Universitat d'Alacant
Universidad de Alicante

TRABAJO FIN DE GRADO
Grado en Ingeniería en Sonido e Imagen

Implementación de un generador de prosodia para un conversor texto-voz

INTRODUCCIÓN

Un **conversor texto-voz** (CTV) genera, a partir de un texto, una voz artificial que emula a la de una persona.

La **prosodia** representa un conjunto de fenómenos fónicos clave para la interpretación de la expresión oral, incluyendo características relativas al estado de ánimo del hablante e información sociolingüística y dialectal.

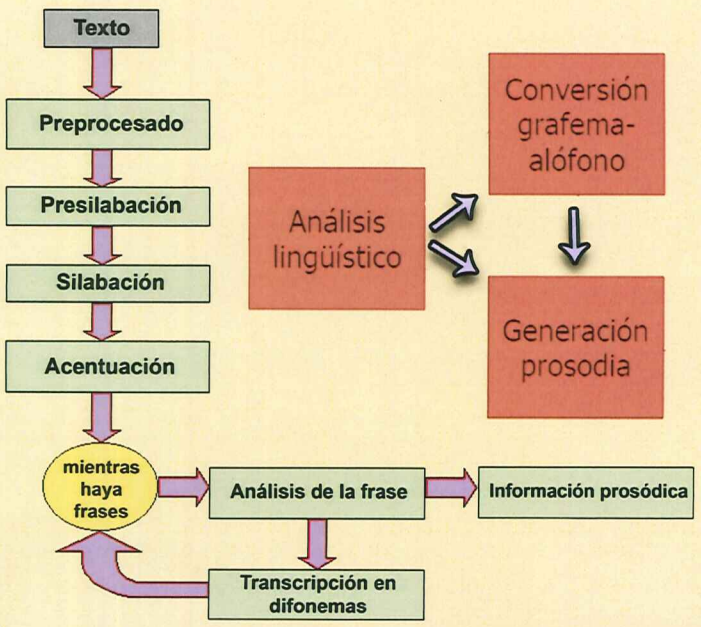
Objetivos de un CTV:
- Inteligibilidad
- Naturalidad

ACENTUACIÓN

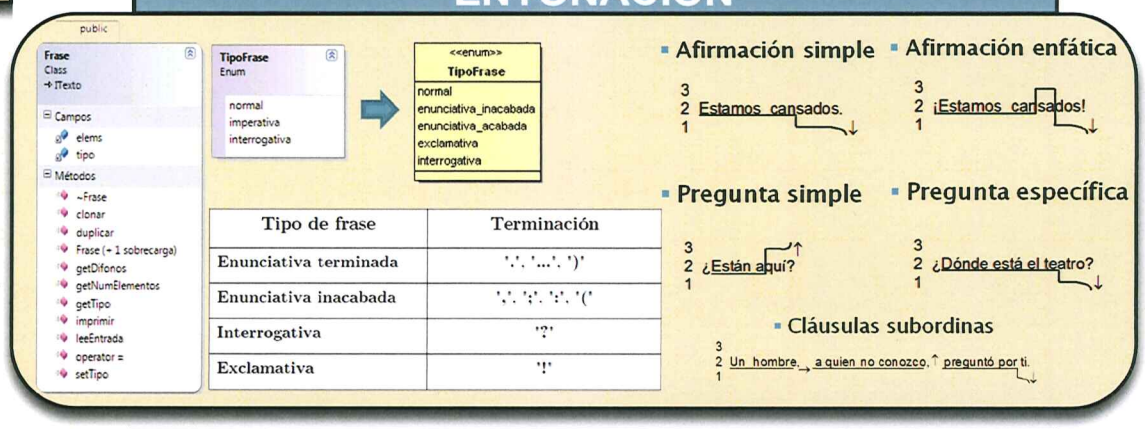
- **Reglas de silabación básicas**
  - V C V → V ¬ C V
  - C1 C2 → ¬ C1 C2
    (C1 C2 = /obstruyente/ + /líquida/)
  - C1 C2 → C1 C2 ¬
    (C1 C2 = /consonante/ + /s/)
  - C1 C2 C3 → C1 ¬ C2 C3
    (C2 C3 = /obstruyente/ + /líquida/)
  - C1 C2 C3 → C1 C2 ¬ C3
    (C1 C2 = /consonante/ + /s/)
  - C1 C2 C3 C4 → C1 C2 ¬ C3 C4
    (C1 C2 = /consonante/ + /s/)
    (C3 C4 = /obstruyente/ + /líquida/)
- **Reglas de silabación avanzadas**
  - Diptongos
  - Triptongos
  - Hiatos
  - Hache intercalada
  - Fronteras silábicas
  - Dobles consonantes
  - Diéresis
  - /tl/ diferente según regiones
  - /u/ muda con /g/
  - /u/ muda con /q/
  - Etc.

