“ALS and Speech synthesis, Voice Banking and Message Banking™: HISTORY, CURRENT AND FUTURE TRENDS”

John M. Costello
Meghan O’Brien
Boston Children’s Hospital
ALS Augmentative Communication Program
Bostonchildrens.org/ALSAugComm

ALS Augmentative Communication Program

John Costello, M.A. CCC-SLP
Peggy Dellea, M.S. OT/L
Meghan O’Brien, M.S. CCC-SLP
Amaliya Silsby, M.A. CCC-SLP
The most important qualities of a speech synthesis system are naturalness and intelligibility (Taylor, 2009)

- While the early electronic speech synthesis was robotic and frequently barely intelligible, contemporary speech synthesis much more closely approximates the sound and clarity of natural speech.
- Despite this, speech synthesis continues to remain distinguishable from authentic human speech.
The ‘Speaking Machine’ of Von Kempelen 1734

The "speaking machine"
ofWolfgang von Kempelen(1734 - 1804)

The VoderHomer Dudley(Bell Telephone Laboratories)1939

First device that could generate continuoushuman speech electronically. The compositionof the many speech sounds had to be done manually in realtime on a special keyboard.
Dennis Klatt  DECTalk

- 1976 sentence level phonology incorporated by Dennis Klatt
- Pioneer of speech synthesis in technology sense
- Provided interface for non experts...created the first text to speech
- Focused on the listener perspective to help understand what qualities people attended to in determining
- a voice as male, female or child
As intelligibility and naturalness of speech synthesis has continued to improve, research and development has shifted to the personalization of voice and the impact on the communication and comprehension of intent with one’s personal voice.

(Higginbotham, 2010)
Voice Banking: What is it?

**Historically:** Voice banking is a process of recording a large inventory of your speech following a pre-determined script designed to capture a sampling of all co-articulated sounds in the language. The completed collection of recordings is then used to create a synthetic voice that approximates your natural voice.

Costello, 2011

---

Costello and O'Brien. Boston Children's Hospital ASL Augmentative Communication Program. © 2018
AssistiveWare Video July 2012

Putting the pieces together from a voice bank to create a new word!
- ModelTalker
- Cereproc (Edinburgh Scotland)
- VocaliD
- Acapela
- Lyrebird

<table>
<thead>
<tr>
<th>Voice Bank Comparison Table. Costello August 2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>ModelTalker Confirmed 7/22/18</td>
</tr>
<tr>
<td>Process</td>
</tr>
<tr>
<td>Time commitment</td>
</tr>
<tr>
<td>Literacy need</td>
</tr>
<tr>
<td>Bank from home</td>
</tr>
<tr>
<td>Premier before Pay*****</td>
</tr>
<tr>
<td>Cost (US dollar)</td>
</tr>
<tr>
<td>Incorporate into SGD</td>
</tr>
<tr>
<td>Requires wireless to access</td>
</tr>
<tr>
<td>Can use Message bank instead of artificial script</td>
</tr>
<tr>
<td>Languages</td>
</tr>
<tr>
<td>Equipment needed</td>
</tr>
</tbody>
</table>

NOTE: time commitment based on strong voice and ease of speaking.
Link to voice bank table:

ModelTalker

Here's how it works:

If you have banked a sufficient amount of speech, we can recreate your voice in digital form. Your Vocal Legacy is wholly created from your own recordings.

Bank your voice

1. Record 5-7 hours of speech on The Human

Your Vocal Legacy

2. VocalID reconstructs your voice to be used

VocalID
Acapela from 30 minutes of non-scripted message bank *(double dipping!)* DNN

**MOV 1**

- John, US English
  - Sample - Original voice: [http://bit.ly/2l3w00]
  - Sample - voice created by Acapela DNN: [http://bit.ly/3r3mmp]

- Stephen, US English
  - Sample - Original voice: [http://bit.ly/2u0P0P8]
  - Sample - voice created by Acapela DNN: [http://bit.ly/3r3mmp]

- Anonymous user, French
MOV DNN (Swedish)

Lyrebird

THIS IS AN ARTIFICIAL VIDEO CREATED BY LYREBIRD. IT DOES NOT CONVEY THE OPINION OF BARACK OBAMA.
Voice vs Message Banking™

Message Banking™ : What is it

“Digitally record and store words, phrases, sentences and personally meaningful sounds and/or stories using one’s natural voice, inflection and intonation.”

