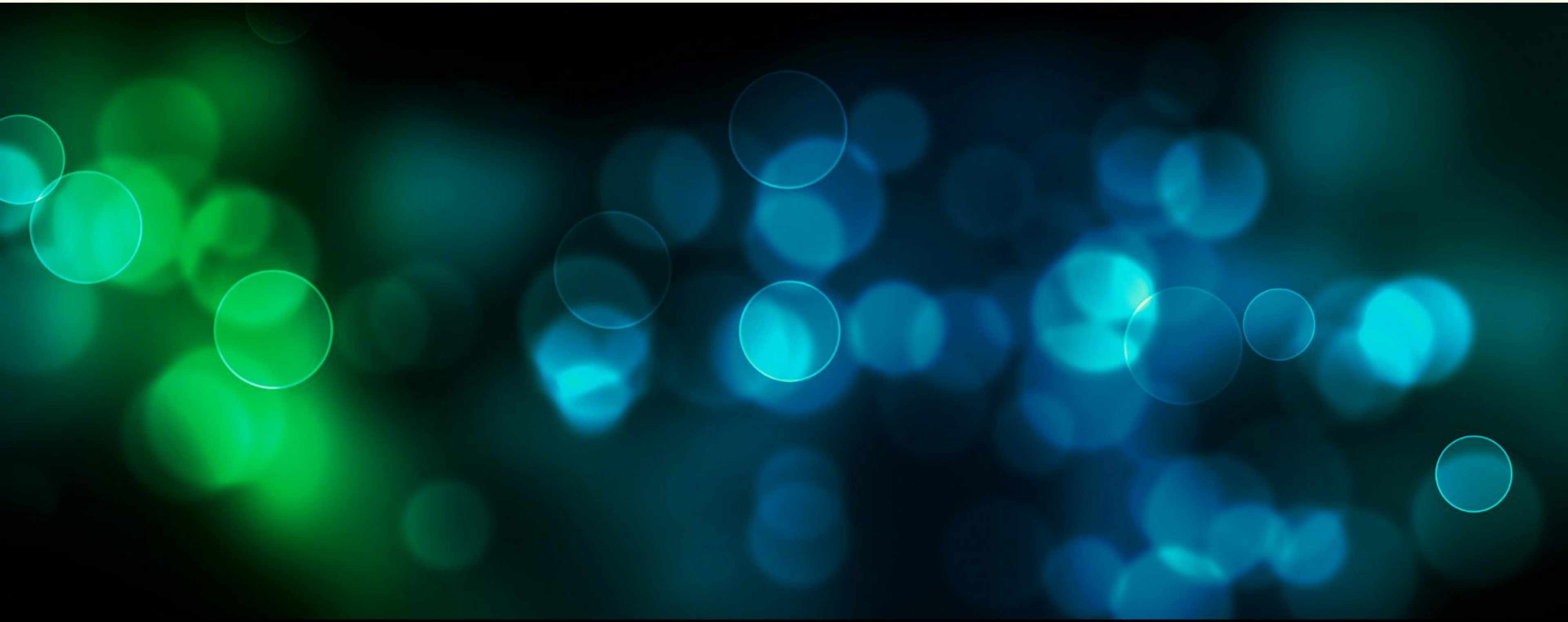
Westat®

Text-To-Speech for Multilingual ACASI Systems



Comparative Survey Design and Implementation: Annual Workshop

March 21, 2013

Jeff Phillips, Brad Edwards, Ed Dolbow

Westat®

PATH
Population Assessment
of Tobacco and Health

This project has been funded in whole or in part with Federal funds from the National Institute on Drug Abuse, National Institutes of Health, and the Food and Drug Administration, Department of Health and Human Services, under Contract No. HHSN271201100027C.

DEPARTMENT OF HEALTH & HUMAN SERVICES • USA

NIH National Institute on Drug Abuse

FDA

Contents

- Overview, Background
- Population Assessment of Tobacco and Health (PATH)
- ACASI Process
- Text-to-Speech (TTS) Technology
- Examples
- Conclusion

Overview, Background

What is ACASI?