Los sonidos acentuados (sílabas tónicas) tienen:

- Mayor intensidad que los sonidos inacentuados
- Un tono más agudo
- Una duración un poco mayor



ENTONACIÓN



Tipo de frase	Terminación
Enunciativa terminada	'.', '...', ')'
Enunciativa inacabada	',', ';', ':', '('
Interrogativa	'?'
Exclamativa	'!'

- **Afirmación simple**

- **Afirmación enfática**

- **Pregunta simple**

- **Pregunta específica**

- **Cláusulas subordinas**

RESULTADOS

```
1  <?xml version="1.0"?>
2  <!DOCTYPE speak PUBLIC "-//W3C//DTD SYNTHESIS 1.0//EN"
3                         "http://www.w3.org/TR/speech-synthesis/synthesis.dtd">
4  <speak version="1.0"
5         xmlns="http://www.w3.org/2001/10/synthesis"
6         xml:lang="en-US">
7
8      <metadata>
9          <title>El ingenioso hidalgo Don Quijote de la Mancha</title>
10     </metadata>
11
12     <voice gender="male" age="30" lang="ES" speech_rate="normal">
13
14         <s tipo="normal">
15             <w>/#É/ /Én/ /n#/ </w>
16             <x>[En]</x>                    <t1>monosilabo</t1>  <t2>aguda</t2>
17
18             <w>/#ú/ /ún/ /n#/ </w>
19             <x>[un]</x>                    <t1>monosilabo</t1>  <t2>aguda</t2>
20
21             <w>/#l/ /lu/ /ug/ /gá/ /ár/ /r#/ </w>
22             <x>lu-[gar]</x>                <t1>bisilabo</t1>    <t2>aguda</t2>
23
24             <w>/#d/ /dé/ /é#/ </w>
25             <x>[de]</x>                    <t1>monosilabo</t1>  <t2>aguda</t2>
26
27             <w>/#l/ /lá/ /á#/ </w>
28             <x>[la]</x>                    <t1>monosilabo</t1>  <t2>aguda</t2>
29
30             <w>/#m/ /má/ /án/ /nc/ /ch/ /ha/ /a#/ </w>
31             <x>[man]-cha</x>               <t1>bisilabo</t1>    <t2>llana</t2>
32
33             <w>/$,/</w>
```

CONCLUSIONES

La generación de prosodia constituye la parte más compleja del procesador lingüístico debido a la dificultad de conseguir la naturalidad del habla humana por sus ambigüedades y características especiales.

✓ Todos los objetivos propuestos para este proyecto han sido conseguidos con resultados satisfactorios

El programa es capaz de reconocer y marcar los grafemas correspondientes a la sílaba tónica de una palabra.

<w>/#A/ /Aé/ /ér/ /re/ /eo/ /o#/ </w>

<w>/#s/ /si/ /ié/ /ém/ /mp/ /pr/ /re/ /e#/ </w>

```
CARACTERISTICAS EN EL FICHERO DE SALIDA:

(1)- DIFONOS
(2)- DIFONOS + SILABIZACION
(3)- DIFONOS + SILABIZACION AVANZADA

OPCION (1-3): _
```

Desde el menú de opciones se puede configurar la cantidad de información que se escribirá en el fichero de salida XML.

Carlos Sepúlveda Rubio
csr14@alu.ua.es