

# ***ModelTalker: Automatically creating personal synthetic voices***

Chris Pennington ('88)  
AgoraNet, Inc.  
Newark, DE

# *Current Project Staff*

- Dr. H. Timothy Bunnell (PI),  
Jason Lilley, James Mantell, James Polikoff  
(*Speech Research Laboratory,  
Alfred I. duPont Hospital for Children, Wilmington DE*)
- John Gray, Chris Pennington,  
Debra Yarrington  
(*AgoraNet, Newark DE*)



AgoraNet, Inc.



Alfred I. duPont  
Hospital for Children  
Nemours  
Children's Clinic

# Overview

- Background and Project rationale
- System components
  - InvTool / BCC recording and analysis
  - ModelTalker synthesis
- Discussion and Future development



# *Project Background*

- 10+ years of research at Speech Research Laboratory headed by Dr. Tim Bunnell
- AgoraNet founded by former staff of UD Center for Applied Science and Engineering
- NIH STTR (Small Business Technology Transfer) Phase I grant awarded in 2003. Phase II awarded / started Jan 1, 2007.



AgoraNet, Inc.

Nemours

Alfred I. duPont  
Hospital for Children

Nemours  
Children's Clinic

# *Focus on AAC*

- Augmentative and Alternative Communication
- Unable to speak intelligibly, use pre-recorded or Text-to-Speech for communication
- Strong desire for personalized voices
  - Especially in groups like ALS (aka Lou Gehrig's disease)
- Minimal funding available



# *Current AAC technology*

- Prerecorded speech
  - very natural sounding and intelligible
  - limited vocabulary, no unique utterances
- Synthetic Speech (most still use DECTalk)
  - able to create novel utterances
  - limited intelligibility and voice selection
- Newer AAC devices support SAPI voices
  - much higher intelligibility, broader voice selection



# *Current Speech Synthesis*

- Unit Concatenation (UC)
  - improvement over formant-based synthesis (e.g., DECTalk)
  - speech is recorded
  - stored in a database with labels and other information
  - “units”, or pieces of speech appended together
  - unlimited synthesis with limited recorded speech

# *Unit Concatenation Approaches*

- Diphones
  - unrestricted speech
  - speech database size reasonable - reasonable amount of recording required
  - individual's voice is present in synthesized speech
  - too limited - sometimes utterances that don't match well acoustically must be appended together – can sound unnatural and jarring

# *Unit Concatenation Approaches*

- Variable-Length Unit Concatenation
  - large database of speech is recorded with labels
  - during synthesis, algorithm searches database for longest stretches of speech that, when appended together, will best match the desired utterance
  - synthetic speech quality
    - *best*: approaches that of recorded speech
    - *worst*: high quality synthetic speech in similar “voice”

# *Phase I Feasibility Study*

- Compared DECTalk with several unit concatenation systems (all female voices) using Semantically Unpredictable Sentences
- All unit concatenation systems significantly more intelligible than DECTalk
- ModelTalker comparable in intelligibility to commercially-available systems (at less than 10% of the cost and recording time)

