Building ACASI Surveys Using State of the Art Text-to-Speech Technology

Gilbert Rodriguez, Emily McFarlane Geisen, Patricia LeBaron, Martin Meyer,

Vorapranee Wickelgren and Hilary Zelko (RTI International), Herman Alvarado (SAMHSA)

Introduction

- Audio Computer-Assisted Self-Interviewing (ACASI) software is used widely in federal surveys to gather sensitive data and to enable participation among special populations (i.e. low literacy levels or sight impairments).

- Conventional ACASI uses human voice recordings. Producing and modifying these recordings can be costly and labor intensive.

- Text-to-Speech (TTS) technology can replace human voice recordings in ACASI software.

- This demonstration showcases ACASI software that integrates Microsoft's freely available TTS technology (MS Speech Platform).

- ACASI with TTS is now implemented on the 2015 National Survey on Drug Use and Health (NSDUH) survey in both English and Spanish.

Study Background

- The National Survey on Drug Use and Health (NSDUH) provides national, state and sub-state data on substance use and mental health in the civilian noninstitutionalized population ages 12 and older.

- Approximately 700 field interviewers (FIs) working in all 50 states & D.C.

- Approximately 140,000 household screenings and 67,500 interviews completed annually.

- NSDUH survey instrument uses ACASI to collect sensitive data about alcohol, tobacco, prescription and illicit drug use, and mental health issues.

- Conducted by RTI under contract with SAMHSA's Center for Behavioral Health Statistics and Quality (CBHSQ).
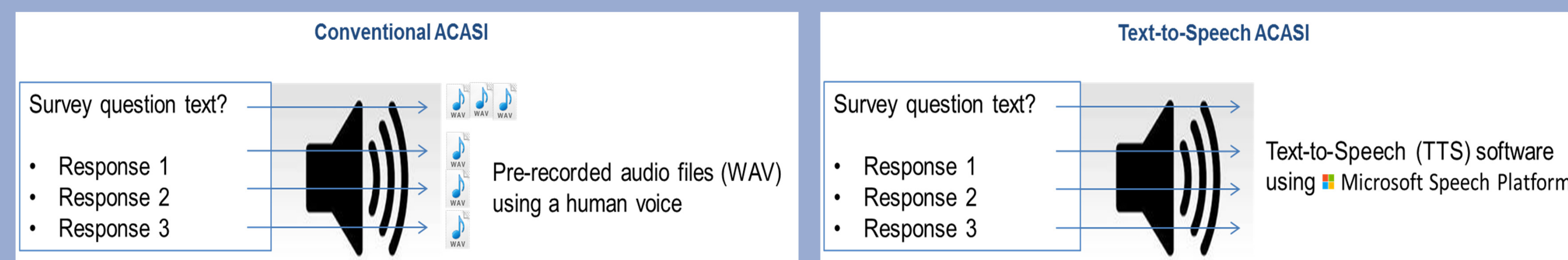
Acknowledgements

Text-to-Speech (TTS) Software

- Currently, a wide variety of TTS software products are available. TTS synthesis technology is currently built into major operating systems from Apple, Microsoft and Android. Proprietary software is also available from third party vendors such as Neospeech, Ivona, AT&T Natural Voices, etc.

- Products vary with respect to the number of different "voices" available. Voices differ by gender, accent, or language.

- Overall, TTS synthesis software works by transforming text input into phonetic representations which are then converted into audio output (synthesized speech).