- Audio Computer Assisted Self Interviewing
- Questions are displayed and read aloud by the computer
- Offers advantages in situations involving:
  - Privacy / Sensitive subject matter
  - Low-literacy subjects
  - Sight Impaired participants

How ACASI Works

- The ACASI instrument is presented using a simplified user interface
- Modern laptops and tablets allow touch screen use
- A tutorial precedes the ACASI session
- Headphones are used for privacy
- Some systems feature a “hide” function to turn off the visual display, also for privacy
- The interviewer assists if needed

Putting the “A” in ACASI

- ACASI yields higher reports of sensitive behaviors compared to paper-and-pencil questionnaires (Turner et al 1998) and CAPI (Tourangeau & Smith 1996)

- BUT, computerization more important in improving data quality than audio (Tourangeau & Smith 1996; Couper, Tourangeau & Marvin 2009)

- Evidence that neither the gender of the voice nor whether it was synthesized (TTS) or a recorded human voice affects responses (Couper, Singer, Tourangeau 2004)

Couper et al 2012

- Recent gains in TTS voice quality, reduction in costs
- National Survey of Family Growth Cycle 7 used recorded voice; NSFG Cycle 8 uses TTS
- Quasi-experimental design suggested TTS had no negative effects on data quality
- Respondents may make more use of TTS audio
- Respondents take less time with TTS ACASI

PATH

PATH

Population Assessment of Tobacco and Health

- Population Assessment of Tobacco and Health
- PATH is funded by NIH through the National Institute on Drug Abuse (NIDA) and the Food and Drug Administration (FDA)
- PATH is a national longitudinal field study of tobacco use and how it affects the health of people in the United States
- Westat has just completed the PATH Field Test and is preparing for main study launch in September 2013

The PATH ACASI Story

- Because of the sensitive nature of some of the questions in the PATH instruments, PATH chose to use ACASI for the main in-person interviews.
- The PATH team began with the standard “voice talent” concept, and a one-hour ACASI instrument.
- Intense schedule pressure and the need for Multi-lingual instruments (Spanish and 5 Asian languages) forced the team to look for efficiencies.
- PATH turned to TTS for field test 2012.
- PATH is currently evaluating field test results and preparing for national study.

ACASI Process

Traditional ACASI Voice

Recorded voice "snippets"

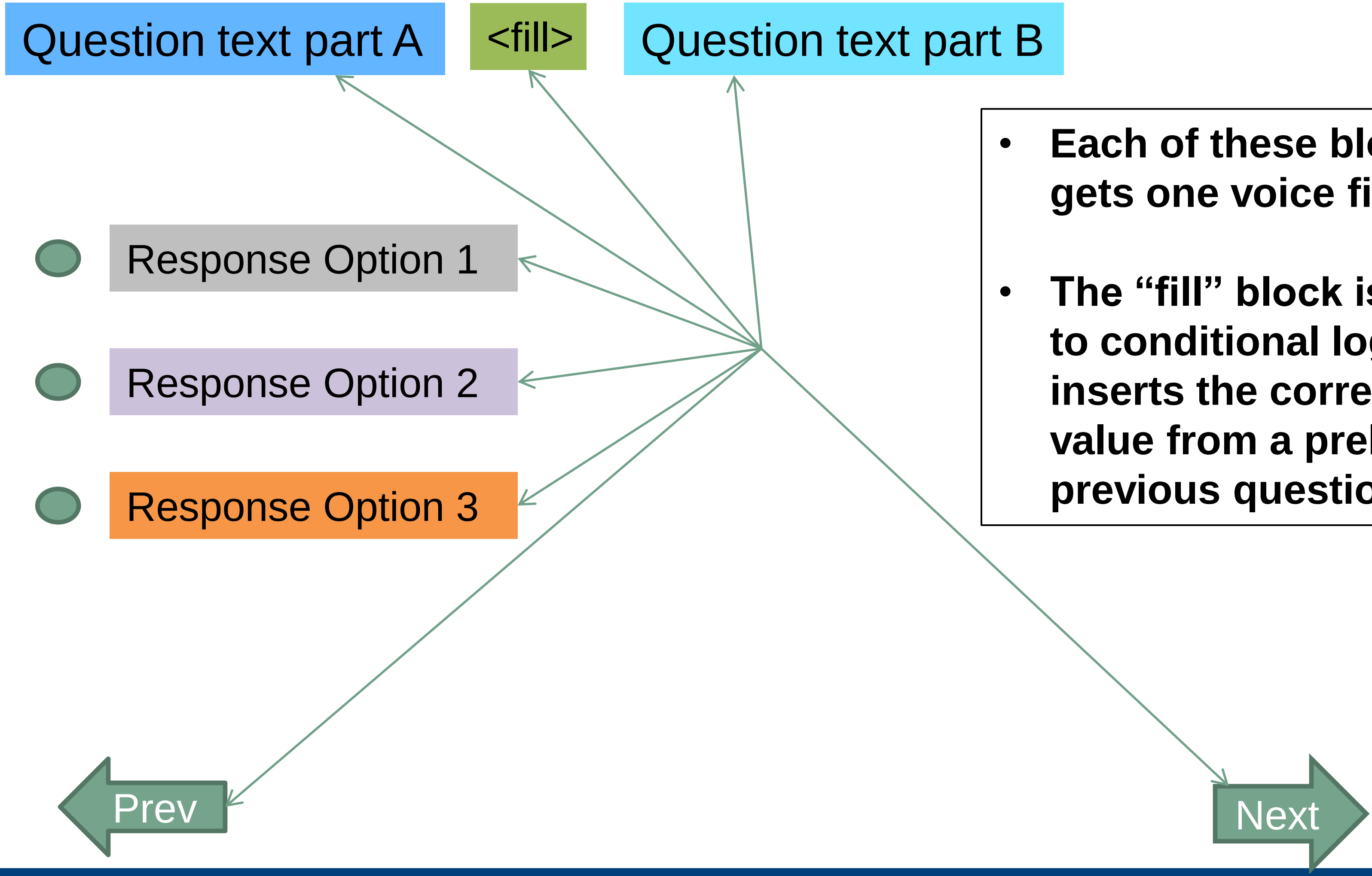
A "voice talent" reads from detailed script, based on final instrument:

