

# ***HANGOS BÖNGÉSZŐ*** ***(és ami hozzá kapcsolódik)***

---

***W3C – BME nyílt nap***

2007. február 22.

**Németh Géza, Zainkó Csaba, Kiss Géza, Olaszy  
Gábor, Fegyó Tibor, Mihajlik Péter, Tóth Bálint**

**BME Távközlési és Médiainformatikai Tanszék  
(BME TMIT)**

**[nemeth@tmit.bme.hu](mailto:nemeth@tmit.bme.hu)**



# Tartalomjegyzék

---

I. Rövid áttekintés

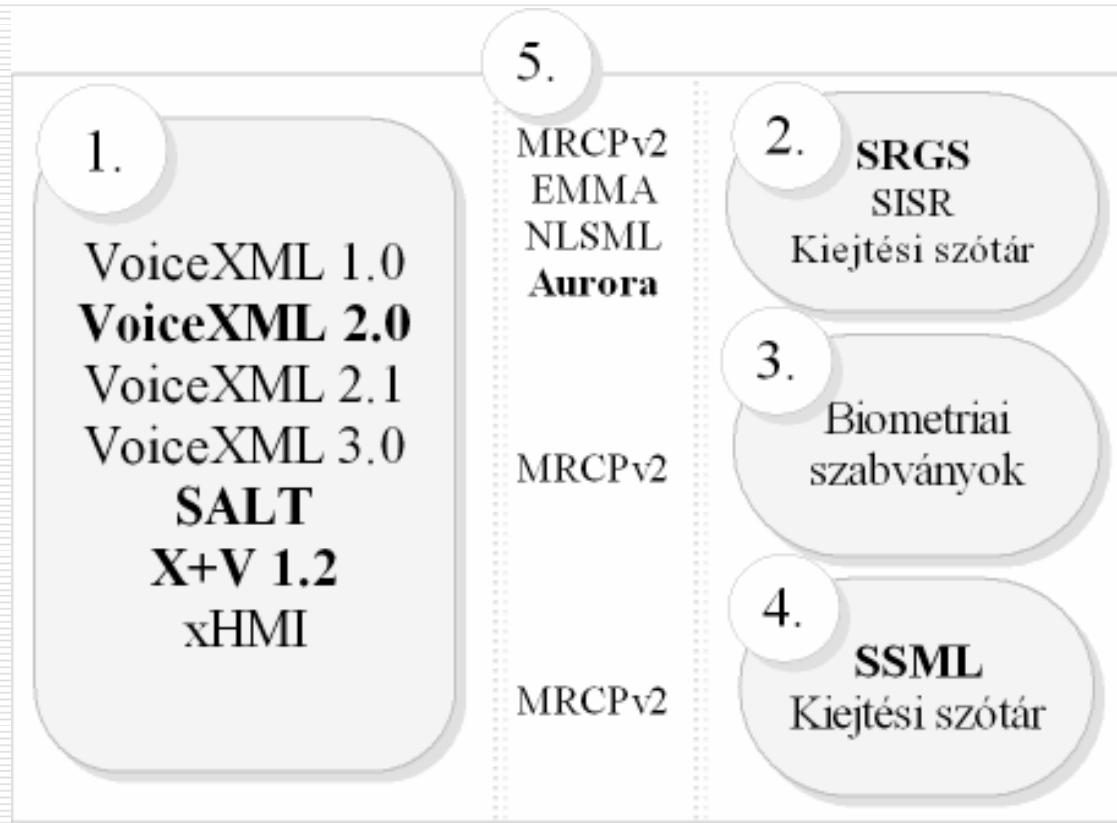
II. Demonstráció

III. Bepillantás egy W3C munkaülésre

IV. Új témajavaslat



# I. Rövid áttekintés <sup>(1)</sup>



Forrás: Abari Kálmán: Hangos jelölőnyelvek, Híradástechnika, 2006/3



# Rövid áttekintés <sup>(2)</sup>

---

- Mi történt nálunk (BME TMIT) eddig?
  - n VoiceXML (említés-szintű) bevezetése a Beszédinformációs rendszerek tárgyban (2001)
  - n Önálló labor, diplomaterv feladatok (2001-)
  - n VoiceXML alapú magyar nyelvű hangos böngésző mintarendszer kifejlesztése OpenVXI alapokon (2002-2003)
  - n Meghívottként részvétel egy W3C SSML munkaülésen (2006). Erőforrások!?!
  - n MRCP implementációs kísérletek



