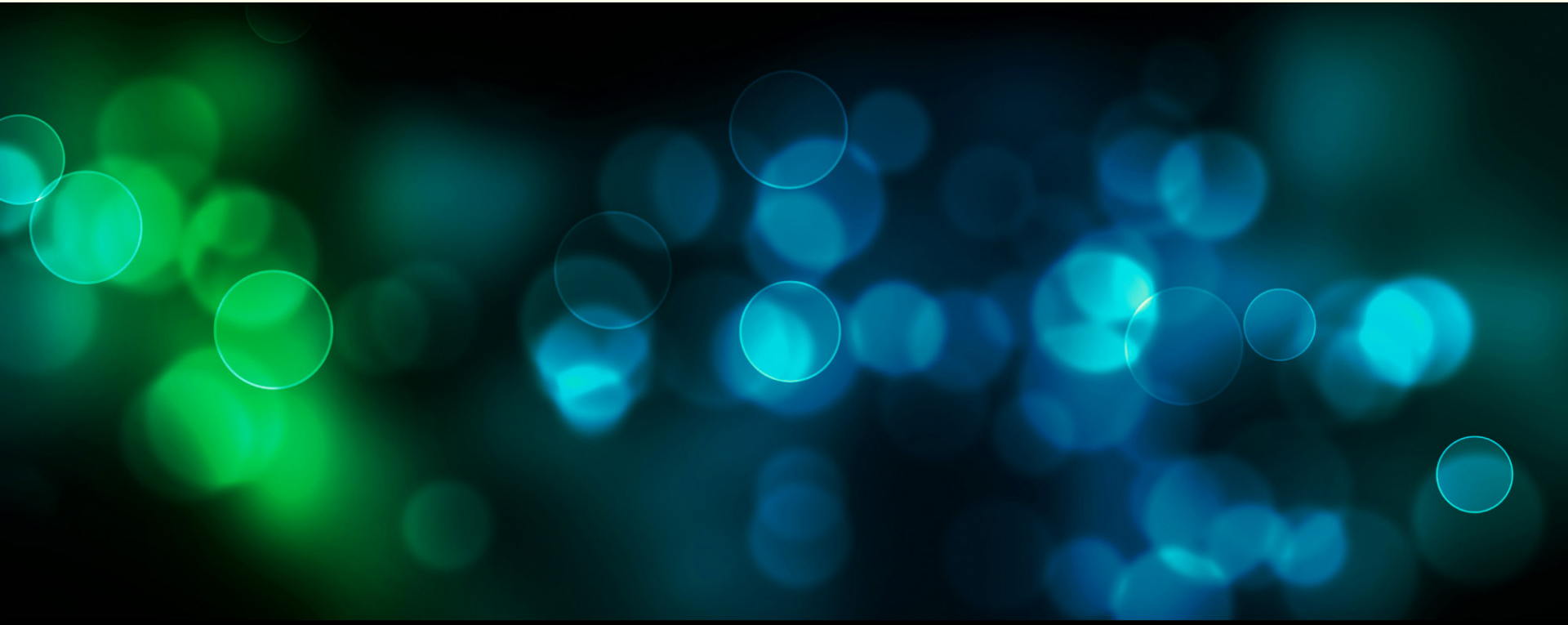
Using Text-To-Speech Software for ACASI



FedCASIC

March 21, 2013

Jeff Phillips, Brad Edwards, Ed Dolbow

Westat

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.

Contents

- Overview of ACASI
- The PATH Study
- The ACASI Process
- Inefficiency of Voice Talent ACASI Development
- TTS Technology
- Examples
- Conclusion

Overview of ACASI

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 usually precedes the ACASI interview
- After the tutorial, the participant may turn the computer away from the interviewer or move to another room
- Headphones can be used to keep the audio private
- The interviewer stands by to assist 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

The PATH Study

PATH
Population Assessment of Tobacco and Health

- Population Assessment of Tobacco and Health
- PATH has been funded 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.
- 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 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.

