Mejora De La Calidad De Voz En Castellano Para El Sintetizador *Festival* Utilizando El Método De Autómatas Adaptativos: Números Arábigos Y Fechas (18 de octubre de 2010)

C. Ruíz, R. Caya , *Member IEEE* y C. Zapata, *Member IEEE*

***Abstract*— The project presented in this article carries on a study of the improvement in the quality of the synthetic voice generated by Festival using the adaptive automata technique. These automatons are designed trying to solve the issues found during the speech synthesis of Arabic numbers and dates, which were identified according to the linguistic rules for Spanish language. Once the development were finished will be possible execute formal tests to evaluate the voice quality generated by the version of Festival that includes the implemented improvements.**

Keywords*— *speech synthesis, natural language interfaces, adaptive automata.

I. Nomenclatura

- AA: Autómata adaptativo.
- TTS: Texto to speech, en español texto a voz.
- Inteligibilidad: la facilidad para comprender la señal de voz producida.
- Naturalidad: un indicador de la semejanza de los sonidos producidos artificialmente con los naturales.

II. Introducción

ES un hecho ampliamente aceptado que por medio de la computadora es posible resolver múltiples operaciones en tiempos muy cortos en comparación con el que le tomaría al cerebro humano. Del mismo modo, es capaz de procesar grandes cantidades de información de acuerdo a criterios específicos, y buscar información en muchos casos de manera altamente eficiente. Sin embargo, es necesario que las computadoras no sólo sean capaces de realizar estas tareas, sino que sean capaces de realizarlas por medio de una comunicación natural y efectiva por medios auditivos. La razón fundamental de ello es la naturaleza social del ser humano, y sobre todo el uso primordial de la comunicación oral y escrita como medio para satisfacer sus necesidades.

Actualmente, la tecnología ha propuesto vías para establecer esta interacción con las computadoras. Una de ellas es la síntesis del habla, por medio de la cual se busca generar artificialmente voz similar a la humana. Sin embargo la diferencia entre la voz sintética y la humana es claramente perceptible y en muchos casos llega a generar rechazo por parte de los usuarios.

FESTIVAL es uno de los sistemas de síntesis disponibles, el cual opera en la plataforma Linux y es de licencia open source. Dentro de él se han desarrollado módulos para diversos idiomas, uno de ellos es el castellano. Ciertamente, la calidad de la voz sintética varía de acuerdo al idioma en cuestión, en inglés y francés se encuentra bastante más avanzada que en el caso del castellano, idioma en el cual se reporta aún una voz de muy poca natural, entendida de acuerdo con **Erro! Fonte e referência não encontrada.** como la similitud lograda entre la voz artificial generada y el discurso humano.

Una de las causas que impiden el desarrollo de una voz natural es la falta de identificación de algunos formatos en el texto a ser sintetizado. Particularmente, los números arábigos y fechas, se presentan en patrones idiomáticamente establecidos y estandarizados para el idioma castellano. No obstante, Festival no es capaz de identificar los patrones que los representan y realiza la síntesis transformando cada carácter de estos números y fechas. Por ello es necesario establecer las reglas gramaticales correspondientes para el castellano que permitan transformar exitosamente las frases que contengan números arábigos y fechas a voz sintética.

III. Marco conceptual

La síntesis de voz es el proceso en el cual se busca la producción artificial de señales de voz muy semejantes al habla humana. Tal como se señala en **Erro! Fonte de eferência não encontrada.** a síntesis de voz es usualmente realizada a partir de texto, por lo que se le denomina, en inglés, text-to-speech (TTS). Sin embargo, también es posible realizar este proceso a partir de alguna representación lingüística simbólica que los sintetizadores de voz son capaces de reconocer.

C. Zapata, Pontificia Universidad Católica del Perú, Lima, Perú, zapata.cmp@pucp.edu.pe.
R. Caya, Pontificia Universidad Católica del Perú, Lima, Perú, rosalia.caya@pucp.edu.pe.
C. Ruiz, Pontificia Universidad Católica del Perú, Lima, Perú, cruizv@pucp.edu.pe.

Asimismo, en **Erro! Fonte de referência não encontrada.** e menciona que actualmente existen dos técnicas o métodos principales que se utilizan para la síntesis: Síntesis concatenativa y síntesis por formantes. Para poder establecer la calidad de la voz obtenida se suelen tener en cuenta dos parámetros fundamentales: la naturalidad y la inteligibilidad **Erro! Fonte de referência não encontrada.**

Para conseguir un sintetizador de voz ideal es necesario que sea natural e inteligible al mismo tiempo. Sin embargo la proporción ideal entre estas características la definirá el contexto y las condiciones en las cuales se requiera aplicar la síntesis.

Tal como se mencionó en el acápite anterior un aspecto en el cual se puede aportar para la mejora de la calidad de la síntesis de voz es la identificación de los números arábigos y las fechas. En general, un sistema de numeración es el conjunto de símbolos y reglas que permiten construir todos los números válidos en dicho sistema. Estas reglas son diferentes para cada sistema de numeración considerado, pero una regla común a todos es que para construir números válidos en un sistema de numeración determinado sólo se pueden utilizar los símbolos permitidos en dicho sistema. En el caso de este proyecto se trabajará bajo las especificaciones del sistema de numeración decimal pues es el más utilizado.

Dentro de todo sistema existen numerales que representan la cantidad de elementos de un conjunto en relación con la serie de números dentro del sistema, o de acuerdo con **Erro! onte de referência não encontrada.** también llamados números cardinales. Y existen los numerales que expresan orden o sucesión en relación con los números naturales e indican el lugar que ocupa, dentro de una serie ordenada, el elemento al que se refieren. Estos últimos se denominan de acuerdo con **Erro! Fonte de referência não encontrada.** úmeros ordinales

Para el caso de las fechas, es necesario seguir un estándar para definir claramente el concepto de día, así como la representación del paso de los mismos en agrupaciones mayores como mes y año. El Calendario Gregoriano, si bien es un calendario originario de Europa, actualmente es utilizado de manera oficial en todo el mundo occidental. Es por esta globalidad que se tendrán como estándar para los conceptos antes mencionados lo que dicta el calendario gregoriano.