# II. Demonstráció

---

- OpenVXI alapú magyar nyelvű kísérleti VoiceXML böngésző (BME – TMIT)



# Internationalizing W3C's Speech Synthesis Markup Language, Workshop II

2006. május 30-31, Kréta

---



# Az SSML és a VoiceXML viszonya

---

- Mindkét nyelv XML alapú W3C specifikáció.
- Míg a VoiceXML dialógus leíró formanyelv, addig az SSML célja a szöveg kiegészítő információkkal való ellátása a felolvasás segítéséhez.
- Az SSML támogatása követelmény a VoiceXML és a SALT platform számára is.



## Suggested word element

```
<w [syllables="...-..."]  
  [POS="..." [number="..." ...]]> ... </w>
```

E.g.

```
<w syllables="hossz-szú"> hosszú </w>
```

```
<w POS="noun" number="plural"  
  case="accusative"> halászsasokat </w>
```





## Suggestion extended from other proposals

```
<w [syllables="...-..."]  
  [POS="..." [number="..." gender="..."  
    case="..." ...]  
  [morph="...+..."]  
  [tone="h+l+..."] ]> ... </w>
```

When not a word, but an expression is labeled:

```
<e [POS="..." [number="..." ...]]> ... </e>
```

E.g. three kilos

```
<e POS="cardinal" number="plural"  
  gender="neutral" case="genitive"]>  
  3 k. </e>
```

Enough to know syllable structure.

Instead of

```
<phoneme alphabet="ipa"  
  ph="&#x25B;ge&#x2D0;#x283;#x283;  
  e&#x2D0;g"> egészség </phoneme>
```

you can write

```
<w syllables="e-gész-ség"> egészség </w>
```

(Note: here you could also write

```
<sub alias="e-gész-ség"> egészség </sub>)
```



- Word forms may have several meanings/pronunciations
- Specifying part-of-speech may help

E.g.

- I will `<w POS="verb" tense="present">` read `</w>` the book
- I have `<w POS="participle">` read `</w>` the book

## Suggested language attribute

```
<phoneme [lang="..." | "x-unknown"]  
  [ph="..." [alphabet="..."] ]> ...  
</phoneme>
```

If both `lang` and `ph` is given, `lang` has priority

If language is "x-unknown", LID (language identification) is used.

We suggest that "x-unknown" can be used with `xml:lang` also.



- Text normalization effectively assisted by `say-as` element.
- The constructs we found appropriate in our practice include: `date`, `time` (including time intervals like `opening hours`), `number`, `currency`, `name`, `address`.
- Additionally suggest as standard values: `acronym/abbreviation`, `web`, `e-mail`, `phone`, `program-code`, `table`, `equation`.



## Suggested `emotion` attribute

- Mentioned here, although prosody is only one of its aspects
- Complementary to speaking-style, therefore separate attribute is suggested
- Can be used where the `xml:lang` element, i.e. `voice`, `speak`, `p`, `s`, `w`
- Possible values: "happiness", "sadness", "anger", "surprise", "disgust", "fear".



- Part-of-speech (POS) of word may affect emphasis and other aspects of prosody
- Not always possible to automatically determine
- More desirable to specify POS than to prescribe prosody (higher level, speaking style can override it)

Example in Hungarian:

- "Mondd, hogy vagy?" ("Tell me, how are you?")  
– interrogative adverb, strong (focus) emphasis
- "Igaz, hogy jól vagy?" ("Is it true that you are alright?")  
– conjunction, reduced emphasis

## Suggestion for extensions to prosody:

- Stress and prosody can be described on a **per-syllable basis**
- Extension to prosody: time can be **syllable position**
  - n decimal fractions can also be used
  - n negative values indicate n<sup>th</sup> position from end
  - n special symbol `syl_end` indicates end of expression

E.g.:

```
<prosody contour="(syl1,...) (syl1.5,...)  
  (syl2,...) ... (syl-1,...)(syl_end,...)">
```