The ACASI Process

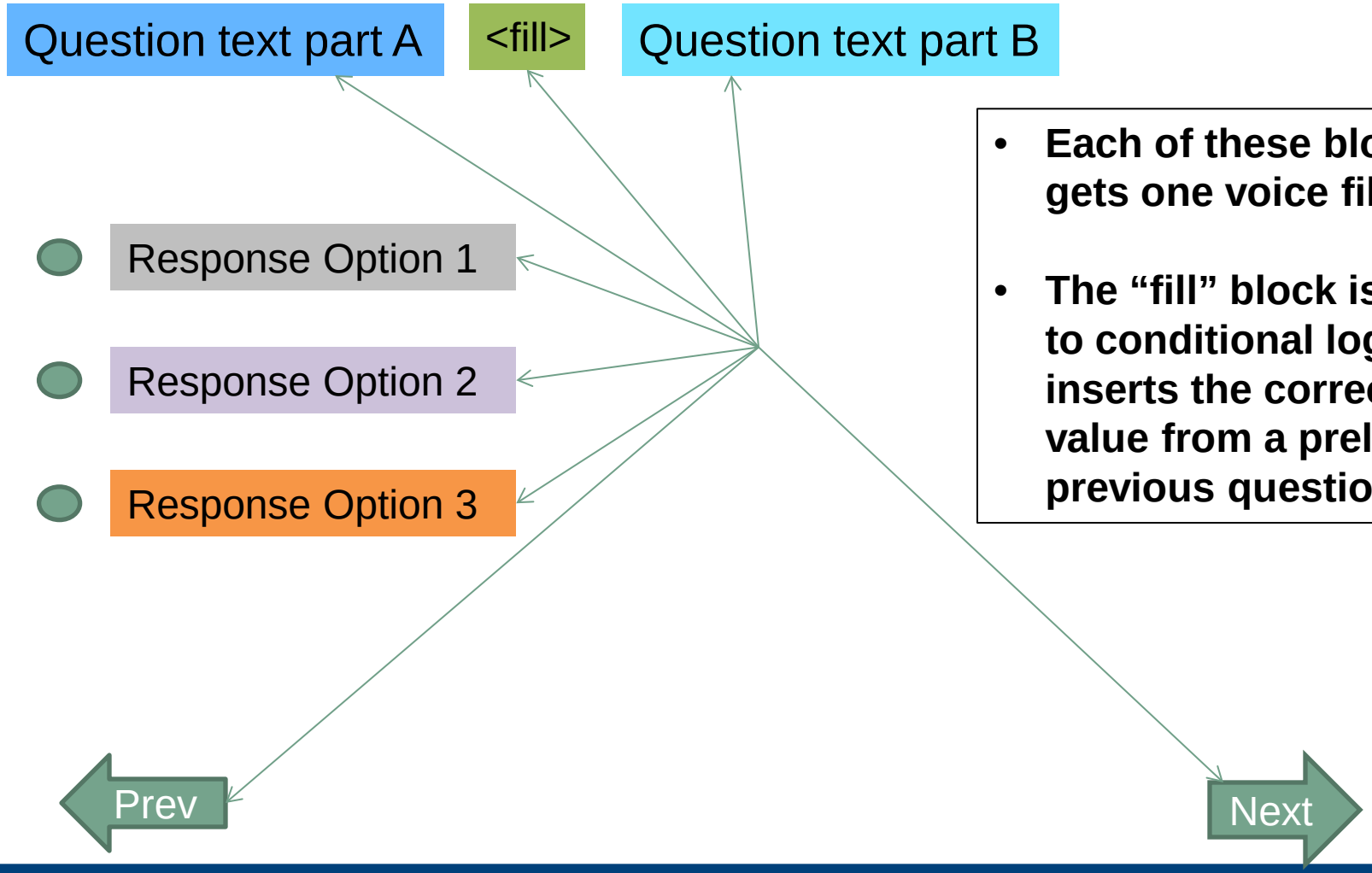
The ACASI Voice

- Traditional ACASI relies on recorded voice snippets
- Once the instrument questions are developed, a "voice talent" is given a detailed script
- The script includes:
  - The questions
  - The response options
  - Standard responses such as "I don't know"
  - Controls words such as "Next," "Erase," or "Clear"
  - Numbers and letters

