Ajuste Subjetivo de Pesos para Selección de Unidades a través de Algoritmos Genéticos Interactivos

Francesc Alías
Dpto. Comunicaciones y Teoría de la Señal
Enginyeria i Arquitectura La Salle
Universidad Ramon Llull
Pg. Bonanova 8, 08022-Barcelona
falias@salleURL.edu

Xavier Llorà
Illinois Genetic Algorithms Laboratory
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign
104 S. Mathews, Urbana, 61801 IL
xllora@illigal.ge.uiuc.edu

Ignasi Iriondo
Dpto. Comunicaciones y Teoría de la Señal
Enginyeria i Arquitectura La Salle
Universidad Ramon Llull
Pg. Bonanova 8, 08022-Barcelona
iriondo@salleURL.edu

Lluís Formiga
Departamento de Informática
Enginyeria i Arquitectura La Salle
Universidad Ramon Llull
Pg. Bonanova 8, 08022-Barcelona
is08148@salleURL.edu

Resumen: este trabajo se sitúa en el marco de los sistemas de síntesis concatenativa del habla basados en selección de unidades. Concretamente, se ha desarrollado una interfaz que permite establecer los pesos que ponderan los parámetros que intervienen en la función de coste del módulo de selección de unidades, mediante la incorporación de algoritmos genéticos interactivos. De este modo, el proceso de selección incorporará el criterio subjetivo de los usuarios finales del sistema. La aplicación se ha desarrollado bajo una plataforma *web* y se ha distribuido en distintos servidores para poder ofrecer un buen rendimiento y una alta portabilidad.
Palabras clave: conversión texto-habla, selección de unidades, función de coste, algoritmos genéticos interactivos

Abstract: the work presented in this paper deals with text-to-speech systems based on unit selection. The quality of the synthesis relies on having an accurate unit selection process. Usually, the quality of this procedure can be tuned by adjusting a set of weights that control the selection process. However, in order to achieve a good quality, the tuning process must take into account some subjective dimensions. Interactive genetic algorithms overcome this issue, allowing the user to take active part in the tuning process. With the fusion of the tuning technique and the final user (by means of a web interface), the unit selection can be adjusted to trap the subjective elements that lead to a high quality synthesis.
Keywords: text-to-speech, unit selection, cost function, interactive genetic algorithms

1. Introducción

La mayor parte de los sistemas de conversión texto-habla (CTH) actuales utilizan una estrategia de síntesis concatenativa basada en la selección de las mejores unidades de un corpus de voz (Black y Taylor, 1997; Beutnagel et al., 1999; Coorman et al., 2000; Quazza et al., 2001). Estos sistemas pretenden mejorar las limitaciones de los CTH basados en difonemas (con una única realización por unidad) al incluir un gran número de realizaciones para cada una de las unidades del corpus. De este modo, se consigue partir de secuencias de unidades más largas y adecuadas a la prosodia deseada para la síntesis, minimizando la necesidad de modificar las unidades del corpus de voz. En general, los sistemas de CTH han conseguido obtener una señal de voz de alta calidad, aunque todavía existen situaciones donde la síntesis presenta problemas (Black y Lenzo, 2000). Esto es debido a que esta tecnología está todavía en desarrollo, siendo aún necesario trabajar en el ajuste de los parámetros que intervienen en el proceso de selección (Black, 2002).

Este trabajo se centra en el estudio del módulo de selección de unidades, elemento fundamental de este tipo de CTH. Concretamente se diseña la función de coste de selección y se entrenan los pesos que la ponderan.

Estos cálculos son los que permiten escoger la secuencia de unidades *óptima* del corpus de voz. Por lo tanto, es esencial definir eficientemente su diseño y entrenamiento. En cuanto a los pesos, sus valores se pueden asignar de forma manual o a través de un proceso automático. Hasta el momento, los métodos automáticos utilizados se han basado en el ajuste de los pesos mediante una función que mide la calidad de la señal de forma objetiva (Hunt y Black, 1996; Meron y Hirose, 1999). En este trabajo se presenta una nueva aproximación para el entrenamiento automático de los pesos de selección, incorporando la valoración subjetiva mediante la inclusión de un algoritmo genético interactivo (Takagi, 2001).