- The questions
- The response options
- Standard responses such as "I don't know"
- Control words ("Next," "Erase")
- Numbers and letters

Traditional ACASI Process

1. Specify and program the instrument (same as CAPI) (and translate)
2. Record the voice(s)
3. Generate voice fragments (.wav or .MP3)
4. Place questions in ACASI framework
5. Integrate voice files with questions
6. Test instrument
7. Test all applications together on device

Where ACASI Voice Files Go



Question text part A

<fill>

Question text part B

○ Response Option 1

○ Response Option 2

○ Response Option 3

Prev

Next

- **Each of these blocks gets one voice file.**
- **The “fill” block is linked to conditional logic that inserts the correct fill value from a preload or previous question.**

The Voice File Math

- A simple ACASI question requires multiple individual audio files to be placed into code to read the question, responses, and controls.
- The previous example required 5 audio files for the static parts of the question and responses.
- The fill could represent another X number of audio files.

Using Audio Files for Fills

Prior Question 1:

What is your preferred tobacco product? *[stored as TobaccoProduct]*

- ○ Cigarettes
- ○ Cigars
- ○ Pipe
- ○ Chewing Tobacco

Prior Question 2:

How many times per day do you use <*TobaccoProduct*> ?

Enter a number: □ *[stored as NumUses]*

Prior Question 3:

When was the last time you used <*TobaccoProduct*>?

Enter a date: □ *[stored as DateUsed]*

Placing the Fills

When you used <TobaccoProduct> <NumUses> times **on**

<DateUsed>, **did you consider that to be "too often"?**

○ **Yes**

○ **No**

Prev

Next

How Many Audio Files Do I Need?

- Current question:
  - —6 files for question text and responses
- Fill from prior question 1 (tobacco type):
  - —4 files (more in a real question)
- Fill from prior question 2 (enter a number):
  - — Depending on how the talent records numbers, it could be 1, 2, or perhaps 3 audio files selected from the individual numbers pre-recorded
- Fill from prior question 3 (enter a date):
  - — 3 files (month, day, and year)

The Stitching Problem

- As an instrument grows in length and complexity
  - —Audio file library becomes huge
  - —Placing individual files (“stitching”) is laborious
  - —Testing for fills is complex, time consuming
- Changes require:
  - —Finding the same voice talent as before
  - —Recording the new text
  - —Re-stitching and re-testing
- Three variables drive cost: instrument length, complexity, volatility

TTS Technology

Definition

Text to Speech (TTS) software converts normal language text into speech.

Other voice synthesis technologies convert linguistic symbols, such as phonetic transcriptions, into speech.

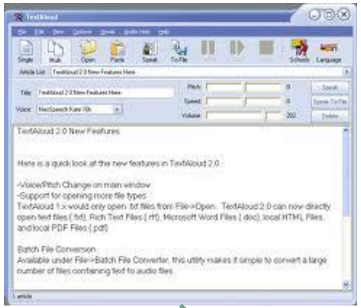
TTS History

- The first TTS device was arguably the *Voder*, demonstrated by Homer Dudley at the 1939 New York World's Fair.