The Traditional ACASI Process

- Spec and program the instrument (same as a CAPI)
- Record the voice
- Generate all the voice fragments as .wav or .MP3 files
- Place the questions in the ACASI framework
- “Stitch” the voice files into the questions

Where ACASI Voice Files Go



- **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.**

Inefficiency of Voice Talent ACASI Development

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
  - The audio file library becomes very large
  - Placement of the individual files (“stitching”) is laborious
  - Testing that the proper files have been used for fills is complex and time consuming
- To change an instrument one must:
  - Find the same voice talent as before
  - Record the new text
  - Re-stitch
  - Re-test
- Three variables drive cost:
  - Volatility of the instrument
  - Length of the instrument (number of quex)
  - Complexity of the instrument

Text to Speech Technology

Definition

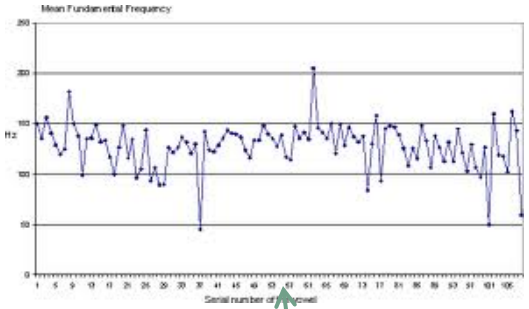
Text to Speech (TTS) software converts normal language text into speech.

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



Text Analysis → Prosody Generation → Speech Signal Generation

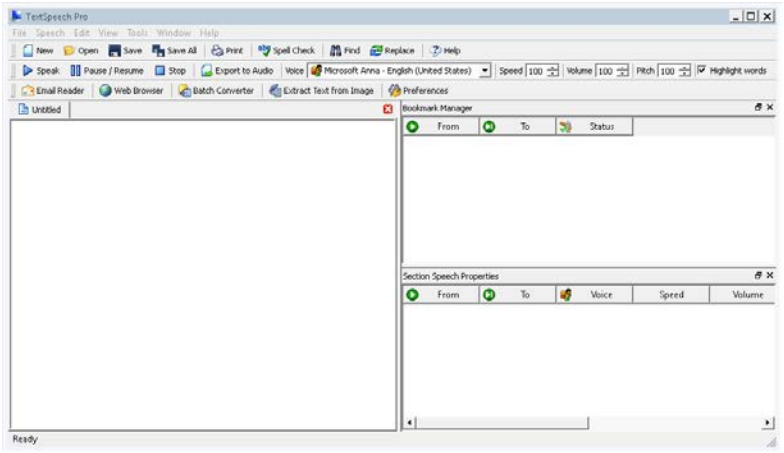
- Text Analysis → Language Dictionary (Phonemes)
- Prosody Generation → Prosody Dictionary (Phrase / Sentence Control)
- Speech Signal Generation → Speech Segment Dictionary