El apartado 2 introduce la aplicación de los algoritmos genéticos para resolver el ajuste de parámetros. El apartado 3 presenta los elementos que forman el módulo de selección de unidades de un sistema CTH. Seguidamente, el apartado 4 describe el algoritmo genético interactivo utilizado en este trabajo. En el apartado 5 se detalla la implementación de la plataforma de *test*. Finalmente, se analizan las pruebas preliminares realizadas (apartado 6), como paso previo a una próxima experimentación masiva, y se presentan las conclusiones de este trabajo (7).

2. Subjetividad, evolución y síntesis del habla

La creación de un CTH de alta calidad comporta afrontar distintas problemáticas. Algunas involucran conceptos puramente objetivos y cuantificables, como pueden ser la trascripción fonética o el análisis del corpus de voz. Otras, por el contrario, involucran parámetros que pertenecen al dominio perceptivo, siendo estrictamente subjetivos y difícilmente medibles. Por ejemplo, la elección de los parámetros de la señal de voz considerados (función de coste) o la importancia que éstos deben tomar (pesos) en el proceso de selección de unidades.

Debido a que la calidad de la síntesis depende en gran medida del valor que tomen estos parámetros, es importante escoger un método que permita el ajuste de su valores (cuantificados numéricamente) a partir de una percepción subjetiva. Aunque este objetivo pueda parecer contradictorio, la evolución artificial (Holland, 1975), y en especial los algoritmos genéticos (Goldberg, 1989; Goldberg, 2002), permiten afrontar este propósito de una forma simple y eficiente.

Estos algoritmos se basan en dos procesos naturales inspirados en la evolución natural de las especies. El primero proviene del principio de la selección natural, por la que los individuos mejor adaptados tienen más probabilidad de sobrevivir. En este caso, sobrevivir significa disponer de una mayor probabilidad de generar descendencia. Por consiguiente, permanece el material genético de estos individuos dentro de la población, guiando así su proceso de adaptación al medio. El segundo proceso es la creación de nuevas soluciones (descendientes) que recombinen aquellos rasgos de los progenitores que han permitido su adaptación y su supervivencia. De este modo, consiguen adaptar (o ajustar) la población de individuos (soluciones potenciales) a un determinado entorno (o problema).

En este trabajo se utiliza este tipo de algoritmos para el ajuste de los pesos que intervienen en la función de coste de selección (descrita en el apartado 3). Entonces, desde el punto de vista evolutivo, se dispondrá de una población de pesos que debe ser adaptada al ámbito de la síntesis de alta calidad. Hasta este punto no hay nada que permita resolver la incorporación de la dimensión subjetiva en el ajuste de dichos parámetros. Este objetivo se puede abordar fusionando el proceso de selección de los individuos mejor adaptados con la incorporación de la evaluación humana (Takagi, 2001) (ver apartado 4). De este modo, es el usuario el que escoge a los mejores individuos, es decir, aquella configuración de los parámetros que le proporciona una mayor calidad de síntesis. Es importante resaltar en este punto que dicha elección se realiza en el dominio psicológico subjetivo del usuario. De esta forma, el algoritmo captura esta subjetividad en el proceso evolutivo de ajuste de los parámetros que controlan el CTH.

3. Selección de unidades

Los dos elementos fundamentales para un CTH basado en selección de unidades son: (i) el corpus de voz, que debe cubrir la máxima variabilidad lingüística y prosódica del ámbito de aplicación del conversor (Black, 2002) (p.e., todo un idioma, para síntesis de propósito general); (ii) el módulo de selección de unidades, encargado de escoger, mediante programación dinámica, el conjunto de realizaciones del corpus de voz que minimizan una función de coste (Hunt y Black, 1996).

Esta función toma en consideración la similitud entre la unidad candidata (u_i) y la objetivo (t_i) mediante el coste de unidad (C^t, ecuación 1) y el grado de continuidad entre unidades consecutivas (u_{i-1}, u_i) a través del coste de concatenación (C^c, ecuación 2).

$$C^t(t_i, u_i) = \sum_{j=1}^{p} w_j^t C_j^u(t_i, u_i) \qquad (1)$$

$$C^c(u_{i-1}, u_i) = \sum_{j=1}^{q} w_j^c C_j^c(u_{i-1}, u_i) \qquad (2)$$

3.1. Subcostes de selección

El coste de selección se obtiene mediante la integración de los costes C^t y C^c, que a su vez se calculan como un sumatorio ponderado de p y q subcostes, tal y como se ha indicado en las ecuaciones (1) y (2), respectivamente. Por lo tanto, el diseño de los subcostes y de sus ponderaciones, es un elemento crítico para conseguir una síntesis de alta calidad.

