

CURSO-TALLER DE VERIFICACIÓN PERCEPTIVA DE CURVAS DE ENTONACIÓN ESTILIZADAS USANDO PHONEDIT

Versión: 15 de marzo de 2005
Sergio Coria

OBJETIVO

El objetivo de este curso-taller es adiestrar a los integrantes del grupo de etiquetación del Corpus DIME para realizar la verificación perceptiva de las curvas de entonación estilizadas mediante el algoritmo MOMEL [HIRST & ESPESSER, 1993], y producidas por los software M.E.S. (Motif Environment for Speech) [ESPESSER, 1999] y Phonedit [S.Q. Lab, 1998], ambos desarrollados en la universidad de Aix-en-Provence, Francia.

Phonedit es una versión comercial de su antecesor M.E.S., que inicialmente fué desarrollado para el funcionar sobre el sistema operativo Solaris. M.E.S. presenta algunas complicaciones para su instalación sobre Linux, y éste a su vez impone algunas restricciones respecto al hardware sobre el cual puede operar. Por ello, la verificación perceptiva se realizará usando Phonedit, que es compatible con Windows 98, lo cual facilita su instalación en las computadoras de nuestro grupo de etiquetación.

ALGUNAS DEFINICIONES

Curva de entonación: representación gráfica de la frecuencia fundamental (f_0) de una elocución. Se produce automáticamente mediante diversos algoritmos que analizan y dibujan la f_0 .

Curva de entonación estilizada: es la curva de entonación que con el mínimo grado de complejidad puede representar visual y auditivamente a la curva de entonación de una elocución. Descarta lo "micro" y conserva lo "macro".

Algoritmo MOMEL: Un método para producir curvas de entonación estilizadas. Basado en búsqueda de valores de f_0 que constituyan máximos locales o mínimos locales (puntos de inflexión o *targets*). Esos valores se unen con parábolas usando ecuaciones que se llaman *Quadratic Splines*. MOMEL es una abreviatura de "Modeling Melody".

INTRODUCCIÓN

Antes de iniciar la verificación perceptiva, es necesario que se tenga una carpeta con el nombre del diálogo que será analizado, y dentro de esa carpeta se deben colocar las carpetas que se especifican a continuación:

- a) *audios*: contiene los archivos de audio en formato WAV de las elocuciones del diálogo.
- b) *momel_default*: contiene los archivos *auto.cb*, que a su vez contienen los valores de tiempo y frecuencia de puntos de inflexión (*targets*), previamente producidos por el software M.E.S. Estos son los *targets* que generalmente serán sometidos a verificación perceptiva, aunque más adelante se explica que la verificación podría hacerse sobre *targets* provenientes de otro origen.
- c) *momel_verificados*: serán los *targets* resultantes del proceso de verificación.

RECOMENDACIONES GENERALES

Hay que ver y escuchar tanto la curva producida por Phonedit como la de momel_default, para después hacer la verificación perceptiva usando la que se vea y oiga más similar a la *f0* original. Las curvas de Phonedit y de momel_default pueden ser diferentes porque, aunque se producen con el mismo algoritmo (MOMEL), éste se ejecuta con diferentes valores de configuración para ciertos parámetros que determinan la selección de targets candidatos al generar la estilización.

Si desde el principio la curva estilizada (ya sea la de Phonedit o la de momel_default) se vieran y se oyeran como la de *f0* original, probablemente *no* será necesaria ninguna modificación de targets, excepto quizá el primero y el último, como se explica más adelante.

La verificación se hace considerando tres zonas de la curva: inicial, media y final. Tomando en cuenta el uso que se dará a los datos de targets dentro del proyecto DIME, para nosotros la parte más importante de la curva es la región final. Por ello, debemos verificarla con especial cuidado. También, porque al generar la estilización, las imprecisiones más comunes de MOMEL ocurren en los targets de la región final.

El procedimiento de verificación consiste en lograr que la curva estilizada (ya sea que se haya elegido la producida por Phonedit, o la de momel_default) suene lo más posible como la curva *f0* original. La restricción principal es que sólo se modifique la mínima cantidad posible de targets. Para determinar si la curva estilizada se parece o no a la original, debemos observar y escuchar ambas, de modo que la vista ayude al oído, comparando cómo se ven y cómo suenan. Lo principal es que suene igual.

Para hacer modificaciones tenemos permitido: mover, insertar o eliminar targets (no demasiados).

