



Machine Learning in Signal Processing

Projects (Speech Synthesis)

Text to Speech without Text

- ◆ *Many languages don't have orthography*
- ◆ *Build a TTS voice without letters*
 - *Use acoustics to derive "phones"*
 - *Uses statistical parameters synthesis*
 - *Test on know language (e.g. English)*
 - *Test on unknown language*
 - *Bonus: can you find "words"?*

Live Mimics

- ◆ *Use voice conversion techniques to convert a voice*
- ◆ *Do it live (as the person speaks)*
- ◆ *Use GMM spectral, intonation conversion*

De-identification of speech

- ◆ *Make your voice not identifiable*
- ◆ *Modify spectrum, pitch duration*
 - *So *no* speaker id system or human can tell*
- ◆ *Preserve: style emotion etc*
 - *But no speaker properties*

Voice Characters

- ◆ *Build various characters*
 - *For a 3D world (Second Life/Alice)*
- ◆ *Different genders, age, styles*

Will this call succeed?

- ◆ *Let's Go Bus Information systems*
 - *Live CMU system for Pittsburgh*
 - *Gives time of next bus*
- ◆ *When can you tell if a call will succeed*
 - *Estimate signal-to-noise, ASR accuracy*
 - *Style of speech etc.*
- ◆ *100,000 calls for training data*

Sing!

- ◆ *Build a singing synthesizers*
 - *(Sinsy.jp)*
- ◆ *Make your voice sing (or others)*

Speech from Gesture

- ◆ *Map 3D movements to speech*
 - *Use Wii (accelerometers)*
 - *Control Articulatory Parameters*
 - *Which Control Speech Output*