De las posibles medidas propuestas para el cálculo de estos subcostes (Coorman et al., 2000; Blouin et al., 2002; Peng, Zhao y Chu, 2002), en este trabajo sólo se toman en consideración aquellas que afectan a la información prosódica de las unidades. De este modo, se simplifica el proceso de selección y se enfatiza la importancia del ajuste de los pesos.

Para C^t, se han calculado las diferencias entre los valores medios de frecuencia fundamental (F_0 o *pitch*), energía y duración de las unidades comparadas, respecto a la desviación del parámetro considerado en el corpus de voz utilizado (ver ecuación 3).

$$C_j^u(t_i, u_i) = \frac{\left|\overline{P_j}(t_i) - \overline{P_j}(u_i)\right|}{\overline{\sigma}_{P_j}} \qquad (3)$$

Para C^c, se han evaluado las diferencias de los valores de F_0, energía y los coeficientes cepstrales en la escala Mel (MFCC) en el punto de concatenación (L, *left*, y R, *right*, en la ecuación 4), normalizadas según su desviación a lo largo del corpus.

$$C_j^c(u_{i-1}, u_i) = \frac{\left|P_j^R(u_{i-1}) - P_j^L(u_i)\right|}{\overline{\sigma}_{P_j}} \qquad (4)$$

3.2. Ponderación de los subcostes

La importancia que toma cada uno de los subcostes en la elección de la secuencia de unidades del corpus se evalúa mediante el peso que lo acompaña en la función de coste (ver ecuaciones 1 y 2). Luego, el entrenamiento eficiente de estos pesos es fundamental para escoger las mejores unidades del corpus de voz, según la prosodia deseada.

Existen distintas aproximaciones para el ajuste de estos pesos, algunas de ellas se basan en un proceso manual supervisado perceptualmente (Coorman et al., 2000; Blouin et al., 2002), y otras parten de un sistema automático. En este segundo contexto, destacan dos aproximaciones presentadas en el trabajo de Hunt y Black (1996): la búsqueda del espacio ponderado o *weighted space search* (WSS) y el cálculo mediante una regresión multilínea (MLR).

WSS realiza una discretización del espacio multidimensional de posibles pesos ($|\mathcal{W}|^{p+q}$), con un conjunto finito de pesos potenciales $\mathcal{W}$. El conjunto óptimo de pesos se obtiene mediante un proceso de análisis (elección de unidades) y síntesis (generación de voz) que evalúa la bondad de los resultados de forma objetiva. En cambio, MLR, aplicado inicialmente para el cálculo de los pesos de C^t, escoge aquellos pesos que mejor mapean, según una regresión multilínea, la función de coste respecto a la distancia que evalúa la similitud de las unidades (distancia cepstral).

Posteriormente, Meron y Hirose (1999) mejoraron las prestaciones de ambos métodos y los adaptaron para el entrenamiento de pares de unidades, considerando la modificación prosódica de la secuencia de unidades posterior al proceso de selección. Recientemente, se ha introducido el ajuste objetivo de pesos mediante algoritmos genéticos en el trabajo presentado por (Alías y Llorà, 2003).

Pero el principal punto débil de estos métodos es su funcionamiento basado en una estimación objetiva de características que el usuario percibe subjetivamente. Por ello, se propone una solución que permita la interacción humana en el proceso de ajuste de los pesos de selección.

4. Algoritmos genéticos interactivos

Los algoritmos genéticos interactivos (IGA) son un modelo de optimización capaz de combinar el ajuste de parámetros cuantitativos con la evaluación subjetiva de los resultados. Este tipo de algoritmos han sido aplicados en gráficos por ordenador, en ingeniería mecánica, o en procesado de la señal, entre otros (Takagi, 2001). Concretamente, han sido am-