Los targets ubicados en zonas inicial o final de ruido/silencio *no* son tan importantes como todos los demás, pero deben analizarse para garantizar que en etapas posteriores del proyecto DIME las etiquetas INTSINT se generen en forma conveniente.

Después de cada cambio de target que se haga a la curva estilizada, hay que oírla para verificar perceptivamente si el cambio fué bueno. Si no lo fué, hay que regresar el target a su situación previa. Si fuera necesario, se puede volver a poner en la ventana MOMEL el conjunto de targets que se haya elegido para hacer la verificación: los de Phonedit o los de momel_default. Es recomendable ir salvando los targets en la carpeta de momel_verificados conforme se van haciendo modificaciones, aún cuando no se hayan terminado de revisar todos los targets de la elocución.

Generalmente cuando una curva estilizada está bien hecha, su contorno toca a la mayoría de las zonas más altas del contorno de la *f0* original.

Para lograr una buena visualización de las curvas original y estilizada conviene que las ventanas MOMEL y *f0* de Phonedit se vean del mismo tamaño entre sí. Para esto puede ser necesario modificar los valores de los parámetros de frecuencia mínima y frecuencia máxima, haciendo que cada uno de éstos sea igual al de la otra ventana.

Phonedit padece algunas deficiencias que hacen que se bloquee al estar analizando elocuciones demasiado ruidosas, muy cortas o con silencio demasiado largo al inicio o al final. Desafortunadamente en esas condiciones no se puede aplicar una solución sencilla. Más adelante se explican éstos y otros casos que imposibilitan hacer la verificación.

PROCEDIMIENTO DETALLADO

Para cada elocución, antes de abrir su audio en Phonedit hay que ver si el archivo de momel_default (*auto.cb*) contiene algún valor negativo (en cualquier renglón o en cualquier columna). El archivo se debe abrir en algún editor de texto como *Notepad* o *Wordpad*. Si existe algún valor negativo, se debe tomar nota del núm. de diálogo y de elocución. Después, se debe intentar crear los targets con Phonedit. Si se pueden crear, usarlos para hacer la verificación perceptiva. Si *no* se pueden crear, entonces omitir la verificación de esa elocución, tomar nota del problema y continuar con otra.

El caso más sencillo.

Una vez elegida la curva estilizada para verificar (de Phonedit o de momel_default), hay que verla y escucharla atentamente comparando contra la curva de *f0* original. Si se ven y suenan exactamente igual (lo cual no es muy frecuente), sólo hay que revisar los targets inicial y final, como se indica en la sección *Regiones inicial y final*, y luego salvar los targets como se indica en la sección *Salvando los targets*.

Regiones inicial y final.

Ver si la curva original tiene ruido o silencio (cualquiera de las dos cosas) en las regiones inicial o final (cualquiera de las dos regiones). Si se cumple cualquiera de estas condiciones, aplicar lo que se indica en esta misma sección; si no, consultar la sección sobre *Target inicial y target final*, y luego la sección sobre *Región media*.

Ver si hay targets ubicados en la región inicial o en la región final (cualquiera de las dos). Si hay un sólo target en la inicial, o uno sólo en la final o uno en cada una, ver la sección *Target inicial y target final*. Si en la inicial hay más de un target, eliminar el segundo target y también el (los) otro(s) que se encuentren después de éste pero antes del primero que esté localizado en una región sonora. Si en la final hay más de un target, eliminar el penúltimo target y también el (los) otro(s) que se encuentren antes de éste pero después del último que esté localizado en una región sonora. Después, aplicar las indicaciones que se dan en la sección *Target inicial y target final*.

Target inicial y target final.

Ver si están en zonas de silencio o de ruido (cualquiera de las dos cosas). Si alguno de los dos está en alguna zona de ese tipo, hay que alinearlos a la misma altura que su target vecino; siempre y cuando la modificación no altere la entonación de la curva.

Phonedit no permite mover horizontalmente los targets inicial ni final, sólo verticalmente; tampoco permite eliminarlos. Más adelante explicamos cómo resolver casos donde el movimiento horizontal sea indispensable.

Si el target inicial o el final están en una zona sonora, puede haber dos posibilidades:

- a) Que el target esté *bien* ubicado y produzca una buena curva, en cuyo caso *no* hay que modificarlo.
- b) Que el target esté *mal* ubicado y esté produciendo una curva deficiente. En este caso hay que moverlo verticalmente hasta ubicarlo donde más convenga. Si fuera indispensable moverlo horizontalmente, hay que modificar el archivo de los targets como se explica en la sección *Modificación del archivo de targets usando editor de texto*.