Text Input → Phonetic Representations → Audio output
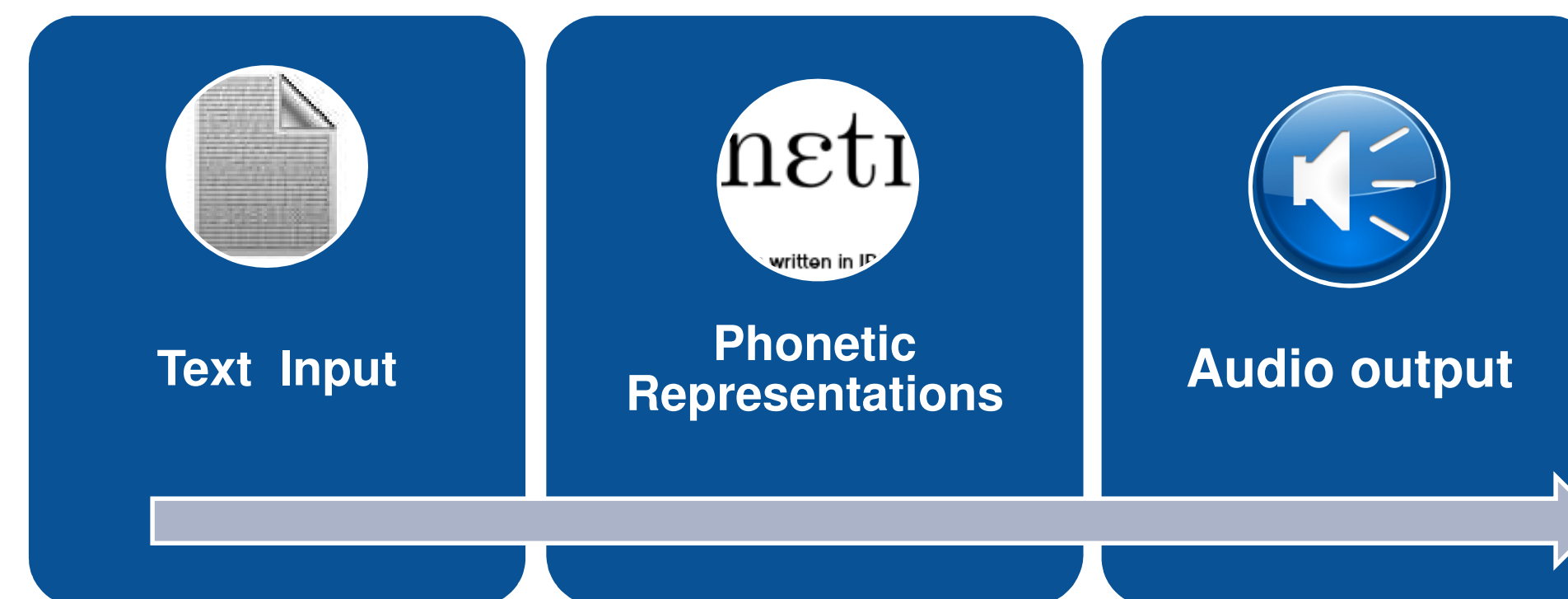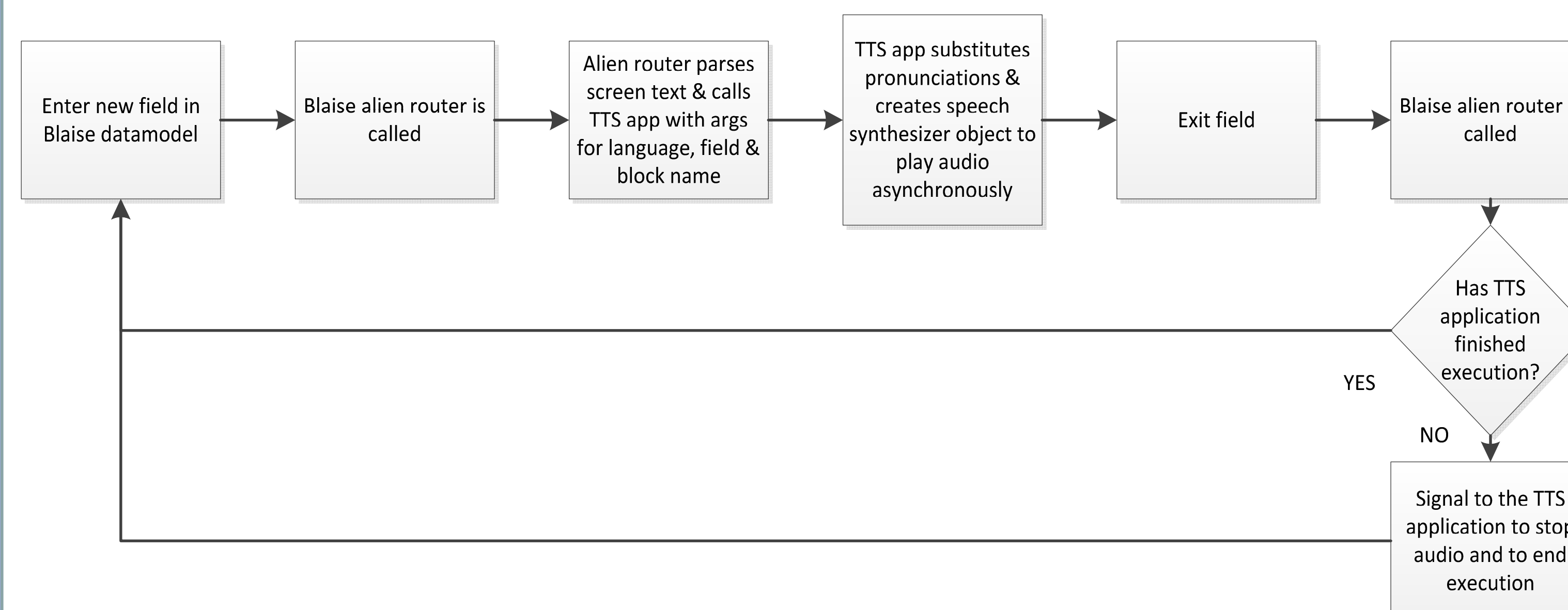
NSDUH Survey Instrument

