

---

# SSML for Indian Languages

## Text to Speech Synthesis

---

Presented by: Vibhu Agarwal

President and co-founder  
Prologix Software Solutions Pvt Ltd.  
( [www.prologixsoft.com](http://www.prologixsoft.com) )

---

# Common difficulties encountered while rendering Indian language texts.

- ◆ Use of mixed language
  - ◆ Encoding schemes vary across fonts
  - ◆ Non standard encoding of text
  - ◆ Use of Roman transliterated text
  - ◆ Multiple transliteration schemes
-

---

# Handling encoding issues in SSML

- ◆ Provide script and encoding explicitly.

- ◆ `<transliterate format URL> ....`  
`</transliterate>`

- ◆ **Example**

```
<transliteration format="myformat"
  url=http://www.prologixsoft.com/transformat/mytrans.map> mai
  hindee me likhe shabdo ko paDh_saktaa hun </transliteration>
```

- ◆ Transliteration rules can be read by the processor at the indicated URL
-

a	
aa	
i	
ee	
u	
oo	
k	
kh	
g	
gh	
ng	
	.
	.
	.
s	
sh	
h	

# Example of a transliteration rule

---

# Handling encoding issues in SSML

- ◆ **Example**

```
<transliteration format="null"
```

```
  url=http://www.prologixsoft.com/encode/encod.map>
```

```
  </transliteration>
```

- ◆ In case of non-standard encoding of characters, rules can be read by the processor at the indicated URL



## Encoding Map

ASCII Character	Legacy Font	Unicode
!	/	!
"	ॐ	ब्
#	ॐ	र
\$	+	+
%	:	०:
:	ॐ	रु
;	य	य
<	ढ	ढ
=	व	त्र
>	झ	झ
?	६	घ्
@	/	/
A	/	।
B	ठ	ठ
C	ॐ	ब्
D	य	क्
E	न	म्
F	६	थ्
G	ळ	ळ
H	४	भ्



# Handling mixed language use

- ◆ Should be suitably indicated to TTS where use of mixed language is expected to affect rendition

- ◆ Example: breaks

< break domain= "interlanguage"> इस software की performance बेकार है< break/ >

The SSML implementation may choose to interpret the above as

[pause] software [pause]      [pause] performance [pause]