IV. DESCRIPCIÓN DE LA SOLUCIÓN

Este proyecto propone implementación de una mejora de la calidad del sintetizador de voz *FESTIVAL*, utilizando el método de autómatas adaptativos en la representación de las normas gramaticales aplicadas para los números arábigos y fechas. Cada una de ellas en el contexto que se presenten y afecten la calidad de la voz sintética. La evaluación de la mejora de la calidad se hará, siguiendo el modelo utilizado en **Erro! Fonte de referência não encontrada.** desde as perspectivas de: naturalidad e inteligibilidad.

Para el alcance de este proyecto es importante recalcar que no se implementará un nuevo sintetizador de voz. Lo que se realizará es la modificación de las reglas que utiliza el sintetizador *FESTIVAL* a fin de adicionar el reconocimiento de los patrones necesarios para solucionar los problemas mencionados anteriormente. Dichas reglas se representarán usando el método de autómatas adaptativos.

Tal como se mencionó anteriormente el sintetizador con el cual se trabaja es *FESTIVAL* Speech Synthesis System, esta plataforma está dedicada por completo a la síntesis de voz, fue creada por el Centro de Investigación de Tecnologías del Lenguaje de la Universidad de Edimburgo **Erro! Fonte de eferência não encontrada.** y actualmente es mantenida por el Instituto de Tecnologías del Lenguaje de la Universidad Carnegie Mellon. Su elección corresponde a que se trata de un software de licencia open source y por lo tanto accesible, cuenta con documentación oficial,, ha sido utilizado anteriormente por otros trabajos de investigación con resultados satisfactorios como en **Erro! Fonte de referência ão encontrada.** y actualmente cuenta con una comunidad de mantenimiento oficial dentro de las principales distribuciones de Linux, Además de ello se han desarrollado múltiples aplicaciones que hacen uso de su voz sintética para fines de brindar accesibilidad, el caso más conocido es la herramienta Orca. Debido a estas razones su uso académico es no solo viable sino también recomendado.

Siguiendo la experiencia anterior, detallada en **Erro! Fonte e referência não encontrada.**, el proyecto ha dividido en dos fases: la investigación de las reglas gramaticales del español para la identificación y lectura de números arábigos y los formatos de fechas, y la implementación del aporte producto de la investigación. Los objetivos establecidos para la primera fase son: identificar los factores que impiden la producción de una voz sintética natural por la presencia de números arábigos y/o fechas; formular las reglas dependientes del contexto usando Autómatas Adaptativos que solucionen algunos de los factores identificados en el punto anterior para el idioma castellano; Analizar la estructura del sintetizador de voz *FESTIVAL* para decidir dónde es más conveniente realizar la modificación. La fase de desarrollo del aporte tiene como objetivos: implementar los autómatas adaptativos en la plataforma de síntesis elegida; diseñar y realizar experimentos con los autómatas adaptativos implementados en casos reales de síntesis de voz para el castellano, y por último la realización de pruebas para obtener indicadores que respalden la mejora de la calidad de la voz sintética. Actualmente el proyecto se encuentra finalizando la primera fase.

A. Sustento de la solución

En los últimos años se han obtenido grandes avances en lo que respecta a la síntesis de voz, ello se debe a que se han desarrollado investigaciones que han dado como resultado nuevas técnicas utilizadas parar la mejora en los resultados de la voz sintética. Sin embargo, aún se evidencian problemas de naturalidad en los resultados. Esto se debe principalmente a los errores en el reconocimiento de frases del castellano por la incorrecta identificación de parámetros y patrones que permitan identificar determinados fragmentos de frases como formas particulares de las mismas, que tendrán una pronunciación distinta a la que puede ser determinada en forma literal por el sintetizador.

En el año 2009 se presentó al WTA la primera fase de este conjunto de proyectos **Erro! Fonte de referência não ncontrada.**, donde se trataba de mejorar la calidad de voz realizada en FESTIVAL modificando los patrones melódicos como efecto de los signos de puntuación, utilizando autómatas adaptativos. Entonces quedaron pendientes el estudio de la síntesis de números arábigos y fechas. El presente proyecto retoma dicha investigación para resolver el problema de la incorrecta lectura de números arábigos y de fechas.

Cuando un sintetizador como FESTIVAL procesa la lectura de números reales en formato arábigo y de fechas muchas veces, y sobre todo al sintetizar castellano, reproduce una síntesis incorrecta. Esto es debido a que los patrones son leídos como un símbolo más dentro del texto. Por ejemplo no se considera el separador decimal ni los separadores de fechas. La forma en la que se leen los números y las fechas depende del formato que se ha utilizado para representarlos por lo tanto son palabras léxicas que para su síntesis dependen del contexto. Esta dependencia justifica el uso de autómatas adaptativos para la solución planteada.

B. Resultados esperados

El aporte práctico del proyecto es la implementación de una mejora de la inteligibilidad en la voz sintética producida por FESTIVAL para el castellano, lograda mediante AA **Erro! onte de referência não encontrada.**. Esta mejora en inteligibilidad por consecuencia da una mejora en la naturalidad del sintetizador al leer números y fechas.

En relación con los objetivos específicos del proyecto los resultados esperados son:

- Elaborar las reglas de detección de frases de acuerdo al contexto que permitan solucionar el problema.