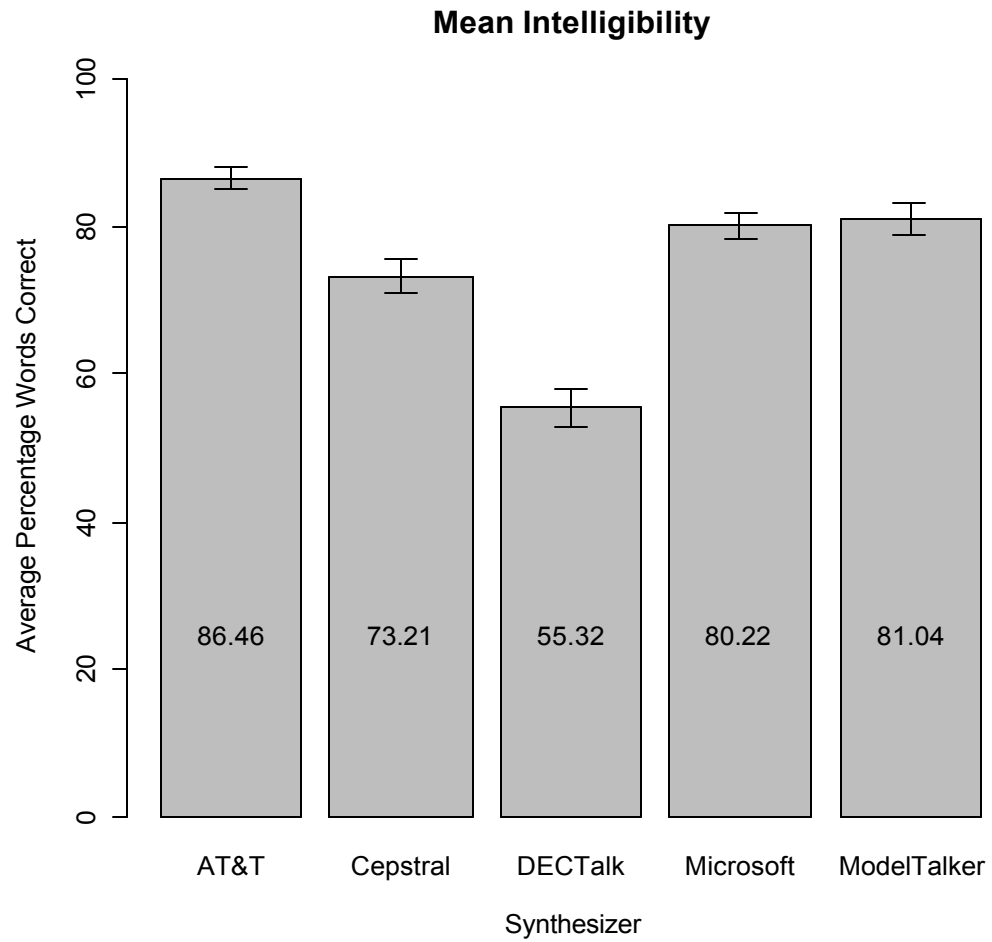


Figure 1 Overall mean intelligibility scores (percentage of words correct) for the five synthetic voices. Error bars are the 95% confidence intervals around each mean.

# Overview

- Background and Project rationale
- **System components**
  - InvTool / BCC recording and analysis
  - ModelTalker synthesis
- Voice Database construction
- Discussion and Future development