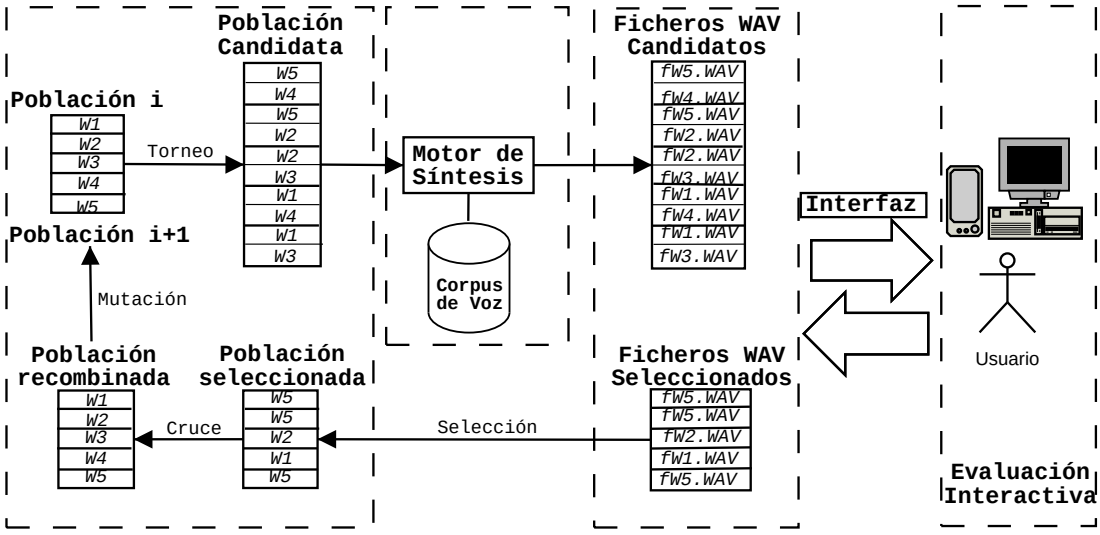


Figura 1: Diagrama de funcionamiento del IGA para el ajuste de los pesos de selección.

pliamente explorados en sistemas de procesado del habla (Watanabe y Takagi, 1995; Sato, 1997; Todoroki y Takagi, 2000). En algunos de ellos, los IGA han sido utilizados para el ajuste de coeficientes en filtros FIR. En otros, han sido empleados para el ajuste de parámetros de control para la incorporación de la emotividad al procesado de la señal.

El IGA utilizado en este trabajo se ha diseñado para el ajuste de los pesos de los subcostes de la función de selección de unidades del CTH. La estructura de dicho algoritmo es la misma que se puede encontrar en un algoritmo genético convencional (Goldberg, 1989; Goldberg, 2002). Se basa en la evolución de un vector de individuos $w = (w_0, \dots w_n)$, que corresponde a los pesos utilizados en la función de coste, hasta hallar su configuración $\mathcal{W} = (w_1^t, \dots w_p^t, w_1^c, \dots w_q^c)$ óptima.

El proceso de evolución se divide en dos fases: (i) la selección de la mejores soluciones contenidas en la población, y (ii) su posterior recombinación para generar nuevas soluciones (ver figura 1). Es en el proceso de selección donde los IGA presentan sus peculiaridades. En cada una de las iteraciones del proceso de ajuste, el algoritmo dispone de un conjunto de pesos w_i para la síntesis del texto estudiado (CTH). El resultado de dicha síntesis es evaluado interactivamente por el usuario al que se le presentan las posibles soluciones de la iteración según un proceso de elección por *torneo* binario. De este modo, el usuario escogerá, después de escuchar varias veces las dos soluciones emparejadas, aquella opción (vector de pesos w) que presente una mejor señal de voz desde el punto de vista subjetivo.

En cuanto al proceso de recombinación en el IGA, éste sigue la misma directiva que en un algoritmo genético tradicional. Este proceso se basa en la recombinación probabilística del material genético (en nuestro caso el conjunto de pesos w). El mecanismo utilizado consiste en el intercambio de fragmentos de material genético procedentes de dos progenitores (*población seleccionada*). En este caso, el operador de cruce utilizado es el clásico de punto de *cruce* único (Goldberg, 1989). Este proceso viene acompañado por la incorporación de posibles errores en dicho proceso de recombinación (*mutación*). Esto se consigue perturbando probabilísticamente fragmentos del material genético, afectando a algunos de los pesos (Goldberg, 1989).