- The first computer TTS engines were created in the 1950's and 60's.

- TTS became common on personal computers in the 1990's and on portable devices soon after.

The TTS Process



Article List: TextAloud 2.0 New Features Here

Title: TextAloud 2.0 New Features Here

Voice: NeoSpeech Kate 16k

Pitch: 0

Speed: 0

Volume: 200

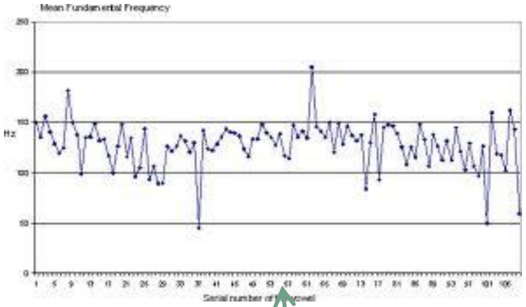
TextAloud 2.0 New Features

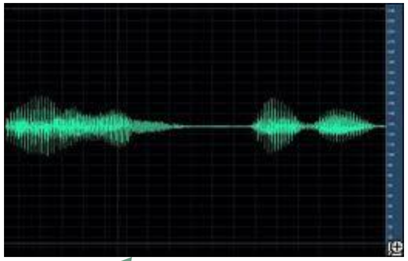
Here is a quick look at the new features in TextAloud 2.0.

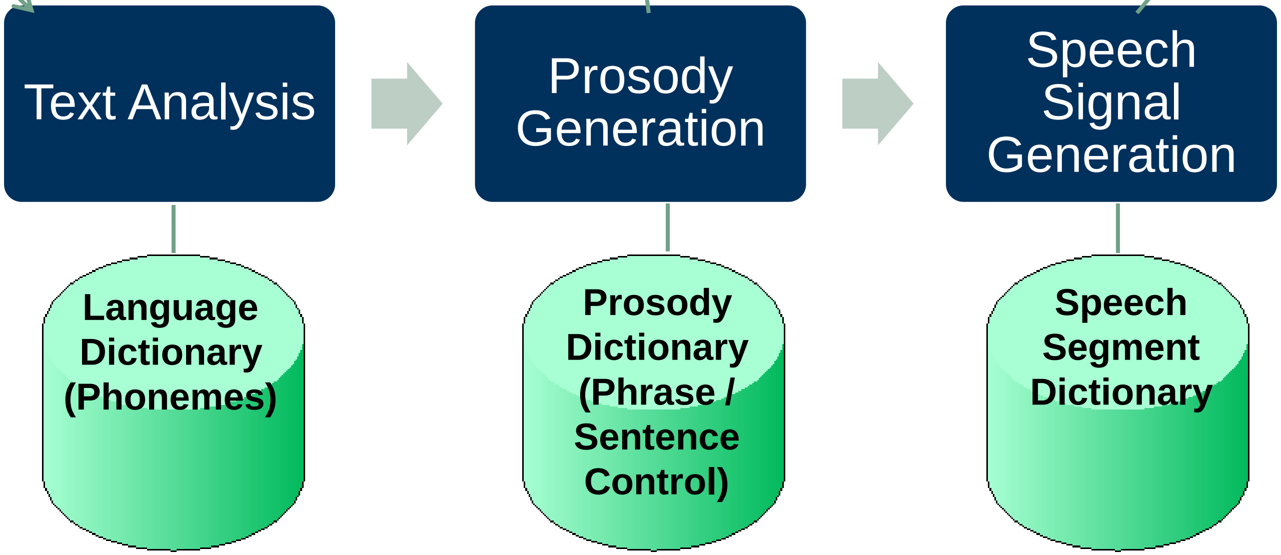
-Voice/Pitch Change on main window
-Support for opening more file types
TextAloud 1.x would only open .txt files from File->Open.. TextAloud 2.0 can now directly open text files (.txt), Rich Text Files (.rtf), Microsoft Word Files (.doc), local HTML Files, and local PDF Files (.pdf)

Batch File Conversion
Available under File->Batch File Converter, this utility makes it simple to convert a large number of files containing text to audio files.