- NSDUH survey instrument has implemented ACASI software using a conventional human audio-file based approach since 1999.

- NSDUH's ACASI software includes approximately 40 modules that include about 3000 question screens.

- Beginning in 2015, NSDUH adopted a new approach to ACASI that integrates Text-to-Speech (TTS) technology to dynamically generate audio as each question screen appears.

Conventional ACASI

Survey question text?
- Response 1
- Response 2
- Response 3

Pre-recorded audio files (WAV) using a human voice

Text-to-Speech ACASI

Survey question text?
- Response 1
- Response 2
- Response 3

Text-to-Speech (TTS) software using Microsoft Speech Platform

Technical Components of NSDUH's ACASI Software using TTS

- Microsoft Speech Platform 11 run-time, voices and SDK
- Blaise 4.8
- VB.NET for alien router and TTS application
- .NET Framework 3.5 or higher
- Microsoft Access for pronunciation databases

Integrating TTS with NSDUH's ACASI Blaise Software

Enter new field in Blaise datamodel → Blaise alien router is called → Alien router parses screen text & calls TTS app with args for language, field & block name → TTS app substitutes pronunciations & creates speech synthesizer object to play audio asynchronously → Exit field → Blaise alien router is called → Has TTS application finished execution? — YES / NO → Signal to the TTS application to stop audio and to end execution

Pronunciation Customization

- TTS pronunciations must be customized due to inherent challenges TTS faces in processing some types of text. Numbers, acronyms or abbreviations, proper nouns, specialized terminology, heteronyms require customization.

- MS Access database used to develop a table of pronunciations sorted by screen name/block name. Each record contains the literal word or phrase and then the phonetic spelling for the desired pronunciation.

- A global pronunciation file can be used to apply standard pronunciations across an entire survey instrument. We selected a Spanish voice that featured a "Castilian" pronunciation but used global pronunciations to minimize the accent.

- Pronunciation of English words in the Spanish instrument may need to be modified to ensure comprehension of the word (for example, words like "gallery", "rush", "hash").

- For some fields where the audio should not be read exactly as the question text appears onscreen, a separate Blaise language text specifically for the TTS audio is used instead (for example, text in columns that should be read vertically).

- Can be extended to use IPA pronunciations.

English Word		TTS Pronunciation
Daiquiri	→	Dackery
Drambuie	→	Drambuoy
Marnier	→	Marynyea
Triple sec	→	Triple seck

Spanish Word		TTS Pronunciation
ácido	→	ásidoh
hacienda	→	asiendo
Marzo	→	Marso

Advantages & Disadvantages of TTS ACASI

ADVANTAGES

- Eliminate dependence on a single human "reader" and enable consistency of voice over time.

- Flexibility in changing the wording/ordering of survey questions and responses, since audio is generated dynamically for each screen.

- Pronunciations can be updated "on the fly" quickly and easily by modifying a pronunciation file (text file).

- Smoother delivery of audio, especially where human voice relies on fills.

- Cost and process efficiencies in ACASI development.

- The MS Speech Platform is freely available and there are no licensing fees required for deployment.

- Eliminate management of audio files, but must manage customized pronunciations.

DISADVANTAGES

- Currently, the voices available for the MS Speech Platform are only female.

- Can be a challenge to generate appropriate inflection for some screens due to the wording used in the screen.

Conclusions

- As part of a pretest study to explore the use of TTS for NSDUH, participants indicated that the TTS voice was pleasant and understandable.

- None of the TTS components that we use require special licensing so getting started is low cost and does not require obtaining licenses from vendors.

- Although it requires sufficient time for reviewing and customizing the audio files, TTS is more flexible than conventional ACASI.

Contact Information
Gilbert Rodriguez
Research & Computing Division
Phone: 919.248.5825
grod@rti.org

RTI International
3040 Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194 USA

www.rti.org