5. *Plataforma de test*

En este apartado se detalla a grandes rasgos la plataforma que se ha desarrollado para llevar a cabo el ajuste de pesos mediante una evaluación subjetiva de los resultados obtenidos. Se ha optado por diseñar una plataforma *web*, como en el trabajo de Jilka y Syrdal (2002), distribuida en distintos servidores para poder ofrecer un alto rendimiento y una alta portabilidad. A continuación se describe el diseño de la plataforma así como las tecnologías utilizadas.

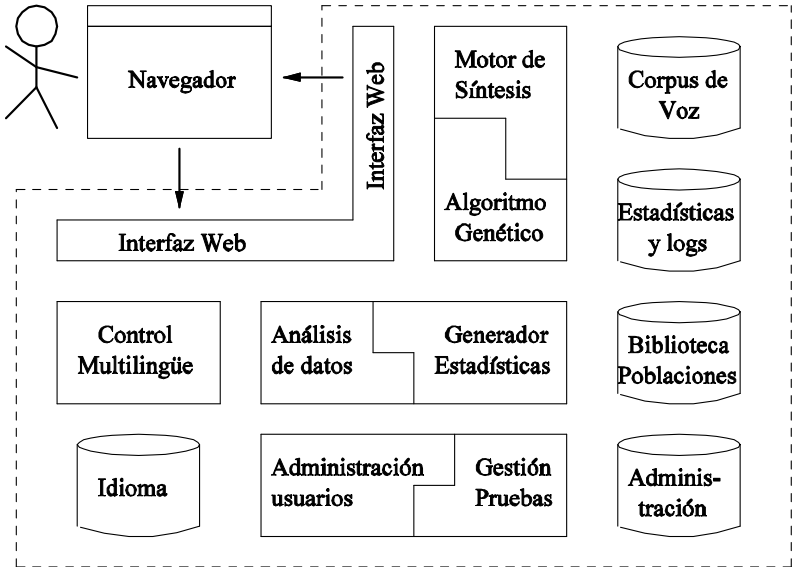


Figura 2: Diagrama de funcionamiento del IGA para el ajuste de los pesos de selección.

5.1. Diseño de la plataforma

Para contemplar todas las necesidades del problema planteado, se ha desarrollado una plataforma a partir de distintos módulos de proceso, acceso y gestión. Entre otros, la plataforma incorpora un interfaz de usuario que permite el acceso remoto (*web*), un conversor texto-habla o un algoritmo genético interactivo. En la figura 2 se presentan todos estos módulos, junto a los distintos corpus (de voz o de datos) que los acompañan.

5.1.1. Interfaz *web*

Se ha desarrollado mediante un conjunto de *scripts* que cargan distintas plantillas *web* en función de los privilegios del usuario (Administrador o Usuario), el idioma con el que se quiere trabajar y la acción que se esté realizando en aquel momento (análisis, realización o administración de pruebas).

5.1.2. Motor de síntesis

Está formado por el bloque de procesado de la señal de un CTH. Parte de la información prosódica del texto de entrada y, mediante la selección de unidades, escoge la secuencia de difonemas y trifonemas del corpus de voz. Dicha información es explorada y recuperada del corpus de voz insertado en la plataforma.

5.1.3. Algoritmo Genético

Este módulo implementa la máquina de estados necesaria para la incorporación de un algoritmo genético en el ajuste de los pesos de selección. Debido al entorno en el que se utiliza (*web*), el estado del proceso evolutivo se almacena en una base de datos relacional. Esta necesidad permite, al mismo tiempo, almacenar la traza del proceso evolutivo seguido bajo la guía del usuario.

5.1.4. Análisis de datos

La plataforma incorpora un bloque que permite consultar de forma gráfica los resultados obtenidos por los usuarios a lo largo de las pruebas. Para presentar los resultados de las pruebas realizadas se han incorporado varios tipos de análisis estadísticos: histogramas, correlaciones, etc.