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

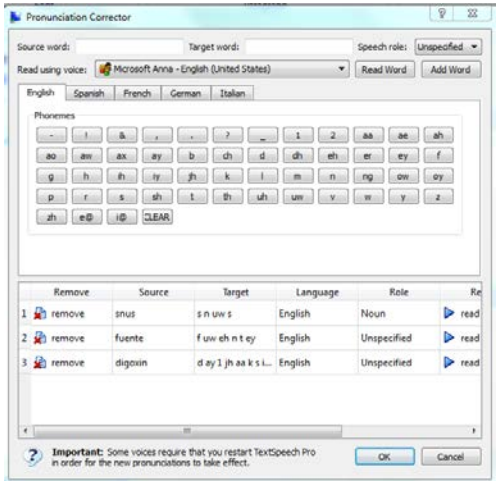
Prosody Adjustment Interface



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

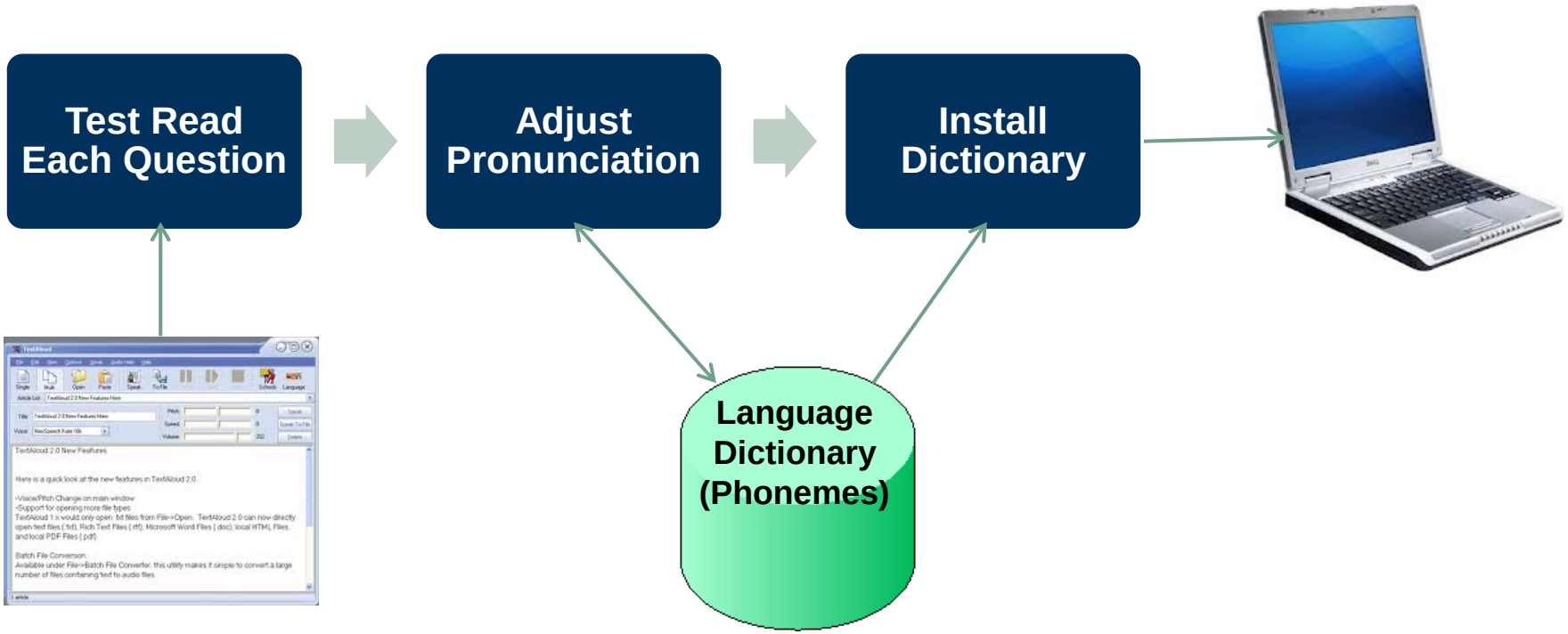
Runtime Voice