1 article





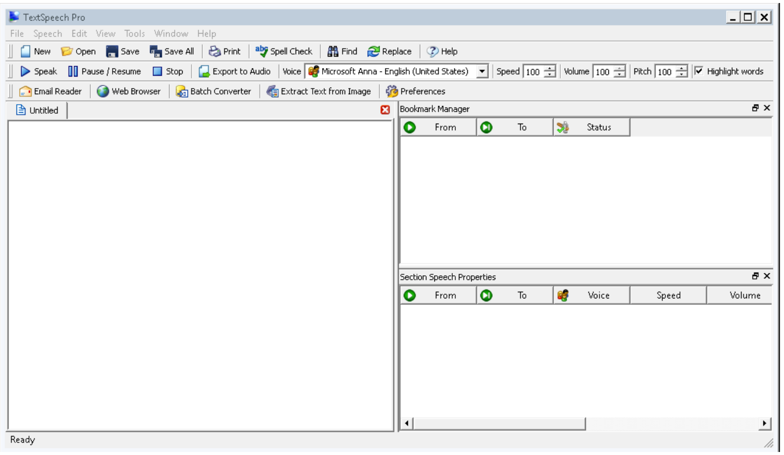

Text Analysis
Prosody Generation
Speech Signal Generation
Language Dictionary (Phonemes)
Prosody Dictionary (Phrase / Sentence Control)
Speech Segment Dictionary

How TTS Supports ACASI

- A runtime TTS engine, called a “voice,” resides on the ACASI computer.
- Instead of executing audio files, the ACASI code calls the voice, which reads the text.
- The instrument platform has already computed the fills, so the voice just reads what’s there.

Components of TTS Software

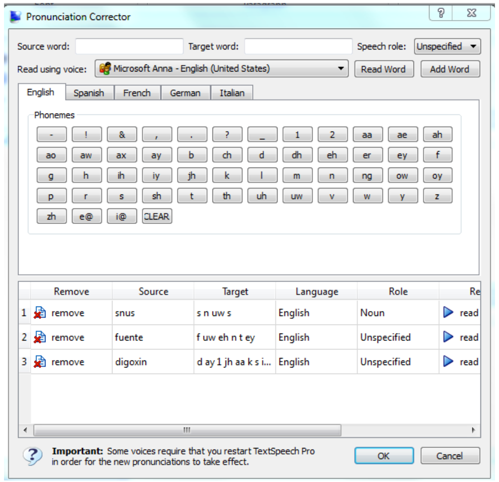
Prosody Adjustment Interface



Voice: Microsoft Anna - English (United States)
Speed: 100
Volume: 100
Pitch: 100
☑ Highlight words

Runtime Voice

Pronunciation Editor



Source word:
Target word:
Speech role: Unspecified
Read using voice: Microsoft Anna - English (United States)

	Remove	Source	Target	Language	Role	Re
1	remove	snus	s n uw s	English	Noun	read
2	remove	fuente	f uw eh n t ey	English	Unspecified	read
3	remove	digoxin	d ay 1 jh aa k s i...	English	Unspecified	read

Important: Some voices require that you restart TextSpeech Pro in order for the new pronunciations to take effect.

- Allows adjustment of:
  - ✓ Speed
  - ✓ Pitch
  - ✓ Pauses between phrases
- Generates SSML markup

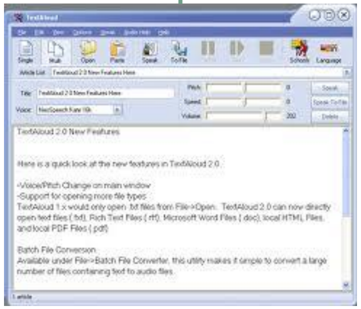
- Allows
  - ✓ Changing the phoneme read when text is interpreted
  - ✓ Changing a syllable’s stress
  - ✓ Inserting between-syllable pauses
- Generates phonemic transcription

Building A TTS ACASI Instance

Test Read Each Question

Adjust Pronunciation

Install Dictionary



Article List: TextAloud 2.0 New Features Here

Title: TextAloud 2.0 New Features Here

Voice: NeoSpeech Kate 16k

Pitch: 0

Speed: 0

Volume: 200

TextAloud 2.0 New Features

Here is a quick look at the new features in TextAloud 2.0.