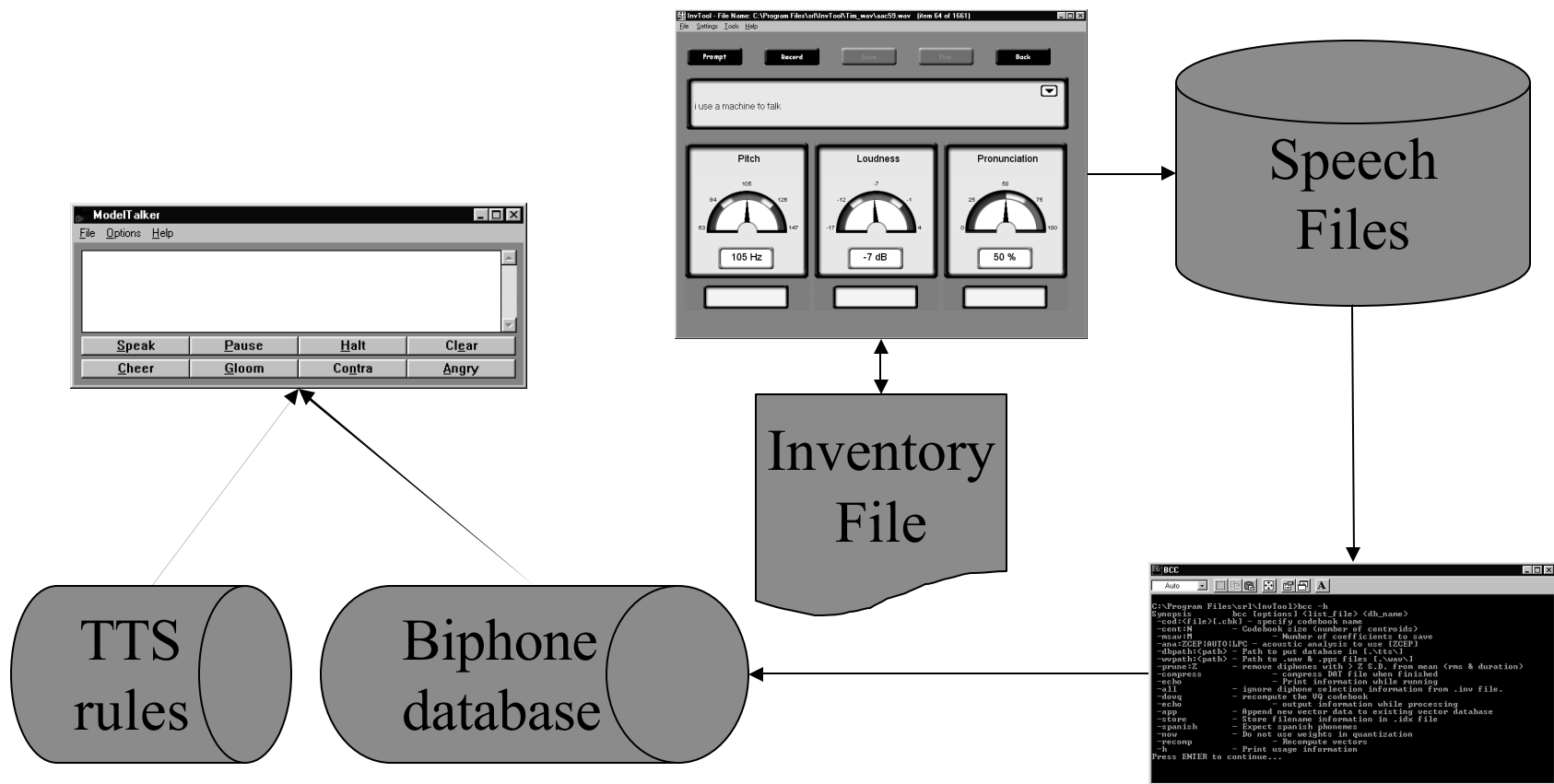


# *System Components*

- Three major applications
  - ***InvTool***: speech recording
  - ***BCC***: speech analysis and database construction
  - ***ModelTalker***: The Text-to-Speech engine
- Recording Inventory
  - Editable list of utterances
  - Ordered by significance to task



# Overall System Design



AgoraNet, Inc.



Alfred I. duPont  
Hospital for Children  
Nemours  
Children's Clinic

# InvTool

The screenshot shows the InvTool application window with the following components and annotations:

- File Name:** C:\invtool\wav\invest9.wav (item 9 of 91)
- Menu:** File, Settings, Tools
- Buttons:** Prompt, Record, Save, Play, Back
- Text Entry:** A text box containing "could be better" with a dropdown arrow.
- Pitch Gauge:** A semi-circular gauge with values 69, 92, 115, 138, 161. The needle points to 107 Hz. Below it is a "Good!" button.
- Loudness Gauge:** A semi-circular gauge with values -17, -12, -7, -1, 4. The needle points to -7 dB. Below it is a "Good!" button.
- Pronunciation Gauge:** A semi-circular gauge with values 0, 25, 50, 75, 100. The needle points to 83%. Below it is a "Good!" button.

Annotations with arrows pointing to specific features:

- Pitch:** Measures the average pitch of your recording and assures that you are within a proper range.
- Loudness:** Assures that your recording is not too loud or too soft.
- Pronunciation:** Compares what you said to what you were supposed to say.
- Performance bars:** let you know if the parameters are within an acceptable range.



AgoraNet, Inc.



Alfred I. duPont  
Hospital for Children  
Nemours  
Children's Clinic

# *Inventory Control File*

```
0 aac2 hello. 0B<a6>HXAX<a1>LLOU0E<c6>{HL-L%} keep
0 aac8 I'm_ok. 0B<a6>TQAIMM<c2>OU<a1>HKEI0E<c6>{HL-L%} keep
2 User0016 vickie 0B<a6>VVIHKK<c0>{HL-L%}II0E<a6> keep
```

- Five fields per entry:

- ? Current status flag
- ? File name
- ? English gloss
- ? Phonetic transcription with prosodic tags
- ? Process flag (NULL, drop, keep, freeze)



# *ModelTalker*

- Laboratory-level research tool for manipulating and outputting synthetic voices
- In Phase I, core code was tested, revised and made more robust (including source control)
- Also, speech engine was made SAPI 5 (Speech Application Programming Interface) compliant for use with any supporting Windows application (e.g. MS Reader)