- Diseñar de algunas reglas de producción de voz natural enfocadas en el problema planteado de acuerdo al método de Autómatas Adaptativos.

- Someterá el sintetizador a las pruebas convenientes para poder observar la mejora.

V. APOYO TEÓRICO

La identificación de distintas formas de fragmentos de frases hace posible que, de acuerdo al contexto en el que se encuentran, determinadas frases representadas gráfica o textualmente de la misma forma puedan tener distinto significado.

Para ello, es necesaria la evaluación del contexto en el que se encuentran las frases para poder determinar el significado que se les dará y, a partir de esto, seleccionar la forma en la que se desarrollará la síntesis de voz.

En el caso específico de números y fechas, estas pueden encontrarse en distintas formas o posiciones dentro de frases de mayor longitud. De acuerdo a su posición en la frase o por la presencia de otros fragmentos de frases en el contexto general se puede reconocer la utilización de estos elementos para un determinado significado, lo que permite determinar la forma en la que el sintetizador de voz trabajará, acercándolo a las formas del lenguaje humano y ganando naturalidad.

Mediante la utilización de máquinas de estados finitos es posible representar las distintas reglas gramaticales de los lenguajes naturales. Además, mediante autómatas pueden representarse subconjuntos de reglas, las cuales pueden aplicarse para la mejora en la detección de los fragmentos de frases que permitirán adaptar la pronunciación hacia formas mejoradas por el análisis contextual mediante técnicas adaptativas.

Como se menciona en **Erro! Fonte de referência não ncontrada.**, la técnica de Autómatas Adaptativos es aplicable principalmente para problemas de lenguajes contexto-dependientes, por lo que la aplicación de esta solución vendría a ser la más adecuada para los problemas presentados en este proyecto.

VI. MÉTODOLOGÍA APLICADA

El proyecto de investigación se encuentra guiado por la metodología propuesta en **Erro! Fonte de referência não ncontrada.** para proyectos de investigación. La adaptación de las líneas generales descritas en **Erro! Fonte de referência não encontrada.** para este proyecto dieron como resultado las siguientes etapas:

- Identificación de los parámetros presentes en la sintaxis que afectan la naturalidad del discurso.
- Identificación de casos en los cuales se encuentra presente la dependencia del contexto.
- Identificación de las reglas que solucionan los casos seleccionados.
- Construcción de los autómatas adaptativos que implementan dichas reglas.
- Selección y realización de pruebas que evalúen la calidad del sintetizador.
- Análisis de resultados.
- Elaboración de conclusiones.

VII. DISEÑO E IMPLEMENTACIÓN DE LOS AUTÓMATAS ADAPTATIVOS

Una vez realizada la identificación de los factores que afectan la mejora de las características de la voz sintética, pueden definirse las reglas de los autómatas adaptativos.

La mejora en la síntesis de voz para números y fechas se debe realizar mediante modificaciones en distintos puntos de la arquitectura de Festival. A continuación se detallan el diseño e implementación de los Autómatas Adaptativos para cada uno de los casos planteados.

A. Autómata Adaptativo para la formación de números arábigos: enteros y reales

El sistema numérico arábigo es un sistema posicional que consta de 10 glifos diferentes para representar los 10 símbolos del sistema. 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.

El sistema "arábigo" se representa utilizando muchos

conjuntos de glifos diferentes. Estos glifos pueden dividirse en dos grandes familias, los numerales arábigos occidentales y los orientales. Los que estudiaremos en este proyecto serán los occidentales, representados por:

0 1 2 3 4 5 6 7 8 9

Para el análisis de los números *Enteros* y *Reales* se han definido los siguientes autómatas que ha definido los siguientes grupos de elementos:

- # = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}
- separadorDecimal = {",”}
- signoPuntuacion = {“.” “,” “;” “:” “”” “’” “)” “]” “}” “!” “?”}

Siguiendo la notación formal descrita en **Erro! Fonte de eferência não encontrada.**, el autómata adaptativo encargado de la construcción de números enteros, nombrado como M_F, se define formalmente como una 8-tupla:

M = (Q, §, q_0, F, δ, Q,¡,¦), donde:

Q: {a_0, a_f}

§: {#, signoPuntuacion, ε}

q_0 : {a_0}

F : { a_f }

δ:

[1] P_0 : a_0, ε, a_f

[2] P_1 : a_0, #, a_f

Q : {a1, a2, a3,...}

¡: (−,a,ε,a_f) (+,a,#,a_I) (+,a_I,#,a_{II}) (+,a_I,ε,a_{IV})
(+,a_I,signoPuntuacion,a_f) (+,a_{II},#,a_{III}) (+,a_{II},ε,a_{IV})
(+,a_{II},signoPuntuacion,a_f) (+,a_{III},ε,a_{IV})
(+,a_{III},signoPuntuacion,a_f) (+,a_{IV},#,a_V)(+,a_{IV},¬#,a_f)
(+,a_V,#,a_{VI}) (+,a_{VI},#,a_{VII}) (+,a_{VII},ε,a_{VIII})
(+,a_{VII},signoPuntuacion,a_f) (+,a_{VIII},#,a_V)
(+,a_{VIII},¬#,a_f)

¦ ((a$_i$,#,a_f)) = {(−,a,ε,a_f), (+,a,#,a_I),
(+,a_I,#,a_{II}), (+,a_I,ε,a_{IV}),
(+,a_I,signoPuntuacion,a_f),
(+,a_{II},#,a_{III}), (+,a_{II},ε,a_{IV}),
(+,a_{II},signoPuntuacion,a_f), (+,a_{III},ε,a_{IV}),
(+,a_{III},signoPuntuacion,a_f), (+,a_{IV},#,a_V),
(+,a_{IV},¬#,a_f), (+,a_V,#,a_{VI}), (+,a_{VI},#,a_{VII}),
(+,a_{VII},ε,a_{VIII}), (+,a_{VII},signoPuntuacion,a_f),
(+,a_{VIII},#,a_V), (+,a_{VIII},¬#,a_f)}

La cadena inicial del AA denotada como ω corresponde al carácter inicial de cada token que ya ha sido separado por Festival en la etapa de tokenizing. Es a partir de este carácter que comienza la validación del autómata. En caso cumpla con las condiciones iniciales se continua con el análisis de los demás caracteres, continuando con la ejecución del autómata hasta que cumpla el estado final o termine en alguno de los estados no finales al no cumplir las condiciones planteadas.