## Suggested extensions

1. `<w [syllables="...-..."]  
[POS="..." [number="..." ...]]  
</w>`
2. `<phoneme lang="..." | "x-unknown"  
[ph="..." [alphabet="..."]]> ...  
</phoneme>`
3. `<voice | speak | p | s | w  
[speaking-style="spelling" |  
"syllabification" | "causal" |  
"news reading" | "story telling" | ...]  
[emotion="happiness" | "sadness" | "anger" |  
"surprise" | "disgust" | "fear"]  
[<xml:lang="..." | "xml-unknown">]  
</voice>`
4. `<prosody contour="(syll1,...) (syll2,...) (syll2.5,...) ...  
(syll-2,...) (syll-1,...) (syll_end,...)">  
optionally: pitch-contour (=contour),  
rate-contour, volume-contour; break, emphasis`

# Az általunk (is) megfogalmazottak közül kiemelésre került

---

<http://lists.w3.org/Archives/Public/www-voice/2006JulSep/0000.html>

- Szükség lehet a szó jelölésére.
- Van érdeklődés a szófaj felvételére.
- Vannak jól kidolgozott mechanizmusok közel-keleti (pl. arab) és kelet(közép!!!)-európai (pl. magyar) TTS-ekben, aminek a belefoglalását érdemes megfontolni.
- Külön címkével jelölik a szöveg nyelvét (a kiejtés nyelvét).



# A műhelymunka tapasztalatai

---

- Van nyitottság a véleményünk befogadására
  - Valódi eredményekhez folyamatos és erőteljes részvétel szükséges
    - n a W3C hangos böngésző munkacsoportja az egyik legaktívabb közösség
    - n hetente min. egy délelőtt távkonferencia
  - Az érdemi magyar részvételhez erőforrás és az érdekeltek együttműködése szükséges
- 



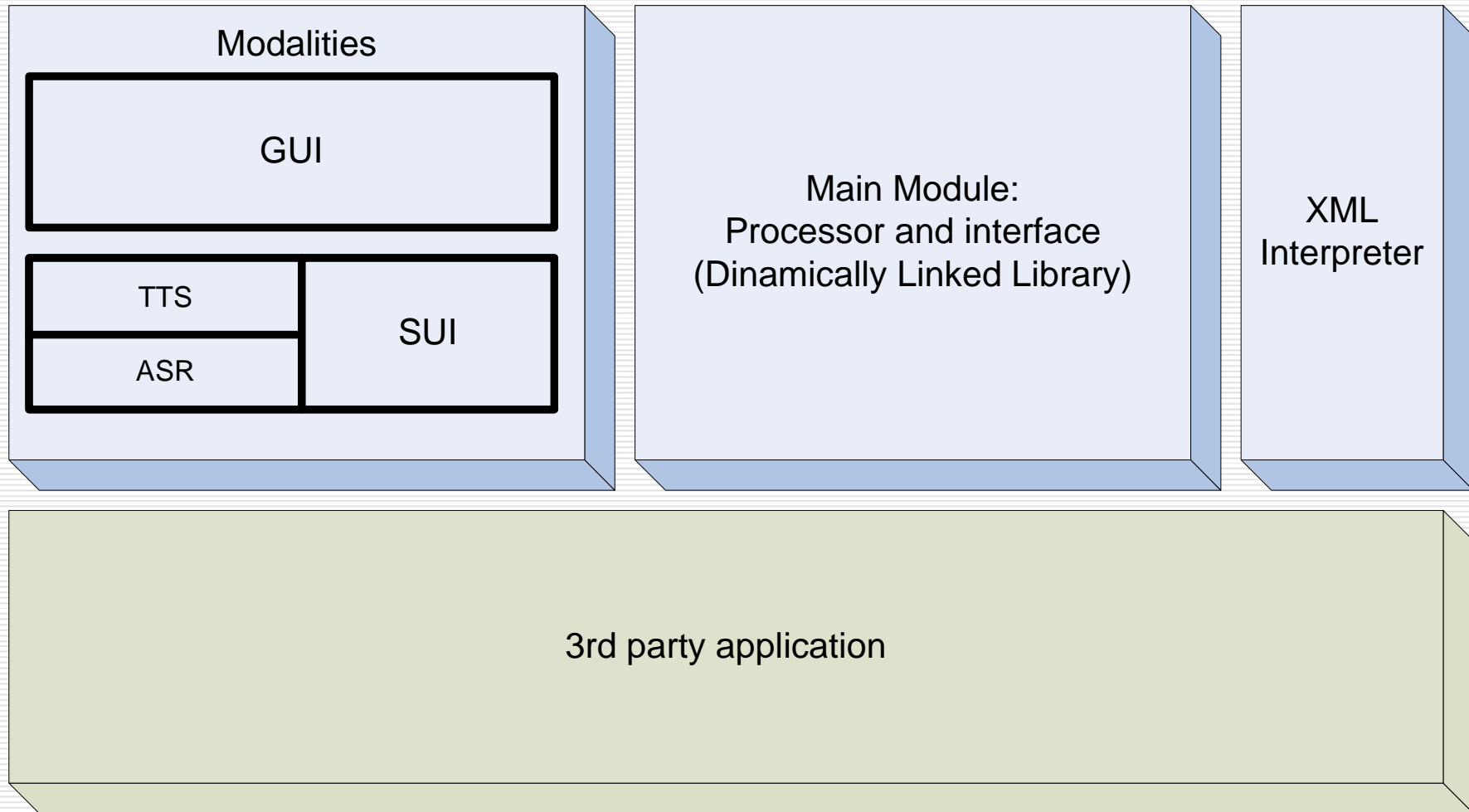
# IV. Új témajavaslat: Multimodális felhasználói felület mobil környezetben

