

Modificaciones al sintetizador Festival para el español con autómatas adaptativos

Investigadora responsable: Claudia Zapata

Coinvestigadores: Rosalía Caya y César Ruiz

Departamento de Ingeniería-Sección Ingeniería Informática

La producción de voz de manera automática por medio de un computador es un campo de profundo interés para muchos centros de investigación, no solo por el gran impacto que causaría lograr una aproximación bastante similar al habla humana, sino porque también implica un trabajo de colaboración exhaustiva de diversas especialidades.

En el proyecto se han realizado dos modificaciones del software Festival para brindar, al español presente en este, características propias del habla en nuestro idioma: pausas, patrones melódicos y lectura adecuada de números (horas, cardinales, monedas y fechas).

Método utilizado

Para diseñar las modificaciones al sintetizador *FESTIVAL*, se utilizaron los autómatas adaptativos que son una secuencia de estados y reacciones para una condición determinada. En nuestro caso, la condición de entrada es el texto, correctamente escrito, que se quiere reproducir, y la secuencia de estados es la forma en la que este texto debe reproducirse oralmente. (Ver figura 2)

Mejoras realizadas

- Pausas por signos de puntuación: la duración de la pausa se determina de acuerdo con el signo de puntuación utilizado.
- Pausas por uso del lenguaje: al leer una oración, no se introducen pausas entre cada palabra escrita, sino que se pronuncia una secuencia de palabras antes de realizar una pausa. Esta secuencia es conocida como palabra fonológica. Por ejemplo: la frase "El tigre se subió en el árbol" tiene 7 palabras y se lee como si fueran tres "Eltigre sesubió enelárbol". (ver figura 1)
- Patrones melódicos: la entonación al inicio o final de las frases depende de si le antecede o prosigue un determinado signo de puntuación. Por ejemplo, si una oración se encuentra entre paréntesis, debe ser leída con una baja de entonación en la primera y última sílaba: "Hacia calor ↓(el verano vino aquel año muy adelantado)↓ y las ventanas estaban abiertas".
- Lectura de fechas: existen distintos formatos de fecha. Se adaptó el software para una correcta lectura de ellos.
- Lectura de números arábigos: el valor de un dígito varía según la posición que ocupa dentro del número multiplicándose por la base elevada a la posición. Así, el primer dígito

comenzando por la derecha tiene el valor que representa su símbolo multiplicado por $10^0 (=1)$. El dígito situado a su izquierda tiene el valor que representa su símbolo multiplicado por $10^1 (=10)$, y así sucesivamente.

- Lectura de monedas: de acuerdo con el símbolo que precede la cantidad, se puede deducir cuál es el sistema monetario al que pertenece y, por lo tanto, la forma correcta de leerla.

Resultados de las pruebas de comprensión

Se utilizó una prueba estándar para medir calidad de voz en dispositivos de comunicación. La prueba se aplicó a 75 personas y se pudo apreciar una mejora significativa en el software modificado.