

“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





For lots of related resources, Join us on Facebook at:

<https://www.facebook.com/ACPCHBoston>



Boston Children's Hospital
ALS Augmentative
Communication Program



HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

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.



Boston Children's Hospital
ALS Augmentative
Communication Program



HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

The 'Speaking Machine' of Von Kempelen 1734

The "speaking machine"
of
Wolfgang von Kempelen
(1734 - 1804)

The Voder

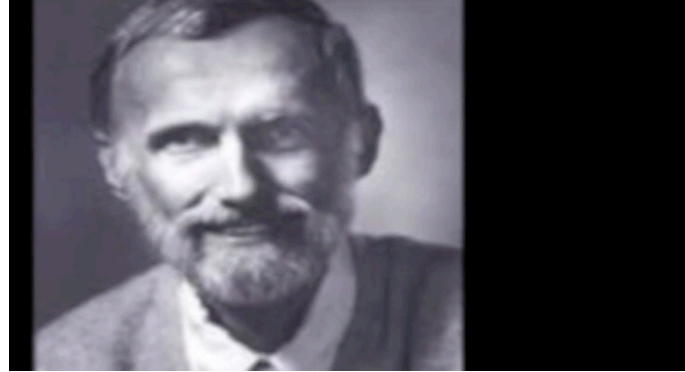
Homer Dudley
(Bell Telephone Laboratories)
1939

First device that could generate continuous human speech electronically. The composition of the many speech sounds had to be done manually in realtime on a special keyboard.

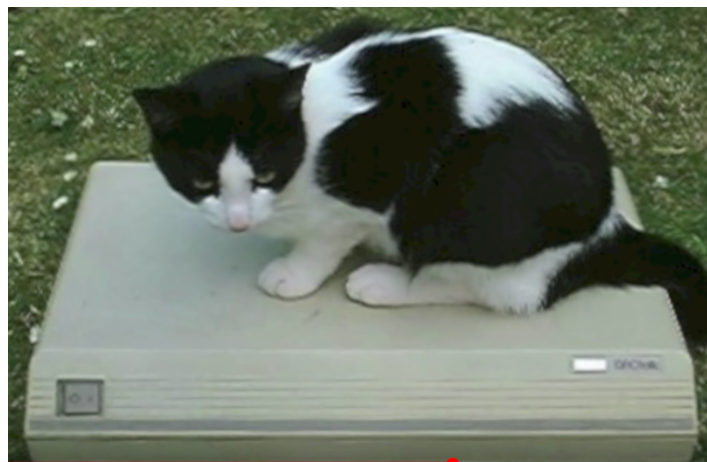


Dennis Klatt DECTalk

- 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



Dennis Klatt: DecTalk



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)

A journey through time: from generic to personal



From: nicolas.mazars@acapela-group.com

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



The screenshot shows the Acapela website interface. On the left, a navigation menu includes 'Home', 'Login', 'Record', and 'Listen'. The main content area features a 'Home' section with the following text:

Acapela group is proud to offer you a unique chance to create your own text-to-speech voice. Usually to your text in a small handful of seconds, our website will create your computer voice that yourself, flavoured with all the personal mood and voice colour that you put into it.

Acapela's **MyOwnVoice.com** is for you if you are:

- a professional voice talent curious to explore the endless possibilities of text-to-speech
- an amateur voice talent willing to show the world how nice your text-to-speech voice sounds
- an individual in need to preserve your voice for later

The text-to-speech interface is highlighted with a blue box in the original image. Below the list, the text reads: "This website will guide you through the whole process of creating and reproducing the sound of your voice from A to Z. All you need to do is get a good quality microphone and record the text materials offered through MyOwnVoice online interface. The recordings will be automatically analyzed on this server using our state-of-the-art unit selection algorithms. You will get access to your own personalized and secure text-to-speech demo, while you enjoy your dinner. By the time you finish the recordings, you will have completed a perfect speech workout having pronounced most possible sound combinations in your language."

On the right side of the screenshot, there is a video player showing a young boy holding a white telephone receiver to his ear, and a sidebar with the heading "Let's Talk :)" and a "voice solutions" link.

AssistiveWare Video July 2012



Putting the pieces together from a voice bank to create a new word!