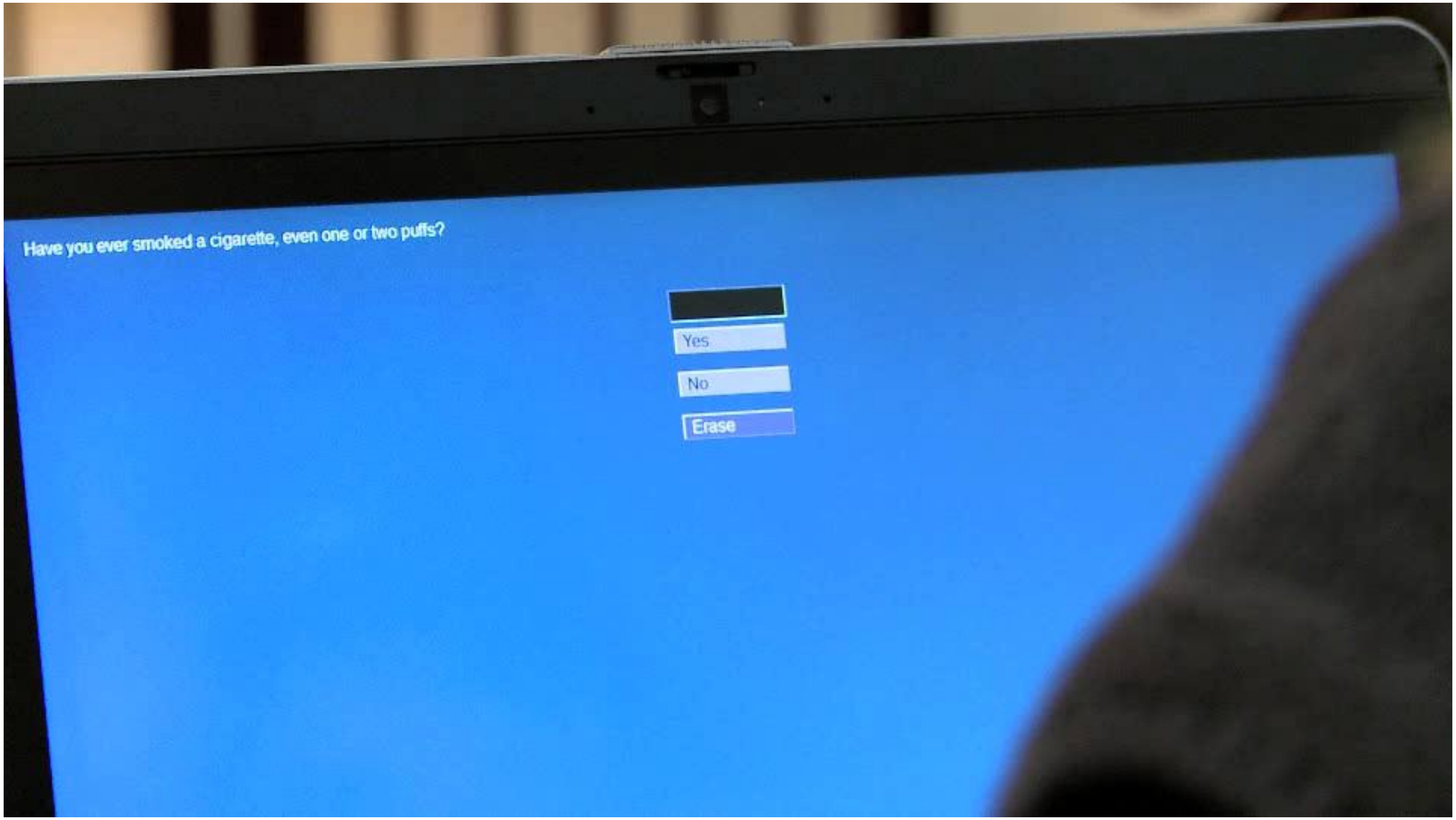
-Voice/Pitch Change on main window
-Support for opening more file types
TextAloud 1.x would only open .txt files from File->Open. TextAloud 2.0 can now directly open text files (.txt), Rich Text Files (.rtf), Microsoft Word Files (.doc), local HTML Files, and local PDF Files (.pdf)

Batch File Conversion
Available under File->Batch File Converter, this utility makes it simple to convert a large number of files containing text to audio files.