5.1.5. Generador de estadísticas

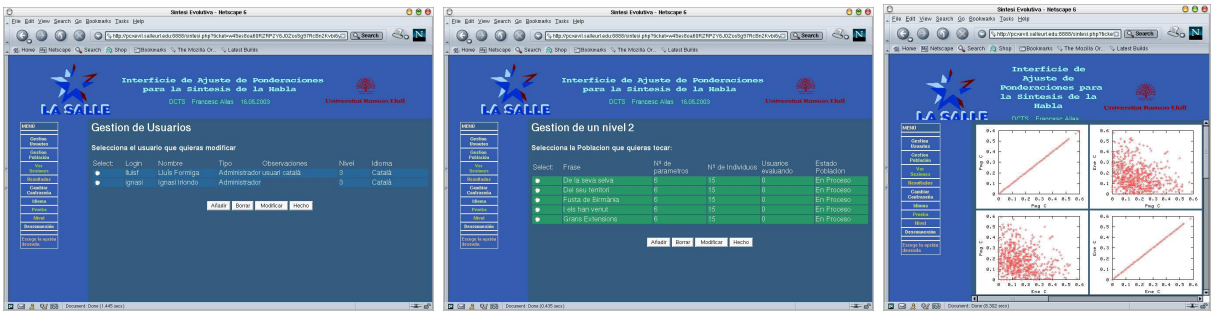
Obtiene las estadísticas de uso de la plataforma, tales como: número de accesos, pruebas que se están realizando, archivos sintetizados, memoria disponible, etc. Información sólo consultable por el administrador.

5.1.6. Gestión de pruebas

Mediante este módulo se gestionan los parámetros de configuración de las pruebas a realizar: el texto a sintetizar, la configuración del nivel de dificultad y los parámetros del algoritmo genético interactivo, entre otros.

5.1.7. Administración de usuarios

Este elemento de la plataforma se encarga de la gestión de los usuarios que pueden acceder a la aplicación. Además, permite asignar el perfil de usuario (idioma, nivel de dificultad...) así como otros parámetros configurables de la interfaz de usuario.



(a) Gestión de usuarios

(b) Gestión de pruebas

(c) Análisis de resultados de las pruebas

Figura 3: Pantallas ilustrativas de la plataforma desarollada.

5.2. Tecnologías utilizadas

Las tecnologías utilizadas para el diseño de los tres módulos principales que componen la plataforma *web* son: (i) CGI (*Common Gateway Interface*) para la interfaz *web* (programada mediante PHP), (ii) una base de datos relacional (MySQL, 2003) que almacena toda la información de la plataforma, y (iii) C ++ compilado en GCC con el que se implementa el bloque de síntesis.

6. Experimentos

Los experimentos se han realizado sobre un corpus de voz en catalán formado por 1520 frases (unas 10000 unidades), que no ha sido diseñado explícitamente para ser utilizado en un sistema de CTH basado en selección de unidades. En este contexto, cabe destacar que no todas las unidades del corpus (en este caso difonemas y trifonemas) presentan un número suficiente de realizaciones para disponer de la variabilidad necesaria para las pruebas.

Por este motivo, los experimentos realizados pretenden (i) validar la plataforma desarrollada y la viabilidad (convergencia) del ajuste subjetivo de pesos mediante IGA y (ii) obtener las conclusiones necesarias para el diseño de la experimentación a gran escala sobre la que se está trabajando actualmente.

Teniendo presente el objetivo de los experimentos desarrollados, la pruebas han sido realizadas por tres usuarios expertos que han evaluado 5 frases extraídas de un documental de televisión. La entrada al sistema de síntesis del habla ha consistido en la trascripción fonética de dichas frases junto con la prosodia obtenida de las frases originales. En cada paso el usuario debe escoger el mejor individuo de dos posibles candidatos, teniendo como patrón de comparación la frase original. Los resultados obtenidos hasta el momento nos han permitido llegar a una serie de conclusiones que serán la base de la futura experimentación así como para el ajuste de algunas funcionalidades de la plataforma.

En relación al proceso de realización de las pruebas, hay que tener en cuenta las consideraciones siguientes, fruto de la puesta en común de ideas por parte de los tres usuarios expertos:

- Resulta complicado mantener el mismo criterio de comparación de individuos a lo largo de toda la prueba. Asimismo, resulta difícil que los criterios entre los usuarios sean del todo comunes.
- La aparición de un error en una palabra de la frase (un pequeño ruido, un fonema erróneo, ...) implica descartar esa realización en beneficio de la otra. Este problema puede no ser debido al vector de pesos sino a un fallo en la segmentación y/o etiquetado del corpus oral.
- En determinadas frases y después de un cierto número de iteraciones, la diferencia entre los candidatos (frases sintetizadas) es prácticamente imperceptible, por lo que la prueba se vuelve tediosa sin aportar nueva información. Esta situación viene motivada por el uso de un corpus de voz no diseñado explícitamente para selección de unidades, disponiendo de unidades con un reducido número de realizaciones.