AssistiveWare Video July 2012

- ModelTalker
- Cereproc (Edinburgh Scotland)
- VOCALiD
- Acapela
- Lyrebird

Voice Bank Comparison Table. Costello August 2018

	Model Talker Confirmed 7.22.18	Cereproc	VocalID Confirmed July 23	Acapela Confirmed July 24	Lyrebird
Process	1541 scripted messages	1000 Scripted messages Choose build for Window or OSX in advance.	1575 scripted messages	-30 minutes of quality message banking – authentic messages -minimum 30 minutes of voice bank up to three hours from script	Scripted messages
Time commitment	Seven hours	Four hours or 40 minutes for lesser quality	Seven hours	Minimum 30-minutes to 90 minutes -total of one's message bank with at least 30 minutes	Three hours
Literacy need	yes	yes	NO, each message can be read aloud by software	No	yes
Bank from home	yes	yes	yes	yes	yes
Premier before Pay*****	yes	no	no	yes	n/A
Cost (US dollar)	100.	675.	1500.	99.00 per year OR 999. once.	free
Incorporate into SGD	yes	yes	yes	yes	NO, only works on the Lyrebird sitte/ with wireless
Requires wireless to access	NO	NO	NO	NO	YES
Can use Message bank instead of artificial script	No **	no	no	yes	no
languages	english	english	English	13 languages	English
Equipment needed	special microphone	special microphone provide @ purchase	Special microphone	special microphone OR handheld .wav recorder	special microphone

Link to voice bank table:

- <http://www.childrenshospital.org/voicebankjuly2018>

CereProc Voice Banking



CereVoice Me Voice Cloning Service



CereVoice Me is a revolutionary online voice cloning tool from CereProc - allowing you to create a computer version of your own voice! Our engineers have simplified CereProc's industry-leading text-to-speech voice creation process, allowing you to carry out recordings in your own home in as little as a couple of hours, for a fraction of the cost of a traditional voice build (currently £499.99).

Typical voice creation methods require a large amount of recorded speech and intensive post-production work. This produces outstanding results, but it is time consuming and expensive. Unfortunately, this can be a barrier for those with most need for a TTS voice that sounds like them. The CereProc team has designed CereVoice Me to make voice cloning accessible to everyone. It is especially useful for voice banking. Degenerative illnesses like MND or ALS, or operations such as a laryngectomy, can lead to the loss of speech - with a CereVoice Me voice, a speech generating device can later be used to talk with your own voice.

CereVoice Me is web based - you can use it via your own computer, without the need for specialist equipment or a studio environment. You will be asked to read out prompts from our specially designed script, and the collected audio data is then used to build a Windows SAPI TTS version of your own voice.

To gain access to CereVoice Me, first create a CereProc website account and then visit the Store.

Below are some examples of our Jess voice created using CereVoice Me.

Jess Example 1:

- Text: *Speech synthesis is the artificial production of human speech*
- Recording: Click for a recording of Jess
- CereVoice Me Synthesis: Click for a CereVoice Me sample of Jess

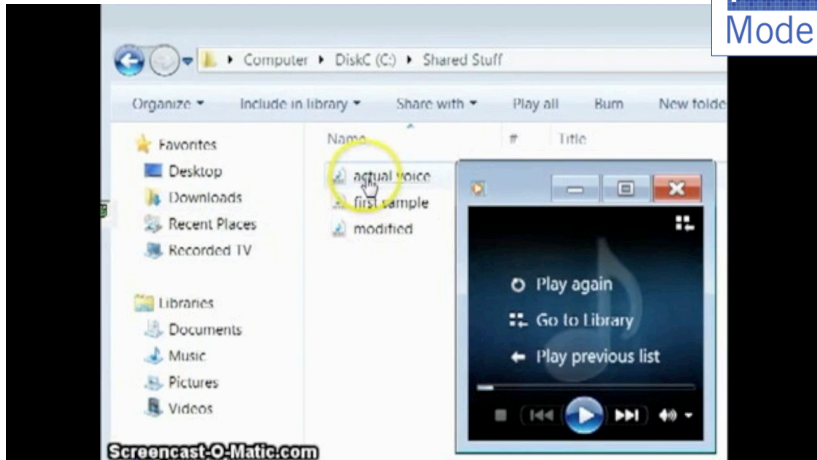


Boston Children's Hospital
ALS Augmentative
Communication Program



HARVARD MEDICAL SCHOOL
TEACHING HOSPITAL

ModelTalker



VocalID

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.

- ▶ Play Alan
- ▶ Play Lisa
- ▶ Play Joseph
- ▶ Play Elisabeth
- ▶ Play Amy

Bank your voice

Your Vocal Legacy



1.


Record 5-7 hours of speech on The Human



2.