Language Dictionary (Phonemes)

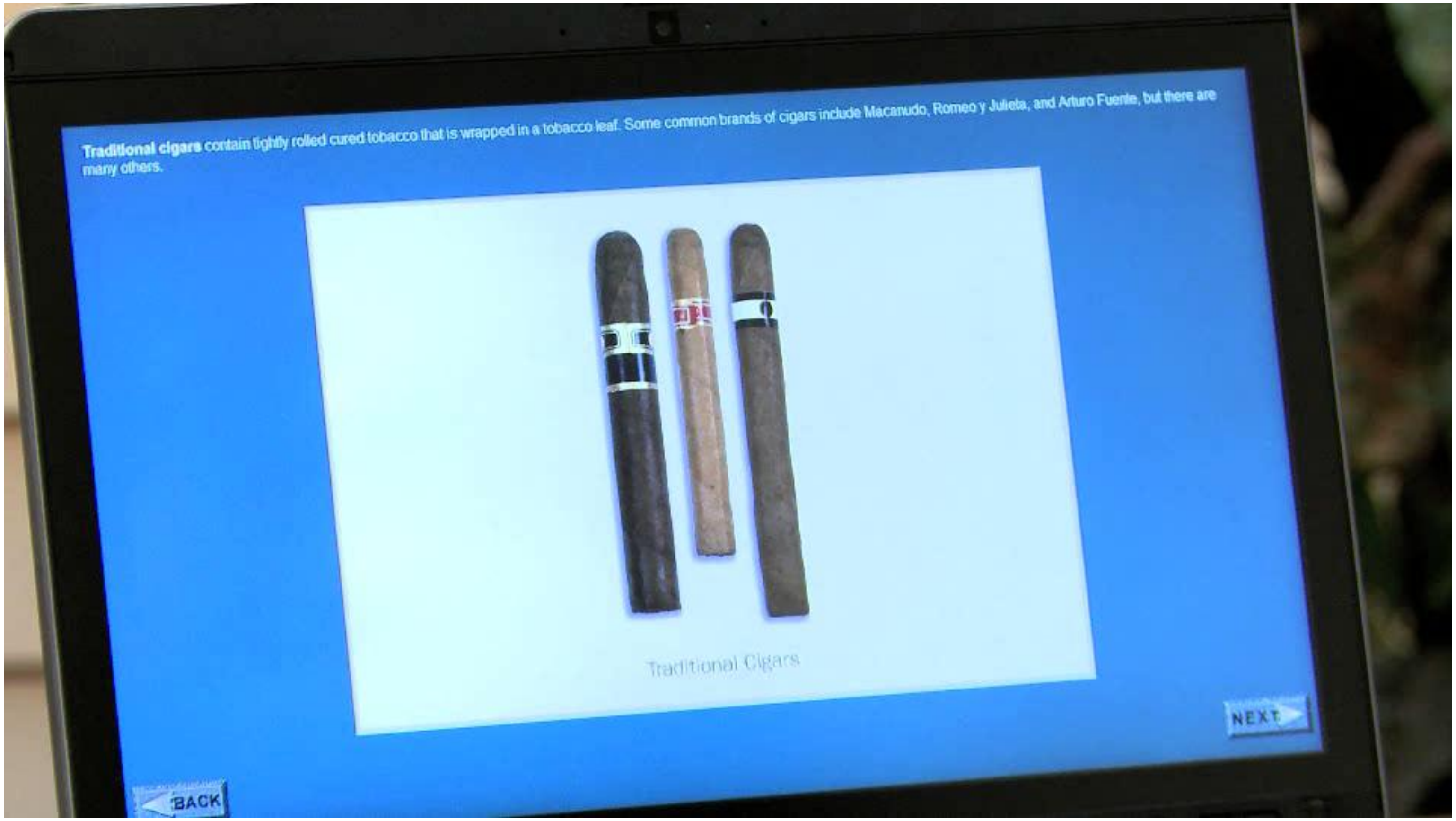
TTS ACASI Examples

TTS/ACASI Example 1



Have you ever smoked a cigarette, even one or two puffs?