La operación de M_F se encuentra descrita en la Fig. 1

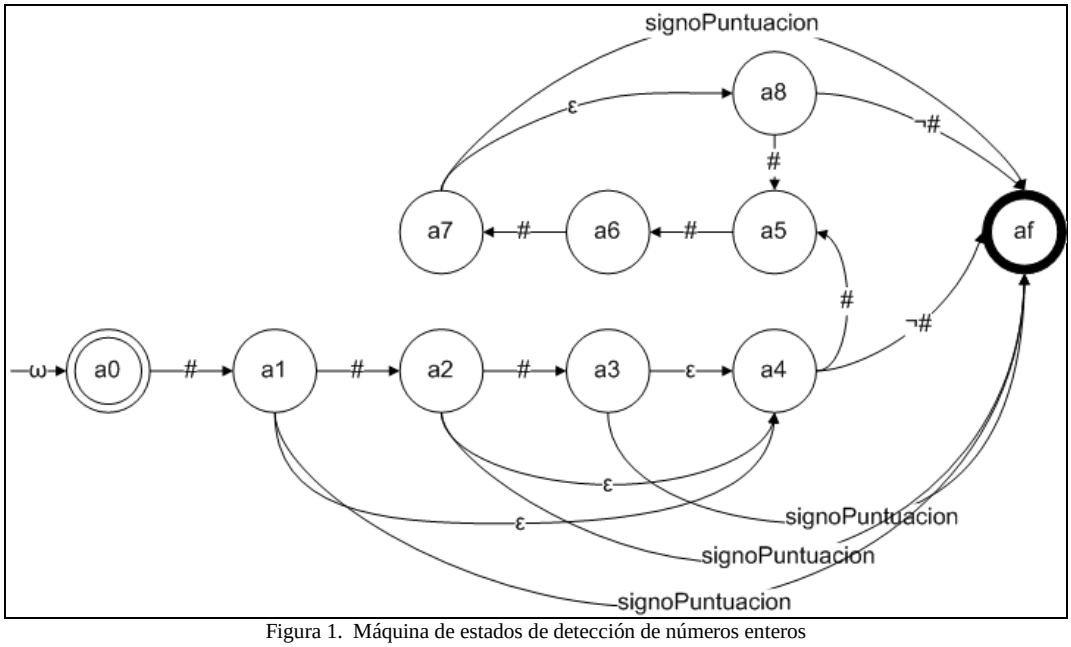

Figura 1. Máquina de estados de detección de números enteros

Para la construcción de números reales se define formalmente la 8-tupla M_f de la siguiente manera:

M = (Q, §, q_0, F, δ, Q,¡,¦), donde:

Q: {b_0, b_f}

§: {#, signoPuntuacion, separadorDecimal, ε}

q_0 : {b_0}

F : { b_f }

δ:

[3] P_0 : b_0, ε, b_f

[4] P_1 : b_0, #, b_f

Q : {b1, b2, b3,...}

¡: (−,b,ε,b_f) (+,b,#,b_I) (+,b_I,#,b_{II}) (+,b_I,ε,b_{IV})
(+,b_I,separadorDecimal,b_{IX}) (+,b_{II},#,b_{III})
(+,b_{II},ε,b_{IV})
(+,b_{II},separadorDecimal,b_{IX}) (+,b_{III},ε,b_{IV})
(+,b_{III},separadorDecimal,b_{IX}) (+,b_{IV},#,b_V)
(+,b_V,#,b_{VI}) (+,b_{VI},#,b_{VII}) (+,b_{VII},ε,b_{VIII})
(+,b_{VII},separadorDecimal,b_{IX}) (+,b_{VIII},#,b_V)
(+,b_{IX},#,b_X) (+,b_X,#,b_X) (+,b_X,ε,b_f)
(+,b_X,signoPuntuacion,b_f))

¦ ((b_i,#,b_f)) = {(−,b,ε,b_f), (+,b,#,b_I),
(+,b_I,#,b_{II}),
(+,b_I,ε,b_{IV}), (+,b_I,separadorDecimal,b_{IX}),
(+,b_{II},#,b_{III}), (+,b_{II},ε,b_{IV}),
(+,b_{II},separadorDecimal,b_{IX}), (+,b_{III},ε,b_{IV}),
(+,b_{III},separadorDecimal,b_{IX}),
(+,b_{IV},#,b_V), (+,b_V,#,b_{VI}), (+,b_{VI},#,b_{VII}),
(+,b_{VII},ε,b_{VIII}),
(+,b_{VII},separadorDecimal,b_{IX}),
(+,b_{VIII},#,b_V), (+,b_{IX},#,b_X), (+,b_X,#,b_X),
(+,b_X,ε,b_f), (+,b_X,signoPuntuacion,b_f)}