VocalID reconstructs your voice to be used

Acapela



**Made with
MOV1**

English (US)	100.0%	Done
English (UK)	0.0%	Not enough recorded sentences
French (Canada)	0.0%	Not enough recorded sentences
North American Spanish	0.0%	Not enough recorded sentences
English (Australia)	0.0%	Not enough recorded sentences




Language: English (US)

Text to pronounce:
This is John. This is my computer voice.
This is John. this is my computer, voice.


Speed: %

Shaping: %

0:00







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



MOV 1

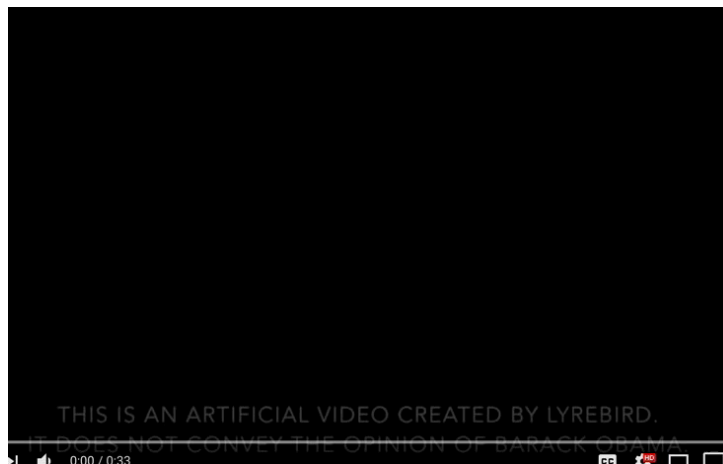
- John, US English
 Sample - Original voice: <http://bit.ly/2t3su03>
 Sample - voice created by Acapela DNN: <http://bit.ly/2trp3mp>
- Stephen, US English
 Sample - Original voice: <http://bit.ly/2u0PXPB>
 Sample - voice created by Acapela DNN: <http://bit.ly/2t3wMVg>
- Anonymous user, French
 Sample - Original voice: <http://bit.ly/2s833X5>
 Sample - voice created by Acapela DNN: <http://bit.ly/2sp9ibe>

MOV DNN (Swedish)



Lyrebird



Voice vs Message Banking™

Text to pronounce:
Yo!

Speed: %
Shaping: %

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.”

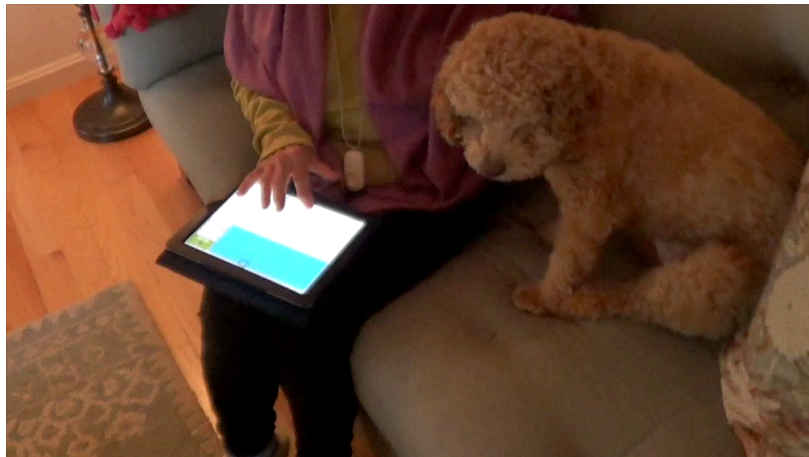
Costello, J.M. 2008

“Our voice is our ACOUSTICAL fingerprint”