Si fuera indispensable insertar un nuevo target antes del primero y/o uno nuevo después del último, se tiene que modificar el archivo de los targets como se explica en la sección *Modificación del archivo de targets usando editor de texto*.

Región media.

Esta región generalmente no presenta complicaciones. Su modificación es relativamente sencilla, excepto cuando alguna secuencia de alófonos suena similar a una voz robotizada. Para esto, ver la sección *Voz “robotizada”*.

Salvando los targets.

Se recomienda salvar los targets conforme se avanza con su verificación, aún cuando no se haya terminado de revisar todos los de la elocución. Se deben salvar en la carpeta de momel_verificados. Si ya se había salvado algo en el archivo, simplemente se debe sobrescribir.

Insertando un nuevo target.

Se puede considerar la inserción de un nuevo target sólo si al mover los targets existentes no se puede obtener una buena curva. Cuando resulte indispensable insertarlo, hay que marcar el lugar de inserción con clic izquierdo, y luego dar clic derecho para elegir la opción de insertar. Hay una excepción a esta indicación: si se requiere insertar un target antes del primero o después del último, hay que hacerlo como se indica en la sección *Modificación del archivo de targets usando editor de texto*.

Eliminando un target.

Se recomienda eliminar un target cuando su existencia *no* sea indispensable para la entonación. Por regla general, no se deben eliminar ni el primero ni el último targets. Esto obedece a dos posibilidades: su eliminación podría alterar la entonación, o bien, su presencia es necesaria para que la posterior generación de etiquetas INTSINT cumpla ciertas convenciones asumidas en el proyecto DIME para procesos subsecuentes.

CASOS COMPLICADOS DE LA CURVA ESTILIZADA

Silencio o ruido muy largos al inicio y/o al final.

La presencia de un silencio o un ruido muy largos al inicio y/o al final de la elocución pueden producirle problemas a Phonedit, haciendo que éste se quede bloqueado y que muestre mensajes de error. Aunque una solución sería usar algún editor de audio (como SpeechView o Praat) para eliminar (cortar) el silencio/ruido que esté causando el problema, hemos decidido que dentro del proceso de verificación perceptiva, no se harán cambios de este tipo. Lo que se debe hacer es omitir la verificación de esta elocución, y tomar nota del núm. de diálogo y de elocución.

Elocución muy breve.

Las elocuciones muy breves, como un simple “no”, un “sí” o un “okey”, causan problemas al algoritmo MOMEL y a Phonedit, haciendo que éste último se bloquee y muestre mensajes de error. Si ésto ocurre, hemos decidido que no se hará nada, simplemente no se hace la verificación de esa elocución y, se debe tomar nota del problema.

Voz “robotizada”.

Aunque ésto puede ocurrir en cualquiera de las 3 regiones de la curva, parece ser más común en la región media.

En los casos vistos, parece que esto se presenta más frecuentemente dentro de los límites de tiempo de un mismo alófono, principalmente en algunas vocales. A veces al intentar corregir el problema, se vuelve peor. Se pueden probar las siguientes alternativas (en cualquier orden) e ir escuchando el resultado que más convenga:

- a) Insertar nuevos targets a lo largo de la zona problemática.
- b) Dejar solamente un target en esa zona, eliminando los demás, y alineando el target a la misma altura de sus vecinos.
- c) Borrar todos los targets que estén en la zona “robotizada”.

Si ninguna de estas soluciones funciona, dejar los targets como estaban inicialmente. El problema sería más importante si apareciera en la región final, en cuyo caso hay que tomar nota del problema, indicando el núm. de elocución y de diálogo.

Problemas de visualización.

A veces al mover verticalmente un target inicial o final, resulta imposible que la curva estilizada pueda verse igual que la original. Sin embargo, al escucharla, se percibe que ambas suenan igual. Pareciera que Phonedit se negara a modificar la curva como lo pide el usuario, y sólo permitiera un ligero cambio.

Esto puede estar causado por las características propias del algoritmo MOMEL para unir los targets con parábolas. Otra posibilidad es que se deba a una deficiencia del software, en cuyo caso los datos numéricos pueden ser correctos pero al transformarlos en pixeles que se ven en pantalla su ubicación parece incorrecta. Lo más importante es cómo suena la curva, así que si la estilizada suena como la original, podemos aceptarla y salvar los targets verificados.