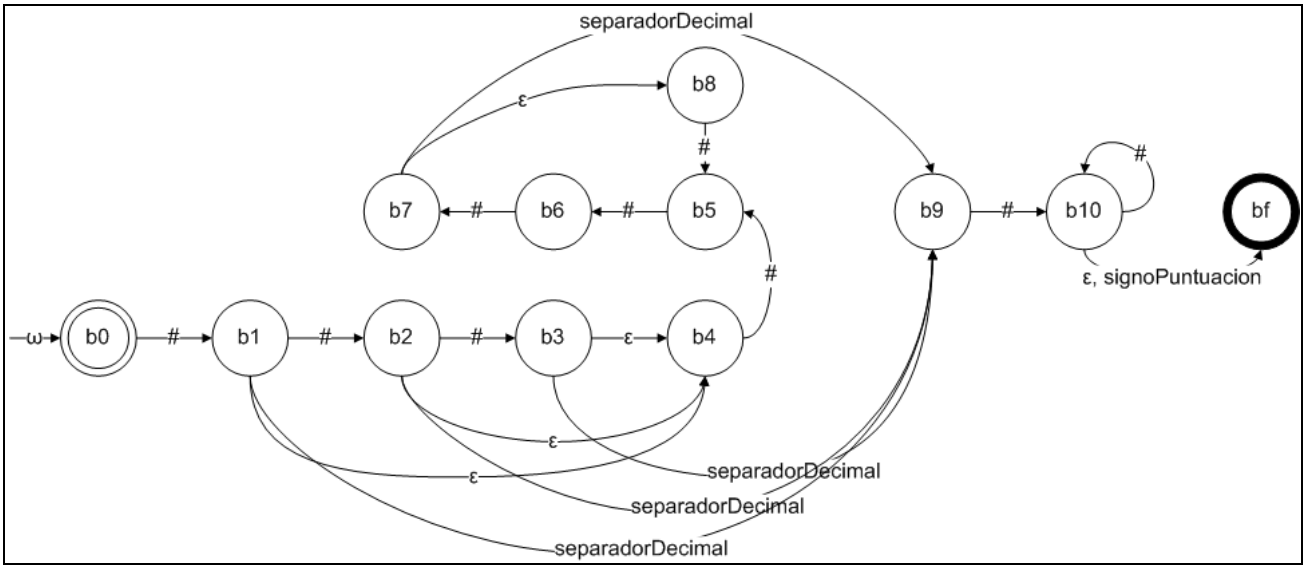

Figura 2. Máquina de estados de detección de números reales

B. Autómata Adaptativo formación de fechas: estándar ISO y personalizadas

Existen distintos formatos de fecha que han sido definidos por ISO o que son de utilización general a nivel de Latinoamérica. Es la norma que define el estándar para la representación de instantes, intervalos e intervalos recurrentes de tiempo evitando ambigüedades. Esta notación facilita la migración entre distintas plataformas. Se especifican en orden primeramente los periodos de tiempo más largos y posteriormente los más cortos.

Para representar fechas de calendario se usarán 4 cifras para el año, dos para el mes y dos para el día del mes, rellenando con ceros si es necesario. Por ejemplo, para representar la fecha 9 de octubre de 2009 se utilizará '2009' para representar el año, '10' para representar el mes de octubre y '09' para representar el día 9 de ese mes.

En la siguiente tabla se muestran las posibles representaciones y formatos recogidos en la norma para representar fechas de calendario, tomando como base para los ejemplos la fecha 9 de octubre de 2009. En las representaciones expandidas se representan los años con 6 cifras.

	formato básico (ejemplo)	**formato extendido (ejemplo)**
representación completa	YYYYMMDD (20091009)	YYYY-MM-DD (2009-10-09)
representación de precisión reducida	YYYY-MM (2009-10) YYYY (2009) YY (20)	no se aplica
representación expandida	±YYYYYMMDD (+0020091009) ±YYYYY-MM (+002009-10) ±YYYYY (+002009) ±YYY (+0020)	±YYYYY-MM-DD (+002009-10-09) no se aplica no se aplica no se aplica

TABLA I: Formatos de fecha según norma ISO 8601

La representación completa permite especificar un día concreto del calendario. La representación de precisión reducida permite especificar un mes (YYYY-MM), un año (YYYY) o un siglo concretos. La representación expandida permite especificar un día (±YYYYYMMDD), un mes (±YYYYY-MM), un año (±YYYYY) o un siglo (±YYY) concretos (en estos tres últimos casos, se trataría de una representación expandida y de precisión reducida). La representación expandida requiere que haya un acuerdo previo de las partes que intercambian la información acerca del número de dígitos que requiere la representación del año, siendo como mínimo cinco.

Además de la forma estándar descrita anteriormente, existen formas de representación de las fechas del calendario de manera personalizada, con variaciones en forma y separadores, aunque tomando como base el estándar ISO.

Estos cambios consisten en:

- Cambio de separador guión ("-") por una barra oblicua ("/").
- Cambio de separador guión ("-") por un punto (".").
- Cambio de separador guión ("-") por un espacio (" ").
- Utilización del formato del tipo "DD-MM-YYYY".

Para el análisis de fechas se presentan los siguientes autómatas que han definido los siguientes grupos de elementos:

- # = {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}
- separadorFecha = {"/", "-", "."}
- signoPuntuacion = {"." "," ";" ":" "”" "’" ")" "]" "}" "¡" "?"}

Para la construcción de fechas siguiendo el estándar ISO y su personalización se define formalmente la 8-tupla M_f de la siguiente manera:

$M = (Q, \S, q_0, F, \delta, Q, ¡, ¦)$, donde:

Q: $\{c_0, c_f\}$

§: {#, signoPuntuacion, separadorFecha ε}

q_0 : $\{c_0\}$

F : $\{ c_f \}$

δ:

[5] $P_0 : c_0, \varepsilon, c_f$

[6] $P_1 : c_0, \#, c_f$

Q : {c1, c2, c3,...}

¡: $(-,c,\varepsilon,c_f)$ $(+,c,\#,c_I)$ $(+,c_I,\#,c_{II})$ $(+,c_{II},\#,c_{III})$
$(+,c_{II},\text{separadorFecha},c_V)$ $(+,c_{III},\#,c_{IV})$
$(+,c_{IV},\text{separadorFecha},c_V)$ $(+,c_V,\#,c_{VI})$ $(+,c_{VI},\#,c_{VII})$
$(+,c_{VI},\text{separadorFecha},c_{VIII})$
$(+,c_{VII},\text{separadorFecha},c_{VIII})$ $(+,c_{VIII},\#,c_{IX})$ $(+,c_{IX},\#,c_X)$
$(+,c_{IX},\varepsilon,c_{VII})$ $(+,c_{IX},\text{signoPuntuacion},c_f)$ $(+,c_X,\varepsilon,c_f)$
$(+,c_X,\text{signoPuntuacion},c_f)$

¦ $((c_i,\#,c_f))$ = $\{(-,c,\varepsilon,c_f)$, $(+,c,\#,c_I)$,
$(+,c_I,\#,c_{II})$, $(+,c_{II},\#,c_{III})$, $(+,c_{II},\text{separadorFecha},c_V)$,
$(+,c_{III},\#,c_{IV})$, $(+,c_{IV},\text{separadorFecha},c_V)$,
$(+,c_V,\#,c_{VI})$, $(+,c_{VI},\#,c_{VII})$,
$(+,c_{VI},\text{separadorFecha},c_{VIII})$,
$(+,c_{VII},\text{separadorFecha},c_{VIII})$,
$(+,c_{VIII},\#,c_{IX})$, $(+,c_{IX},\#,c_X)$, $(+,c_{IX},\varepsilon,c_{VII})$,
$(+,c_{IX},\text{signoPuntuacion},c_f)$, $(+,c_X,\varepsilon,c_f)$,
$(+,c_X,\text{signoPuntuacion},c_f)\}$

De manera similar a los autómatas anteriores, la cadena inicial del AA denotada como ω corresponde al carácter inicial de cada token que ya ha sido separado por Festival en la etapa de tokenizing. Es a partir de este carácter que comienza la validación del autómata. En caso cumpla con las condiciones iniciales se continua con el análisis de los demás caracteres, continuando con la ejecución del autómata hasta que cumpla el estado final o termine en alguno de los estados no finales al no cumplir las condiciones planteadas.

La configuración de Mp y su cambio con cada transición adaptativa se representa en la Fig. 3.

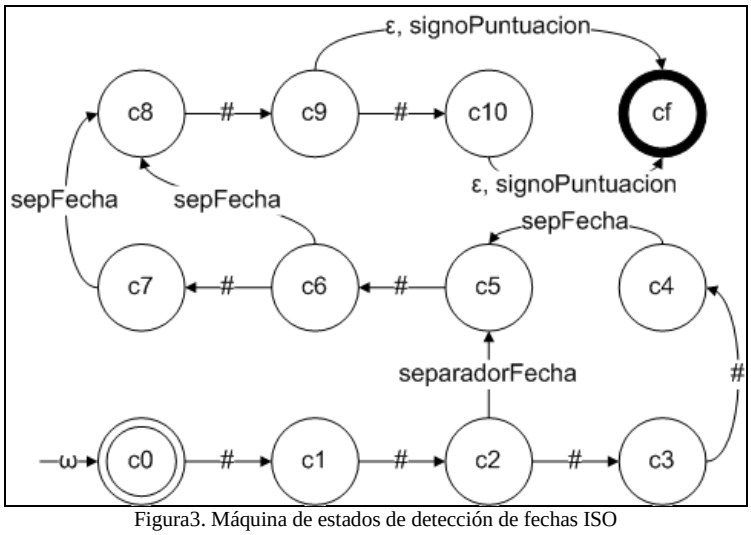

Figura3. Máquina de estados de detección de fechas ISO

Para la construcción de fechas personalizadas se define formalmente la 8-tupla Mf de la siguiente manera:

M =(Q, §, q_0, F, δ, Q,¡,¦), donde:

Q: {d_0, d_f}

§: {#, signoPuntuacion,separadorFecha ε}

q_0 : {d_0}

F : { d_f }

δ:

[7] P_0 : d_0, ε, d_f

[8] P_1 : d_0, #, d_f

Q : {d1, d2, d3,...}

¡: (−,d,ε,d_f) (+,d,#,d_I) (+,d_I,#,d_{II})
(+,d_I,separadorFecha,d_{III}) (+,d_{II},separadorFecha,d_{III})
(+,d_{III},#,d_{IV}) (+,d_{IV},#,d_V) (+,d_{IV},separadorFecha,d_{VI})
(+,d_V,separadorFecha,d_{VI}) (+,d_{VI},#,d_{VII})
(+,d_{VII},#,d_{VIII}) (+,d_{VIII},#,d_{IX}) (+,d_{VIII},ε,d_f)
(+,d_{VIII},signoPuntuacion,d_f) (+,d_{IX},#,d_X) (+,d_X,ε,d_f)
(+,d_X,signoPuntuacion,d_f)

¦ ((((d_i,#,d_f)) = {(−,d,ε,d_f), (+,d,#,d_I),
(+,d_I,#,d_{II}), (+,d_I,separadorFecha,d_{III}),
(+,d_{II},separadorFecha,d_{III}), (+,d_{III},#,d_{IV}),
(+,d_{IV},#,d_V), (+,d_{IV},separadorFecha,d_{VI}),
(+,d_V,separadorFecha,d_{VI}),(+,d_{VI},#,d_{VII}),
(+,d_{VII},#,d_{VIII}), (+,d_{VII}l,#,d_{IX}), (+,d_{VIII},ε,d_f),
(+,d_{VIII},signoPuntuacion,d_f),(+,d_{IX},#,d_X), (+,d_X,ε,d_f),
(+,d_X,signoPuntuacion,d_f)}}

La descripción de su funcionamiento se presenta en la Fig. 4.