Pronunciation Editor



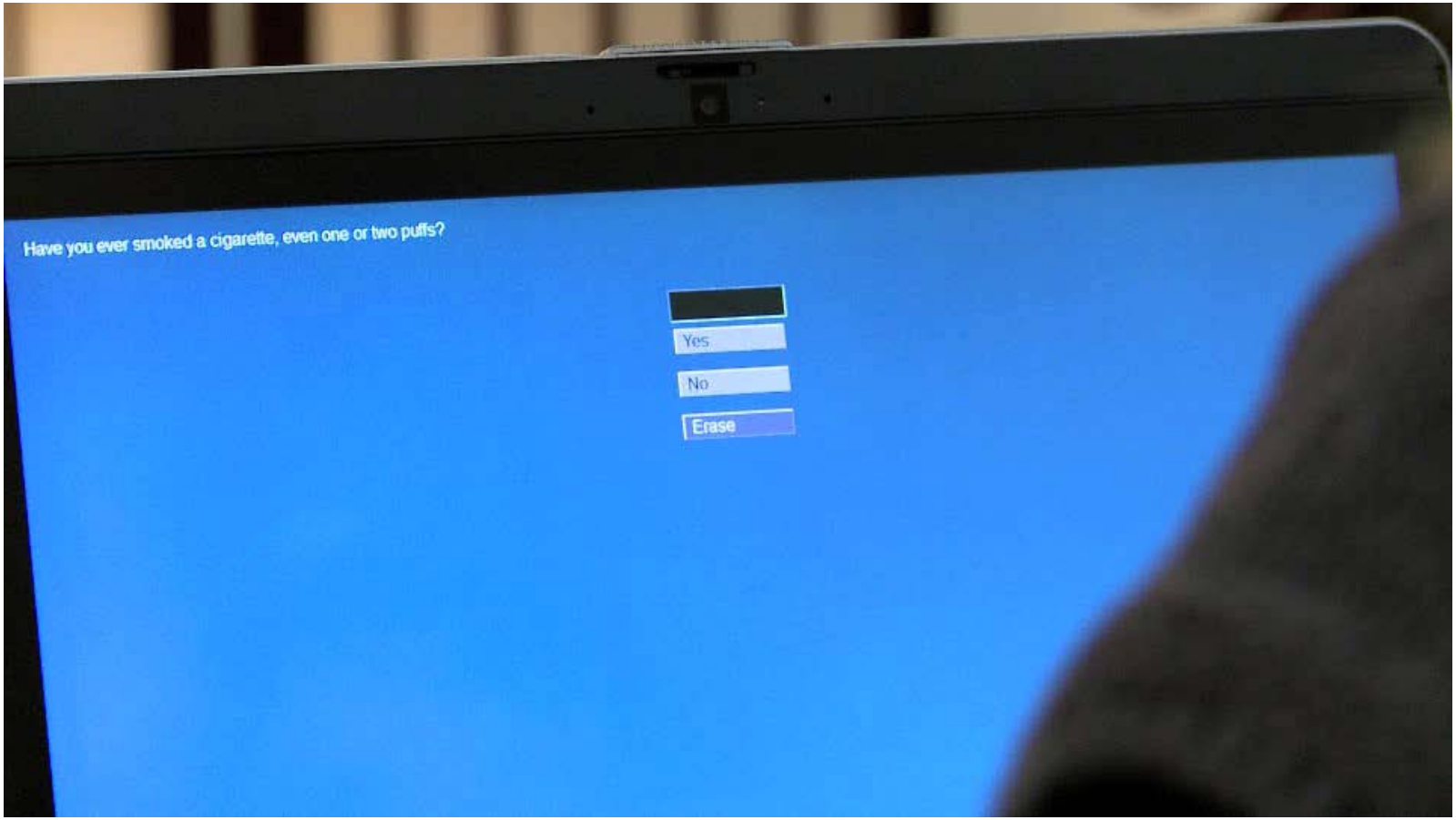
- Allows
  - ✓ Changing the phoneme read when text is interpreted
  - ✓ Changing a syllable's stress
  - ✓ Inserting between-syllable pauses
- Generates phonemic transcription