Target “fantasma”.

Denominamos target “fantasma” a aquel que no resulta visible dentro de la ventana MOMEL de Phonedit pero cuya existencia es evidente porque la curva estilizada se prolonga hasta algún punto más allá del límite superior de la ventana MOMEL. En este caso, lo deseable es que el target se haga visible, para lo cual se pueden modificar los parámetros para valores máximo y/o mínimo de visualización de frecuencia de la ventana MOMEL y de la ventana de f_0 . Este cambio es sólo de la visualización, ya que *no* se está haciendo ningún cambio a los datos. Si el target en cuestión se encuentra en un pico alto y breve, consultar la sección de *Picos altos y breves*.

Generalmente el problema se resuelve fácilmente; pero si el “fantasma” es un target inicial o final, podría ser necesario modificarlo con algún editor de texto, como se explica en la sección de *Modificación del archivo de targets usando editor de texto*.

Picos altos y breves.

Puede ocurrir que existan picos de frecuencia demasiado alta (respecto al resto de la curva estilizada) que duran muy pocos milisegundos (menos de 70 mseg). Éstos pueden estar causados por ruido de la grabación o por algún error en la extracción de la f_0 o en la generación de targets. Se sugiere eliminar el (los) targets que estén en el pico, y escuchar el resultado. Generalmente su eliminación *no* altera la entonación original.

MODIFICACIÓN DEL ARCHIVO DE TARGETS USANDO EDITOR DE TEXTO

Phonedit *no* permite que se eliminen ni el primer target ni el último, tampoco permite moverlos horizontalmente (tiempo); sólo permite que se modifique el valor de su frecuencia (altura). Tampoco permite que se inserten targets nuevos antes del primero, ni después del último. Si resultara indispensable hacer algún cambio como éstos para corregir la curva estilizada, se tendría que hacer fuera de Phonedit, usando algún editor de texto como Notepad o Wordpad para corregir el archivo *tgt* en la carpeta de *momel_verificados*.

Para definir los valores de tiempo y frecuencia del target que se modificará (o se creará) en el editor de texto, primero hay que estimarlos viendo la curva y los puntos que ya existan en Phonedit. Se puede hacer clic izquierdo sobre el lugar de la curva donde se desee colocar el target para ver en Phonedit los valores de tiempo y frecuencia correspondientes.

Antes de modificar con editor de texto los targets, éstos deben estar salvados en su correspondiente archivo (*.tgt*) en el directorio *momel_verificados*, ya que en ese archivo se harán los cambios con el editor.

Antes de cargar el archivo en el editor de texto, hay que cerrar Phonedit. No es conveniente dejarlo minimizado con los targets en pantalla porque se podrían producir inconsistencias en los datos.

Una vez cargados los targets en el editor de texto, hay que identificar en cuál renglón se insertarán (o modificarán) los datos de tiempo y frecuencia del target en cuestión. A cada renglón (tiempo, frecuencia) del archivo le corresponde un target que se ve en Phonedit. Si se desea insertar un nuevo target antes del primero, hay que escribir los valores de tiempo y frecuencia antes del primer renglón; si se desea después del último, tiene que ser después del último renglón. Para modificar el tiempo del target inicial o del final, sólo hay que modificar el valor de la primera columna en el renglón que corresponda. Las modificaciones tienen que salvarse (sobreescribirse) en el archivo (*.tgt*) en *momel_verificados*. Después, abrir Phonedit, cargar el audio, crear f_0 , y crear MOMEL. Ahora los targets se deben cargar del archivo que fué modificado en el editor de texto. Se debe verificar si el cambio fué bueno, y continuar hasta concluir y salvar.

REFERENCIAS

[ESPESSER, 1999]. ESPESSER, Robert. Mes Signaux package. Speech Signal processing tools(v 1.0).

Disponible en: www.lpl.univ-aix.fr/ext/projects/mes_signaux.htm

[HIRST & ESPESSER, 1993]. HIRST, Daniel & ESPESSER, Robert, 1993. Automatic Modelling of Fundamental Frequency Using a Quadratic Spline Function.

[S.Q. Lab, 1998]. Software Phonedit.
Disponible en:
www.lpl.univ-aix.fr/~sqlab/phoneditRootUK.htm