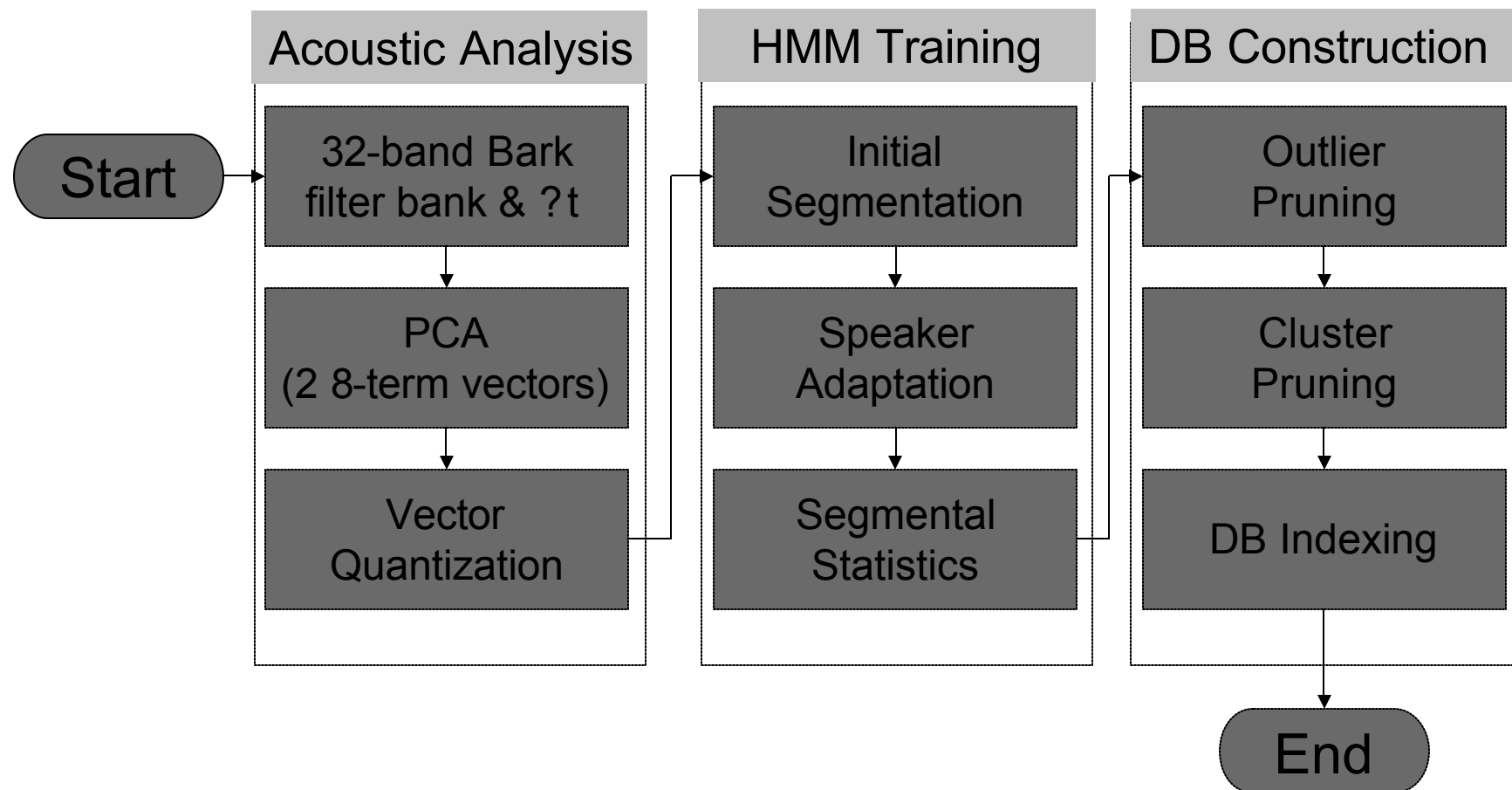


# Overview

- Background and Project rationale
- System components
  - InvTool / BCC recording & analysis
  - ModelTalker synthesis
- **Voice Database construction**
- Discussion and Future development



# BCC Processing



# *ModelTalker Example Voices*


- Field-recorded ALS patients

- Male (actual)  (synthetic) 

- Female (actual)  (synthetic)  (actual)  (synthetic) 

- Non-ALS speakers

- Female (actual)  (synthetic)  (pitch control) 

- Male (actual)  (synthetic)  (actual)  (synthetic) 

# Overview

- Background and Project rationale
- System components
  - InvTool / BCC recording & analysis
  - ModelTalker synthesis
- Voice Database construction
- **Discussion and Future development**



# *Discussion and Future*

- Research Project Goals
  - segmentation working well, but may still benefit from some improvement.
  - signal processing is slated for complete revision: better pitch and intonation control
  - continuing to examine how to balance prosodic control / naturalness / inventory size (currently requires 4 to 6 hours to record)
  - field study of prototype voice creation systems deployed at 5 ALS clinics in U.S.

