M	0%	0%	0%	0%	50%	10%	20%	20%

Tabla 4: Frases con elección libre

[6] E=Enfado, S=Sorpresa, A=Alegría, T=Tristeza, M=Miedo, V=Vergüenza, C=Culpa, EN=Envidia

	Enfado	Sorpresa	Alegría	Tristeza	Miedo
Enfado	90%	10%	0%	0%	0%
Sorpresa	20%	70%	5%	0%	5%
Alegría	20%	0%	80%	0%	0%
Tristeza	0%	0%	0%	100%	0%
Miedo	25%	35%	0%	0%	40%

Tabla 5: Cuento con elección forzada

	Enfado	Sorpresa	Alegría	Tristeza	Miedo
Enfado	70%	0%	30%	0%	0%
Sorpresa	0%	50%	20%	0%	30%
Alegría	10%	30%	50%	0%	10%
Tristeza	0%	0%	0%	100%	0%
Miedo	20%	20%	0%	0%	60%

Tabla 6: Frases con elección forzada

La primera afirmación que podemos realizar a la vista de los resultados es que las emociones siempre se diferencian del texto neutro, no ha habido confusión en este sentido. También podemos afirmar que existe una pequeña diferencia entre los resultados de los cuentos y los de las frases: los aciertos en los cuentos son más elevados que en las frases lo que nos lleva a la conclusión de que el significado influye de manera positiva en la asociación de una emoción a unas determinadas características de la voz. En cuanto al reconocimiento de cada una de las emociones los resultados que se extraen de las pruebas anteriores son los siguientes.

El *enfado* es una de las emociones mas reconocidas, tan solo se confunde, con un porcentaje significativo, con la alegría. A la hora de evaluar el resto de emociones, todas ellas, excepto la tristeza, se confunden en algún momento de las pruebas con el enfado. Esto quizá es debido a que la voz neutra que se ha elegido como base es una voz muy seria, con una frecuencia fundamental baja, lo que conlleva que cualquier emoción tenga una percepción de seriedad.

La *sorpresa* es una de las emociones menos logradas, en los experimentos realizados tan solo obtiene un porcentaje de acierto elevado en el experimento de elección forzada con el cuento, es decir, en el reconocimiento de la sorpresa intervienen tanto el significado como la existencia de un conjunto restringido de emociones.

La *alegría* es una de las emociones menos reconocidas, obteniendo un reconocimiento especialmente bajo en el experimento de elección libre con las frases, esto es debido a que cuando existen muchas posibilidades y el significado no ayuda a determinar la emoción, en el caso de las frases alegres se realiza la asignación de emoción de manera aleatoria, ya que se confunde con todas las emociones ofrecidas como posibles.

Sin duda alguna la *tristeza* es la emoción más lograda, obteniendo un 100% de reconocimiento en los experimentos de elección forzada.

La media de acierto en el caso del *miedo* ronda en todos los experimentos el 50%, las confusiones mas importantes se dan con el enfado y la sorpresa.

3 Conclusiones

Una vez realizadas todas las pruebas llegamos a la conclusión de que habría que estudiar más a fondo la sorpresa y la alegría ya que no son fácilmente identificables. El enfado, la tristeza y el miedo consiguen buenos resultados, con unos ligeros retoques podrían llegar a ser reconocidas sin ninguna duda. De todos modos, obtener un 100% de reconocimiento es muy difícil, ya que el reconocimiento de emociones es tarea subjetiva y bastante complicada. Los evaluadores han comentado la dificultad que tienen las personas para determinar las emociones que percibimos.

