Text-to-Speech Application in Audio CASI

Evaluation of Implementation and Deployment

Jeremy Kraft and Wes Taylor
International Field Directors & Technologies Conference 2006
May 21 – May 24

University of Wisconsin Survey Center

www.uwsc.wisc.edu

Outline – Text to Speech in ACASI

- Aspects of Traditional Approach to ACASI
  - Using live voice talent to record items
  - Pros and Cons
- Aspects of Text to Speech Approach (TTS)
  - Using voices generated artificially via software
- Examples of Text to Speech Recordings
  - Different Suites
  - Different Styles
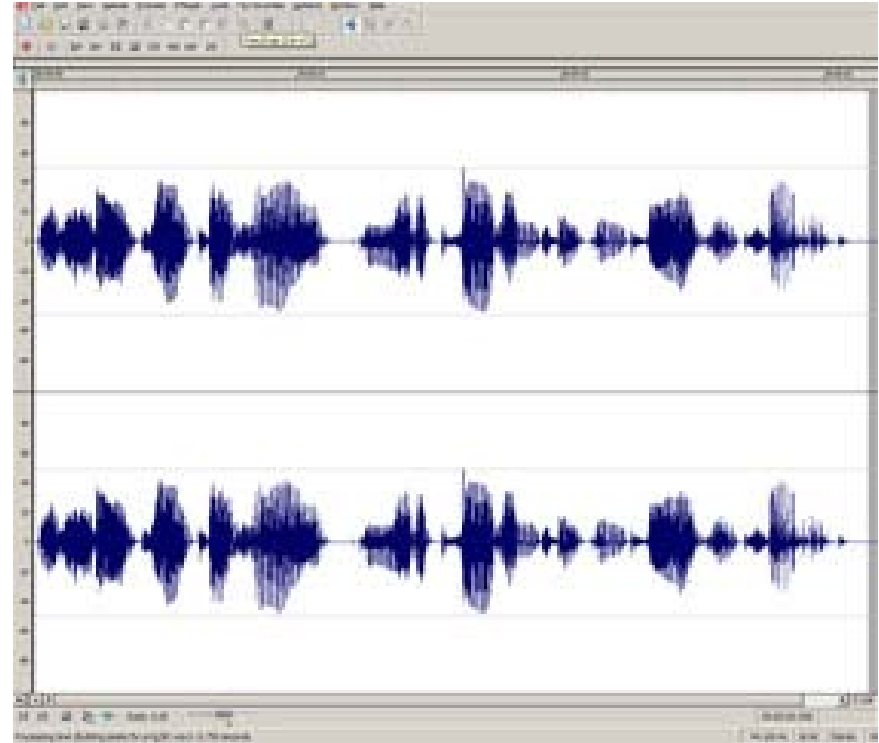
UWSC ACASI in CAPI settings

- Audio Computer-Assisted Self-Interviewing (ACASI)
  - Offers privacy to respondents
  - Increases reporting of socially undesirable behaviors
  - Benefits populations with low literacy rates
    - Respondent see both text and hears audio files played through headphones

University of Wisconsin Survey Center

ACASI Section Development Process

- Development of script
- Hire of voice talent and recording studio
  - Use interviewers or students on campus
- Record
  - Access to professional studio
- Edit individual sound files

Traditional UWSC Approach to ACASI

- Pros of using natural voice
  - Human voice talent is 'coachable'
  - Pronunciation can be fine-tuned
  - Human voice may build more rapport
- Cons of using natural voice
  - Unconscious inflection in wording
  - Time-consuming
  - Last minute changes are impractical

New Approach to ACASI Recording?

- Text To Speech (TTS)
  - Bypass voice talent, recording, and editing steps
- Pros of TTS
  - Greater flexibility for edits late in development
  - Less development time
  - Impersonal approach to sensitive questions
- Cons of TTS
  - Occasional unclear pronunciation
  - May seem too impersonal
  - Hard to duplicate accents (if desired)

New Approach to ACASI Recording?

- ## Previous Research

  - Couper, Mick P., Eleanor Singer, R. Tourangeau. "Does Voice Matter? An Interactive Voice Response (IVR) Experiment." *Journal of Official Statistics,* 20 (3): 551-570. 2004.

    - Telephone setting

    - No difference found between a recorded human voice, a human-like TTS, and a machine-like TTS

    - Gender had no affect on responses to sensitive items

    - Confirms significant differences between responses elicited by live interviewers (CATI) and those elicited by an automated system (IVR)

  - Other research suggests respondents prefer improved, female voice

UWSC

University of Wisconsin Survey Center

Text To Speech (TTS)

- Software
  - AT&T Natural Voices (www.research.att.com/~ttsweb/tts)
    - US & UK English, Spanish, German, French
    - $295 per voice (base system comes with male & female English voice)
  - Loquendo (www.loquendo.com)
    - Italian, Dutch, US & UK English, Spanish (4 dialects) German, French, Greek, Portuguese, Swedish & Chinese?
    - $750 for 30 audio pay-as-you-go minutes, also available for purchase
  - Festival (www.cstr.ed.ac.uk/projects/festival)
    - Open Source – developed by University of Edinburgh
    - US & UK English, Welsh, Spanish

Text To Speech (TTS)

- ## Examples – What does TTS sound like?

How many times did you deliberately damage property that didn't belong to you during the past 12 months?

	Female	Male
- AT&T:	🔊 🔊	🔊
- Loquendo:	🔊	🔊
- Festival:	🔊	🔊
- Human:	🔊	🔊

Text To Speech (TTS)

- ## AT&T Examples (cont'd)

  How many times did you deliberately damage property that didn't belong to you during the past 12 months?

- Crystal 🔊 (tempo) 🔊 (emphasis) 🔊

- Lauren 🔊

- Claire 🔊

- Mike 🔊

- Rich 🔊

- Rosa 🔊 (human) 🔊 (¿Qué es la tasa de cambio?)

- Arnaud 🔊 (Voulez-vous cesser de me cracher dessus pendant que vous parlez!)

Future Plans

- ## More research

  - ### In-house experiment with face to face setting with TTS software and recorded human voice

  - ### Wait? TTS software will only improve

Contact Information

Jeremy Kraft

jkraft@ssc.wisc.edu

608.262.5261

Wes Taylor

wtaylor@ssc.wisc.edu

608.263.3349