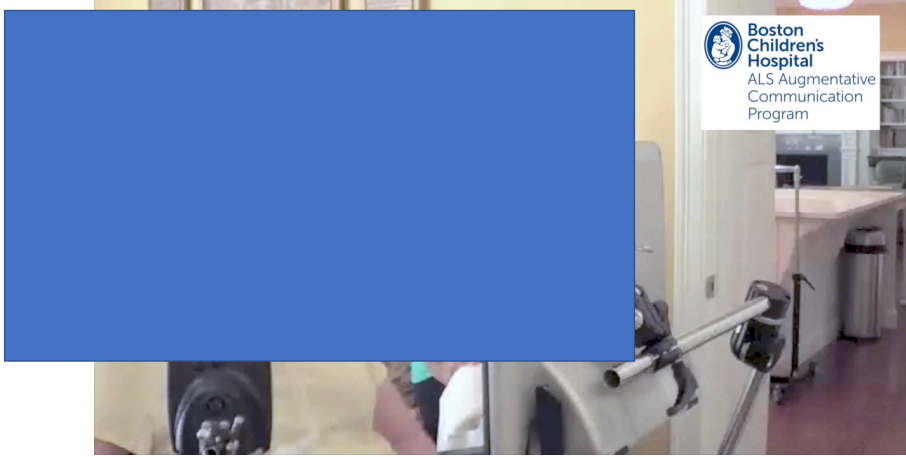



Costello, J.M. 2010


Our Message Banking™ Model



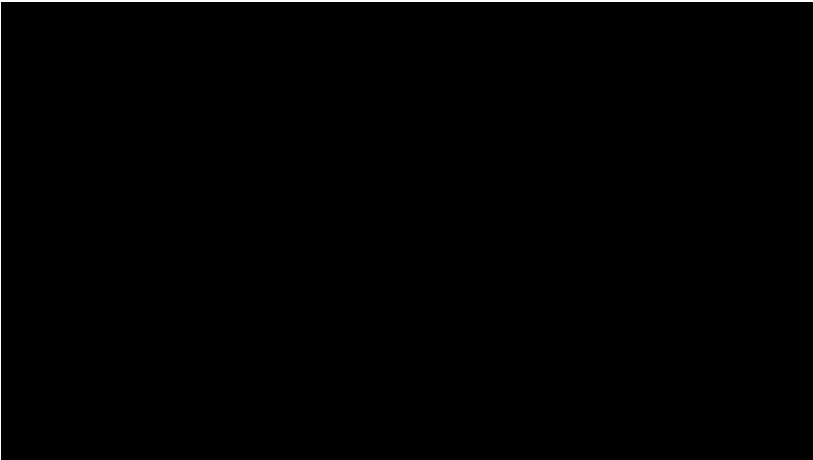
Our Message Banking™ Model





 **Boston Children's Hospital**
ALS Augmentative
Communication Program

 **HARVARD MEDICAL SCHOOL**
TEACHING HOSPITAL

Our Message Banking™ Model



 **Boston Children's Hospital**
ALS Augmentative
Communication Program

 **HARVARD MEDICAL SCHOOL**
TEACHING HOSPITAL

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

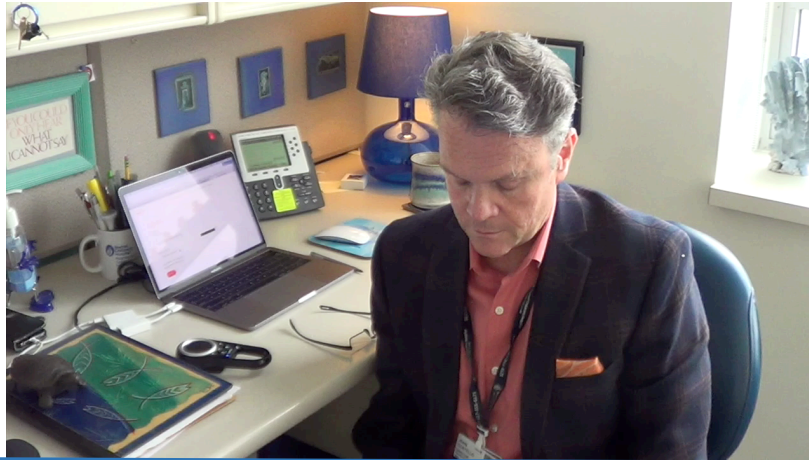
	Model Talker Confirmed 7.22.18	Cereproc	VocalID Confirmed July 23	Acapela Confirmed July 24	Lyrebird
Process	1541 scripted messages	1000 Scripted messages Choose build for Window or OSX in advance.	1575 scripted messages	-30 minutes of quality message banking – authentic messages -minimum 30 minutes of voice bank up to three hours from script	Scripted messages
Time commitment	Seven hours	Four hours or 40 minutes for lesser quality	Seven hours	Minimum 30-minutes to 90 minutes -total of one's message bank with at least 30 minutes	Three hours
Literacy need	yes	yes	NO, each message can be read aloud by software	No	yes
Bank from home	yes	yes	yes	yes	yes
Premier before Pay*****	yes	no	no	yes	n/A
Cost (US dollar)	100.	675.	1500.	99.00 per year OR 999. once.	free
Incorporate into SGD	yes	yes	yes	yes	NO, only works on the Lyrebird sitte/ with wireless
Requires wireless to access	NO	NO	NO	NO	YES
Can use Message bank instead of artificial script	No **	no	no	yes	no
languages	english	english	English	13 languages	English
Equipment needed	special microphone	special microphone provide @ purchase	Special microphone	special microphone OR handheld .wav recorder	special microphone



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