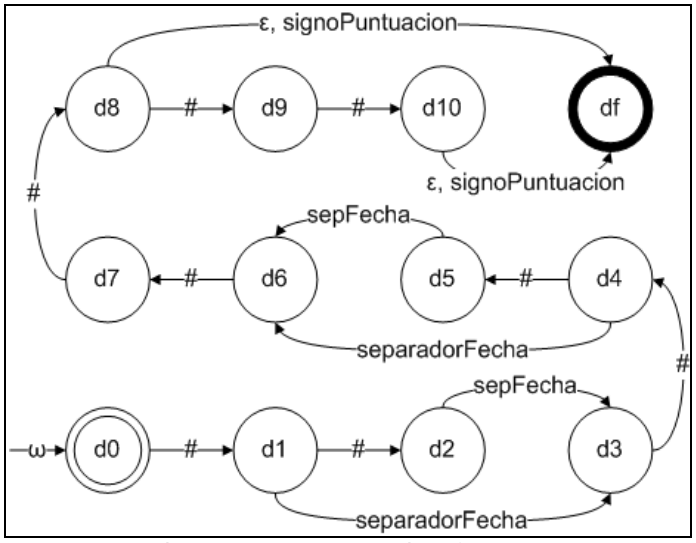

Figura4. Máquina de estados de detección de fechas personalizadas.

VIII. Implementación

Con respecto a la implementación de la mejora, se espera desarrollar librerías dinámicas que permitan la selección en tiempo de ejecución, y de acuerdo a parámetros establecidos, de los autómatas que se aplicarán durante el desarrollo de la síntesis de voz, además del orden de ejecución.

Las configuraciones antes mencionadas contendrán, además, definiciones hacia los elementos del autómata, haciendo posible que puedan definirse más adelante nuevos casos o variantes a los estados posible del autómata, que permitirán mayor dinamicidad a la ejecución de los autómatas.

Este no es un trabajo concluido pues aún no se ha terminado la implementación y las pruebas.

IX. Evaluación de la Calidad

La calidad de voz es evaluada por su inteligibilidad y su naturalidad. Además, dado que se han seleccionado casos específicos para frases que se encuentren en un entorno mayor, se puede definir una nueva perspectiva de evaluación en base a la adaptación de la voz sintética generada al contexto general.

En vista que la calidad es un concepto subjetivo, en el que el resultado de la evaluación depende de la persona que responda a la prueba planteada, se tendrán valores aproximados para la explotación de los datos que permitan obtener características generales sobre el producto desarrollado y la mejora obtenida luego de la implementación del autómata, ya que se desarrollarán pruebas que permitan al usuario escoger entre un determinado número de rangos de calidad, disminuyendo la variedad de respuestas y agrupando los resultados de acuerdo a características similares para un grupo de personas encargadas de resolver las pruebas planteadas.

Como la evaluación de la calidad será subjetiva, lo que la mayoría de personas califiquen lo que oyen como "perfecto" tendrá mayor calidad, lo que nos llevará a la síntesis de conclusiones en base a las respuestas de los evaluados.

Los métodos de evaluación planteados servirán para evaluar la calidad de la voz sintética en base a sus características generales: naturalidad e inteligibilidad. Cada una de las características mencionadas anteriormente tendrá métodos de evaluación diferentes, los cuales se definen a continuación.

La prueba MOS (Mean Opinion Score) **Erro! Fonte de referência não encontrada.** es de aplicación general para obtener índices que permitan validar la calidad de la voz sobre telefonía IP. En este caso, aplicaremos las pruebas MOS para poder obtener estos resultados relacionados a la calidad de la voz sintética de manera más ordenada y que haga posible la medición, a través de valores numéricos, de indicadores generales relacionados a las características de la voz sintética que el evaluado escucha. Con los datos obtenidos podremos obtener estadísticas más elaboradas, gracias a que con el cuestionario se pudo dejar de lado la calificación subjetiva por una

Esta prueba MOS consiste en preguntas que relacionan determinadas preguntas del cuestionario, que se puede encontrar en el anexo D, con la voz sintética que se le brinda al usuario. Como alternativas de respuesta presenta 5 niveles, que van del 1 (peor) al 5 (mejor).

La naturalidad, en el contexto de la mejora en la síntesis de números y fechas, estará relacionada con la forma en la que la voz sintética producida con la mejora pueda ser similar a la voz humana. Estas pruebas están más relacionadas con el

aspecto supra-segmental del texto, debido a que es necesario ubicar la naturalidad en un contexto general, el cual se da al desarrollar un texto completo y donde se muestren las dependencias de las frases que contengan números arábigos y fechas en el texto completo a sintetizar. Para la evaluación de esta característica se aplicará el método de evaluación MOS explicado anteriormente. Sin embargo, las preguntas a evaluar serán las relacionadas a la voz sintética generada y si esta fue de agrado del oyente.

La inteligibilidad del texto presentado al evaluado será una característica muy importante a evaluar, ya que lo que se desea es que se pueda comprender perfectamente el texto que se planteó sintetizar. Estas pruebas están más relacionadas con el ámbito segmental del texto, ya que para poder entender el mensaje es necesaria la audición correcta de los números arábigos o fechas, según corresponda, para que se comprenda específicamente la porción del texto que los representa.

Nuevamente la evaluación MOS será aplicable para esta prueba, esta vez dirigiendo las preguntas al aspecto de entendimiento de los números o fechas de manera concreta.

X. Conclusiones

Como resultado del trabajo realizado hasta hoy, se pueden obtener las siguientes conclusiones:

1) La síntesis de voz como medio de interacción entre el humano y el computador es cada día más utilizada por lo que las investigaciones sobre síntesis de voz están concentradas en la calidad de la voz producida para garantizar que los sintetizadores sean usados por las personas y no rechazados.