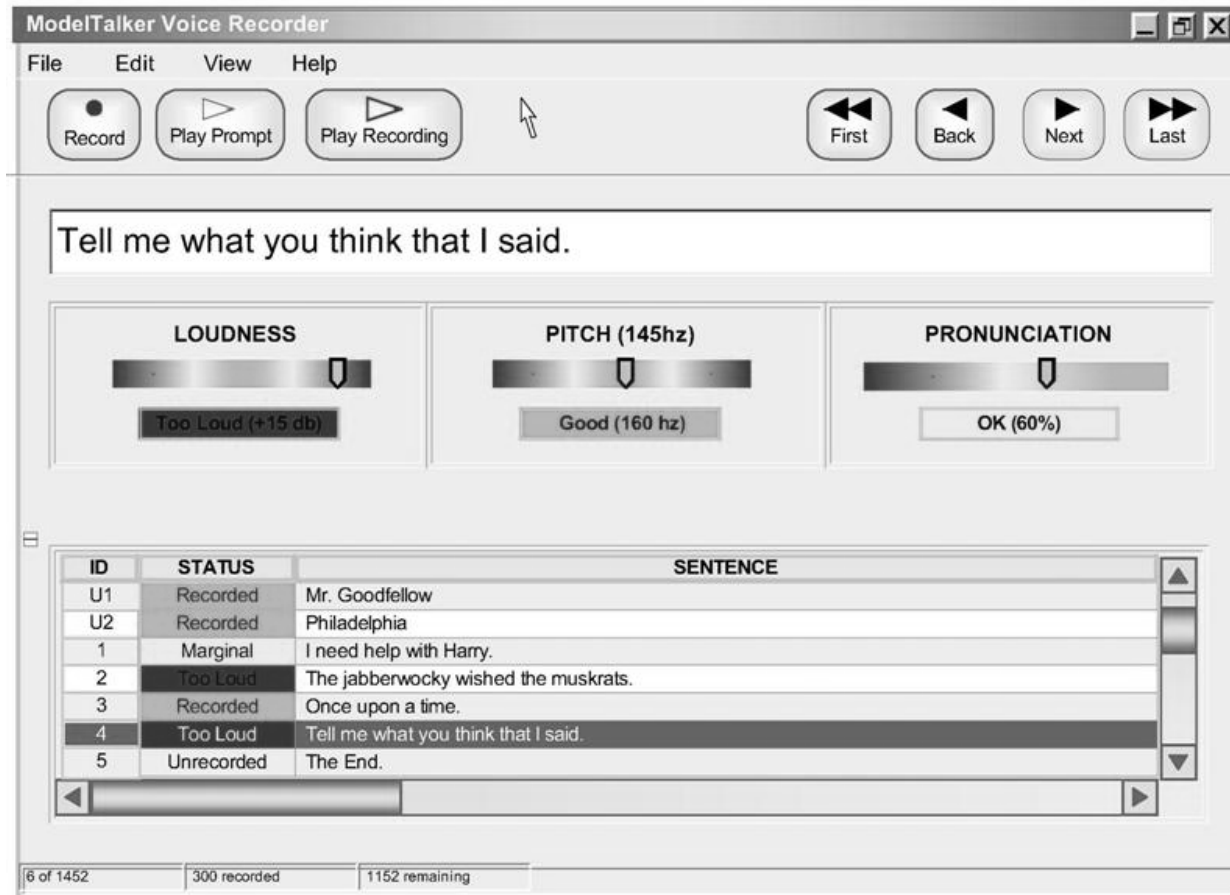


# *Discussion and Future*

- Software Development Goals
  - Additional SAPI testing and feature refinement (most were implemented in Phase I)
  - Complete rewrite of voice recording tool interface including full accessibility for people with visual impairments
  - Porting at least the speech synthesis engine to Windows Mobile (several AAC devices use this OS), Linux and MacOS
  - Possible web/server-based speech synthesis



# Proposed Recording Interface



# *Discussion and Future*

- Commercialization Challenges
  - Trademark, patent, IP and naming issues
  - Broader applications of technology (e.g., screen readers, gaming/avatar voices, literacy and educational support)
  - Product line determinations
  - Secure funding for marketing research, sales, production and technical support



AgoraNet, Inc.

Nemours

Alfred I. duPont  
Hospital for Children

Nemours  
Children's Clinic

# *Acknowledgments*

This work was supported by STTR grants R41/R42-DC006193 from NIH/NIDCD and by Nemours Biomedical Research.

Thanks to the many ALS patients, AAC specialists in clinics, and other interested individuals who have used this software and provided valuable feedback.



AgoraNet, Inc.

Nemours

Alfred I. duPont  
Hospital for Children

Nemours  
Children's Clinic

# Questions?

<http://www.asel.udel.edu/speech/>

<http://www.modeltalker.com>

[mtinfo@modeltalker.com](mailto:mtinfo@modeltalker.com)

Chris Pennington ([penningt@agora-net.com](mailto:penningt@agora-net.com))



AgoraNet, Inc.



Alfred I. duPont  
Hospital for Children

Nemours  
Children's Clinic

# *ModelTalker Demonstration*