Costello, J.M. 2008
“Our voice is our ACOUSTICAL fingerprint”

Costello, J.M. 2010

Our Message Banking™ Model
Our Message Banking™ Model
Montage of message banking

Our Message Banking™ Model

• Technology/platform agnostic

• Bank with as much or as little structure/’hand holding’ as you would like

• Continue over time for as long as you would like. As life and the disease progress, new messages become evident.

• You are never ‘finished’ but also you have no quota. Focus on what is important to you.
Web-based video tutorials

Bostonchildrens.org/ALSMessageBanking

Demo of Mymessagebanking.com
‘Double Dip’ Message Bank to create SYNTHETIC VOICE with Voice Banking

- Creating custom synthetic voice historically requires recording hours of nonsense sentences
- This is a poor use of time for someone facing a life threatening illness such as ALS
- BCH has worked for nearly three years with industry leaders in voice banking to ‘double dip’; using a person’s authentic MESSAGE banked messages to create custom synthetic voice.
- BCH partnered to created a free website that manages the files for patients (Mymessagebanking.com)
- Acapela is the first industry leader to modify their process to work with the BCH model

---

**Voice Bank Comparison Table. Costello August 2018**

<table>
<thead>
<tr>
<th>Model Talker</th>
<th>Cereproc</th>
<th>VocalID</th>
<th>Acapela</th>
<th>Lyrebird</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process</strong></td>
<td>1541 scripted messages</td>
<td>1000 scripted messages</td>
<td>1575 scripted messages</td>
<td>30 minutes of quality message banking – authentic messages</td>
</tr>
<tr>
<td></td>
<td>Choose build for Window or OSX in advance.</td>
<td></td>
<td></td>
<td>-minimum 30 minutes of voice bank up to three hours from script</td>
</tr>
<tr>
<td><strong>Time commitment</strong></td>
<td>Seven hours</td>
<td>Four hours or 40 minutes for lesser quality</td>
<td>Seven hours</td>
<td>Minimum 30-minutes to 90 minutes -total of one’s message bank with at least 30 minutes</td>
</tr>
<tr>
<td></td>
<td>Three hours</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Literacy need</strong></td>
<td>yes</td>
<td>yes</td>
<td>NO, each message can be read aloud by software</td>
<td>No</td>
</tr>
<tr>
<td><strong>Bank from home</strong></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Premier before Pay</strong>**</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Cost (US dollar)</strong></td>
<td>100. 675. 1500. 99.00 per year OR 999.00 once.</td>
<td>free</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Incorporate into SGD</strong></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>NO, only works on the Lyrebird sit/ with wireless</td>
</tr>
<tr>
<td><strong>Requires wireless to access</strong></td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
</tr>
<tr>
<td><strong>Can use Message bank instead of artificial script</strong></td>
<td>No **</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Languages</strong></td>
<td>English</td>
<td>English</td>
<td>English</td>
<td>13 languages</td>
</tr>
<tr>
<td><strong>Equipment needed</strong></td>
<td>special microphone</td>
<td>special microphone</td>
<td>Special microphone</td>
<td>special microphone OR handheld .wav recorder</td>
</tr>
</tbody>
</table>

*NOTE: time commitment based on strong voice and ease of speaking.*
25-minutes of message banking used to create a synthetic voice (available as of Oct 4, 2018)

Natural message from message banking™ compared to voice created with double dipping/Acapela MOV
Natural message from message banking™ compared to voice created with double dipping/Acapela MOV

Other Message Banking Models

- Record directly to augmentative communication software on a device
- Record directly to an augmentative communication software on app following a script

*** We have concerns and have seen some difficult situations/less than optimal outcomes when someone has been advised to
1. acquire a device or software and message bank prior to knowing it is the right match to their authentic needs OR
2. to message bank from a script
Future

• *Ability to ‘double dip’ and create custom voice through message banking
• Ability to create high quality customized voices with extremely small sample
• Focus on infusing personality with intonation, appropriate pitch change and cadence based on content of message
• Google voice
• New software in Saudi that will read the message and apply contextual intonation (coming soon).

Future

• *More options to ‘double dip’ and create custom voice through message banking
• Ability to create high quality customized voices with extremely small sample
• Focus on infusing personality with intonation, appropriate pitch change and cadence based on content of message