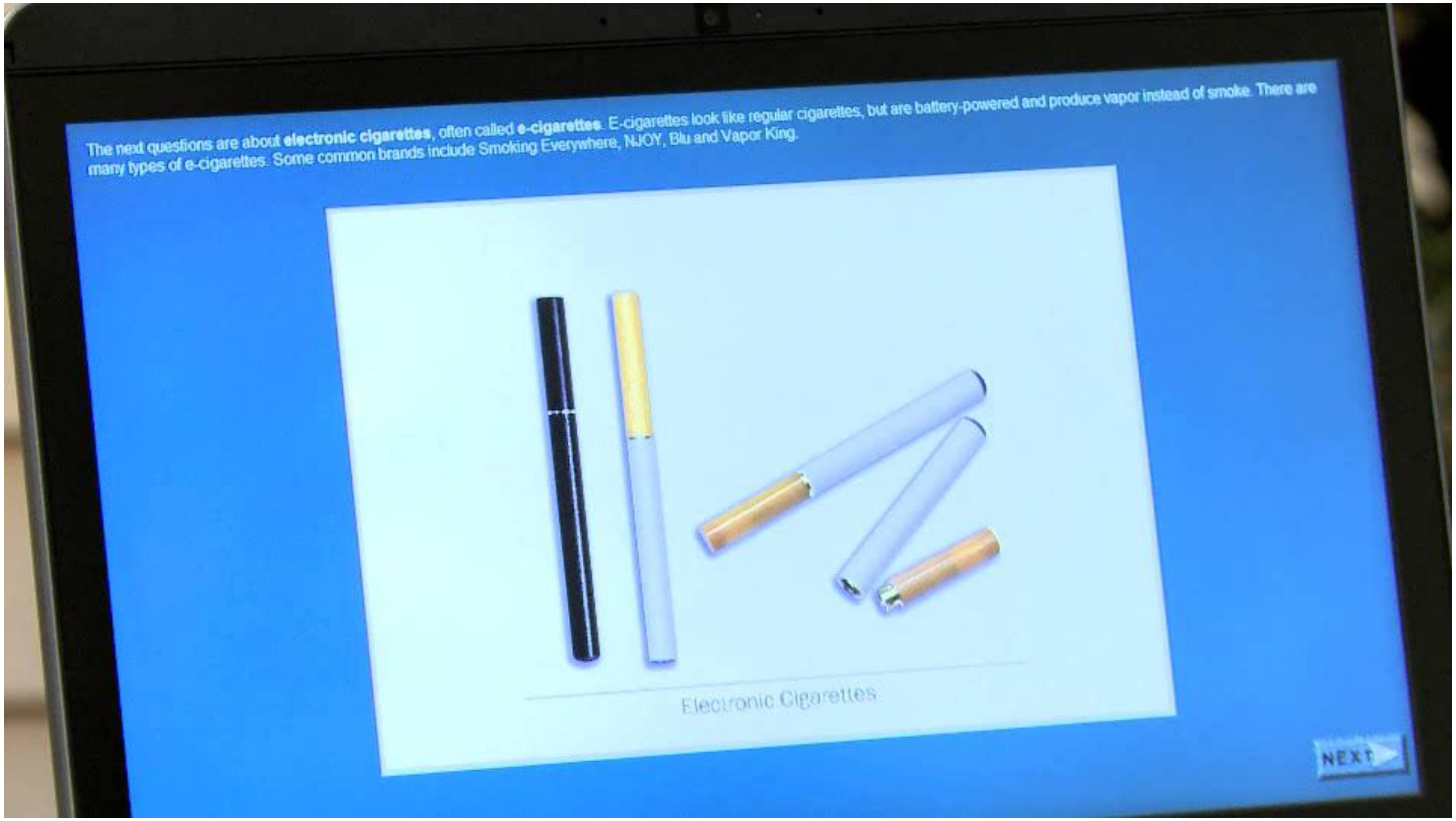
TTS/ACASI Example 2



Traditional cigars contain tightly rolled cured tobacco that is wrapped in a tobacco leaf. Some common brands of cigars include Macanudo, Romeo y Julieta, and Arturo Fuente, but there are many others.

Traditional Cigars

TTS/ACASI Example 3



The next questions are about **electronic cigarettes**, often called **e-cigarettes**. E-cigarettes look like regular cigarettes, but are battery-powered and produce vapor instead of smoke. There are many types of e-cigarettes. Some common brands include Smoking Everywhere, NJOY, Blu and Vapor King.

TTS Pros and Cons

PROS

- The stitching process goes away.
- Question text changes require little or no change to the TTS voice.
- There is no dependence on the availability and schedule of a human voice talent.
- Adding languages is easy.

CONS

- Voice quality is not as high as a carefully stitched human voice.
- The voices require licensing, usually per computer.
- Less common languages are not always available.

Evolving Technology

Microsoft
“Anna”

Natural Voices
“Crystal”

NeoSpeech
“Kate”

Conclusion

TTS Is Good, Getting Better

- No negative effect on data quality
- Major cost and schedule advantages over voice talent
- Even greater advantages for multilingual ACASI, beginning with pretesting
- Unfortunately, PATH baseline interviews will be limited to English and Spanish

Thank You

bradedwards@westat.com