- Las frases originales (extraídas del documental), de las que se obtiene la prosodia objetivo, pertenecen a un locutor distinto al del corpus de voz utilizado para la síntesis. Sería conveniente que ambos corpus fuesen grabados por el mismo locutor para poder llevar a cabo una comparación más precisa en cuanto al ritmo y a la entonación del habla.

En la figura 4 se muestran los valores medios obtenidos para los pesos según cada usuario después de siete generaciones. Analizando los valores obtenidos al promediar los resultados de los tres usuarios, se obtiene una primera valoración numérica de la tendencia que siguen los pesos mediante el método de ajuste presentado en este trabajo. Del análisis de este estudio preliminar se deduce lo siguiente:

- El peso con mayor valor es el Mel Frequency Cepstrum de Concatenación (MFC C) que sirve para garantizar una continuidad espectral entre las unidades.
- Seguidamente tenemos el Pitch Medio de Concatenación (PMG C) que nos indica la importancia que tiene la continuidad de F_0 entre unidades.
- El peso de unidad más importante es la duración (DUR T) que intenta considerar la velocidad del habla.
- El peso de Pitch Medio de Unidad (PMG T) parece tener menos importancia, si bien ya hemos comentado que la comparación de entonaciones no es sencilla.
- Los pesos con menos importancia son los asociados a la energía (ENE C y ENE T) debido a que el proceso de síntesis realiza un ajuste de la energía para adaptarla a la deseada. Por lo tanto no se descartan locuciones por su poca variabilidad energética y los pesos asociados tienden a disminuir. Estos pesos sirven para controlar el funcionamiento de la prueba y se podrían descartar para las nuevas pruebas.

7. *Conclusiones y trabajo futuro*

En este trabajo se ha presentado una plataforma para el entrenamiento de los pesos que intervienen en el proceso de selección de unidades de un CTH. Esta plataforma integra un módulo de síntesis del habla, encargado de generar los ficheros de voz comparados por el usuario, junto a un algoritmo genético interactivo, que evoluciona la población de pesos candidatos según el criterio subjetivo introducido por el usuario.

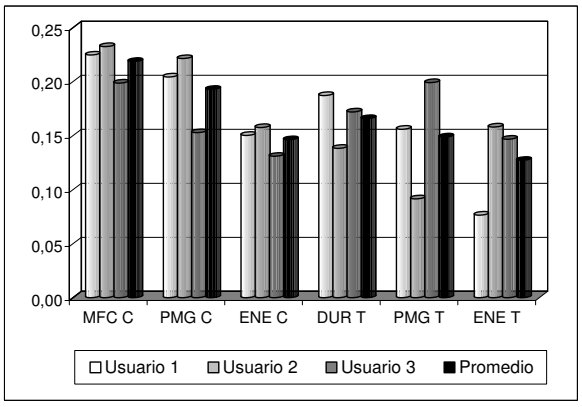


Figura 4: Promedio del valor de los pesos finales para cada usuario.

Además, la plataforma incluye una interfaz *web* que permite un acceso remoto a las pruebas (tanto para los usuarios como para el administrador), junto a la automatización del análisis de los resultados obtenidos. De este modo se flexibiliza la generación, supervisión, realización y evaluación de las pruebas para el análisis de la calidad de un CTH.

Asimismo, se ha validado la convergencia del método de ajuste de pesos a partir de la incorporación de algoritmos genéticos interactivos a este. Este método permite la integración de la subjetividad en el proceso de ajuste automático de los elementos que intervienen en la función de coste del módulo de selección de unidades de un CTH.

Los resultados obtenidos a partir de las pruebas preliminares realizadas han permitido depurar los módulos de la plataforma y descubrir las necesidades de mejora de las pruebas. Entre otras, el diseño de un *buen* corpus de voz, la configuración de las pruebas (frases, parámetros, etc.), el grupo de usuarios considerados o la dimensión y cobertura de las pruebas. Todas estas cuestiones deberán ser analizados en profundidad como base de la siguiente generación de experimentos, para disponer de unas pruebas más extensas para confirmar las conclusiones recabadas hasta el momento.