---

- **Cél:** Grafikus felhasználói felület (Graphical User Interface, GUI) és hangos felhasználói felület (Speech User Interface, SUI) összekapcsolása különböző típusú mobil eszközökön.
- **Probléma:**
  - n Két fő ág: Symbian és Windows Mobile (PDA, Smartphone).
  - n A különböző típusú mobil eszközök kód és bináris szinten sem kompatibilisek.
  - n Ezért sok idő és energia az alkalmazások kifejlesztése külön-külön az egyes platformokra.



# Architektúra



# GUI koncepció

---

## o Grafikus felhasználói felület (GUI)

- n Olyan leírásra van szükség, mely segítségével a különböző platformokon (közel) azonos megjelenést és viselkedést érhetünk el.
- n A felhasználói vezérlők és azok viselkedése nem egységes a különböző platformokon.
- n Megoldás:  
XML Style Sheet Transformation: egy adott platformra elkészített leírás alapján generál egy másik platformon működő leírást.
- n Így a fő feladatok:
  - o Az XML leírást kezelő modul megírása a különböző platformokra.
  - o Felhasználói interakciók kezelése a „3rd party” alkalmazásban (gomb megnyomása, szövegdozoz tartalmának megváltozása, stb.).



# Példa

## ○ Grafikus felhasználói felület (GUI) - Példa

```
<platform type="PocketPC">
<object type="textbox" id="1">
  <layout enabled="true" visible="true" layer="1">
    ... általános tulajdonságok.
  </layout>
  <usercontrol>
    ... a felhasználói vezérlő specifikus tulajdonságai.
  </usercontrol>
</object>
<object ...>
  ...
</object>
...
</platform>
```



# SUI koncepció

---

- Hangos felhasználói felület (SUI)
  - n BelAmi keretein belül, VoiceXML alapján készült leírást vettük alapul. (A VoiceXML a kutatási/fejlesztési stádiumban túl bonyolult)
  - n A felhasználói vezérlőkhöz dialógusok rendelhetőek (TTS, ASR).
  - n A mobil készülékeken nincs SAPI, saját interfészt definiálunk, amihez TTS/ASR fejlesztők tudják csatolni a motorjukat.
  - n Fő feladatok:
    - ASR, TTS átírása a különböző rendszerekre.
    - A SUI-t kezelő modul implementálása a különböző platformokon.





# Példa

## ○ Hangos felhasználói felület (SUI) - Példa

```
<platform type="desktop">
<SUI_settings>
    ... globális beállítások.
</SUI_settings>
<object type="textbox" id="1" >
    <dialog dialog_id="Input" next="Input">
        <SUI_dialog>
            <SUI_settings>
                ... helyi beállítások.
            </SUI_settings>
            <waiting_for_userinput>
                ... ASR/TTS rész.
            </waiting_for_userinput>
        </SUI_dialog>
    </dialog>
    ...
</platform>
```



# Tervek

---

- Jelenleg Windows Mobile alapú készülékeken folyik a fejlesztés.
- Következő lépések:
  - partner keresés
  - Komplex példaalkalmazások készítése az XML leírás segítségével.
  - Szerver (streaming) alapú ASR/TTS bevezetése.
  - Symbian implementáció
  - Teszt és konformancia specifikáció és környezet kialakítása