Building A TTS ACASI Instance

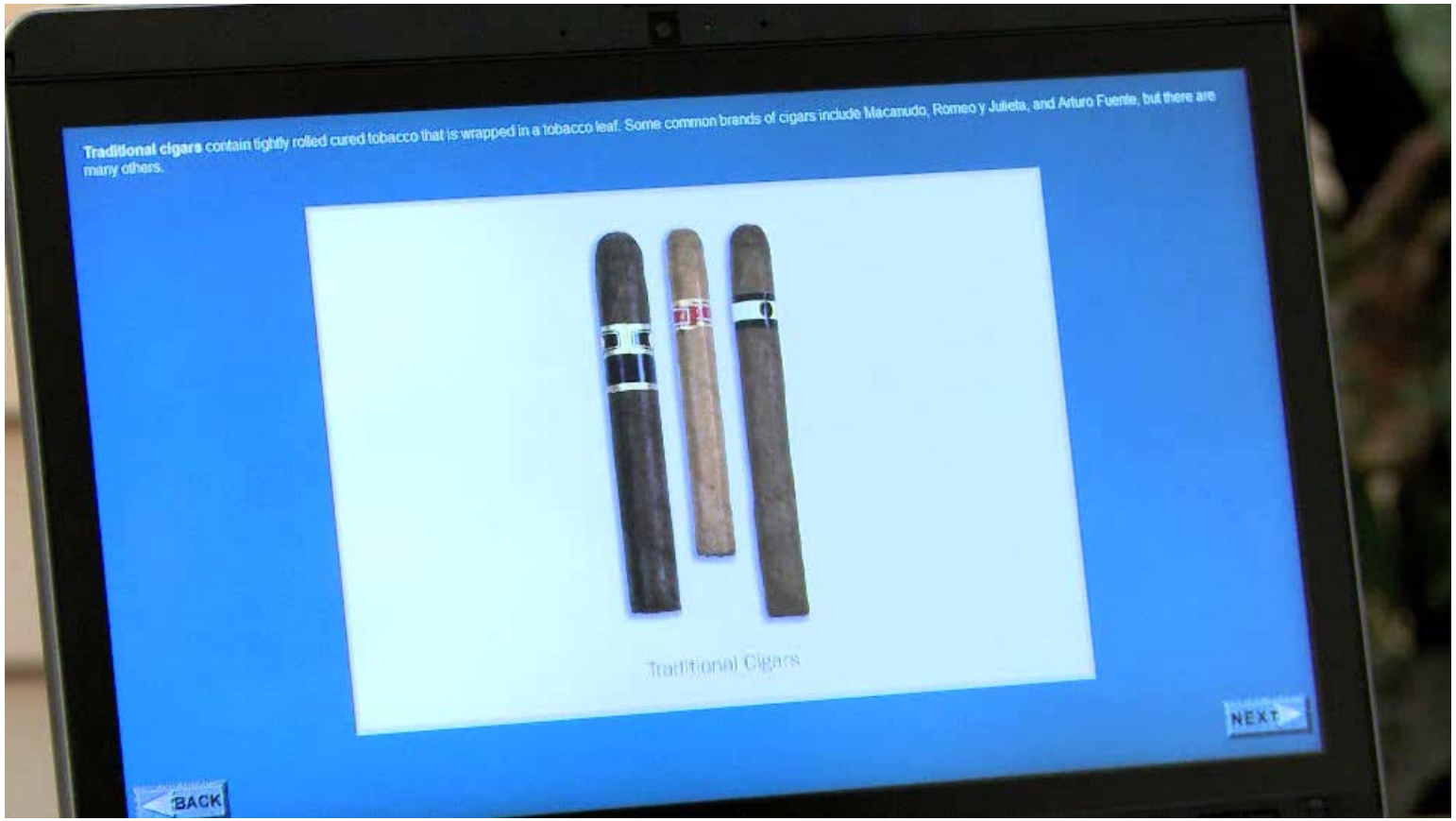


TTS ACASI Examples

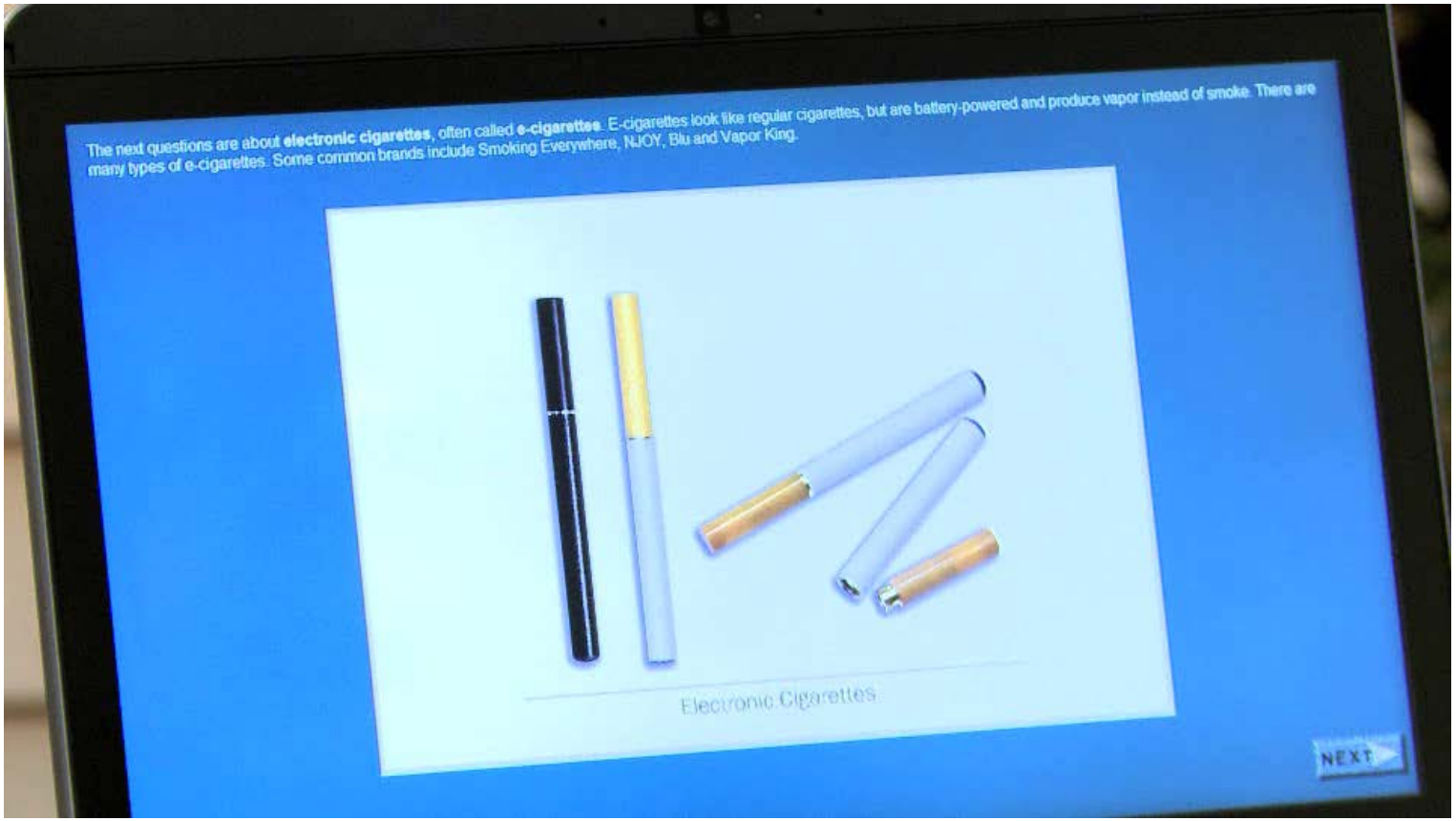
TTS/ACASI Example 1



TTS/ACASI Example 2



TTS/ACASI Example 3



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”

Questions?