2) Las técnicas de autómatas adaptativos siguen siendo una de las más simples y prácticas formas para resolver aspectos de la síntesis de voz que son dependientes del contexto.

3) No sólo la melodía que utiliza un sintetizador de voz es importante para la naturalidad de la voz pues si un texto es mal reproducido existirá mala inteligibilidad y por consiguiente el receptor detectará que la voz es robótica al no obtener la salida esperada.

4) Existen distintos formatos, como los de fechas, que son de uso común pero no son parte de un estándar o simplemente son una adaptación de uno. Es importante considerar estos formatos para la creación de sintetizadores más naturales.

Agradecimientos

Los autores reconocen las contribuciones de la profesora Beatriz Mauchi y al profesor Jorge Pérez del Departamento de Humanidades de la Pontificia Universidad Católica del Perú, por su aporte en el aspecto lingüístico del conjunto de proyectos del cual forma parte el aporte descrito en este artículo. Asimismo, al Doctor Joao José Neto de la Universidad Politécnica de Sao Paulo (Brasil) por su colaboración respecto a los Autómatas Adaptativos, y al Doctor Joaquim Llisterri de la Universidad Autónoma de Barcelona (España) por su colaboración en el tema de métodos de evaluación de TTS.

Referencias

[1] J.J. Neto, "Adaptive Automata for Context-Sensitive Languages", *ACM Sigplan Notices*, vol. 29, n° 9, pp. 115-124, 1994.

[2] A. Black, P. Taylor y R. Caley, *Festival Speech Synthesis System*, 1.4 ed., University of Edinburh, 2002. [Online] Available: http://www.cstr.ed.ac.uk/projects/festival/

[3] E. Da Silva and E. Muszkat, *Metodologia da Pesquisa e Elaboração de Dissertação*, 3th ed., UFSC - Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2001.

[4] D. Jurafsky and J. H. Martin, *Speech and language processing: An Introduction to Natural Language Processing*, Computational Linguistics, and Speech Recognition, 2nd ed. New York: Prentice-Hall, 2000, p.p. 30-50.

[5] Real Academia Española y Asociación de Academias de la Lengua Española, *Diccionario Panhispánico de dudas*, Madrid: Santillana, 2005.

[6] R. Caya. "Estudio de la mejora de la calidad de un Sintetizador de voz para el castellano utilizando El método de autómatas adaptativos". Trabajo de fin de carrera de la especialidad de Ingeniería Informática de la Pontificia Universidad Católica del Perú. Perú, 2009.

[7] Sapiensman, Matemáticas: Números Decimales. [En línea]. http://www.sapiensman.com/matematicas/matematicas9.htm

[8] Nadeem Unuth, Mean Opinion Score (MOS) - A Measure Of Voice Quality. [En línea]. http://voip.about.com/od/voipbasics/a/MOS.htm

[9] R. Caya, C. Zapata. "Estudio de la mejora de la calidad de voz para un sintetizador en idioma castellano usando el método de Autómatas Adaptativos" en la memoria de Workshop de Tecnologías Adaptativas. Sao Paulo, Brasil. Enero, 2009

[10] R. Caya, C. Zapata. "Results of the improvement on synthesis system's speech quality for Spanish using Adaptive Automatas", en la memoria de the Third International Conference on Advances in Computer-Human Interactions - ACHI 2010. St. Maarten, 2010.

[11] S. Lemmety, "Review of Speech Synthesis Technology", Master's thesis, supervised by M. Karjalainen, Helsinki University of Technology, University, Helsinki, March 1999.

[12] H. Pistori, "Tecnologia adaptativa en Engenharia de computação: estado da arte e aplicações", Dissertação de doutorado, orientada por J. J. Neto, Departamento de Engenharia de Comparação e Sistemas Digitais, Escola Politecnica da Universidad de Sao Paulo, 2003.

[13] International Organization for Standardization, ISO – FAQs – Date and time format. [En línea]. http://www.iso.org/iso/support/faqs/faqs_widely_used_standards/widely_used_standards_other/date_and_time_format.htm

[14] Schollaris, Aritmética - Números naturales - Conceptos. [En línea]. http://schollaris.com.mx/020101nnconceptos.php

Rosalia Edith Caya Carhuanina(M'2010) nació en Lima, Perú, el 27 de febrero de 1986. Se recibió de Bachiller en Ciencias con mención en Ingeniería Informática de la Pontificia Universidad Católica del Perú. Obtuvo el título profesional de Ingeniero en Informática mediante trabajo de tesis. Ejerció profesionalmente en BCTS y actualmente es profesor TPA de la Pontificia Universidad Católica del Perú. Entre sus campos de interés están el procesamiento de lenguaje natural y las interfaces humano-computador.

Claudia Zapata Del Río (M'2007) nació en Lima, Perú, el 03 de octubre de 1978. Se recibió de Bachiller en Ciencias con mención en Ingeniería Informática de la Pontificia Universidad Católica del Perú y concluyó en la misma los estudios de Maestría en Ciencias de Computación.

Obtuvo el título profesional de Ingeniero en Informática mediante trabajo de tesis y es miembro del Colegio de Ingenieros del Perú.

Ejerció profesionalmente en Synopsis S.A. y actualmente es profesor auxiliar de la Pontificia Universidad Católica del Perú

Ruiz Vergara, César Augusto nació en Ica, Perú, el 26 de setiembre de 1988. Se recibió de Bachiller en Ciencias con mención en Ingeniería Informática de la Pontificia Universidad Católica del Perú.

Ejerce profesionalmente en Novatronic S.A.C y actualmente es asistente de docencia de la Pontificia Universidad Católica del Perú.