El siguiente paso sería realizar un análisis más exhaustivo de los valores de los parámetros de la voz que intervienen en las emociones. La asignación de las emociones en los cuentos tomados como base para el análisis debería ser realizada por un conjunto más extenso de personas, de modo que esta asignación sea más general y menos subjetiva. Sería bueno que estas mismas personas fuesen en la última fase las que realizasen la evaluación. Otro punto importante a tener en cuenta en un trabajo futuro sería el de buscar más voces de base, tanto de mujeres como de hombres, y con ellas llegar a unas emociones más reconocibles, sobre todo en lo que a la alegría y la sorpresa se refiere. Una vez conseguido este objetivo, la siguiente meta será el modelado de emociones subyacentes.

Bibliografía

Alter, K., Rank, E., Kotz, S. A., Toepel, U., Besson, M., Schirmer, A., and Friederici, A. D. (2000). Accentuation and emotions – two different systems? In Proceedings of the

ISCA Workshop on Speech and Emotion, pages 138–142, Northern Ireland. http://www.qub.ac.uk/en/isca/proceedings

Cahn, J. E. (1990). The generation of affection synthesized speech. Journal of the American Voice I/O Society, 8:1-19.

Campbell, N. (2000). Databases of emotional speech. In Proceedings of the ISCA Workshop on Speech and Emotion, pages 34-38, Northern Ireland. http://www.qub.ac.uk/en/isca/proceedings

Cowie, R. and Cornelius, R.R. (2003) Describing the emotional states that are expressed in speech. Speech Communication Special Issue on Speech an Emotion, 40(1-2): 5-32.

Fairbanks, G. and Pronovost, W. (1939). An experimental study of the pitch characteristics of the voice during the expression of emotion. Speech Monograph, 6:87-104.

Gervás, P., Díaz-Agudo, B., Peinado, F. and Hervás, R. (2004) Story plot generation based on CBR. In A.Macintosh, R. Ellis, and T.Allen, editors, 12th Conference on Applications and Innovations in Intelligent Systems, Cambridge, UK. Springer, WICS series.

Montero, J.M. (2003) Estrategias para la mejora de la naturalidad y la incorporación de variedad emocional a la conversión texto a voz en castellano. Tesis Doctoral, ETSI Telecomunicación, UPD.

Mozziconacci, S. J. L. (1998). Speech Variability and Emotion: Production and Perception. PhD thesis, Technical University Eindhoven.

Murray, I. R. and Arnott, J. L. (1995). Implementation and testing of a system for producing emotion-by-rule in synthetic speech. Speech Communication, 16:369-390.

Peinado, F., Gervás, P., Díaz-Agudo, B.: "A Description Logic Ontology for Fairy Tale Generation". In Veale T., Cardoso A., Camara Pereira, F., Gervás, P. (Eds.): Procs. of the Workshop on Language Resources for Linguistic Creativity, LREC'04, 56-61. 29th May, Lisbon, Portugal. LREA, 2004.

Ortony, A., Clore, G.L., and Collins, A. (1988). The Cognitive Structure of Emotion. Cambridge University Press, Cambridge, UK.

Scherer, K. R. (1979). Personality markers in speech by K. R. Scherer and H. Giles (eds): Social markers in speech. Cambridge: Cambridge University Press.

Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. Review of Personality and Social Psychology, 5:37-63.

Scherer, K. R., Ladd, D. R., and Silverman, K. (1984). Vocal cues to speaker affect: Testing two models. Journal of the Acoustic Society of America, 76(5):1346-1356.

Schröder, M. (1999). Can emotions be synthesized without controlling voice quality? Phonus 4, Research Report of the Institute of Phonetics, University of the Saarland, pages 37-55. http://www.dfki.de/~schroed

Smith, C.A. (1989) Dimensions of appraisal and physiological response in emotion. Journal of Personality and Social Psychology, 56(3):339-353.

Stallo, J. (2000). Simulating emotional speech for a talking head. Honour's thesis, School of Computing, Curtin University of Technology, Australia. http://www.computing.edu.au/~stalloj/projects/honours

Wundt, W. (1896). Grundriss der Pshychologie. Verlag von Wihelm Engelmann, Leipzig.