Agradecimientos

Este trabajo se ha realizado con el apoyo del Departament d'Universitats, Recerca i Societat de la Informació de la Generalitat de Catalunya mediante la beca 2000FI-00679. Además, se ha recibido el apoyo del Technology Research, Education and Commercialization Center, un programa de la Universidad de Illinois at Urbana-Champaign, administrado por el National Center for Supercomputing Applications (NCSA) y patrocinado por el Office of Naval Research (N00014-01-1-0175). Asimismo, agradecer el apoyo de la Air Force Office of Scientific Research, Air Force Materiel Command, USAF (F49620-00-0163), y la National Science Foundation (DMI-9908252).

Bibliografía

Alías, F. y X. Llorà. 2003. Evolutionary weight tuning based on diphone pairs for unit selection speech synthesis. En *EuroSpeech*, A celebrarse en Ginebra (Suiza).

Beutnagel, M., A. Conkie, J. Schroeter, Y. Stylianou, y A. Syrdal. 1999. The AT&T Next-Gen TTS system. En *Joint Meeting of ASA, EAA, and DAGA2*, páginas 18–24, Berlin, Germany.

Black, A.W. 2002. Perfect Synthesis for all of the people all of the time. En *IEEE TTS Workshop 2002 (Keynote)*, Santa Monica, USA.

Black, A.W. y K. Lenzo. 2000. Limited Domain Synthesis. En *ICSLP*, Beijing, China.

Black, A.W. y P. Taylor. 1997. Automatically clustering similar units for unit selection in speech synthesis. En *EuroSpeech*, páginas 601–604, Rodes, Greece.

Blouin, C., O. Rosec, P.C. Bagshaw, y C. d'Alessandro. 2002. Concatenation cost calculation and optimisation for unit selection in TTS. En *IEEE TTS Workshop*, Santa Monica, USA.

Coorman, G., J. Fackrell, P. Rutten, y B. Van Coile. 2000. Segment selection in the L&H RealSpeak laboratory TTS system. En *ICSLP*, volumen 2, páginas 395–398, Beijing, China.

Goldberg, D.E. 1989. *Genetic Algorithms in Search Optimization and Machine Learning.* Addison-Wesley.

Goldberg, D.E. 2002. *The Design of Innovation: Lessons from and for Competent Genetic Algorithms.* Kluwer Academic Publishers.

Holland, J.H. 1975. *Adaptation in Natural and Artificial Systems.* Univesity of Michigan Press.

Hunt, A. y A.W. Black. 1996. Unit selection in a concatenative speech synthesis system using a large speech database. En *ICASSP*, volumen 1, páginas 373–376, Atlanta, USA.

Jilka, M. y A. K. Syrdal. 2002. The AT&T German Text-to-Speech System: realistic linguistic description. En *ICSLP*, volumen 1, páginas 113–116, Denver, USA.

Meron, Y. y K. Hirose. 1999. Efficient weight training for selection based synthesis. En *EuroSpeech*, volumen 5, páginas 2319–2322, Budapest, Hungary.

MySQL. 2003. http://www.mysql.com.

Peng, H., Y. Zhao, y M. Chu. 2002. Perpetually optimizing the cost function for unit selection in a TTS system with one single run of MOS evaluation. En *ICSLP*, Denver, USA.

Quazza, S., L. Donetti, L. Moisa, y P. L. Salza. 2001. ACTOR©: a multilingual unit-selection speech synthesis system. En *The Fourth ISCA Workshop on Speech Synthesis*, Perthshire, Scotland.

Sato, Y. 1997. Voice conversation using evolutionary computation of prosodic control. En *Intelligent Processing of Manufacturing of Materials '97*, páginas 342–348.

Takagi, H. 2001. Interactive evolutionary computation: fusion of the capabilities of the ec optimization and human evaluation. *Proceedings of the IEEE*, 89(9):1275–1296.

Todoroki, Y. y H. Takagi. 2000. User interface of an interactive evolutionary computation for speech processing. En *6th International Conference on Soft Computing (IIZUKA2000)*, páginas 112–118.

Watanabe, T. y H. Takagi. 1995. Recovering sytem of the distorted speech using interactive genetic algorithms. En *IEEE, International Conference on Systems, Man and Cybernetics (SMC'95)*, volumen 1, páginas 684–689.